A Quantitative Framework For Large-Scale Model Estimation and Discrimination In Systems Biology

by

Hoda Eydgahi

M.S., Massachusetts Institute of Technology (2008)
B.S., Virginia Commonwealth University (2006)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Author .....
Department of Electrical Engineering and Computer Science
May 17, 2013

Certified by .....
Peter K. Sorger
Otto Krayer Professor of Systems Biology
Thesis Supervisor

Certified by .....
John N. Tsitsiklis
Clarence J. Lebel Professor of Electrical Engineering
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Chairman, Department Committee on Graduate Students

A Quantitative Framework For Large-Scale Model Estimation and Discrimination In Systems Biology

by

Hoda Eydgahi

Submitted to the Department of Electrical Engineering and Computer Science
on May 17, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Using models to simulate and analyze biological networks requires principled approaches to parameter estimation and model discrimination. We use Bayesian and Monte Carlo methods to recover the full probability distributions of free parameters (initial protein concentrations and rate constants) for mass action models of receptor-mediated cell death. The width of the individual parameter distributions is largely determined by non-identifiability but co-variation among parameters, even those that are poorly determined, encodes essential information. Knowledge of joint parameter distributions makes it possible to compute the uncertainty of model-based predictions whereas ignoring it (e.g. by treating parameters as a simple list of values and variances) yields nonsensical predictions. Computing the Bayes factor from joint distributions yields the odds ratio (~20-fold) for competing "direct" and "indirect" apoptosis models having different numbers of parameters. The methods presented in this thesis were then extended to make predictions in eight apoptosis mini-models. Despite topological uncertainty, the simulated predictions can be used to drive experimental design. Our results illustrate how Bayesian approaches to model calibration and discrimination combined with single-cell data represent a generally useful and rigorous approach to discriminating between competing hypotheses in the face of parametric and topological uncertainty.

Thesis Supervisor: Peter K. Sorger
Title: Otto Krayer Professor of Systems Biology

Thesis Supervisor: John N. Tsitsiklis
Title: Clarence J. Lebel Professor of Electrical Engineering

Acknowledgments

In this thesis, I use a random walk to explore parameter space for a model of cell death. Looking back at my time at MIT, my journey through graduate school can also be somewhat compared to a random walk.

My journey as a Ph.D. candidate initially began under the supervision of Prof. Michael Feld, whose dedication to science I found inspiring. He was the person who convinced me that I should pursue a Ph.D. instead of joining the work force after obtaining my Masters degree – and for that I am thankful. Unfortunately, Michael passed away and in the nine months that followed, my random walk took a turn into many regions of low probability as I explored professor space trying to find a new Ph.D. advisor. Prof. Terry Orlando, who was the Graduate Counselor at the time, came to the rescue. He acted as my Metropolis-Hastings – accepting or rejecting my proposal advisors. At any given time, I could walk into Terry's office and he would be available to chat; he even managed to find me some financial support. I am forever indebted to his infinite patience and positivity as I struggled to find my way through the region of nonidentifiability: there are hundreds of potential professors at MIT, but not all of them provide a good fit!

I finally converged on my current amazing two advisors: Prof. Peter Sorger and Prof. John Tsitsiklis. The biggest thank you hand down goes to the two of them for agreeing to take me on as a student and supporting me throughout the years. For their sense of humor, intellectual support, moral support, and trust, both in an academic and non-academic setting, I am forever grateful. They had faith in me even when I was beginning to lose faith in myself. Their constant encouragement, flexibility, and guidance have been amongst the most important reasons for the successful completion of this thesis. The two of them took a genuine interest in my success and I consider myself very fortunate to have had the opportunity to learn from them.

Will Chen and I worked together since my first day in the Sorger Lab. From helping me improve my public speaking skills to helping me think about research problems, he helped me grow intellectually and professionally. He was almost like

a third advisor to me. I am very thankful for all his constructive feedback and the many clarifying discussions we had through the years. I'd also like to thank my other collaborators in the Sorger Lab: Jeremy Muhlich, for help with the Harvard cluster, Carlos Lopez, for help with PySB, John Albeck and Sabrica Spencer, for experimental data, and John Bachman and Josh Sims, for everything to do with the biological end of the spectrum. I'd like to thank Prof. Doug Lauffenburger and Prof. George Verghese for agreeing to serve on my committee and for their guidance, feedback, and continuous interest in my research. I'd also like to thank NIH Grant GM68762 and the NSF Graduate Research Fellowship for financial support.

I'd like to thank everyone in both the Sorger Lab, Tsitsiklis Lab, Harvard Systems Biology, and LIDS with whom I interacted and for creating a friendly environment. In particular, I'd like to thank Lee Albacker, Ammar Ammar, Verena Becker, Chris Bird, Marina Chan, Sameer Chopra, Lynne Dell, Jennifer Donovan, Kimon Drakopoulos, Debbie Flusberg, Saman Honarnejad, Annie Jenney, Brian Jones, Doug Jones, Laura Kleiman, Pascal Kramer, Christina Lee, Jimmy Li, Hector Lucero, Michael Menden, Nate Moerke, Mario Niepel, Jeremie Roux, Pat Sampattavanich, Katie Szeto, Tariq Warsi, Jeff Werbin, Rob Yang, Yuan Zhong, and Spyros Zoumpoulis. I'd like to incredible officemates for all the fun and the laughs throughout the years: Arshed Al-Obeidi, Will Chen, Mohammad Fallahi, Ali Kazerani, David Liffman, Mihalis Markakis, Bjorn Millard, Jagdish Ramakrishnan, Julio Saez-Rodriguez, Kuang Xu, and Yunjian Xu. I can't possibly forget to thank all the EECS Headquarters staff as well as Leo, Pearl, and Eleni from the Stata Cafe, who, like true parents away from home, always made sure I had enough to eat.

My friends in Cambridge have made my time at MIT unforgettable: Omid Abari, Kostas Bimpikis, Demba Ba, David Bradwell, Vince Constanzo, Santiago Cuellar, Jennifer Fang, Stephanie Gil, Jennifer Gil, Samantha Green-Atchley, Sara Hernandez, Yola Katsargyri, Narges Kaynia, Shahriar Khushrushahi, Julia Ma, Miriam Makhlouf, Carlo Mannino, Usman Masood, Golbarg Mehraei, Shervin Mehryar, Antonio Molins, Sahand Negahban, Grinia Nogueira, Uzoma Orji, Nina Panagiotidou, Nikolas Pyrgiotis, Arman Rezaee, Samira Salari, Deniz Sevinc, Yannis Simaiakis, Danielle Smith,

Fei Siu, Seb Uzel, and John Yazbek.

My random walk through graduate school also got me involved in the world of entrepreneurship. To my Bluelight teammates—John Ikeda, Mustafa Khalifeh, Manoah Koletty, and Blaize Wallace—thanks for your patience and understanding as I finished up my last semester.

Finally, I would like to express my deepest gratitude to my parents, Ali Eydgahi and Maryam Taabodi, for their unconditional love and support. This thesis is dedicated to them.

به نام گشاینده کارها

ز نامش شود سهل دشوارها

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Signaling networks are the fundamental means by which cells transmit information used to regulate a wide range of cellular functions, such as proliferation, migration, differentiation, survival, and death [69]. Signaling networks therefore play an important role in the development and maintenance of multicellular organisms. Information is transmitted through the cell through a series of protein modifications and conformational changes. Causality in the network is particularly important since the sequence of the events forms the basis of the signal which leads to a particular response in the cell. The identity of many individual components within these signaling networks has been known for some time and recent advances in experimental methods have uncovered inhibitory or activating relationships among specific biochemical components within the network. Traditionally, experimentalists think about network dynamics in an intuitive way based on experiments that target specific proteins one at a time. While this methodology works on a small scale, it becomes more difficult to implement on a larger scale. Due to the large, nonlinear behavior of signaling networks, it is difficult to use intuition alone to extend knowledge of individual proteins into a thorough understanding of the entire network [85].

Mathematical models have the potential to combine information about each protein and interactions with other proteins to generate testable predictions and lead to new insights about the system as a whole. Mathematical models also make it possible to study the system under experimental conditions that are impossible to reproduce

at the bench. For instance, by analyzing the probability distributions of rate constants, most of which can not be measured, one can obtain a better temporal and mechanistic understanding of the whole system. Furthermore, mathematical models help explain experimental observations and develop hypotheses about the underlying structure and connectivity of the signaling network [69]. In this thesis, we demonstrate how mathematical models can abate ambiguities about the underlying biochemistry. Specifically, we present a quantitative framework for using mathematical models to calibrate models, estimate parameter values, make predictions, determine model structure, and guide experimental design.

1.1 Previous Modeling Methods

A wide variety of graphical and mathematical methods currently exist for modeling biological networks to better understand the relationship between input signals and phenotypic outputs, each of which varies with respect to its level of specificity. A particular method can be selected based on a few metrics:

- *Data*: Is it quantitative or qualitative?
- *Research question*: Is it mechanistic or phenotypic?
- *Prior knowledge*: To what detail do we have information about network structure and network components?

Graphical Methods

On the most abstract end of the spectrum lie graphical methods which help visualize large numbers of proteins and the interactions between them as means of presenting the network topology in a compact manner. The first of these graphical methods is a protein interaction network (PIN), which uses nodes to represent proteins; undirected edges are used to represent the binding interactions between two proteins, whether it be correlated expression, direct binding, or probabilistic interactions occurring with some minimal level of confidence [73, 84, 97]. Edges do not contain any information

regarding the flow of information or mass between nodes [90]. Similar to PINs, protein signaling networks (PSNs) also use nodes to represent proteins; however, edges are signed and directional. Edges represent direct causal effects: signed edges indicate activator or inhibitory relationships; directed edges indicate enzyme-substrate relationships [90, 101]. Unfortunately, since PINs and PSNs do not take into consideration network dynamics, input-output relationships can not be predicted.

Logic-Based Methods

Similar to PINs and PSNs, logic-based models are also typically depicted as edge-node graphs. Logic-based methods, however, are a more specific modeling method than PINs and PSNs and are particularly useful when modeling networks for which there only exists limited or incomplete information on network dynamics. In this case, limited information refers to networks for which the concentration is not known for each species, but causality can be assigned to the system. For instance, if protein A is activated, then protein B binds protein C to form a complex. These models are capable of encapsulating the basic interactions between a large number of proteins, whether it be in a model describing signaling pathways [8, 52, 57, 58, 82] or gene regulation [65]. Logic-based models use a discrete approach in that there is only a finite set of states (i.e., number of molecules for a particular protein) that the model's components can assume at each node. The regulatory interactions between nodes is governed through gates that have logical functions, which are typically either Boolean [52, 57, 66, 82] or fuzzy [8, 58] in nature. Prior knowledge about the network along with experimental data typically dictates the type of gate used to connect a specific set of nodes.

Differential Equation Models

At the most specific end of the spectrum of modeling methods lies the continuous-state approach of differential equations, which relies heavily on prior knowledge about the network topology and its components. Ordinary differential equation (ODE) models describe the kinetics and dynamics of key molecular interactions in a continuum

approximation, in which protein concentrations are represented in a continuous, deterministic fashion [4, 17, 25, 27]. This level of specificity enables these models to capture temporal and spatial dynamics for individual biochemical reactions in considerable detail. The underlying biochemistry can be accurately captured when mass-action kinetics hold and the number of molecules in the system is large. Partial differential equation (PDE) models are used to account for spatial gradients when the two compartmentalization assumptions of ODEs are violated. That is, the compartment is well-mixed and that between two compartments, the transfer of species has an observable rate [6]. Since differential equation models take into consideration network dynamics and kinetics, model-based predictions and input-output relationships can be accurately captured. Nevertheless, as the size of the network and the number of components increases, modeling it with a system of differential equations becomes more challenging. This difficulty is twofold: first, there is an increase in the number of free parameters, which are typically initial protein concentrations and rate constants; and second, the model output is heavily dependent on the values of these free parameters, which are typically estimated.

1.2 ODE Modeling of Extrinsic Recepter-Mediated Apoptosis

The specific signaling pathway for which we want to create a mathematical model is extrinsic receptor-mediated apoptosis. Apoptosis, or programmed cell death, plays a central role in the development of multicellular organisms, but mutations and other lesions in the signaling networks regulating apoptosis result in cancer and autoimmune diseases [51, 113]. During our life span, over 99.9% of our cells undergo apoptosis [111]. Many of the proteins that mediate receptor-mediated apoptosis have been identified but the dynamics and regulation of cell death in diverse cell types is poorly understood. Mathematical models of the biochemistry of apoptosis can lead to a better mechanistic and quantitative understanding of apoptosis regulators, which, in

turn, can potentially be used for designing a new generation of drugs to regulate programmed cell death [10, 115].

Apoptosis can be triggered via two independent mechanisms: the intrinsic pathway which is initiated by intracellular events such as DNA damage and oxidative stress; and the extrinsic pathway, which is initiated by the binding of death ligands such as *T*umor necrosis factor-*R*elated *A*poptosis *I*nducing *L*igand (TRAIL), Fas, or *T*umor *N*ecrosis *F*actor (TNF) to transmembrane receptors. This binding triggers a series of downstream reactions that activate effector caspases leading to proteolysis of essential cellular substrates and activation of CAD nucleases, which ultimately kill cells [49, 67]. In most cell types, activation of effector caspases requires mitochondrial outer membrane permeabilization (MOMP) but a MOMP-independent cascade is sufficient to trigger apoptosis in some immune cells [40, 102]. Multiple steps in both the MOMP-independent (Type I) and MOMP-dependent (Type II) cascades are subject to positive and negative regulation by factors whose abundance and activities vary with cell state and type [50].

Since the identity of the biochemical components underlying apoptosis is relatively well known, to date, apoptosis has been modeled primarily using networks of compartmental ODEs. The topology of these ODE networks is obtained from literature data and is reasonably well established. In the model being studied, apoptosis in response to TRAIL has been modeled with a network of 69 stiff nonlinear ODEs assuming the law of mass action [4, 5, 108].

Modeling the biology of extrinsic cell death using ODE networks has several justifications. EARM1.3 and similar models have been validated in multiple publications and have been shown to describe the dynamics of apoptosis at a single cell level under a range of conditions involving RNAi-mediated protein depletion or protein over-expression [5, 6, 7, 11, 23, 41, 56, 75, 93, 108, 109]. The use of mass action kinetics appears justified since the relatively high abundance of most species will result in a well-mixed system. Mass-action models also make it straightforward to translate findings obtained with purified components to the context of multi-protein networks in living cells [29, 80]. Finally, deterministic ODE networks have proven

surprisingly effective at modeling cell-to-cell variability in apoptosis [94, 108]. When a uniform population of cells is exposed to death ligands such as TRAIL, substantial differences in the timing of death are observed from one cell to the next: some cells die soon after ligand exposure (< 1hr) while others wait 12 hr or more and some cells survive indefinitely [92]. ODE networks can model this variability because it arises from transiently heritable differences in protein concentrations (extrinsic noise) rather than inherently stochastic reaction kinetics (intrinsic noise). Modeling variability with ODEs involves sampling initial protein concentrations from experimentally measured distributions [42], although understanding the origins rather than the consequences of extrinsic noise still requires stochastic simulation [48, 54, 116].

1.3 Overview

The remainder of this thesis is organized in the following manner:

- Chapter Two addresses the problem of parameter estimation for a mass-action model of receptor-mediated apoptosis calibrated to dynamic, live-cell data. A Bayesian framework is introduced and used to obtain statistically complete joint parameter distributions. Despite non-identifiablility and model sloppiness, the approach described returns probabilistic predictions for cell death dynamics that have tight confidence intervals and match experimental data.

- Chapter Three addresses the problem of model discrimination when two or more models are indistinguishable using maximum likelihood approaches. Thermodyanmic integration is used to calculate the Bayes Factor, thereby enabling model comparison within a Bayesian framework. The method was applied to two competing mechanisms of MOMP, the direct and the indirect methods. It was determined that, given the available data, the direct method is approximately twenty times more plausible than the indirect.

- Chapter Four extends the methodology presented in previous chapters to the comparison of eight models describing the MOMP control mechanism. We

demonstrate how despite the uncertainty in the topology of these eight models, we can use predictions to guide experimental design. Furthermore, we propose that by generating thousands or tens of thousands of models topologies, accurate probability distributions for model predictions can be obtained through ensemble modeling.

- Chapter Five summarizes the work in this thesis and discusses suggestions for future work in the field.

- Finally, Appendix A illustrates Monte Carlo Markov Chain Bayesian parameter estimation using synthetic data generated from the simple three species Robertson model.

Chapter 2

A Bayesian Framework For Parameter Estimation

Dynamical models that capture the biochemistry of cell signaling processes are an effective means to investigate the operation of networks involving many interacting components and concurrent reactions [34]. However, even the best characterized networks are ambiguous with respect to the sequence of reactions and the ways in which molecules interact (i.e. reaction topology). Determining parameter values for dynamical models is also problematic: parameter estimates are required for simulation and hypothesis testing but quantitative data on the rates of most biochemical reactions are sparse and data from *in vitro* studies are of unknown relevance to rates in cells. In this chapter we focus on the problem of parameter estimation for a previously validated ordinary differential equation (ODE) model of receptor-mediated (extrinsic) apoptosis in human cells (the extrinsic apoptosis reaction model, EARM1.3) [5, 23, 41, 55, 88, 108]. We address the problem of parameter estimation by implementing a Bayesian MCMC random walk rather than the typical maximum likelihood approach. Using the joint and marginal parameter distributions obtained from the random walk, we make high-confidence predictions on two hallmark characteristics of the EARM1.3 model. We then demonstrate that the joint parameter distribution captures important information about the nonlinear relationship between parameters that should not ignored. To preserve this information, we suggest calibrating models

and reporting all of the parameter vectors obtained in the random walk in matrix format instead of merely reporting the *maximum a posteriori* estimates of parameter values in a table format. EARM1.3 has undergone sufficient literature validation, and is sufficiently similar to extant models of other signaling pathways with respect to the numbers of parameters and amount of training data [68, 87, 103] that it serves as an effective test case in which to explore parameter estimation strategies.

2.1 Background

2.1.1 Structure of The Model and The Data

EARM1.3 models the biochemistry of apoptosis in a continuum approximation as a series of mass-action reactions represented by ODEs. All biochemical transformations are depicted as unimolecular or bimolecular reactions and rate laws are therefore expressed as $r = k[A]$, for a reaction involving one copy of protein A, $r = k[A][B]$, for a bimolecular reaction of A and B, or $r = k[A][A]$, for dimerization of A. No complex algebraic forms such as Hill functions are used. Transport between cellular compartments is modeled as an elementary unimolecular reaction, and the assembly of multiprotein complexes as a series of bimolecular reactions. However, the model also combines proteins with related activities into a single species as a means to reduce the number of free parameters. For example, both caspase-3 and caspase-8 are represented in EARM1.3 by C3, Bax and Bak are represented by Bax, and Bcl2 and BclX are represented by Bcl2.

The EARM1.3 ODE network has 69 dynamical variables representing the concentrations of proteins and protein complexes. In this chapter we focus on estimating the 78 free parameters that control these variables. These free parameters consist of on- and off-rates and catalytic constants as well as lumped parameters used to model protein synthesis or degradation in a coarse-grained approximation. The 16 nonzero initial protein concentrations in the model were assumed to be identical to previously reported values, many of which have been measured experimentally [108]

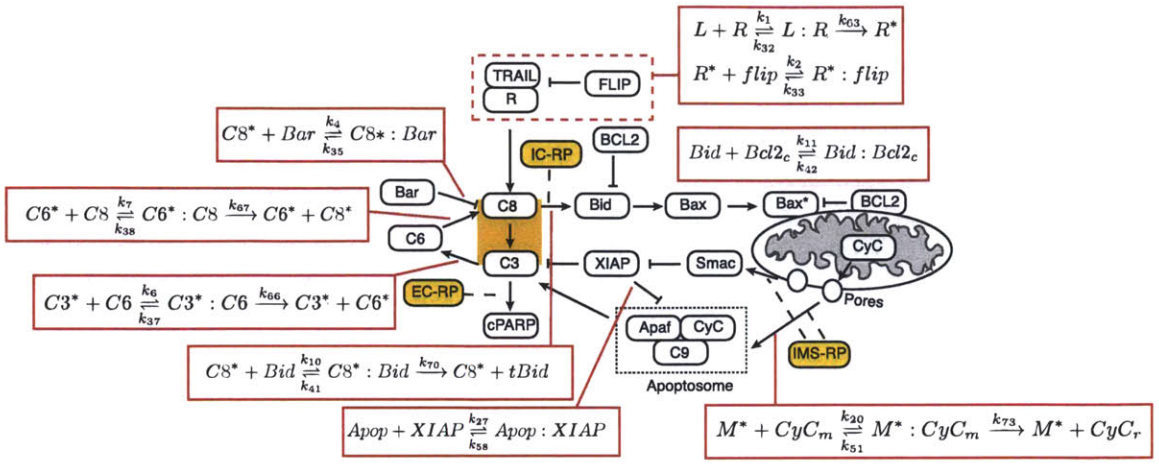

Figure 2-1: Schematic representation of the extrinsic apoptosis model EARM1.3. Binding of death ligand TRAIL, formation of Death-Inducing Signaling Complex (DISC), cleavage of caspases 3, 6 and 8 (C3, C6 and C8), formation of mitochondrial pores, assembly of the apoptosome, and cPARP cleavage are shown. Activating interactions such as caspase cleavage and induction of conformational changes are shown as sharp-tipped arrows; inhibitory interactions by competitive binding of proteins such as FLIP, BCL2 and XIAP are shown as flat-tipped arrows. The three fluorescent reporters IC-RP, EC-RP, and IMS-RP used in experiment are denoted as yellow lozenges. Specific sets of reactions are called out in red boxes and are keyed to features discussed in subsequent figures.

(initial protein concentrations can also be estimated, but adding parameters to the procedure makes the calculation more time-consuming; Figure 2-1)

In EARM1.3 apoptosis is triggered by the binding of death ligands such as TNF and TRAIL to transmembrane receptors (Figure 2-1). This leads to activation of initiator pro-caspases-8/10 (C8 in the model) causing cleavage and activation of effector pro-caspases 3/7 (C3). In HeLa cells, the line in which experimental data for this paper were collected, activation of C3 requires mitochontrial outer membrane permeabilization (MOMP), which is positively regulated by the Bcl2-family member tBid (a cleavage product of Bid generated by active C8) and negatively regulated by Bcl2. tBid binds to and activates the pore forming protein Bax but active Bax (Bax*) is unable to form pores due the presence of the Bcl2 protein. Only when Bax* reaches a threshold level is negative regulation by Bcl2 overcome, leading to rapid pore formation and sudden translocation of Smac and cytochrome c from mi-

tochondria into the cytosol. In HeLa cells, Smac is particularly important because it binds to and inactivates XIAP (the X-linked inhibitor of apoptosis protein), thereby relieving negative regulation of C3 activity and releasing active caspases, causing cell death.

Calibration data on initiator and effector caspase activities were collected from single cells using two cleavage-sensitive reporter proteins, as described previously [4]. Initiator caspase report protein (IC-RP) measures the activity of initiator caspases-8/10 and is a good metric for formation of tBid; effector caspase reporter protein (EC-RP) measures the activity of caspases-3/7 and is a good metric for cleavage of proteins such as PARP [4, 31]. The amount of time between the addition of TRAIL and the activation of caspases varies from one cell to the next [108], and the 40 single cell trajectories used in this thesis were therefore aligned by the time of MOMP to eliminate most cell-to-cell variability [4]. Thus, training data effectively represent the behavior of a typical single cell. A time-dependent value for the variance of the data (σ^2_{data}) was obtained by comparing 40 single-cell trajectories for each reporter protein.

2.1.2 Optimization vs Sampling

Free parameters in mass-action ODE models such as EARM1.3 include initial protein concentrations and forward, reverse, and catalytic rate constants (or in some models, composite constants such as Hill coefficients, Michaelis-Menten constants, or lumped rates that consolidate many elementary reactions into a single number). Initial protein concentrations can often be measured, with reasonable precision, using quantitative Western blotting and mass spectrometry and in the current work we set them at previously determined values [5]. Some information on kinetic parameters can be gleaned from *in vitro* experiments or the literature, but rate constants are generally much less certain than protein concentrations either because no *in vitro* data are available or because the peptidyl substrates used *in vitro* are poor mimics of the large protein complexes found *in vivo*. It is therefore necessary to estimate parameter values [83].

Parameter estimation conventionally relies on minimization of the sum of squared

differences between model and data (using a least squares difference or chi-squared objective function) [83] which, at first glance, may seem like a simple optimization problem. If that were the case, one might contemplate using a variety of unconstrained optimization approaches, such as Steepest Descent, Newton, and Conjugate Direction method [14]. All the of the aforementioned algorithms are classic unconstrained optimization methods that terminate when they reach a stationary point. It is important to note that the goal of parameter estimation in our work is not to optimize a single parameter set but to address the issue of nonidentifiability by finding all the parameter sets that fit the model within experimental error.

2.1.3 The Problem of Parameter Estimation And Model Non-identifiability

ODE-based models of complex biochemical processes such as EARM1.3 are usually nonidentifiable, given available data, so estimation returns many parameter sets having equally good fits to experimental data [26, 77]. Sethna and colleagues have pointed out that even a complete set of time-course data on the concentrations and states of all species in a biochemical model is usually insufficient to constrain the majority of the free parameters in such models, a property known as sloppiness [21, 19, 38, 53].

The problem of parameter identifiability has been tackled in four conceptually distinct ways (leaving aside algorithmic specifics). The first is to consider only simple processes or small reaction networks for which identifiable models can be constructed [12, 70]. Alternatively, for non-identifiable models, a single set of best-fit parameter values can be used [5, 99, 106, 114]. A third approach is to identify a large ($\sim 10^2 - 10^3$) family of fits whose discrepancy from a best fit is less than or equal to estimated experimental error. Properties of the model that are invariant across sets of parameters are assumed to be of the greatest interest [27, 33, 60]. A fourth and more rigorous approach involves sampling the complete probability distribution of parameters, accounting for both experimental error and model non-identifiability, and then using the distribution in model-based prediction or model discrimination

[36, 96].

We next describe a Bayesian approach to parameter estimation (the fourth approach) that moves through parameter space in a manner that assembles a large collection of parameter vectors for EARM1.3 calibrated against time-course and weighted by likelihood; this collection serves as a representation of an uncertain posterior parameter distribution. We approach sampling from a Bayesian rather than a maximum-likelihood perspective to incorporate available prior knowledge (e.g., rates from in vitro studies). Bayesian estimation is well-established in fields ranging from climate control to economics [16, 30] and its use has also been explored for dynamical modeling of biochemical pathways particularly by Klinke and colleagues [13, 71]. However, convergent Bayesian estimation of large, nonidentifiable biochemical models is rare and the properties of the resulting parameter distributions have not been explored in any detail.

2.2 Methods

2.2.1 Experimental Data

All data were obtained by live-cell fluorescence microscopy of HeLa cells stably transfected with vectors expressing IC-RP and EC-RP as reported previously [4]. Apoptosis was initiated by adding media containing 50 ng/ml TRAIL and 2.5 g/ml cycloheximide. All experiments were performed by John Albeck and Sabrina Spencer.

2.2.2 Model Calibration

Since IC-RP and EC-RP currently cannot be measured from the same cell, the IC-RP and EC-RP trajectories obtained from two different cells having the same time of death (80 min) using existing experimental data were chosen for model calibration. Measurement error was accounted for by inserting a variance σ^2_{data} in likelihood function of the Bayesian formulation. For each of the trajectories, this variance was obtained by aligning 40 single-cells trajectories according to the time of death and

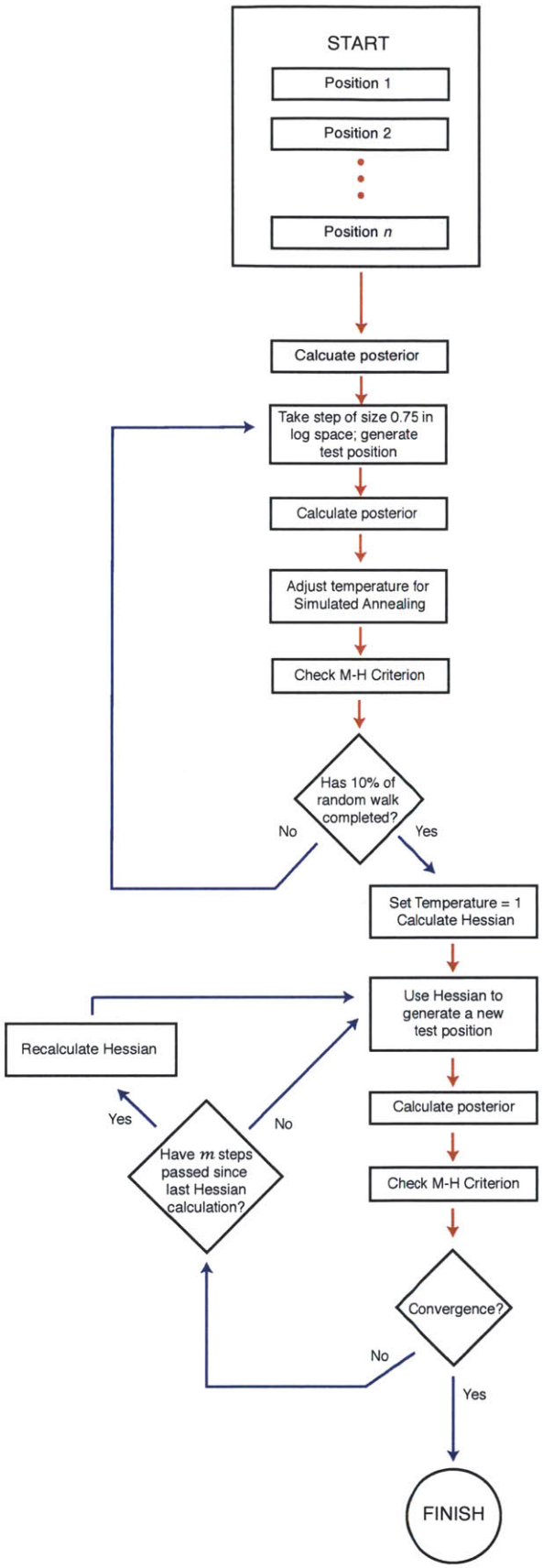

Figure 2-2: A flowchart illustrating the algorithm described in this thesis. The letter n denotes the number of independent, parallel chains that each start at random positions in parameter space. For the first 10% of the MCMC random walk, simulated annealing is performed and the steps sizes are 0.75 in log space. After the completion of the first 10% of the random walk, an MCMC walk is applied, guided by a local Hessian. The algorithm ends once the n chains reach convergence according to the Gelman-Rubin criterion.

measuring the variance at each time point.

2.2.3 The Model

The model in this paper, EARM1.3, was first described by Albeck et al [4, 5]. EARM1.3 as used in the current work differs from the original model in its inclusion of synthesis and degradation reactions for all species and its use of different nominal parameter values [108]. A summary of the algorithm is found in Figure 2-2 and is explained in detail below.

2.2.4 The Algorithm

Parameter estimation using a Bayesian Markov Chain Monte Carlo walk

In this chapter we confine ourselves to estimating 78 free parameters corresponding to rate constants in EARM1.3; the 16-nonzero initial protein concentrations in the model were assumed to be identical to previously reported values, many of which have been measured experimentally [108]. All calculations are performed in log space and we define the following variables:

$$\theta_i \equiv \log_{10}(k_i)$$

$$\Theta \equiv (\theta_1, ...\theta_{78})$$

where Θ denotes a parameter vector in log space and $(k_1, \cdots, k_{78})$ are the rate constants in the model. The deviation between data and model for a particular Θ is computed using the sum of squared differences as follows:

$$\chi^2 = \sum_t \sum_i \frac{1}{2\sigma^2_{data}(t)} [x^i_{model}(t;\Theta) - x^i_{data}(t)]^2 \tag{2.1}$$

The index i runs over all experiments and the index t runs over all times at which measurements are made. The χ^2 function is a conventional objective function and also the negative log of the likelihood that the data will be observed for a given

set of parameters if measurement error (with variance σ^2_{data}) is assumed to have a Gaussian distribution. In the current case, a time-dependent value for σ^2_{data} was estimated by comparing 40 single-cell trajectories for each of the EC-RP and IC-RP reporter proteins (σ^2 ~0.03 to 0.12). According to Bayes' rule, given a set of data the conditional distribution of a parameter vector Θ is given by:

$$\mathrm{P}\left(\Theta|\mathrm{data}\right) = \frac{\mathrm{P}\left(\mathrm{data}|\Theta\right)\mathrm{P}\left(\Theta\right)}{\mathrm{P}\left(\mathrm{data}\right)} \tag{2.2}$$

where the symbol P indicates probability density functions rather than probabilities since parameters are treated as continuous variables. The term P(Θ|data) is commonly known as the posterior or post(Θ), P(data|Θ) the likelihood(Θ), and P(Θ) the prior(Θ). The term P(data) on the right hand side of Equation (2.2) (also known as the evidence for the model) is usually difficult to compute (in the current work we tackle this issue using thermodynamic integration in the next chapter) but in Markov Chain Monte Carlo (MCMC) sampling only the ratios of posterior values are needed, not posterior values themselves. We therefore treated P(data) as a normalization constant yielding:

$$\mathrm{P}\left(\Theta|\mathrm{data}\right) \propto \mathrm{P}\left(\mathrm{data}|\Theta\right)\mathrm{P}\left(\Theta\right) \tag{2.3}$$

We have explored several priors but the most effective (see below) is one in which rate constants k_i are independent log-normal random variables, so that the θ_i are independent and normal, and

$$-\ln\left(\mathrm{prior}\left(\Theta\right)\right) = \sum_{i=1}^{78} \frac{1}{2\sigma_i^2}[\theta_i - \langle\theta_i\rangle]^2 \tag{2.4}$$

where $\langle\theta_i\rangle$ and σ_i^2 are the mean and variance, respectively, of the log of the distribution of θ_i . The value of the log posterior for a particular parameter vector is then obtained by combining the log likelihood and the log prior Equation (2.3):

$$-\ln\left(\mathrm{post}\left(\Theta\right)\right) \propto -\ln\left(\mathrm{likelihood}\left(\Theta\right)\right) - \ln\left(\mathrm{prior}\left(\Theta\right)\right) \tag{2.5}$$

This framework is commonly used to return single good fit vectors Θ that are *maximum a posteriori* (MAP) probability estimates for parameters. Instead, we seek to generate a rich set of vectors that sample the posterior distribution of Θ. To accomplish this, we implemented a random walk in 78-dimesional parameter space using a multi-start Markov Chain Monte Carlo algorithm (MCMC). Such a walk has the important property that the number of steps at a particular position in parameter space is proportional to the posterior probability, allowing parameter vectors Θ to be sampled with the correct statistical weight [28]. At the j^{th} step of each MCMC walk, a Metropolis-Hastings (M-H) criterion was employed as follows:

$$\Theta_{j+1} = \begin{cases} \Theta_{test} & \text{with probability } \alpha \\ \Theta_j & \text{with probability 1-}\alpha \end{cases} \tag{2.6}$$

$$\alpha = \min\left\{a, \frac{\text{post}\,(\Theta_{test})}{\text{post}\,(\Theta_j)}\right\} \tag{2.7}$$

Here, Θ_j is the current position in parameter space and Θ_{test} is the putative next position, generated according to a distribution that is guided by certain Hessian calculations. A test position is accepted based on whether a randomly and uniformly chosen number between 0 and 1 is less than α ($\alpha \leq 1$).

Initiating MCMC Chains

3-5 independent MCMC chains were run simultaneously on a cluster computer. Each chain started at a random initial position in parameter space. These initial positions were obtained by multiplying the log of the nominal parameter values by a random number drawn from a uniform distribution between -1 and 1, in effect yielding a set of parameters 10-fold lower or higher than the nominal values. In log parameter space, the starting position is a point randomly chosen within a box of dimension 78 and sides of length 2. The box is centered at the nominal values, which are those reported in work on the original EARM model [4]. The acceptance rate averaged over all chains was approximately $\sim$0.15-0.19. Prior work on optimal jumping rates suggests the optimal rate is $\sim$0.2 for certain asymptotic conditions and assumptions

of the target distribution [45]. However the same work noted that 0.5 is "reasonable" and achieves 75% maximal efficiency. We have not determined the degree to which our system and procedures satisfy the various conditions of the theoretical optimum. Improvement in this aspect of the algorithm is therefore possible.

Simulated Annealing

For the first 10% of the MCMC algorithm, simulated annealing (SA) was used to bring the chains from random initial starting points to points having high posterior values. The temperature of annealing was lowered according to the exponential function $T = T_0 e^{(-DecayRate * StepNumber)}$, where T_0 is the initial temperature (set to a value of 10), *StepNumber* refers to the MCMC step number, and *DecayRate* is the rate of exponential decay, chosen so that the time constant of the decay is 30% of the number of steps between Hessian calculations, which are used to guide the random walk (25,000 steps; see below). The temperature is reduced until it reaches a value of 1. During SA, new parameter vectors are chosen by taking a step lying on the unit sphere with radius of size 0.75, centered on the current position. The radius size was heuristically determined by systematic exploration of different sizes and choosing the one that showed the most rapid and greatest success in posterior maximization. After SA is complete, the Hessian guides the determination of new parameter vectors.

Adaptive MCMC walks

Among the adaptive approaches we tried, the one that showed the greatest improvement with EARM1.3 involved taking large steps in flat directions and small steps in steep directions based on calculating the curvature of the local landscape using the Hessian matrix of the log-posterior function (represented by yellow ellipses in Figure 2-3); see Table 2.1.

Hessian Directed Search

To improve the performance of the MCMC search algorithm, we developed a procedure for taking large steps in directions in which the local landscape is flat and small

steps in directions in which the landscape has large curvature as determined by a Hessian decomposition at selected positions in parameter space. These positions are defined by the parameter vector $\Theta = (\theta_1, \cdots, \theta_{78})$ (we performed all calculations in $\log_{10}$ space). The Taylor series expansion around a position Θ_{hess} is the following:

$$\ln(\text{post}(\Theta)) = \ln(\text{post}(\Theta_{hess})) + (\Theta - \Theta_{hess})^T \Delta \ln(\text{post}(\Theta)) + \frac{1}{2}(\Theta - \Theta_{hess})^T H (\Theta - \Theta_{hess}) + O(\delta^3) \quad (2.8)$$

Here, Θ is a position in parameter space close to Θ_{hess}; $\Delta \ln(\text{post}(\Theta))$ and H are the gradient vector and the Hessian matrix respectively, evaluated at Θ_{hess}; and δ is the magnitude of $\Theta - \Theta_{hess}$. To determine whether this expansion is a good approximation

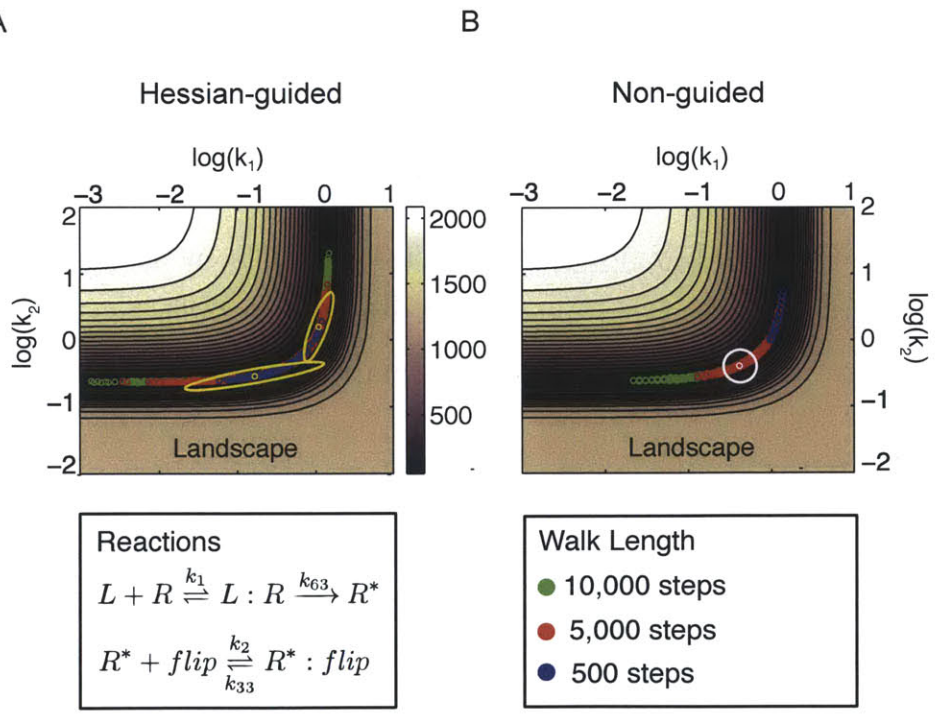

Figure 2-3: Two-dimensional slices of the ln(posterior) landscape computed by numerically evaluating the posterior over a dense set of grid points on a two-dimensional space. A comparison of MCMC random walks with varying step numbers and performed with and without Hessian-guidance. (A) With Hessian-guidance, the longer chains of 5,000 and 10,000 steps respectively (red and green,) converged by the Gelman-Rubin criterion. (B) Without Hessian-guidance, none of the three chains converged. The other 76 parameter values were kept constant for this analysis.

Table 2.1: Improved chain convergence using Hessian-guided MCMC searches

	Fraction of parameters that converged[a]					
MCMC Steps[b]	1,150,000	1,000,000	750,000	500,000	100,000	50,000
Hessian-guided[c]	78/78	77/78	73/78	70/78	40/78	21/78
Non-guided[d]	65/78	62/78	55/78	50/78	24/78	12/78

[a] Average number of parameters for which convergence was achieved for a series of ten MCMC walks after the indicated number of steps
[b] Number of MCMC steps in a chain.
[c] MCMC chains guided by the Hessian.
[d] MCMC chains not guided by the Hessian.

to $\ln(\text{post}(\Theta))$, we calculated the correlation coefficient between Δ_{true} and $\Delta_{predicted}$, where these quantities are defined by

$$\Delta_{true} = \ln(\text{post}(\Theta)) - \ln(\text{post}(\Theta_{hess}))$$

$$\Delta_{predicted} = (\Theta - \Theta_{hess})^T \Delta \ln(\text{post}(\Theta)) + \frac{1}{2}(\Theta - \Theta_{hess})^T H (\Theta - \Theta_{hess})$$

The Hessian at the position Θ_{hess} is decomposed into the form $H = U\Delta U^T$, where Δ is a diagonal eigenvalue matrix and U is a corresponding orthonormal eigenvector matrix. By sampling points on an ellipsoid with major axes that are the eigenvectors of the Hessian and with length 10% of the corresponding inverse square root of the eigenvalues around the central point, we observed a Pearson correlation coefficient of 0.887. This suggests that the Hessian and gradient matrices represent a good estimate of the true posterior, justifying our expansion of the objective function only to second order Figure 2-4.

While simulated annealing is running, the Hessian matrix is not calculated and test positions are generated as follows:

$$\Theta_{test} = \Theta_j + \frac{\Delta_\Theta}{\|\Delta_\Theta\|}, \quad \text{where } \Delta_\Theta = \{\eta_1, \cdots, \eta_{78}\} \text{ and } \eta_i : N(0,1)$$

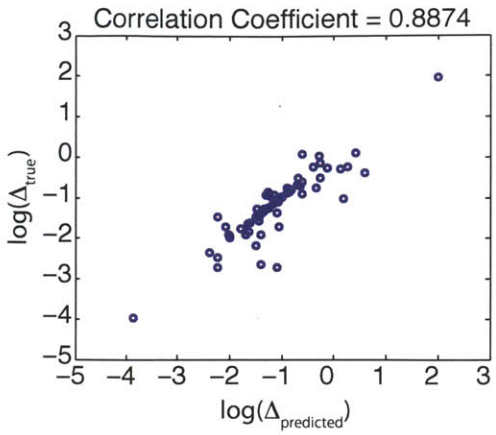

Figure 2-4: Approximating parameter space using the Hessian, which is obtained from a second-order Taylor Series expansion, is justified by a high correlation coefficient value of 0.8874 between the true value of the posterior landscape, Δ_{true}, and the Hessian-based prediction $\Delta_{predicted}$.

Here Δ_Θ, is a 78-dimensional vector with independent and normally distributed components, η_i, normalized so that the magnitude of the step size in log space is one. Once the periodic calculation of the Hessian matrix is initiated during the MCMC walk, we use its eigenvectors to direct the walk onto a new set of orthogonal axes by repeatedly obtaining new test positions of the following form:

$$\Theta_{test} = \Theta_j + \sum_{i=1}^{78} N\left(0, \frac{1}{\sqrt{\lambda_i}}\right) u_i$$

where u_i and λ_i are the i^{th} eigenvector and eigenvalue respectively, and $N(0, \sigma)$ is a random value drawn from a normal distribution with mean 0 and variance σ. Since the landscape is flat in most directions, many of the eigenvalues are much less than one. To prevent the algorithm from taking steps that are too large in any particular direction, all eigenvalues less than 0.25 were set to 0.25, so that the variance of the Gaussian distribution from which the new step size was chosen was limited to a value of 2 in log space.

The entries of the transition matrix $T_{(x \to y)}$ in our MCMC algorithm are composed of a product of two terms: the probability of selecting a particular transition between

two states: $m_{(x\to y)}$ and the probability $A_{(x\to y)}$ of accepting it (the Metropolis-Hasting criterion):

$$T_{(x\to y)} = m_{(x\to y)} A_{(x\to y)}$$

The move is symmetric in that $m_{(x\to y)} = m_{(y\to x)}$ since it is guided by a Hessian (kept constant for a window of 25,000 moves) centered on the current position. These qualities ensure that (as long as the Hessian is kept constant) the posterior distribution is the stationary distribution of the MCMC chain. To test for convergence of chains, we rely on the Gelman-Rubin criterion.

Gelman-Rubin Convergence Criterion

To obtain accurate probabilistic distributions, independent MCMC chains must reach convergence, which can be assessed by a Gelman-Rubin test [18, 46]. The Gelman-Rubin test is conducted by calculating the potential scale reduction factor (PSRF) for two chains. The PSRF value is given by the following expression:

$$PSRF = \sqrt{\frac{\frac{n-1}{n}W + \frac{M+1}{nM}B}{W}}$$

Here B is the inter-chain variance, W is the intra-chain variance, and M is the number of parallel MCMC chains each of which have run for n steps. In other studies, typically a PSRF value of less than 1.2 was used to indicate convergence. In this work we defined convergence as attaining a PSRF value of 1.1 or less. Table 2.1 shows that the parameters in three parallel Hessian-guided MCMC random walks consistently reach convergence before those in the classical MCMC random walk.

2.3 Results

2.3.1 Sampling Parameter Values Using MCMC Walks

To sample the posterior distribution of the EARM1.3 parameter space we implemented a multi-start MCMC walk, using non-uniform priors and imposing a Metropolis-

Hastings (M-H) criterion at each step. Such a walk has the important property that the number of visits to a particular position in parameter space is proportional to the posterior probability, allowing parameter vectors to be sampled with the correct statistical weight [28]. In Appendix A we illustrate how parameter distributions can be recovered from an MCMC walk in the auto-catalytic three-parameter chemical kinetic system of Robertson, a classic example from the CVODES/Sundials suite [98] and by examining 2-dimensional slices of the complex landscape of EARM1.3. In both cases, axes for the landscapes correspond to parameters and elevation corresponds to the negative log of the posterior probability (the posterior is simply the likelihood weighted by the prior). For example, for parameters k_1 and k_2, which describe the binding of receptor R to ligand (to form active receptor R*) and binding of the anti-apoptotic FLIP protein to R*:

$$L + R \underset{k_{32}}{\overset{k_1}{\rightleftarrows}} L : R \xrightarrow{k_{63}} R^*$$

$$R^* + FLIP \underset{k_{33}}{\overset{k_2}{\rightleftarrows}} R^* : FLIP$$

we observed an L-shaped valley in the objective function with a nearly flat bottom bounded by steep walls (Figure 2-3).

The MCMC walk samples this landscape by making a random series of movements along the valley floor and then estimating the posterior probability of each position based on a sum of squares error criterion. Estimated marginal distributions for the parameters k_1 and k_2 can be recovered from the walk by integrating out all other dimensions. We observe k_1 and k_2 to be well-constrained relative to many other parameters in the model, probably because the IC-RP reporter lies immediately downstream of reactions controlled by k_1 and k_2 (Figure 2-1). The two parameters balance each other out in a subtle way: k_1 is the forward rate of the ligand-binding reaction (which promotes cell death) and k_2 is the forward rate of the FLIP-binding reaction (which inhibits cell death). As a consequence, the individual parameter distributions (marginal distributions) do not capture all of the information from the

walk: when the value of k_1 is high, activated receptor R^* is produced more rapidly and this can be balanced by having k_2 at a high value so that formation of inactive $R^* : FLIP$ complexes is rapid. Thus, a good fit to the IC-RP trajectory can be achieved for a range of k_1 and k_2 values as long as their ratio is roughly constant. In general we observed that ratios of parameters (or sums of the logs of parameters) were better constrained than single parameters. This was particularly obvious in the case of Robertson's system, in which the k_1 to k_3 ratio is well estimated but individual parameters are not (Appendix A). The phenomenon is also related to the fact that in simple catalytic systems such as those studied by Michaelis-Menten, it is the ratio of k_f and k_r, (i.e. K_M) that is well-estimated under standard conditions, not forward and reverse rates themselves [26]. Examining other 2D slices of the posterior landscape revealed a wide range of topographies and different degrees of parameter constraint. True ellipsoidal minima were relatively rare but they gave rise to the expected Gaussian marginal parameter distributions; more common were distributions in which one parameter was constrained and the other not (in Figure 2-5 we have assembled a gallery of typical 2-D landscapes and the reactions they represent, along with marginal distributions for all estimated parameters). Many marginal posterior distributions were narrower than the prior and were therefore well estimated (k_8 to k_{12} for example) but others resembled the prior, a phenomenon we analyze in greater detail below. Relative to values previously used for EARM1.3 [5]. Bayesian sampling yielded 33 parameters with modal values differing by ~10-fold and 11 by ~100-fold from previous estimates.

Boot-strapping [91] is a more conventional and widely-used method for putting confidence intervals on individual model parameters. In boot-strapping, statistical properties of the data are computed and "re-sampling" to generate additional sets of synthetic data with similar statistical properties. Deterministic fits are performed against the resampled data to give rise to a family of best fits. Bootstrapping therefore returns a vector of optimum fits, together with confidence intervals, consistent with error in the data whereas MCMC walks used in Bayesian estimation return the family of all possible parameter values that lie within the error manifold of the data.

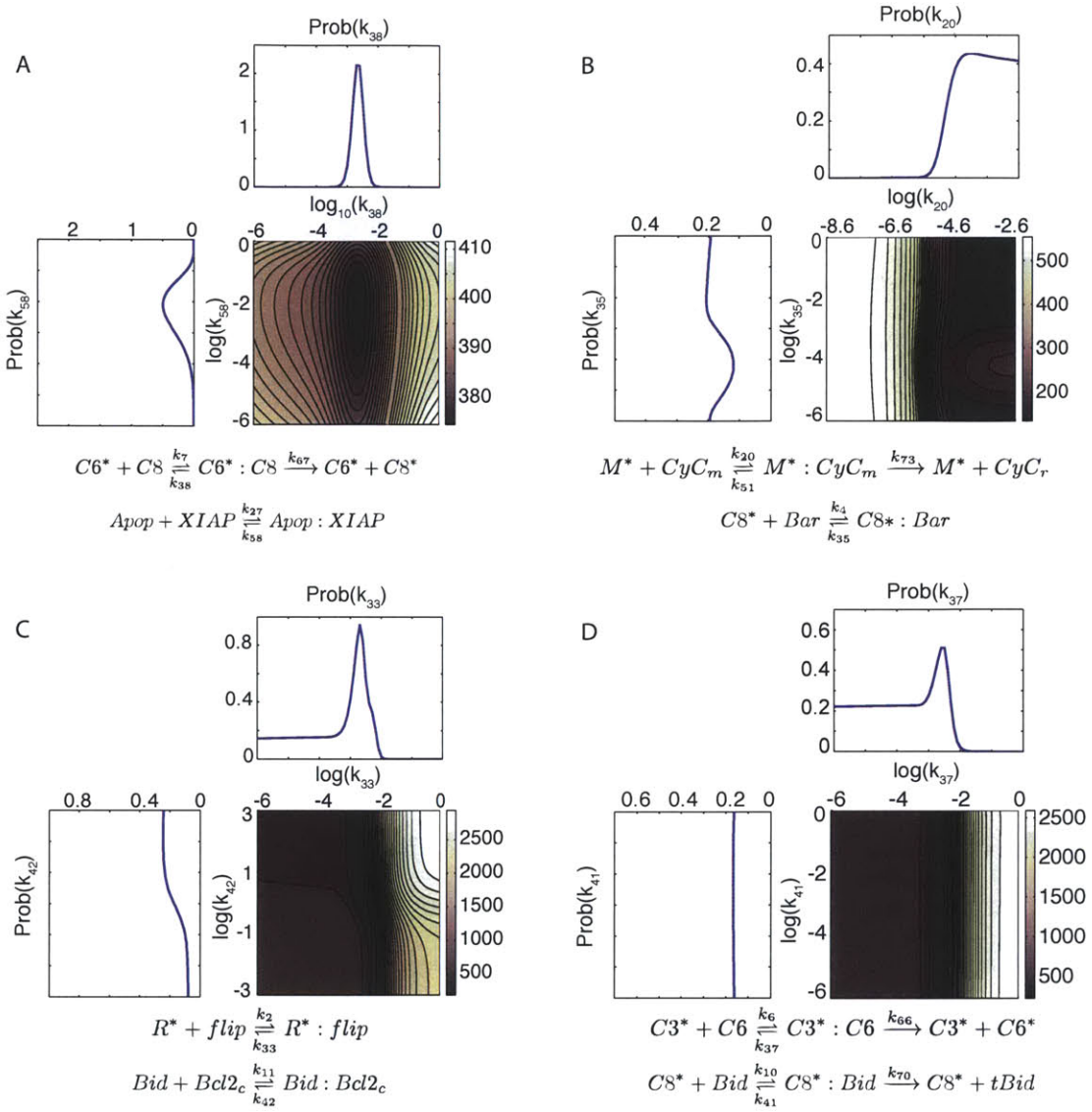

Figure 2-5: A selection of 2D slices of the -ln(posterior) landscape for pairs of parameters with marginal posteriors. (A) A landscape of k_{38} vs k_{58} shows a minimum (corresponding to a maximum in the posterior) with both parameters described by approximately Gaussian distributions. (B) A landscape of k_{20} vs k_{35} shows two minima separated by a hump. The marginal of one parameter is bounded on only one side and the other parameter is largely non-identifiable. (C) A landscape of k_{33} vs k_{42} shows a canyon- like topography; the non-identifiable region runs mostly along the k_{42} axis. Marginal distributions show how the individual parameters are bounded on one side and unbounded on the other. (D) A landscape k_{37} vs k_{41} exhibits nonidentifiability in one parameter (k_{41}) and the other (k_{37}) has a minimum near the central region but plausible values also three orders of magnitude lower.

It is possible that the family of fits obtained through bootstrapping will identify some non-identifiable parameters but in contrast to Bayesian estimation, there is no guarantee that parameter distributions or their point-by-point covariation are completely sampled.

2.3.2 Properties of MCMC Walks

Performing MCMC walks across many parameters is computationally intensive and we observed that walks through the landscape of EARM1.3 proceeded slowly for either of two reasons: at the start of most walks, the landscape was flat in many directions, making it difficult to detect gradients pointing toward minima. Later in the walk, when minima were found, they were often valley-like with many flat and few steep directions. In this case, the MCMC walk was inefficient, because many steps were attempted in directions of lower probability (this is represented by a circle of proposed moves in Figure 2-3). MCMC sampling adequately captures an unknown distribution only if independent chains starting from random points converge to the same distribution. Convergence was assessed using the Gelman-Rubin test, which compares inter-chain to intra-chain variance: failing the test proves non-convergence, although passing the test does not necessarily guarantee it [18, 46]. The importance of convergence is illustrated by the difference in parameter distributions recovered by convergent and non-convergent walks (Figure 2-6).

To improve convergence, a wide variety of "adaptive" methods have been developed based on varying step size and biasing walks in certain directions [47]. A drawback of some adaptive MCMC approaches is that they alter the proposal distribution (which determines how the next step is taken) over the course of a walk and therefore have the potential to violate the stationarity requirement of Metropolis Hastings sampling. We sought a middle ground between stationarity and efficiency by performing MCMC walks in which "Hessian-guided" adaptive moves were performed once every 25,000 steps. Under these conditions all parameters in EARM1.3 reached convergence by Gelman-Rubin criteria. We also attempted to reach convergence without Hessian guidance by increasing the number of steps in a conventional

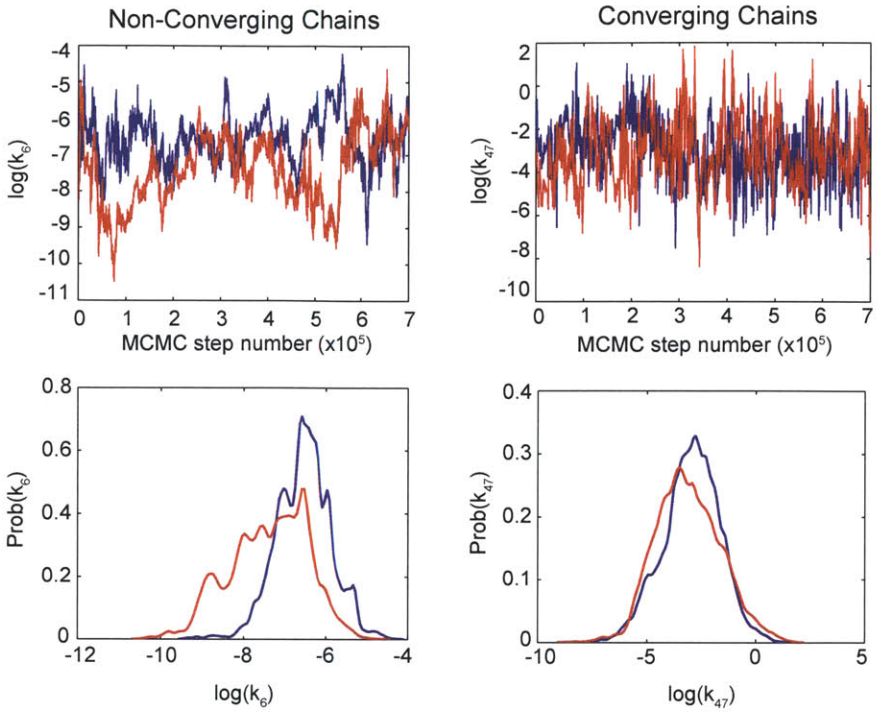

Figure 2-6: Time-series of two chains that converge for parameter k1 (top right) and two chains that do not converge for parameter k2 (top left). Marginal distributions for convergent chains are much more similar (bottom right) than those for non-convergent chains (bottom left).

MCMC walk to $> 1.5 \times 10^6$; in this case 70/78 parameters converged (Table 2.1). We discuss the technical but very important issues associated with normal and adaptive walks in the materials and methods. Future users of our methods should note that approaches for achieving and demonstrating convergent sampling in MCMC walks remains an active area of research and improvements are likely.

2.3.3 Choosing Priors

Pre-existing knowledge about parameters is incorporated in prior distributions that bias sampling of posterior landscapes to values observed in earlier work or otherwise thought to be reasonable. With a biochemical system that is well-studied, relatively narrow priors derived from in vitro data might make sense. However, in the current work we took a more conservative approach and used broad priors derived from

	Forward, First Order (s^{-1})	Forward, Second Order (M^{-1} s^{-1})	Reverse, First Order (s^{-1})	Catalytic (s^{-1})
Mean	−1.53	−5.99	−1.77	0.02
Variance	1.44	1.32	1.17	1.24

Table 2.2: Prior of log of parameter values derived from literature

physiochemical constraints likely to pertain to most biochemical reactions (use of narrower, specific priors would only serve to make calculations easier). In general we favor soft constraints involving Gaussian priors over hard constraints. Although rate constants in a biochemical model pertain to physical process, many are actually "lumped" or "effective" rates in the physicochemical sense: the reactions they describe are comprised of a series of elementary association-dissociation reactions that cannot be distinguished. For an effective rate, a hard constraint is overly restrictive. We picked a prior for the on-rate of protein-protein binding centered at $\sim 10^7 M^{-1} sec^{-1}$, which does not violate diffusion limits and is ~10-fold higher than theoretical values estimated by discrete simulation of linearly and rotationally diffusing bodies [89]. Other plausible priors, corresponding to the mean values and variance for on-rates and off-rates, protein-protein binding constants and catalytic rate constants were obtained from a literature search (Table 2.2). Because these include a mix of *in vitro* and *in vivo* values, they represent conservative estimates for possible parameter values (standard deviations were $\sim 10^2$) and should be generally useful for other models in the absence of more specific prior information.

To evaluate the impact of priors on parameter estimation, we compared five independent Hessian-guided MCMC searches that incorporated either of two priors. The first was a uniform prior in which log(prior) was a constant (the actual value is not significant); this is equivalent to sampling in proportion to the likelihood. The second was a prior involving 78 independent Gaussian distributions, each having a

mean and standard deviation based on a literature value appropriate for that type of parameter. With a uniform prior we observed that only a subset of parameters converged whereas all converged with the literature-based prior. The literature-based prior had the added benefit of minimizing the frequency with which EARM1.3 ODEs failed to integrate. MCMC walks with a uniform prior often ventured into regions of parameter space where numerical integration was not possible, presumably because the system of equations was too stiff. Conversely, we speculate that integration problems are minimized when parameter values at the extremes of the distribution are de-emphasized via the use of a log-normal prior, a potentially significant benefit.

2.3.4 Properties of the Posterior Landscape

It is a basic property of Bayesian estimation that when posterior distributions are unchanged relative to prior distributions, little information is added by data. Inspection of individual (marginal) posterior distributions for EARM1.3 revealed that many were similar to the prior, implying that model calibration did not add significant new information. However, calibrated parameters exhibited significant covariation whereas distributions in the prior were independent. How much information is contained in this pattern of co-variation?

To begin to address this question, we selected a parameter vector $(k_1, \cdots, k_{78})$ from the set of best fits arising from the joint posterior distribution and then projected the values of the individual component parameters to form the corresponding marginal distributions. In Figure 2-7 we see that the first parameter (indicated in red) from a typical best fit vector has a value near the mean of the marginal posterior distribution for k_1 (green; this must be true because the $(k_1, \cdots, k_{78})$ vector was selected based on this property) but other well-constrained parameter values such as k_{36}, k_{64} and k_{73} lie well away from the means of their marginal distributions This makes clear that the peak of the 78-dimensional joint posterior distribution does not project onto the peaks of the marginal distributions. This is also true of the mean of the joint posterior and the means of the marginal distributions. The key point is that the best parameter estimates lie at the peak of the joint distribution and we cannot tell where

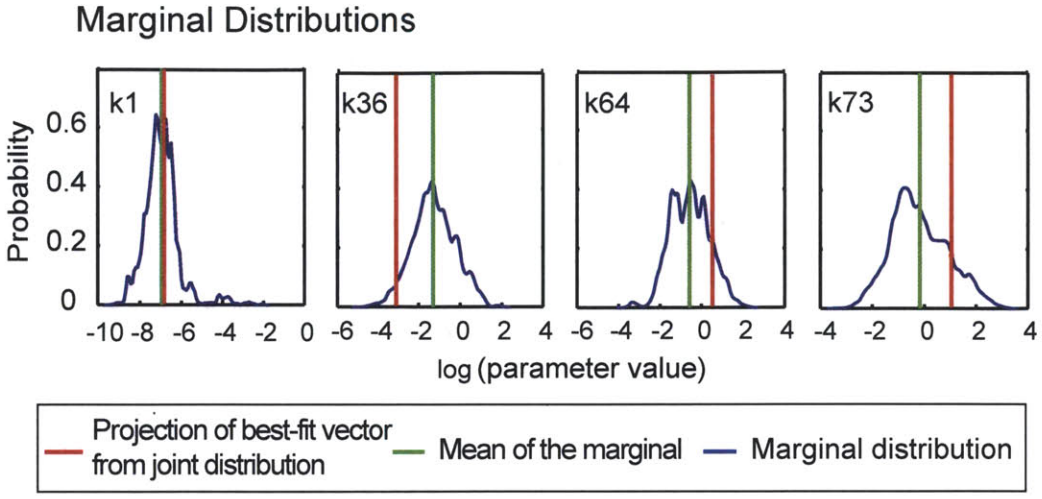

Figure 2-7: Parameter co-variation is an important type of information recovered by model calibration. A parameter vector drawn from the peak of the joint posterior distribution (corresponding to a best-fit parameter vector) does not always have components whose values correspond to the peaks of the corresponding marginal distributions (as illustrated for relatively identifiable parameters k_1, k_{36}, k_{64} and k_{73}). In this case the k_1 component of the best-fit vector matches the mean value of the marginal distribution for k_1, but this is not true of k_{36}, k_{64} and k_{73} or in general.

this lies based on looking at marginal distributions individually.

To investigate the impact of parameter covariation on model-based prediction, EC-RP trajectories were simulated using parameter vectors derived from the posterior distribution using different sampling procedures. Vectors sampled from the manifold of the joint posterior distribution yielded a good match to experimental data as expected ("manifold sampling"; Figure 2-8). However, parameters sampled independently from marginal distributions (i.e. ignoring covariation) yielded a poor fit to experimental data ("independent sampling"; Figure 2-8). To assess whether the observed co-variation could be captured in a compact manner, we computed a 78 x78 covariance matrix for pairs of EARM1.3 parameters and generated a corresponding multivariate Gaussian distribution [71] ("covariance matrix sampling"; Figure 2-8). In this case, simulated EC-RP trajectories had as poor a fit to data as trajectories generated by independent sampling.

These findings imply that a significant fraction of the information added to the posterior by calibration against data involves the discovery of non-linear co-variation

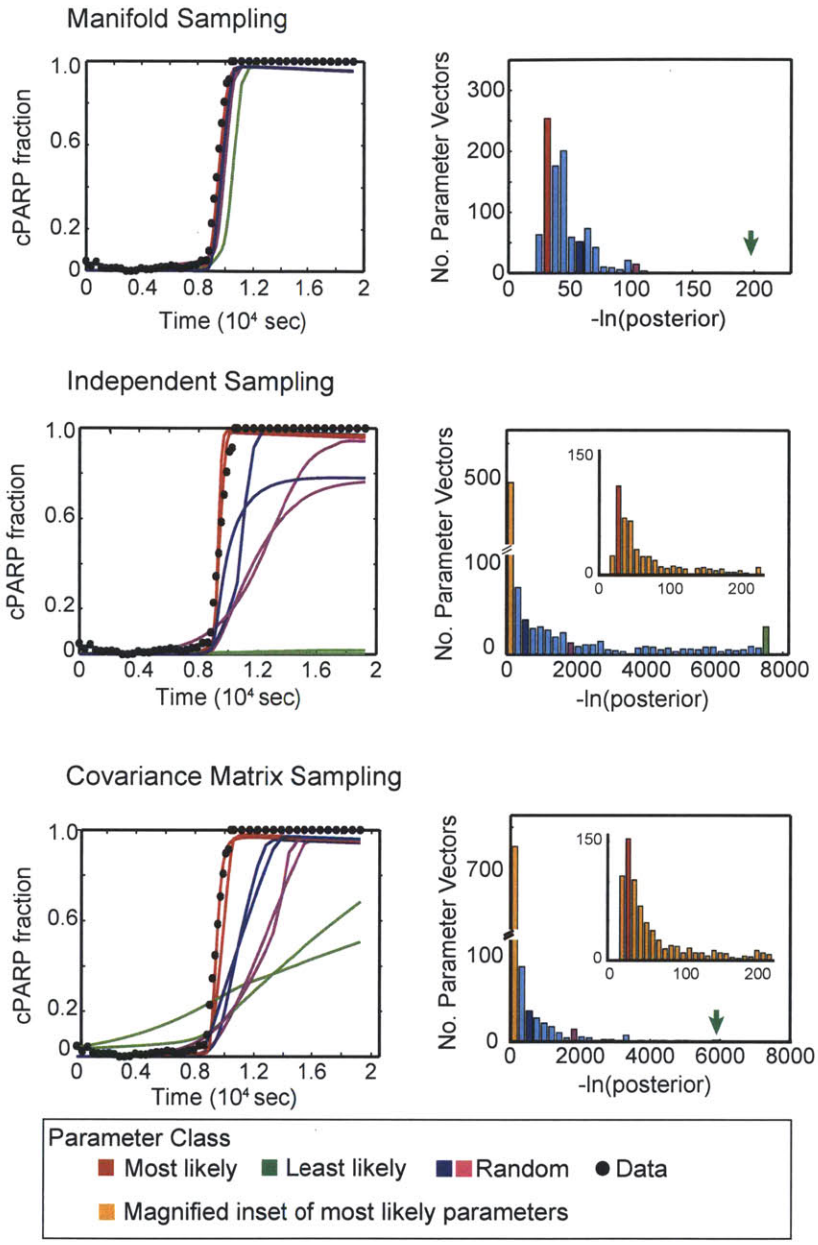

Figure 2-8: Predicted trajectories arising from three different types of parameter sampling. In each set of panels, pink, blue and green denote representative good, fair and poor fits. *Manifold sampling*: Left panel shows goodness-of-fit to experimental data of a few simulated EC-RP trajectories to show as as trajectories and right panel shows the distribution of 1000 ln(posterior) values sampled from the joint posterior distribution (right; joint sampling). *Independent sampling*: similar plots but for parameters sampled from independent marginal distributions ignoring co-variation. The inset panel expands the distribution for the smallest values of ln(posterior), *Covariance matrix sampling*: similar plots but sampling from a multivariate log normal distribution with mean and covariance computed from the MCMC walk.

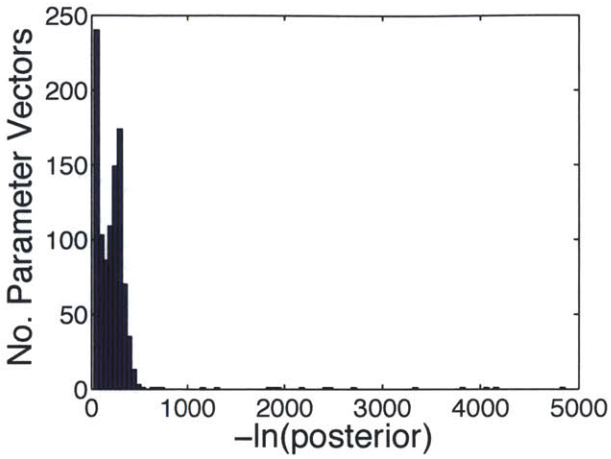

Figure 2-9: $-\ln$(posterior) values for trajectories generated from independent sampling of 10 most insensitive parameters: 5 for which the means of the marginal posterior deviated the least from the prior (ascertained by a t-test), and 5 for which the variance deviated the least (ascertained by a chi-squared variance test).

among parameters and this cannot be captured by a covariance matrix. To further illustrate that this information is important, we selected 10 parameters for which the difference between prior and posterior marginal distributions was the least significant (based on a t-test for the means and variances). The 10 selected parameters would conventionally be flagged as ones in which calibration had added little or no information. We then fixed the other 68 parameters at their MAP values and generated 103 vectors by sampling the 10 selected parameters from independent marginal distributions. EC-RP trajectories were simulated and the $-\ln$(posterior) of values computed. The resulting values for the posterior were dramatically lower than the values of the posterior resulting from parameter vectors obtained by sampling from the complete 78-dimensional posterior (Figure 2-8 and Figure 2-9). This demonstrates that even when the prior and posterior distributions appear nearly identical, calibration adds essential information on the relationships among parameters.

From these observations we conclude that: (i) non-linear covariation in parameters, as captured in the joint posterior distribution, contains critical information, (ii) the most probable values for individual parameters do not correspond to values in best-fit vectors, and (iii) treating parameters as independent values, as in a ta-

ble, or approximating their relationships linearly, as in a covariance matrix, destroys information necessary for accurate model-based prediction. We can understand this conceptually by referring to the landscape of k_1 and k_2 (Figure 2-3): it is evident that the true relationship between the parameters varies across the landscape in a complex non-linear manner. These considerations seem rather technical at first, but they have profound implications for the ways in which model parameters are recorded and used.

2.3.5 Using Parameter Distributions In Simulation And Prediction

Estimation of parameter distributions makes it possible to account for both measurement error and parameter non-identifiability when making model-based predictions. For cells exposed to a range of TRAIL concentrations we computed two descriptors of apoptosis known to be physiologically significant for many cell types [5, 7]: (i) the time interval between the addition of TRAIL and half-maximal cleavage of the caspases substrates whose proteolysis accompanies cell death (that is, the mean and variance in T_d) and (ii) the interval between initial and final cleavage of effector caspases (C3*), which captures the rapidity of death (the mean and variance in T_s [5]). EC-RP trajectories for cells treated with 50 ng/ml TRAIL were used for model calibration and T_s and T_d values were then predicted for 10, 250 and 1000 ng/ml TRAIL. Simulations were performed by sampling 1000 parameter vectors Θ from the posterior distributions arising from two independent MCMC chains and computing trajectories for each Θ. These predictions comprised probability density functions rather than single values and we therefore calculated 60% and 90% confidence intervals. We observed that the mean value of T_d fell with increasing TRAIL concentrations while T_s remained essentially constant, in line with experimental data (Figure 2-10A).

Moreover, distributions had the satisfying behavior of having narrow confidence intervals at the training dose and progressively wider intervals at higher and lower doses. This illustrates two closely related points: first, quite precise predictions can be

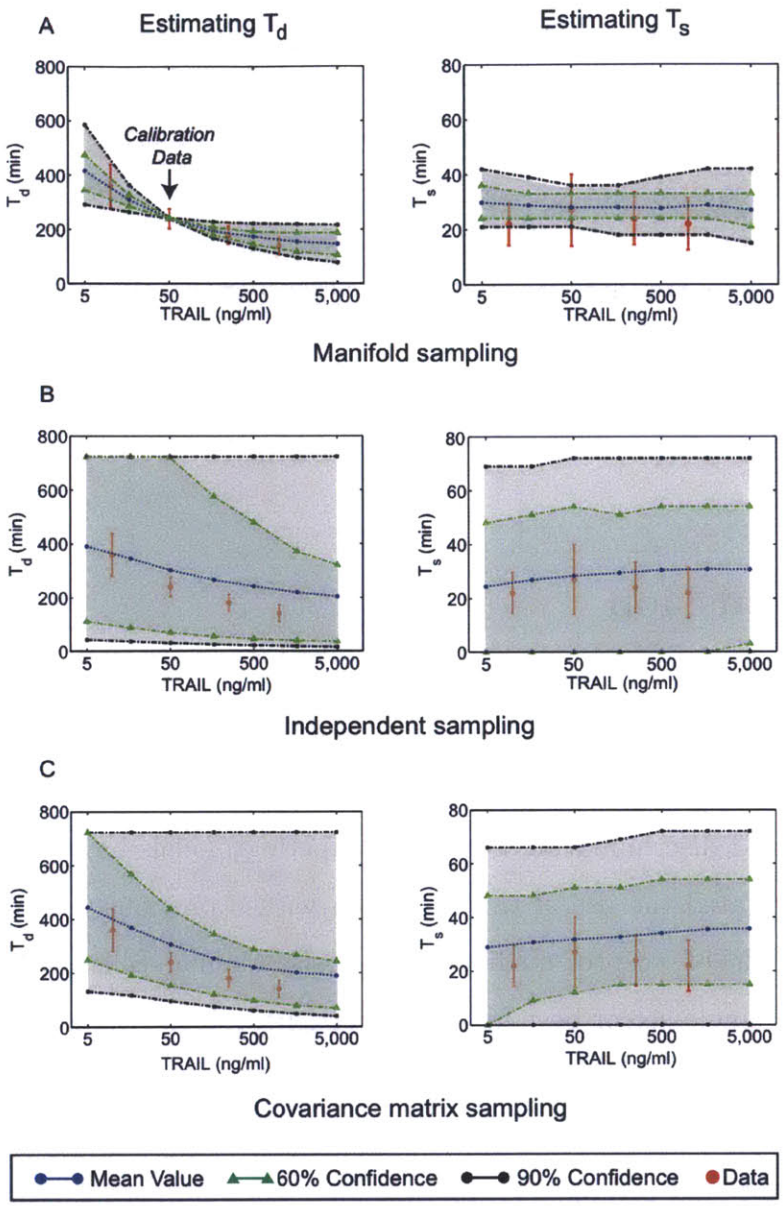

Figure 2-10: Using parameter vectors obtained by three different sampling methods to make model-based predictions of the time between ligand exposure and caspase activation (T_d) or between initial and complete PARP cleavage (T_s;) computed using parameter vectors sampled from (A) the joint posterior distributions obtained from the MCMC walk; (B) a multivariate log normal distribution with mean and covariance computed from the MCMC walk; (C) independent log normal distributions with means computed from the MCMC walk. Mean values (blue dotted line) and estimated 90% (black dotted lines, gray area), and 60% confidence intervals are shown (green dotted lines, light green area along with experimental data (red). PSRF values obtained via the Gelman-Rubin convergence test for these predicted model features ranged between 1.0001 and 1.0442 for T_d and 1.0015 to 1.0343 for T_s.

made from models despite parameter non-identifiability [53, 71] and second, Bayesian sampling makes it possible to compute rigorous confidence intervals for predictions that account for experimental error and our lack of knowledge about parameters. However, this requires that we correctly account for co-variation in parameter estimates: independent and covariance matrix sampling of parameters dramatically impaired our ability of EARM1.3 to predict accurate values for T_s and T_d (Figure 2-10B, C).

2.4 Discussion

In this chapter we described a Bayesian framework for estimating free parameters in ODE-based biochemical models, making probabilistic predictions about dynamical variables and discriminating between competing models having different topologies. We illustrated the use of this approach with a previously validated and non-identifiable model of receptor-mediated apoptosis in human cells (EARM1.3). Rather than return a single set of best fit parameters, Bayesian estimation provides a statistically complete set of parameter vectors k that samples the posterior parameter distribution given a set of experimental observations (time-lapse data from live cells in the current case) and a value for experimental error. Estimation starts with a best-guess initial distribution (the prior) which is then modulated by a sum of squares log-likelihood function that scores the difference between model and data. Recovery of the posterior parameter distribution makes it possible to compute confidence intervals for biologically interesting properties of the model (time and rapidity of apoptosis in the current case). These confidence intervals correctly accounting for measurement noise and parametric uncertainty and can be remarkably precise in the face of substantial non-identifiablility [71]. Simulations that include confidence intervals represent an advance on the prevailing practice of relying on error-free trajectories computed using a single set of maximum likelihood parameter values.

2.4.1 Properties of the posterior distribution and implications for reporting parameter values

With EARM1.3, Bayesian estimation reveals substantial differences from one parameter to the next in the degree of identifiability (as reflected in the widths of the parameter distributions). This is expected given previous work showing that biochemical models are sloppy [53] even when calibrated against complete data on all dynamic variables (which is not the case in the current work). A basic property of Bayesian estimation is that the posterior will resemble the prior when data provides little or no additional information. Conversely, when the data are informative, the shape of the posterior will differ substantially from that of the prior. In the case of EARM1.3, modal values for posterior distributions differed from the priors for about one-third of all parameters while still falling within a biophysically plausible range (with rate constants below the diffusion limit, for example). The exact shape of the prior did not appear to be critical in achieving convergent sampling, a fortunate outcome since we used general-purpose priors applicable to all cellular functions rather than priors derived from specific analysis of apoptosis proteins. One mistake we learned to avoid was constraining the MCMC walk to a fixed interval around nominal parameter values; such hard limits result in artificially truncated marginal distributions. Gaussian priors also had the benefit of improving the fraction of parameter sets for which the EARM1.3 ODE system could be integrated.

It is incorrect to judge the impact of parameter estimation (i.e., what we learn by comparing models to data) simply by examining the shapes of individual parameter distributions: several lines of evidence show that marginal distributions contain only part of the information. Non-linear covariation among parameters accounts for the rest; it is necessary for accurate model-based simulation and cannot be approximated by a covariance matrix. The reasons for this are evident from inspection of the landscape of the objective function. The landscape is steep (the eigenvalues of the Hessian matrix are high) in directions that do not point directly along raw parameter axes (Gutenkunst et al, 2007). Thus, identifiable features of the systems correspond

to ratios of rate constants (this is the basis of parameter co-variation) but the value of the ratio varies through parameter space (this gives raise to curved high-probability valleys that are not well approximated by lines). By direct analogy, the identifiable parameter in a Michaelis-Menten treatment of a simple catalytic reaction is K_M, a ratio of rate constants, rather than k_f or k_r themselves ([26] see also Appendix A). When parameters in a model are highly covariant, it is almost always the case that the system can be described with a simpler model involving a smaller set of more identifiable parameters. In many applications it is desirable to use such reduced models but in the case of biochemistry, parameter non-identifiability and high co-variance appear to be the cost of representing systems as sets of reversible mass-action reactions. Under the assumption that mass-action kinetics (and also stochastic kinetics obtained by solving the chemical master equation) are uniquely appropriate as a means to describe the physics of biochemical systems, we are forced to use models such as EARM1.3. However, there is no reason, from a physical perspective, to believe that proteins in a network that do not actually bind to each other alter each others rate constants: the presence of highly co-variant parameters in best-fit vectors is not a property of the underlying biochemistry. Instead, it represents a limitation on our ability to infer the properties of complex reaction networks based on the time-course data we typically acquire.

One consequence of parameter co-variation in EARM1.3 is that parameter values in best-fit vectors do not correspond to the means of marginal parameter distributions and sampling the means of marginal distributions does not result in a good fit. It is common practice in biochemical modeling to report parameters as a table of single values (with no allowance for non-identifiablility) or as a list of means and ranges. If these parameters are derived from calibration, critical information on co-variation is lost. It is therefore necessary to report the actual vectors recovered by sampling the posterior parameter distribution. In principle, this is an array of size $CM\,(N+1)$ where C is the number of MCMC chains, M the number of steps, and N the number of parameters (N+1 appears because we record a posterior value for each N-dimensional vector) corresponding to $\sim$1.5x10^8 entries for EARM1.3. However, steps in MCMC

chains have characteristic "decorrelation lengths" over which parameter values vary relatively little ($\sim 10^2 - 10^4$ steps, depending on the parameter). Thinning by this amount yields an array of $\sim 10^4 - 10^6$ entries, still a much more complex representation of parameters than the simple tabular summary assumed by current standards such as SBML. It is also important to note that the posterior landscape needs to be revisited repeatedly when adding new data or extracting new hypotheses. In this sense, parameter estimates are best viewed as computational procedures and sets of possible values rather than fixed information.

2.4.2 Limitations of the approach

A conceptual concern with the current work involves the way in which MCMC walks sample the posterior landscape. To establish that sampling is correct it is necessary to show that chains starting at independent positions converge. Convergent sampling is not an abstruse point because probability distributions can differ in shape and modal value when sampling is convergent as opposed to non-convergent. With EARM1.3 we observed that convergence was not possible in a reasonable amount of time (e.g., a week-long cluster-based calculation) using a conventional MCMC walk. We therefore used an adaptive walk involving periodic re-calculation of the local landscape as a means to guide MCMC walks and improve convergence. However, this approach may violate the detailed balance requirement of Metropolis-Hastings sampling. With large models and existing methods we are therefore in the position of having to choose between convergent Hessian-guided chains and partially non-convergent, conventional chains ([71] chose the latter alternative). Moreover, using the Gelman-Rubin test to judge convergence has the weakness that it is one-sided: failing the test demonstrates non-convergence but passing the test does not guarantee it. Analysis of posterior distributions for EARM1.3 computed in different ways suggests that we are on relatively solid ground in the current work (we did not observe significant differences in posterior distributions using different sampling approaches) but the development of methods for analyzing MCMC walks represents an active area of research in applied mathematics and it is necessary to be aware of future developments.

For reliable, probabilistic model-based simulation we also need to consider the fact that the sufficiency of sampling is contingent not only on the model structure and available training data, but also on the types of predictions being sought. Assuming convergence, the posterior landscape sampled by multi-start MCMC walks represents a reasonable approximation to the true but unknown posterior distribution of the parameters, but the same is not necessarily true for predictions or simulated trajectories based on these parameters: the posterior landscape may be poorly sampled in regions of parameter space that have a significant impact on certain simulations. In the current work we show that MCMC chains used to predict T_s and T_d satisfy the Gelman-Rubin test, but this is a weak criterion and importance sampling using umbrella, non-Boltzmann or other methods [9] will generally be necessary to revisit regions of the landscape that have low posterior values but contribute strongly to the distribution of a prediction. This suggests a workflow in which MCMC walks based on calibration data (as described here) are only the first step in model calibration. Inclusion of any new training data mandates a new round of estimation. Additional sampling should also be performed as required by importance sampling to reliably inform predictions. Finally, the use of formal methods for modeling experimental error [61] should make it possible to distinguish errors arising from photon noise, infrequent sampling, incorrect normalization, etc., thereby improving the comparison between data and simulation.

Chapter 3

Application of Thermodynamic Integration To Network Topology Discrimination

In principle, ambiguities in biochemical network reaction topologies can be resolved by constructing alternative models and then determining which ones exhibit the best fit to data. Such comparisons are usually performed using a set of parameters thought to fall near the global optimum for the original model (based on goodness-of-fit to data). However, given the kinds of data that can be collected from cells, parameters in realistic biochemical models are often non-identifiable and re-fitting alternative models often uncovers a new set of parameters having an indistinguishably good fit. In this case, it is not clear whether the models being compared are equally valid. Uncertainty about parameters arises from non-identifiablility, whose ultimate origins are a dearth of quantitative data on the rates of biochemical reactions. Sethna and colleagues have pointed out that even a complete set of time-course data on the concentrations and states of all species in a biochemical model is usually insufficient to constrain the majority of rate constants, a property known as sloppiness [19, 21, 38, 53]. In addition, models with different reaction topologies often differ in the numbers of free parameters. Therefore, a scheme for rigorously comparing competing biochemical models must account for parametric uncertainty and potential differences

in parameter numbers.

3.1 Background

One area in which the existing EARM1.3 model is particularly simplistic is in its treatment of MOMP regulators. The Bcl2 proteins that regulate MOMP can be divided into three families: (1) pro-apoptotic BH3-only proteins, such as Noxa, tBid, and Bad that promote pore formation, (2) effectors proteins such as Bax and Bak that form trans-membrane pores, and (3) anti-apoptotic proteins, such as Bcl2, Mcl1 and BclXl which inhibit pore formation. Conflicting hypotheses exist in the literature about whether MOMP is controlled in a "direct" or "indirect" fashion [29, 80]. The direct model postulates that BH3-only "activators," such as tBid and Bim, bind to Bax and Bak and induce pore-promoting conformational changes (a second class of BH3-only "sensitizers," such as Bad, are postulated to function by binding to and neutralizing anti-apoptotic proteins such as Bcl2). The indirect activation model postulates that Bax and Bak have an intrinsic ability to form pores but are prevented from doing so by association with anti-apoptotic proteins; the sole role of BH3-only proteins in this model is to antagonize anti-apoptotic proteins, thereby freeing up Bax and Bak to assemble into pores. Considerable subtlety exists with respect to the specifics of indirect and direct mechanisms implying that it will ultimately be necessary to compare multiple versions of each model.

We compute the Bayes factors for MOMP models that have different topologies (competing indirect and direct models EARM1.3I and EARM1.3D) thereby estimating their relative likelihood while accounting for different numbers of non-identifiable parameters. The fact that Bayesian methods developed for relatively small models in the physical sciences [22] are effective with large biochemical models opens the door to rigorous reasoning about cellular mechanisms in the face of complexity and uncertainty.

3.2 Methods

3.2.1 Comparing two EARM1.3 models using Bayes factors

Although the scheme presented in the previous chapter represents a principled way to manage parametric uncertainty, it does not account for uncertainty in the structures of reaction networks. We focus on uncertainty involving pore formation by Bcl2-family proteins during MOMP [29, 76, 80]. EARM1.3D instantiates a "direct" model in which MOMP activators such as tBid (Figure 3-1, red lozenges) positively regulate Bak/Bax pore-forming proteins (green), and Bcl2, BclXl and Mcl1 inhibitors (yellow) block this activation (these proteins are themselves antagonized by the sensitizers Bad and NOXA). EARM1.3I instantiates an "indirect" model in which Bak/Bax are always active but are held in check by Bcl2-like inhibitors, whose activity in turn is antagonized by tBid, Bad and NOXA (Figure 3-1). These models represent only two of several possibilities for direct and indirect regulation of MOMP, but the important point for the current work is that they have distinct topologies and different numbers of parameters (88 for EARM1.3I and 95 for EARM1.3D).

When we compared simulated EC-RP trajectories using EARM1.3I or EARM1.3D to experimental data, we observed equally good fits, meaning that the models cannot be discriminated based on a maximum likelihood approach (Figure 3-2). To compare

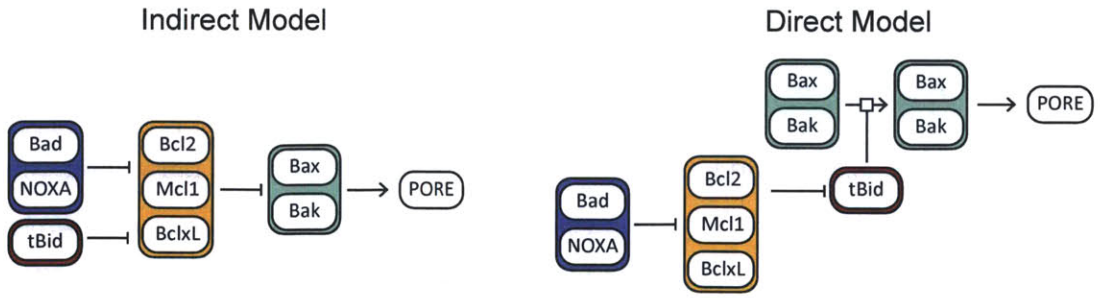

Figure 3-1: Graphical depictions of potential indirect and direct mechanisms controlling pore formation by Bax and Bak. See text for details.

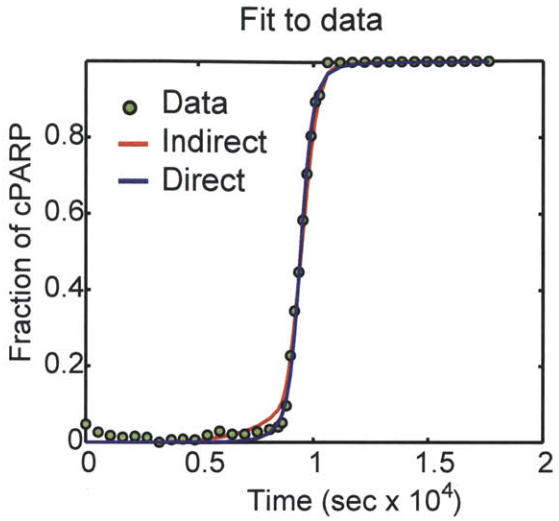

Figure 3-2: Both EARM1.3I indirect (red) and EARM1.3D direct (blue) models exhibited an excellent fit to experimental EC-RP trajectories. Thus, the models cannot be distinguished by simple maximum likelihood criteria. For simplicity, simulations were based on a single best-fit parameter vector.

the models in a Bayesian framework, we applied Bayes theorem at the level of models:

$$\mathrm{P}\left(M_i|\mathrm{data}\right) = \frac{\mathrm{P}\left(\mathrm{data}|M_i\right)\mathrm{P}\left(M_i\right)}{\mathrm{P}\left(\mathrm{data}|M_1\right)\mathrm{P}\left(M_1\right) + \mathrm{P}\left(\mathrm{data}|M_2\right)\mathrm{P}\left(M_2\right)} \quad \text{where } i = 1, 2 \tag{3.1}$$

where "data" refers to experimental measurements against which the objective function was scored (EC-RP trajectories in the current case). M_1 refers to the direct model and M_2 to the indirect model and since both have literature support we assumed that the models are a priori equally plausible (this represents the most conservative assumption). Then, Eq. (3.1) simplifies to:

$$\frac{\mathrm{P}\left(M_1|\mathrm{data}\right)}{\mathrm{P}\left(M_2|\mathrm{data}\right)} = \frac{\mathrm{P}\left(\mathrm{data}|M_1\right)}{\mathrm{P}\left(\mathrm{data}|M_2\right)} = \frac{\int d\Theta_1 L_1\left(\mathrm{data}|\Theta_1\right)\pi\left(\Theta_1|M_1\right)}{\int d\Theta_2 L_2\left(\mathrm{data}|\Theta_2\right)\pi\left(\Theta_2|M_2\right)} \tag{3.2}$$

where $\Theta_1 = (\theta_1, \cdots, \theta_{95})$ and $\Theta_2 = (\theta_1, \cdots, \theta_{88})$ are respectively the parameter vectors for models M_1 and M_2, $\theta_i \equiv \log_{10}(k_i)$ where k_i is the the i^{th} parameter of a particular model (all calculations are performed in log space), $L_i\left(\mathrm{data}|\Theta_i\right)$ is the likelihood function, and $\pi\left(\Theta_i|M_i\right)$ is the prior for the parameters of model M_i. This ratio is known as the Bayes factor and represents the odds ratio of one model being correct

over another [44, 63], and has been used for discriminating alternate models of cross-talk in ERK and cAMP signaling [62]. Both the numerator and the denominator comprise two high-dimensional integrals that represent the overlap volume between a likelihood ($L\left(\text{data}|\Theta\right)$) and the prior for each model is $\pi\left(\Theta|M\right)$. This overlap integral is also known as the evidence. The Bayes factor not only accounts for all plausible parameters based on their likelihood, it also has a built-in "Occam's razor" that implicitly accounts for the possibility that the two models have different numbers of parameters [81].

3.2.2 Computing the Bayes factor by thermodynamic integration

Models were compared using the Bayes factor, a ratio of integrals that can be computed in low dimensions fairly easily, for example using Gauss-Hermite quadrature [43, 64]. However, in high dimensions, quadrature is expensive and we therefore turned to thermodynamic integration [22, 39, 44, 74]. Thermodynamic integration relies on a constructed relation known as the power posterior, which resembles the overlap integral in the numerator or denominator in Eq. (3.2) except for the introduction of a fictious "temperature" t, a variable power to which the likelihood function is raised. Let us define $z(t)$ by

$$z\left(t\right) \equiv \int d\Theta \; L^{t}\left(\text{data}|\Theta\right)\pi\left(\Theta\right) \tag{3.3}$$

At $t = 1$, we recover the evidence (numerator or denominator in Eq. (3.2)) whereas at $t = 0$ we obtain a value of 1 because the prior distribution integrates to 1. The temperature factor serves to flatten the likelihood function so that it resembles the likelihood and prior at $t = 1$ and 0, respectively. Then, by the fundamental theorem of calculus, we obtain:

$$\ln \mathrm{P}\left(\text{data}|M\right) = \ln[z\left(1\right)] - \ln[z\left(0\right)] = \int_{0}^{1} d\tau \frac{d}{d\tau} \ln[z\left(\tau\right)] \tag{3.4}$$

The integrand has an explicit derivative with respect to the temperature variable:

$$\begin{aligned}\frac{d}{d\tau}\ln[z(\tau)] &= \frac{1}{z(\tau)}\frac{d}{d\tau}z(\tau) \\ &= \frac{1}{\int d\Theta\ L^t(\text{data}|\Theta)\,\pi(\Theta)}\frac{d}{d\tau}\left[\int d\Theta\ L^t(\text{data}|\Theta)\,\pi(\Theta)\right] \\ &= \frac{\int d\Theta\ \ln\left(L(\text{data}|\Theta)\right)L^\tau(\text{data}|\Theta)\,\pi(\Theta)}{\int d\Theta\ L^\tau(\text{data}|\Theta)\,\pi(\Theta)} \\ &= \langle \ln\left(L(\text{data}|\Theta)\right)\rangle_\tau\end{aligned}$$

The derivative is therefore a weighted average of the log-likelihood function. The bracket average is obtained by sampling from a temperature-dependent distribution (a normalized distribution which we denote as Q), which can be simulated, in the same way as the posterior, via convergent MCMC sampling. In particular, for any function f, we let:

$$\langle f(\Theta)\rangle_\tau \equiv \int d\Theta\ Q(\Theta;\tau)\,f(\Theta)$$

$$\text{where} \qquad Q(\Theta;\tau) = \frac{L^\tau(\text{data}|\Theta)\,\pi(\Theta)}{\int L^\tau(\text{data}|\Theta)\,\pi(\Theta)\,d\Theta}$$

In this way, calculation of a high dimensional volume is converted into a one dimensional integral of bracket averages over a fictitious temperature. The integrand must be estimated at each temperature via MCMC sampling. There is an appealing physical interpretation to this integral: the temperature factor "flattens" the likelihood function while the bracket averaging calculates the likelihood function at different values of the "flatness." When the temperature-based likelihood function is flattened, the sampled likelihoods will be poor (low), whereas when it is sharp and similar to the original posterior, the sampled likelihoods will be good (high). If the overlap volume is large, then the switch from poor to good will occur at low temperature (higher flatness). Conversely if the overlap volume is small, then comparatively the switch will occur at higher temperature (lower flatness). The evidence term is simply the exponential of the one-dimensional integral:

$$P(\text{data}|M) = e^{\left(\int_0^1 d\tau \langle \ln L(\text{data}|\Theta)\rangle_\tau\right)}$$

The value of the bracket average $\langle \ln L\,(\text{data}|\Theta)\rangle_\tau$ was estimated by running three independent MCMC chains for 1,000,000 steps at each temperature τ. All chains achieved convergence as per the Gelman-Rubin criterion (although only the latter half of the chains were used, to allow a burn-in period for the MCMC algorithm). The integral necessary to calculate the value of $\ln \mathrm{P}\,(\text{data}|M)$ in Eq. (3.2.2) was then discretized over the interval $\tau \in [0, 1]$. The temperatures used to evaluate $\langle \ln L\,(\text{data}|\Theta)\rangle_\tau$ were divided into three segments: $\tau \in [0, 0.01]$, $\tau \in (0.01, 0.1]$, and $\tau \in (0.1, 1]$ comprising 11, 9, and 9, evenly spaced points respectively. These values for the temperatures were chosen so that the smooth transition from poor likelihood values to good ones was accurately captured. The trapezoidal rule was applied to evaluate the integral:

$$\ln P\,(\text{data}|M) \approx \sum_{i=1}^{n-1} \frac{1}{2}\,(\tau_{i+1} - \tau_i)\left[\langle \ln L\,(\text{data}|\Theta)\rangle_{\tau_{i+1}} + \langle \ln L\,(\text{data}|\Theta)\rangle_{\tau_i}\right]$$

3.3 Results

3.3.1 Model discrimination based on computing the Bayes Factor

We used Bayesian procedures to discriminate between competing direct and indirect models of mitochondrial outer membrane permeabilization, a key step in apoptosis. Both models fit experimental data equally well and thus, cannot be distinguished on a maximum likelihood basis. Discrimination involves estimating the evidence for the indirect model $\mathrm{P}\,(M_I|\text{data})$ divided by the evidence for the direct model $\mathrm{P}\,(M_D|\text{data})$, a ratio known as the Bayes factor [62, 63, 64]. We computed the integrals in Eq. (3.2) using the thermodynamic integration method that we just described. This transforms the problem of evaluating high-dimensional integrals into a problem involving a one-dimensional sum over quantities sampled from a series of MCMC walks. Sampling is weighted by a power posterior that depends on a fictitious temperature factor ranging from a value of 0 to 1 [39, 44, 74]. For each model, three MCMC chains were run

at 29 temperatures between 0 and 1. The quantity ln(likelihood) was averaged with respect to the power posterior at each temperature and over three chains, resulting in two curves, one for each model (Figure 3-3). The ratio of the areas under each curve converges to the logarithm of the Bayes factor. Because thermodynamic integration is a sampling method, the computed Bayes factor is subject to sampling error and must be expressed as a confidence interval. We computed the uncertainty on the areas returned by thermodynamic integration by estimating the variance at each point of the curve to generate a two-sided confidence interval.

Computation of the Bayes factor revealed the direct model to be $\sim 16 - 24$ times more probable than the indirect model with 90% confidence, reflecting the greater range of parameter values compatible with the direct model (Figure 3-4). This formalization of a "robustness" criterion for preferring the direct model is consistent with recent experiments obtained from engineered cell lines [29]. With respect to the biological significance of this finding, however, it is important to note that published indirect and direct "word models" are compatible with many different ODE networks. Thus, it will ultimately be necessary to distinguish among extended sets of competing models, not just the two presented here. With improvements in computational speed, methods for calculating the Bayes factor using thermodynamic integration is

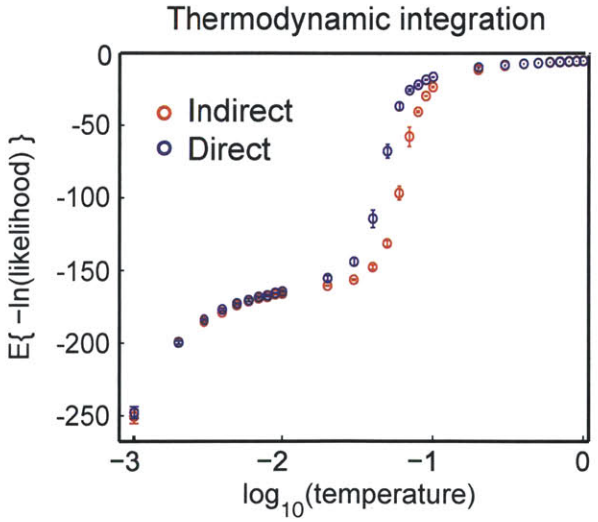

Figure 3-3: Thermodynamic integration curves for the direct and indirect model.

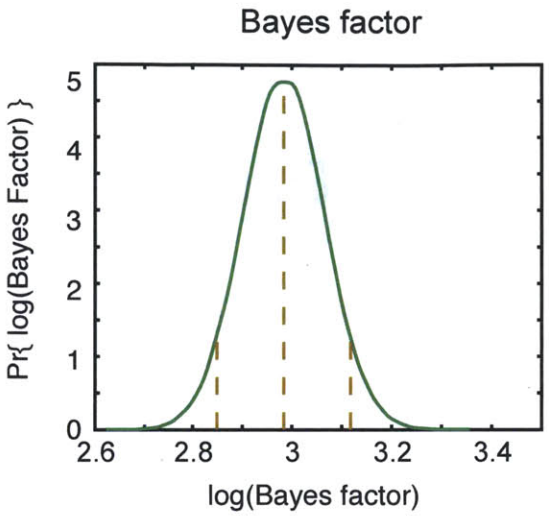

Figure 3-4: Exponentiation of the differential area in the two thermodynamic curves from Figure 3-3 provides an estimate of the Bayes factor for direct and indirect models along with the uncertainty in the estimate. Based on the distribution of the Bayes factor estimate (reflected in the error bars in Figure 3-3) the direct model is preferred to the indirect by a weight of 20, with the 90% confidence interval spanning a range from 16 to 24.

well suited to this task.

Bayesian approaches to model discrimination account not only for uncertainty in parameters values but also for differences in the numbers of parameters. This is important because models that instantiate different hypotheses about biochemical mechanism usually have different numbers of parameters even when the number of unique model species is the same. The complexity penalty embedded in the Bayes factor represents a generalization of the Akaike or Bayesian Information Criteria (AIC and BIC) commonly used to score model likelihoods [64].

We can better understand how the Bayes factor scores models by examining the landscape of the objective function. We approximated the landscape as an n-dimensional ellipsoid (where n refers to the number of parameters in each model) by using a Taylor series at a best-fit position in parameter space. This makes it possible to describe the landscape in terms of an ellipsoid the length of whose axes are inversely proportional to the square roots of the eigenvalues of the second order term of the Taylor expansion (i.e. the Hessian). The direct model had more small

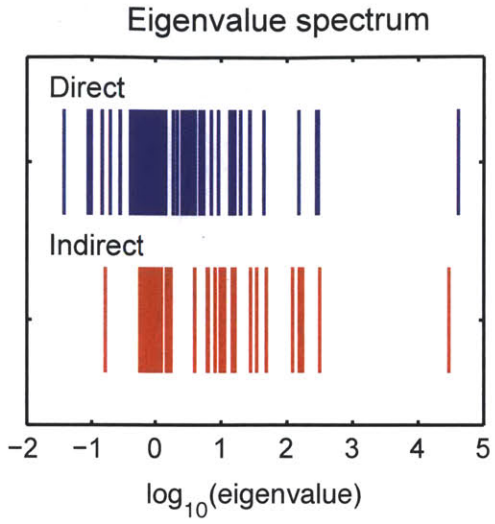

Figure 3-5: Eigenvalue analysis of the landscape around the respective maximum posterior fits shows the direct model (blue) has multiple smaller eigenvalues, suggesting that it is consistent with the data over a larger volume of parameter space and therefore exhibits greater statistical weight.

eigenvalues than the indirect model (i.e., longer axes) and thus a greater volume of equally good parameters (Figure 3-5). The notion that a model is more probable if more parameters give a good fit is frequently if informally applied when models are ranked based on their robustness with respect to parameter variation [24]. It is also intuitively appealing: a model that performs well only over a narrow range of parameter values which are otherwise unknown is less probable than one that is tolerant of variation. Such reasoning is also related, conceptually, to maximum entropy and minimum information approaches.

3.3.2 Validation of thermodynamic integration

We compared six very simple models thought to simulate *in vitro* experiments describing membrane permeabilization. In the experiment, liposomes were used to mimic mitochondria; ANTS dye was placed inside the liposome along with a quencher molecule DPX (liposomes will not fluoresce in the presence of the quencher protein). Measuring fluorescence is indicative of whether permeabilization has occurred or not. If the liposome is intact, no fluorescence occurs; when the liposome is permeabilized,

because the effective volume of the solution is considerably larger, DPX no longer represses ANTS, and fluorescence is observed. Purified Bax and tBid were added to a liposome solution as means of understanding the mechanics of MOMP in a purified setting without any of the complexity that occurs in a natural state.

These models consist of only two proteins, Bid and Bax, and have incremental complexity so that we can determine the most plausible relationship between these two proteins. Six different scenarios were considered:

Translocate Bax:

$$\begin{aligned} Bax_s + Vesicles_f &\rightarrow Bax_f + Vesicles_f \\ Bax_f &\rightarrow Bax_s \\ Bax_s + Vesicles_e &\rightarrow Bax_e + Vesicles_e \\ Bax_e &\rightarrow Bax_s s \end{aligned} \tag{3.5}$$

Translocate Bid:

$$\begin{aligned} Bid_s + Vesicles_f &\rightarrow Bid_f + Vesicles_f \\ Bid_f &\rightarrow Bid_s \\ Bid_s + Vesicles_e &\rightarrow Bid_e + Vesicles_e \\ Bid_e &\rightarrow Bid_s \end{aligned} \tag{3.6}$$

tBid activates Bax

$$tBid + Bax \rightleftharpoons tBid : Bax \rightarrow tBid : Bax^* \rightarrow tBid + Bax^* \tag{3.7}$$

This equation can also be depicted as a simple enzyme-substrate reaction:

$$E + S \rightleftharpoons E : S \rightarrow E : P \rightarrow E + P$$

Bax inhibits tBid

$$tBid + Bax^* \rightarrow tBid : Bax^* \tag{3.8}$$

Continuing with the enzyme-substrate reaction comparison:

$$E + P \rightarrow E : P$$

tBid reverses Bax

$$tBid : Bax^* \rightarrow tBid : Bax \tag{3.9}$$

Continuing with the enzyme-substrate reaction comparison:

$$E : P \rightarrow E : S$$

Dimerization

$$Bax^* + Bax^* \rightarrow Bax2 \tag{3.10}$$

Dye Release

$$Bax_f2 + Vesicles_f \overset{lipo_{efflux}}{\longrightarrow} Bax_f2 + Vesicles_e \tag{3.11}$$

$$Bax_f2 \overset{k_{efflux}}{\longrightarrow} Bax_e2 \tag{3.12}$$

Dimer Dissociation

$$Bax2 \rightarrow Bax^* + Bax^* \tag{3.13}$$

Here, s represents the reaction occurring in solution, e and f indicate empty and full liposomes, respectively. No subscripts imply that the reaction occurs for both e and f liposomes, but that it does not occur in solution.

In order for MOMP to occur, an equilibrium between Bax in solution and Bax on the liposome's membrane needs to be established (Eq. (3.5)). The same equilibrium also needs to occur for tBid (Eq. (3.6)). Once the equilibriums are established, tBid acts as an enzyme and activates Bax (Eq. (3.7)). To accurately mimic a solution of billions of liposomes, the liposomes were divided into two categories: those that contain dye (f), and those that don't (e). The activation of Bax by tBid (Eq. (3.7)) occurs on both empty and full liposomes since the proteins should not selectively

	Increasing Complexity →					
	m1	m1r	m2	m2r	m3	m3r
Translocate Bax	✓	✓	✓	✓	✓	✓
Translocate Bid	✓	✓	✓	✓	✓	✓
tBid activates Bax	✓	✓	✓	✓	✓	✓
Bax inhibits tBid	✗	✗	✓	✓	✓	✓
tBid reverses Bax	✗	✗	✗	✗	✓	✓
Reversible dye release pores	✗	✓	✗	✓	✗	✓

Table 3.1: Six models describing the basic interactions between Bid and Bax

migrate to full liposomes to permeate them. As the number of empty liposomes accumulate, many reactions serve to be unproductive and do not contribute to the observed fluorescence signal. Once Bax is activated, it forms a dimer (Eq. (3.10)). If the dimer is formed on a full liposome, then two reactions should be accounted for: first, the dimer goes from being bound to a full liposome to an empty one (Eq. (3.11)). The rate of this reaction, $lipo_{efflux}$, is mediated by the amount of full vesicles. The half-life of the enzyme Bax_f is determined by k_{efflux}. If k_{efflux} is very fast, then $lipo_{efflux}$ will occur more slowly as there will be less Bax_f2 present in the system. Second, the dimer can now disassociate and bind to other liposomes (accounted only in "r" models, Eq. (3.13)).

If Bax inhibits tBid, then the third reaction describing the activation of Bax by tBid becomes reversible (Eq. (3.7)); otherwise, it is irreversible. Similarly, if tBid reverses Bax, then then the second reaction describing the activation of Bax by tBid becomes reversible (Eq. (3.7)); otherwise, it is reversible. Table 4.1 summarizes the differences between the six model topologies.

In this case, the purpose of model comparison was mainly validation: all models were calibrated to synthetic data generated from m1. The model m1 can be recap-

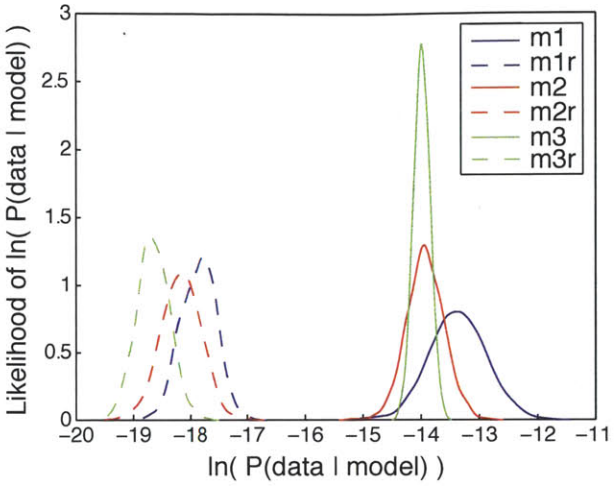

Figure 3-6: Comparison of six simple models describing membrane permeabilization. Synthetic data was generated from the simplest model, m1. Bayes Factor not only determines m1 as the most plausible model, but also segregates the models into two groups: those with reversible dye release pores and those with irreversible ones.

tured from any of the other models by setting all the extra parameters in the other models to zero. We expect that thermodynamic integration would choose m1 as the most likely model since it is set as "ground truth." As illustrated in Figure 3-6, our algorithm does indeed choose m1 as the most likely method. Interestingly, it also divides the networks into two groups: those with reversible dye release pores and those without. These results are a first step in a more subtle and interesting point: model discrimination is seemingly insensitive to most variations in network topologies; certain changes in topology space, such as the existence of reversible dye release pores, significantly change the plausibility of models. Experimental design should be used to further study these particular network connectivities.

3.4 Discussion

3.4.1 Model discrimination in the Bayesian framework

A solid theoretical foundation and large body of literature speaks to the value of Bayesian frameworks for analyzing models having different numbers of uncertain pa-

rameters [62, 64, 81]. The Bayes factor described here, the odds ratio for competing models (i.e. the ratio of the evidence), is computed from an overlap integral between a likelihood and a prior (that is $L(\text{data}|\Theta)$ and $\pi(\Theta|M)$). At first glance it might seem that making a model more complex would increase the number of dimensions and always increase the evidence, but a simple example shows that this is not the case.

Consider a pair of 1- and 2-parameter models of the same hypothetical physical process and a function f that is the ratio of relevant likelihoods: $f(k_1, k_2) = \frac{L(k_1,k_2)}{L(k_1)}$. The evidence for the 1-parameter model is the overlap integral between its likelihood and a normalized prior $\int dk_1\ L(k_1)\,\pi(k_1)$ and for the 2-parameter model it is $\int dk_1\ dk_2\ L(k_1, k_2)\,\pi(k_1)\,\pi(k_2)$. In the case where $f(k_1, k_2) < 1$ for all k_1, k_2 the likelihood of the 2-parameter model is no better than that of the simpler 1-parameter model (note that the evidence for the 2-parameter model is $\int dk_1 L(k_1)\,\pi(k_1)\,g(k_1)$ where the function $g(k_1) = \int dk_2\ \pi(k_2)\,f(k_1, k_2)$ must be less than 1 everywhere, as the priors are normalized to 1). The evidence for the 2-parameter model will therefore be less than the evidence for the 1-parameter model, meaning that it will lose out in a Bayes factor comparison, as it should. When the function $f(k_1, k_2) > 1$ then it must be true that introduction of a second parameter "rescues" or "lifts" the likelihood function by improving the fit to data. In this case, the more complex model will have greater evidence. In the special but interesting case where $f(k_1, k_2) = 1$ for all k_1, k_2, model 2 is completely insensitive to the new parameter. The presence of a parameter with respect to which a model is completely insensitive has no impact in model assessment (the Bayes factor is one). Finally, in the general case where $f(k_1, k_2)$ has values both above and below 1, explicit integration is needed to determine which model is favored, precisely what we do in this paper.

In the simple example presented in Section 3.3.2, the Bayes Factor recovered m1 as the most plausible model since it is simple and sufficient. The Bayes factor, however, is not unique as a means to balance goodness-of-fit and model complexity. The most commonly used metrics are the Akaike information criterion (AIC) and the Bayes information criterion (BIC) [1, 104]:

$$AIC = -2\log(ML) + 2n \qquad BIC = -2\log(ML) + n\log(N)$$

where n is the number of parameters, ML is the maximum likelihood value, and N is the number of data points (ML is simply the highest value achieved by the likelihood function). AIC and BIC do not explicitly account for parameter non-identifiability and the two metrics are therefore good for comparing models only in the "asymptotic" limit where the number of experimental data points becomes "large" and identifiability is "greater" [2, 3, 64]. It is rare in the field of biochemical modeling to explicitly check whether the conditions for computing the AIC and BIC are valid and, in our experience, they are frequently violated. In contrast, the Bayes factor is applicable even with limited data and reduces to the BIC and, under some conditions, to the AIC in the asymptotic limit [2, 3, 64]. Moreover, whereas the AIC or BIC compare models solely on the basis of goodness-of-fit, Bayesian methods allows formal introduction of a prior degree of belief in each model. An arbitrary model (i.e. a physically impossible model) exhibiting a better fit to data might get a better AIC or BIC score than a more realistic mechanistic model, but in a Bayesian approach it would receive a low prior value. We therefore consider evaluation of the Bayes factor to be a better way than AIC or BIC to compare models when models have different numbers of non-identifiable parameters and data are limited.

3.4.2 Limitations of the approach

The computational approach described here has several practical and algorithmic limitations, albeit ones that can be mitigated with further work. A practical concern is that current methods for computing the Bayes factor are too slow to incorporate the full range of data we have collected from cells exposed to drugs, siRNA-mediated protein knock-down and ligand treatment. Using only a subset of available training data, computing the Bayes factor for EARM1.3 required $\sim 6 \times 10^4$ CPU-hr (four weeks on a 100 core general-purpose computer cluster). It should be possible to improve this by optimizing the code (e.g., porting it from MatLab to C/C++;) and performing

multiple analyses in parallel. It also remains to be determined how inclusion of more calibration data will alter the topology of the posterior landscape. It may become more rugged, decreasing the requirement for Hessian-guidance during MCMC walks but increasing the need for simulated annealing to move out of local minima.

3.5 Conclusions

The ubiquity of Bayesian methods in other scientific fields has motivated multiple, parallel efforts to apply the approach to biochemical models [22, 37, 62, 71], but significant challenges remain with respect to development of widely available methods for discriminating between competing models. The algorithm described in this chapter uses the Bayes factor to distinguish between models with different topologies and numbers of parameters in a rigorous manner.

It is our opinion that application of rigorous probabilistic analysis of biochemical models will advance the long-term goal of understanding complex biochemical networks in diverse cell types and disease states [62, 71]. Preliminary application of Bayesian reasoning suggests that some long-standing disputes about cell signaling can be laid to rest, (e.g., direct versus indirect control of MOMP), whereas others cannot be truly discriminated based on available data.

Chapter 4

An Extension To Multiple Models At A Different Scale

Although mathematical models are a useful tool for studying biochemical signaling pathways, they are also ever changing due to new discoveries or new hypotheses (that are constantly formed due to the uncertainty involved with the sheer size of the models and relatively small amount of data). These changes, which typically consist of model species and biochemical reactions being inserted, removed, or modified, are simply part of an ongoing, iterative model improvement process. In this chapter, we show that, despite uncertainty in both model topologies and model parameters, the methodology presented in previous chapters can be leveraged to make predictions on hallmark characteristics of the apoptosis signaling pathway.

4.1 Mini-models Analyzed

The receptor-mediated apoptosis signaling pathway can be divided into three general modules [79]:

1. The DISC module, which begins with the binding of TRAIL to receptor and ends with the assembly of DISC components

2. The MOMP module, which begins with caspase-8 cleaving Bid and ends with

mitochondrial pore formation; and

3. The PARP module, which consists of all the biochemical reactions involving the series of caspase-8 and caspase-3 regulators. Activation of caspase-3 leads to cleavage of PARP and other cellular substrates typically leading to cell death.

In this chapter, we focus on exploring eight previously published models [5, 24, 32, 79] describing alternative hypotheses for the MOMP module. Each of these mini-models proposes a slightly different control mechanism for MOMP regulation by the Bcl2 family of proteins in apoptosis. These simple models are variations on three conflicting hypotheses: the direct, indirect and embedded mechanisms [105]. The direct and indirect mechanisms were described in Chapter Three; the embedded mechanism [79], which incorporates a combination of elements from the direct and the indirect mechanisms, is based on recent experimental findings from various research groups [15, 76, 78]. In the embedded method of MOMP regulation, tBid activates Bax both by direct binding and by inhibiting Bcl2, the inhibitor of activated Bax. Similarly, Bcl2 inhibits apoptosis by binding to both tBid, an activator, and to activated Bax.

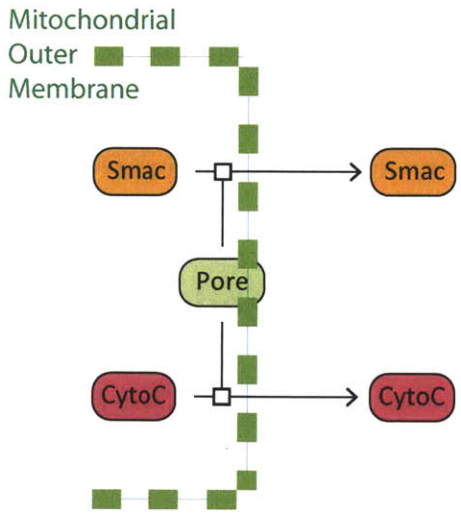

Figure 4-1: An illustration of Smac and CytoC translocation from the mitochondria to the cytosol upon pore formation on the mitochondrial outer membrane.

Model Name	Number of ODEs	Number of parameters	Reference
Albeck 11e	23	34	[5]
Chen Direct	15	23	[24]
Chen Indirect	15	21	[24]
Cui Direct	19	38	[32]
Cui Direct II	20	46	[32]
Lopez Direct	35	62	[79]
Lopez Embedded	43	82	[79]
Lopez Indirect	37	68	[79]

Table 4.1: Eight mini-models describing the control mechanism behind MOMP

Bax auto-activation is also permitted in the embedded model. Last, all of the reactions involving the Bcl2 family of proteins occur at the mitochondrial membrane and not in the cytosol [80].

The eight mini-models incorporate a varying number of ODEs and parameters to describe the underlying biochemistry as evident in Table 4.1; model topologies are illustrated in Figures 4-2 and 4-3. Pore formation results in the translocation of Smac and CytoC from the mitochondria to the cytosol in all of the models as depicted in Figure 4-1.

4.1.1 Chen and Cui Models

The Chen Indirect model [24] proposes an indirect mechanism in which tBid does not activate Bax; rather, Bax is naturally active and able to form oligomers. Additionally, Bcl2 can bind both tBid and Bax. The Chen Direct model is suggested by the same researchers [24] but involves a direct activation method in which tBid activates Bax through a reversible, one-step mechanism. In this model, Bcl2 can bind tBid, but not Bax. The topology of the Cui Direct and Cui Direct II models are both derived from the Chen Direct model; each of these models is incrementally more complex. All of

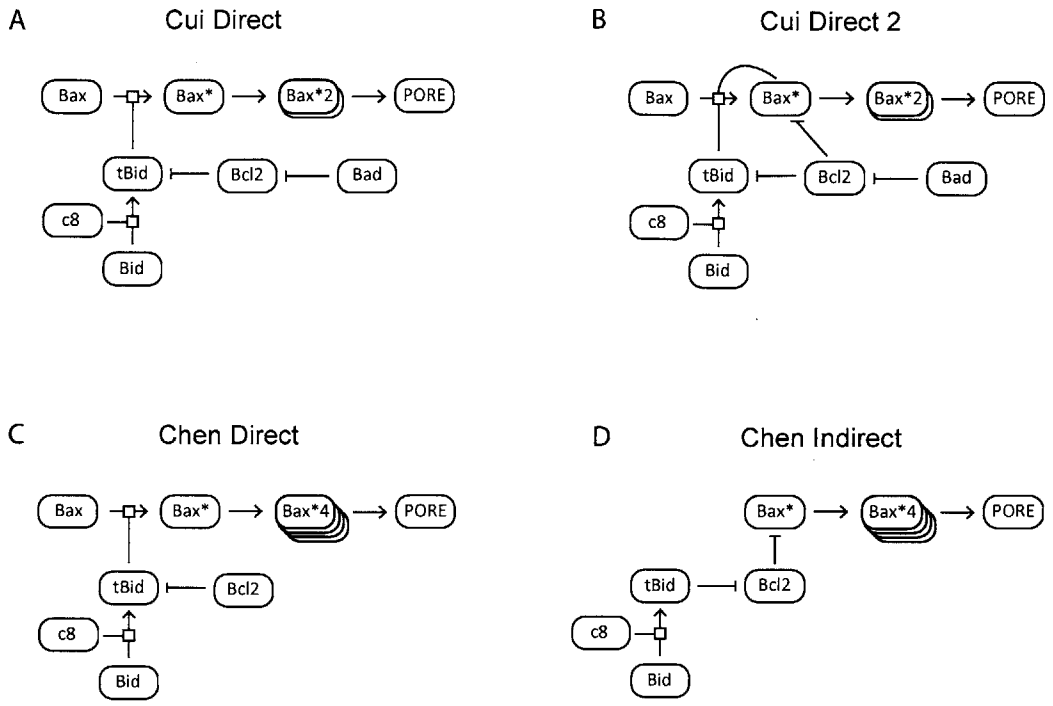

Figure 4-2: Model topologies for the (A) Cui Direct , (B) Cui Direct II, (C) Chen Direct, and (D) Chen Indirect mini-models. The Cui Direct and Cui Direct II models are incrementally more complex models that are both derived from the Chen Direct model.

these models incorporate a pseudo-first-order interpretation of the enzyme-substrate reaction, in which the formation of a complex is completely bypassed, so that the reaction is simplified from $E + S \longleftrightarrow E : S \longrightarrow E + P$ to $E + S \longrightarrow E + P$. This simplified view of the kinetics reduces the number of parameters needed to describe the underlying biochemistry. It also implies a strict assumption: that the enzyme only operates within a linear range and is never saturated [79]. Though largely derived from Chen Direct, the Cui models differ in that they incorporate displacement, synthesis, and degradation reactions; in addition, mitochondrial pores are formed by Bax dimers and not Bax tetramers. As the name suggests, the Cui Direct model proposes a direct mechanism for MOMP regulation while Cui Direct II proposes an embedded one. In addition to all the reactions present in Cui Direct, Cui Direct II also accounts for Bax auto-activation and the inhibition of activated Bax by Bcl2.

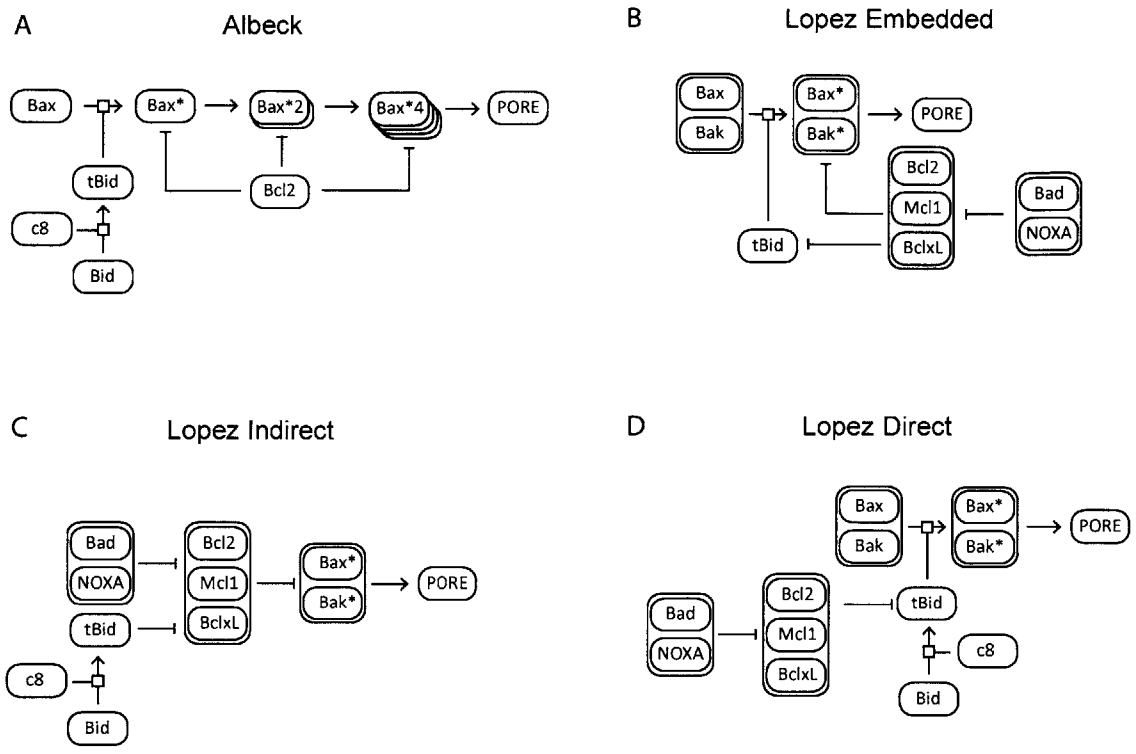

Figure 4-3: Model topologies for the (A) Albeck, (B) Lopez Embedded, (C) Lopez Indirect, and (D) Lopez Direct mini-models. The Lopez Embedded model is a combination of the Lopez Indirect and Lopez Direct models.

4.1.2 Albeck and Lopez Models

Instead of the pseudo-first-order interpretation in the Chen and Cui models, the Albeck and Lopez models incorporate the typical two-step approach to the enzyme-substrate reaction: $E + S \longleftrightarrow E : S \longrightarrow E + P$. The Albeck model [5] is a hybrid of the direct and indirect mechanisms. This model postulates that Bid activates Bax, Bax dimerizes and its dimers dimerize again to form tetramers; Bcl2, in turn, is capable of binding to all forms of Bax: monomers, dimers, and tetramers. Pores on the mitochondrial membrane are formed by Bax tetramers only.

Relative to other models, the Lopez models [79] describe MOMP regulation in more detail and as a result have a higher number of both species and parameters. The Lopez models account for one activator (Bid), two sensitizers (Bad and Noxa), two effectors (Bax and Bak) and three anti-apoptotics (Bcl2, Bcl-xL, and Mcl-1);

the Albeck, Chen, and Cui models only account for one protein from each of the aforementioned categories. The Lopez Embedded model describes MOMP regulation by allowing tBid to activate both effector proteins, which can also auto-activate. The anti-apoptotic proteins can bind tBid, effectors, and sensitizers. A detailed description of the Lopez Direct and Indirect models can be found in Chapter Three.

4.2 Methods

4.2.1 Using PySB to generate model ODEs

Creating mathematical models using the traditional method of writing ODEs to describe biochemical reactions is a major barrier to model creation and modification. To generate the ODE models of the eight mini-models, we used PySB [79], which allows for modularization and quick generation of model variants. Pre-MOMP events were written in the form of a program and common elements between the models were written as macros to prevent repetition and simplify model construction. As an example, Figure 4-4 displays the *add_caspase8* macro written to add the ODEs corresponding to Bid translocation by C8 to the models; this macro was common to all the mini-models as it was used to drive them with the same IC-RP trajectory.

4.2.2 Driving the models with an IC-RP trajectory

The full models, which represent the EARM model in its entirety, constrain the tBid trajectory to match IC-RP. In order for the mini-models, which represent only the MOMP module, to accurately mimic the dynamics of the full models, the same tBid trajectory to which the full models were calibrated was used as the input stimulus to the mini-models. As a result, the mini-models see tBid as an input that has been forced to match the IC-RP trajectory. Since Bid participates in other reactions in the mini-models (i.e. it forms complexes with Bax and Bcl2), we cannot control its dynamics with only one ODE. Instead, we calculated the derivative of the IC-RP trajectory and used it to fit the C8 trajectory using a function that consisted of

```
from pysb import *
from pysb.util import alias_model_components

def add_caspase8():
    # Add caspase 8
    Monomer('C8', ['state', 'bf'], {'state': ['pro', 'A']})

    alias_model_components()

    # Add caspase 8 initial condition (with placeholder value 1)
    Initial(C8(state='A', bf=None), Parameter('C8_0', 1))

    # Add rules C8 + Bid <-> C8:Bid -> Bid + C8*
    kf = Parameter('bind_C8A_BidU_to_C8ABidU_kf', 1e-7)
    kr = Parameter('bind_C8A_BidU_to_C8ABidU_kr', 1e-3)
    kc = Parameter('catalyze_C8ABidU_to_C8A_BidT_kc', 1)

    Rule('bind_C8A_BidU_to_C8ABidU',
         C8(state='A', bf=None) + Bid(state='U', bf=None) <>
         C8(state='A', bf=1) % Bid(state='U', bf=1),
         kf, kr)

    Rule('catalyze_C8ABidU_to_C8A_BidT',
         C8(state='A', bf=1) % Bid(state='U', bf=1) >>
         C8(state='A', bf=None) + Bid(state='T', bf=None),
         kc)
```

Figure 4-4: The PySB macro describing Bid translocation by C8.

the sum of five Gaussian distributions. The function was then added to the ODE corresponding to the C8 dynamic trajectory in the mini-models. Since C8 is solely responsible for converting all of Bid to tBid in the mini-models, the addition of the ODE describing C8 kinetics to the model controlled the kinetics of tBid such that it matched the IC-RP trajectory.

4.2.3 Calibrating mini-models

The mitochondrial intermembrane space reporter protein (IMS-RP) trajectories for cells treated with 50 ng/ml TRAIL were obtained from experiments performed by John Albeck. The IMS-RP trajectory obtained from a single cell was used for model calibration under both normal conditions (Bcl2 concentration of 20 nM) and Bcl2 20-fold over expression, a condition under which cell death is typically blocked, using the Bayesian MCMC random walk approach introduced in Chapter Two. The Gelman-Rubin criterion was used to check for the convergence of the MCMC chains. In the interest of computational efficiency, instead of using the Hessian matrix, we used the inverse of the covariance matrix. That is, at every (user-defined) m steps, the inverse

of the covariance matrix obtained from the last m steps in the random walk was used to guide the selection of the new proposal position in parameter space:

$$C^{-1}_{(\Theta^{t-m},\cdots,\Theta^{t})} = U\Lambda U^{T}$$

$$\Theta^{t+1}_{test} = \Theta^{t} + \sum_{i=1}^{n} N\left(0, \frac{1}{\sqrt{\lambda_i}}\right) u_i$$

As before, Θ^t is the t^{th} position in parameter space, Θ^{t+1}_{test} is the putative next position, n is the number of parameters being estimated, C is the covariance matrix of the last m positions in parameter space, u_i and λ_i are the i^{th} eigenvector and eigenvalue, respectively, of the inverse covariance matrix, C^{-1}, and $N(0, \sigma)$ is a random value drawn from a normal distribution with mean zero and variance σ.

4.3 Results

We calibrated eight mini-models to the IMS-RP trajectory of a single cell treated with 50 ng/ml TRAIL. In the IC-RP trajectory used for calibration, T_d occurs on average 177 minutes after the addition of TRAIL. We then perturbed each network and predicted how T_d (see Chapter Two) changes relative to its original value under the following conditions: (i) Bax and Bak knockout, (ii) Bcl2 knockdown, and (iii) Bid dose response. As before, these conditions were simulated by sampling 1000 parameter vectors Θ from the posterior distributions arising from three independent and convergent MCMC chains and computing IMS-RP trajectories for each Θ. T_d was calculated as the time at which the IMS-RP trajectory reaches its half-maximal point. The IMS-RP data is a time-series beginning at time 0 and ending at 267 minutes. In making predictions, if cell death did not occur within this time frame, T_d was set as occurring at the last point in the time series, i.e. 267 minutes. We calculated 90% confidence intervals for the T_d predictions since they comprise samples from probability density functions rather than single values.

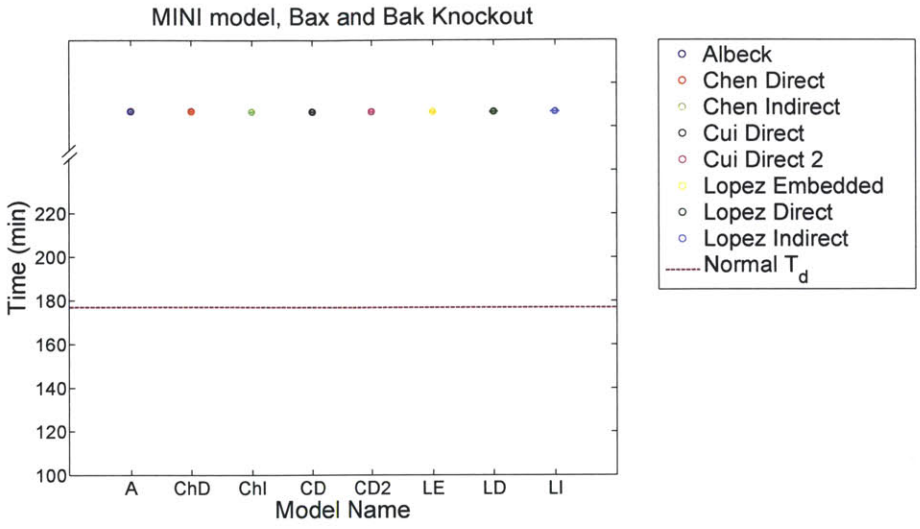

Figure 4-5: Effect of a Bax and Bak knockout on T_d predictions in eight mini-models describing MOMP: none of the mini-models exhibit cell death

4.3.1 Predictions: Bax and Bak Knockout

We first predicted T_d while removing all effector proteins, i.e. Bax and Bak, from the model. Without the effectors, transmembrane pores cannot be formed on the mitochondrial membrane and MOMP should not occur. As illustrated in Figure 4-5, all of the models accurately predict that the cells do not die.

4.3.2 Predictions: Bcl2 knockdown

Next, we knocked down Bcl2's concentration 10-fold, from 20 nM to 2 nM. Bcl2 is an anti-apoptotic protein which, depending on the model topology, inhibits the activation of the pore-forming effector proteins either by binding to the effectors or by inhibiting tBid. A Bcl2 knockdown implies that T_d should occur sooner than under normal conditions. Although we do not have experimental data to predict the exact time that T_d should occur under these knockdown conditions, all the of the models accurately predict the earlier occurrence of T_d; see Figure 4-6.

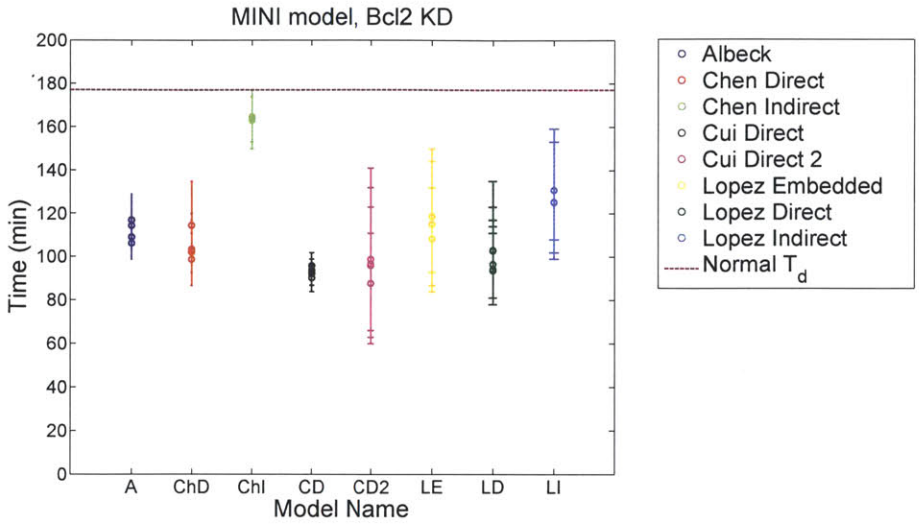

Figure 4-6: Effect of a Bcl2 10-fold knockdown on T_d predictions in eight mini-models describing MOMP: all models predict that cell death should occur more quickly.

4.3.3 Predictions: Bid dose response

Last, we were interested in predicting how the cells would behave under a Bid dose response. The models were calibrated to a Bid concentration of 40nM; T_d was predicted for Bid concentrations of 0 nM, 0.1 nM, 0.3 nM, 1 nM, 3 nM, 10 nM, 30 nM, and 100 nM. As evident in Figure 4-7, the difference in the dynamics of the indirect mechanism from the direct and embedded mechanisms becomes apparent. The indirect method suggests that the effector proteins are naturally in an active, pore-forming state. The role of Bid in the model is to inhibit pore formation by directly binding to Bid. It is therefore not surprising that the Chen Indirect model illustrates that the occurrence of MOMP in all of the cells is independent of the presence of Bid in the system. The Lopez Indirect methods illustrates this independence for only some of the cells in the population, which could be as a result of a having the extra biochemical components present in the system. T_d predictions made by the Albeck, Chen Direct, Lopez Embedded, and Lopez Direct, which propose a form of either an embedded or a direct mechanism, exhibit the same trend: at low Bid concentrations the cells never undergo MOMP since without enough Bid in the cell, the effectors are

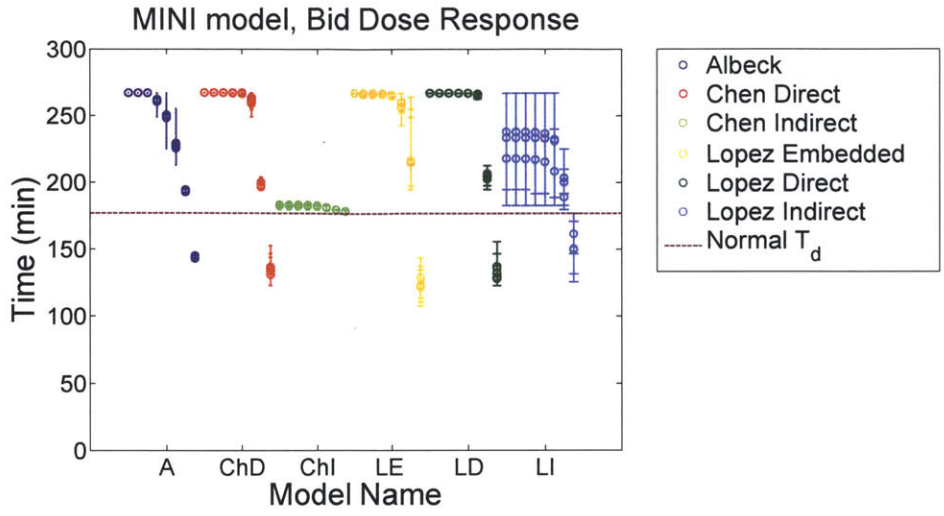

Figure 4-7: Effect of a Bid dose response (0 nM, 0.1 nM, 0.3 nM, 1 nM, 3 nM, 10 nM, 30 nM, and 100 nM) on T_d predictions in eight mini-models describing MOMP.

never activated; as Bid concentrations surpass 10 nM, the average T_d decreases such that at the final Bid concentration of 100 nM, T_d consistently occurs sooner than at the calibrated 40 nM.

The predictions made by the Cui models are different than the remainder of the models mainly due to the presence of synthesis and degradation rate constants in the models. We calibrated the Cui models with nonzero synthesis and degradation rate constants and proceeded to predict T_d under a Bid dose response both with zero and nonzero synthesis and degradations. The effect of the synthesis and degradation rates can be observed by the difference in the predictions made in Figure 4-8. We are currently not certain about the exact biochemical reasoning as to why the presence or absence of these rate constants produces these particular predictions, but feel that it merits further study from the experimental side to better understand the underlying biochemistry.

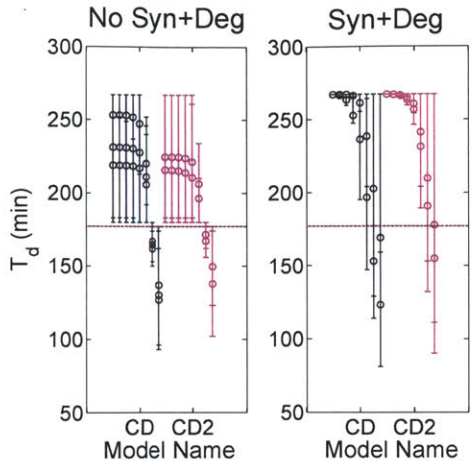

Figure 4-8: Effect of a Bid dose response (0 nM, 0.1 nM, 0.3 nM, 1 nM, 3 nM, 10 nM, 30 nM, and 100 nM) on T_d predictions in the Cui Direct and Cui Direct II models. We calibrated the models while accounting for synthesis and degradation rate constants and then made the T_d predictions while removing all synthesis and degradation rate constants (left panel) and keeping them at their original, nonzero values (right panel).

4.4 Discussion

4.4.1 Experimental Design

We calibrated eight mini-models describing the control mechanism behind MOMP and predicted T_d under various conditions: Bax and Bak knockout, Bcl2 knockdown, and Bid dose response. Although we did not have experimental data to validate said predictions, the results can be used to guide experimental design. These specific perturbations were chosen precisely because they represent cellular conditions that can be replicated as part of an experiment in a feasible manner. The results of the experiments could, in turn, aid in resolving competing hypotheses. For instance, if the results of an Bcl2 10-fold knockdown experiment indicate that T_d consistently occurs at after approximately 140 minutes, we could then claim that the correct topology is either the Chen Indirect, Cui Direct II, Lopez Embedded, or Lopez Indirect model. A Bid dose response experiment could then determine the correct topology depending on the values obtained for T_d.

4.4.2 Ensemble Modeling

Due to model nonidentifiability, we used the MCMC Bayesian framework introduced in Chapter Two to calibrate a biochemical model of apoptosis and generate marginal parameter distributions, which captured the uncertainty in parameter values. The variance in parameter values led to a variance in model output as each parameter vector Θ generated a different trajectory describing protein dynamics. The variance in the dynamical trajectories generated by the model, in turn led to a variance in model predictions for conditions to which the model was not calibrated. As summarized in Figure 4-9, despite only having general prior knowledge about the plausible ranges in which parameter values could fall, the Bayesian framework transformed parameter uncertainty into accurate T_d and T_s predictions; the uncertainty in parameter values is translated into confidence intervals on the predictions. In this case, the values of the parameters are not of interest per se; rather, the information that we can obtain about the system is of interest.

The same idea can be applied on a topological level, to deal with uncertainty in the underlying dynamics. In this chapter we presented eight competing hypotheses for MOMP regulation by the Bcl2 family of proteins. The uncertainty in network topology led to variance in the model output, as each one of the models had a different prediction for the various perturbations in protein concentration. Aggregating these predictions generates a probability distribution for model predictions as before (Figure 4-9)

This idea of aggregating mathematical models despite the presence of large uncertainties in both parameter values and model topology and leveraging it to learn something about model dynamics is typically referred to as ensemble modeling. Ensemble modeling has proven to be quite useful in systems biology [20, 72, 110, 112] due to limited quantitative data, insufficient mechanistic knowledge, and conflicting topological hypotheses [72]. Ensembling modeling has also been used in applications for gas emissions[95], protein folding [59], public health [107], climate change [86], and weather forecast predictions [100]. Applying ensemble modeling to the apoptosis sig-

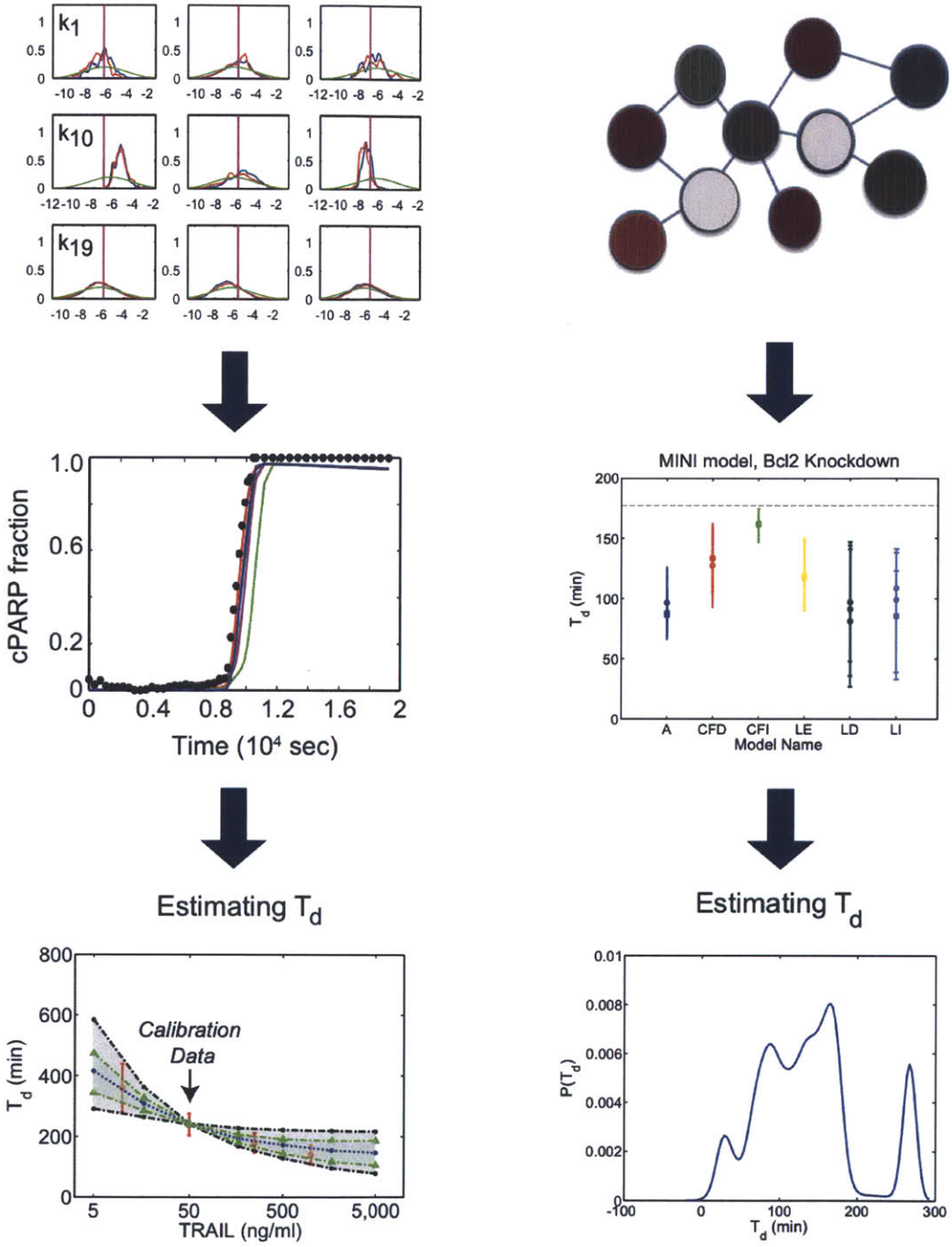

Figure 4-9: Despite parameter uncertainty, the Bayesian framework enabled accurate T_d and T_s predictions by translating the uncertainty in parameter values into confidence intervals on the predictions. In the same manner, ensemble modeling can be used to transform topological uncertainty into variance in model predictions.

naling pathway could serve as a method from which reliable model-based prediction can be obtained despite uncertainty in both parameter values and model topology.

4.4.3 Comparison of full models and mini-models

It would be interesting to explore whether mathematical models capturing the biochemistry of the apoptosis signaling pathway can be replaced by the models describing modules in the system. One approach would be to determine whether the full model or the modules is probabilistically more likely, using the methodology introduced in Chapter Three. Comparing the full and mini-models using the Bayes Factor method is incorrect for two reasons. First, prior knowledge dictates that $\mathrm{P}\left(M_{full}\right) \gg \mathrm{P}\left(M_{mini}\right)$. The full models are a more realistic mechanistic model as they capturee more of the underlying biochemistry. Second, the experimental data to which model output is compared is not the same in the two models. The mini-models are driven by tBid trajectory and are calibrated to IMS-RP data whereas the full models are not driven by any reporter protein and are calibrated to IMS-RP as well IC-RP and EC-RP.

Since the full and mini-models cannot be compared on the basis of probabilistic likelihoods, it is helpful to address the question of whether the full models can be replaced by the modules from a different angle: to determine whether both sets of models produce the same model predictions. This verification is particularly important since modules do not necessarily generate the same dynamics in isolation as when they are integrated into a larger network [27, 35, 79]. This exercise will remain for future work.

4.5 Conclusion

Systems biology is an iteratively evolving field in that the collection of experimental data help create mathematical models that describe the dynamics and kinetics of the biochemistry. These mathematical models, in turn, drive experimental design and further data collection. Due to limitations in both quantitative data and biological experiments that can be performed in a feasible manner, there exist con-

flicting hypotheses on the structure of these models, which are also nonidentifiable. As demonstrated in this chapter, despite uncertainty in both parameter values and model topology, the methods we have introduced in previous chapters can be used to make predictions both for individual models and for ensembles. These mechanistic predictions can shed light on the effects of different topologies on model output (such as the indirect mechanism of MOMP control and its predictions for the Bid dose response in comparison to that of the direct and embedded mechanism or the role of synthesis and degradation rate constants). By driving experimental design, these methods have the potential to reduce ambiguities in the mechanisms of the apoptosis signaling pathway and in other systems biology problems with a similar flavor.

Chapter 5

Conclusion

In this chapter, we summarize the contributions of this thesis and point to directions that could be further explored.

5.1 Concluding remarks

In this thesis, we have introduced a quantitative approach to analyzing large-scale biochemical signaling pathways in systems biology. We focus on the TRAIL-mediated extrinsic apoptosis pathway and model it with a system of ODEs. A typical problem with ODE models of biochemistry is that they are nonidentifiable, meaning that many different parameter values fit the model equally well. Despite the issue of nonidentifiability, in modeling these biochemical systems, it is typical that only a single set of parameter values, based on literature values or *in vitro* experiments, are chosen to fit the model. These parameters can only describe the dynamics and kinetics of the underlying system under a limited number of conditions—those for which either experimental data or some prior intuition exists. In order to accurate capture the full dynamics of the model, we introduce a Bayesian approach to replace the current maximum likelihood approach used to select parameter values.

5.1.1 Model calibration and prediction

We use a Hessian-guided Bayesian MCMC random walk to characterize the region of nonidentifiability in parameter space. The periodic calculation of the Hessian matrix enables the random walk to efficiently explore parameters space as it adjusts step size in parameter space according to the curvature in the posterior landscape so that step size is inversely proportional to the square root of the matrix eigenvalues. From the random walk, we can obtain both the marginal and joint parameter distributions. The marginal parameter distributions were used to demonstrate that the nonlinear co-variation in the parameters contains important information. Two different hallmark characteristics of the apoptosis signaling pathway, T_d and T_s, were predicted under different death-inducing drug doses using parameters obtained from the marginal distributions via three different sampling methods: independent sampling, covariance sampling, and manifold sampling. Independent sampling was used to mimic the current method of reporting parameter values: summarized in a table with a list of parameter mean and standard deviation values. Covariance sampling took into account a linear co-varying relationship between parameter values. Manifold sampling represented the MCMC walk and took into consideration the nonlinear relationship between the parameters. Manifold sampling was the only method that produced accurate, high-confidence predictions, indicating that although many parameters' marginal distributions are typically highly unconstrained and do not significantly change from their imposed prior distribution, important information lies in the nonlinear relationship between parameters that cannot be captured in table format. For the purposes of model calibration and prediction, we suggest storing the parameter vectors sampled throughout the random walk in their entirety in matrix format.

5.1.2 Model discrimination

Due to limited data, there exists relatively high degrees of freedom in biochemical models of signaling pathways. During model calibration, parameter values can be

selected so that model output fits experimental data quite well. This parameter non-identifiability often results in two or more models fitting the same set of experimental data equally well. Even though the models become indistinguishable using simple maximum likelihood methods, they are not equally probable. In order to quantify the plausibility of one model topology compared to another, we calculate the Bayes factor, which is merely an extension of the Bayesian framework proposed for model calibration. The Bayes factor approach encodes three qualities. First, good-ness of fit: models are ranked based on how well model output fits experimental data. Second, robustness: models are ranked based on their robustness to parameter variation. Finally, complexity: the approach has a natural Occam's razor, which reduces to BIC in the asymptotic limit, that penalizes models by their number of parameters. Calculating the Bayes factor is challenging due the high-dimensional integral that is the evidence in Bayes' equation. We used thermodynamic integration, which is a path sampling method, to approximate the log of the evidence. Thermodynamic integration requires minimal additional calculation as it only requires tracking the ln(likelihood) values in the Bayesian MCMC random walk.

This method was applied to the discrimination of two conflicting hypotheses, the "direct" mechanism and the "indirect" mechanisms, which describe how the Bcl-2 family of proteins control MOMP. The Bayes factor enabled the discrimination of the two model topologies in a quantitative manner: with 90% confidence, the direct model was deemed to be approximately 16-24 times more plausible than the indirect one.

5.1.3 Overview

Systems biology is an iterative field in which the creation and modification of mathematical models and the collection of experimental data occur continuously and propel the field forward. There exist many competing hypotheses about model topologies due to limited data sets and insufficient knowledge about network components and their relationships with one another. We chose to analyze eight competing models, two of which are a simple expansion of another, that describe the MOMP module

in the apoptosis pathway. These models also represented three different schools of thought about the mechanism by which the Bcl-2 family of proteins control MOMP: the direct mechanism, the indirect mechanism, and the embedded mechanism. Using the methodology introduced in this thesis, we make predictions the time of cell death while varying initial protein concentrations. Even without experimental data, the results of these predictions can be used to drive experimental design and narrow down the number of plausible model topologies. Despite topology uncertainty, ensemble modeling can be applied to the aggregate of these models in order to make predictions. In this thesis we illustrate that using the proposed Bayesian framework, we can leverage uncertainty in both model parameter values and model structure to calibrate models and make accurate predictions, discriminate topologies, and drive experimental design.

5.2 Future work

5.2.1 Calibration and ensemble of $\sim 10^3$ models

In Chapter Four we calibrated eight different models describing the series of biochemical reactions in the apoptosis pathway's MOMP module. It could be interesting to generate on the order of $\sim 10^3$ models and to form an ensemble of models. Finding the computational resources to calibrate all of the models could prove to be difficult. Perhaps it would be useful if calibration was performed using cloud computing resources. While we did not have experimental data to validate our T_d predictions, those specific perturbations were chosen because they are conditions that can be replicated *in vitro* in a feasible manner. The results of experiments could be interesting for analysis, validation of predictions, and for guiding experimental design.

5.2.2 Calibration of a population of cells

In this thesis, all the models were calibrated to a single cell's time-series trajectories; whether it be EC-RP, IC-RP, or IMS-RP, all the trajectories were obtained from

a cell with a particular MOMP time. In performing model calibration, we also restricted the algorithm to only estimating rate constants while holding initial protein concentrations constant. Rate constants are typically assumed to be the same across various cells while initial protein constants vary. It is this variation in initial protein concentrations that leads to the variation in cell death time in a population of cells. Instead of calibrating the models to only time-series data obtained from a single cell, it would be interesting to calibrate them to time-series data obtained from the entire population of cells. The biggest challenge in this task would yet again be the computation complexity of the entire MCMC walk. At each step in the random walk, the system of ODEs describing the model would need to be solved nm times, where n is the number of cells in the populations and m is number of perturbations, such as knock outs, knock downs, and over expressions, to which we are calibrating. Since solving the ODE system is the rate-limiting step in the algorithm, this approach could prove quite computationally expensive.

Appendix A

A Simple Example of Bayesian Estimation

To illustrate how MCMC-dependent Bayesian parameter estimation works, consider an ODE model of three species ($A - C$) that interact via three reactions (with rates k_1 to k_3) corresponding to a classical example of an autocatalysis developed half a century ago by Robertson [98]:

$$A \xrightarrow{k_1} B \qquad 2B \xrightarrow{k_2} B + C \qquad B + C \xrightarrow{k_3} A + C \tag{A.1}$$

First we generate a set of synthetic data corresponding to the time-dependent concentrations of species A to C by choosing reaction rates and setting the initial concentration of reactant A to A_0=1.0 (in non-dimensional units). The resulting trajectory is sampled periodically assuming a theoretical measurement error of 10% which enters into the term of objective function (red bars, Figure A-1A). We then hide knowledge of the parameter values and attempt to infer them from the synthetic measurements on all three species. In the treatment of Sethna [53] this would correspond to a situation with "complete" knowledge. We perform an MCMC search giving rise to a joint posterior distribution whose marginal values and point by point co-variation is shown in Figure A-1B. We see that the parameters of the system are non-identifiable despite complete data and that k_1 and k_3 co-vary: neither parameter

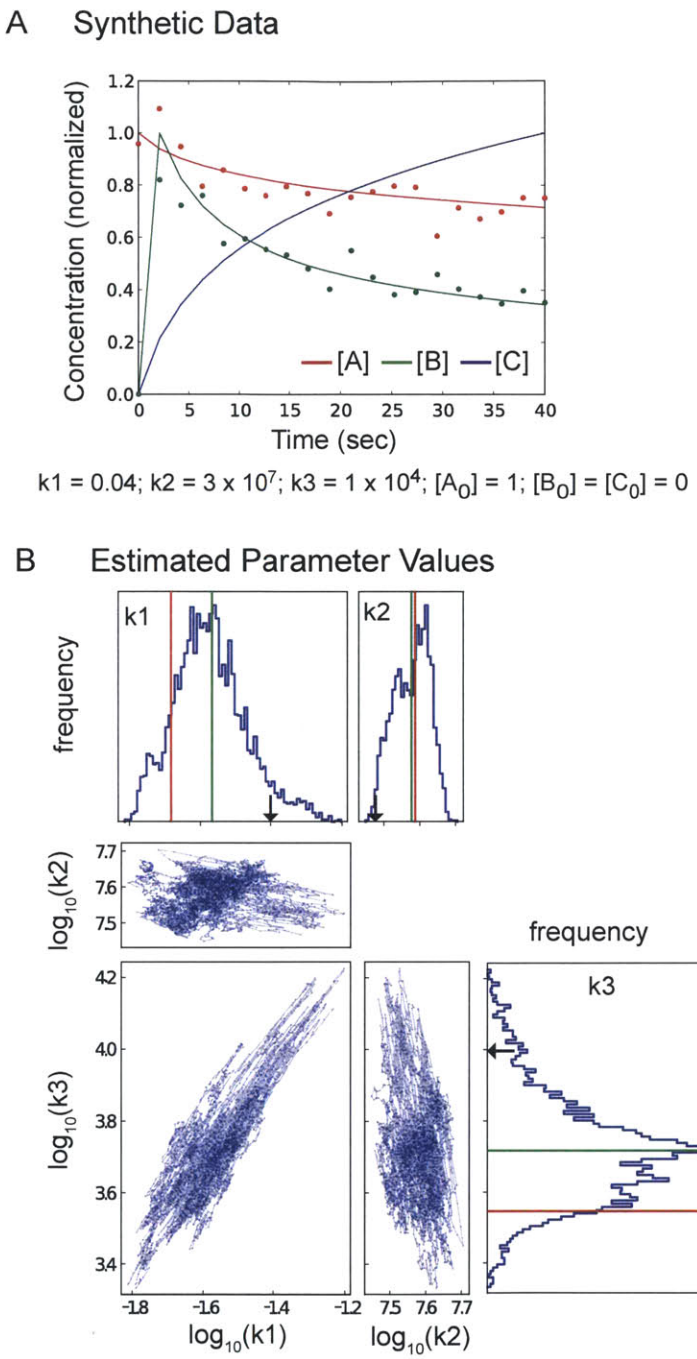

Figure A-1: A) Synthetic data generated from an ODEs corresponding to the Robertson system in A.1 with the parameter values shown immediately below. B) Results of Bayesian parameter estimation using data on species A and B, independent Gaussian priors and showing marginal distributions above and to the right of correlation plots for the three parameters. Red lines correspond to the best-fit parameter set and green lines to the mean of the individual marginal distribution, as in Figure 2-7.

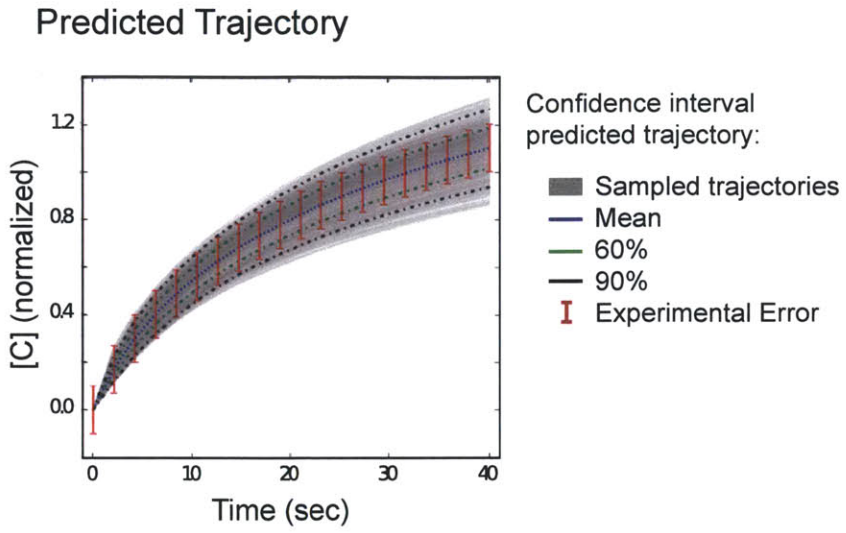

Figure A-2: Predicting the trajectory of species C given the posterior distribution in Figure A-1B. Red error bars denote one standard deviation around the mean of the data (assuming 10% error; data on C was withheld from the estimation) and the 60% and 90% confidence intervals of the prediction.

is particularly well estimated but their ratio is well determined. This ratio plays an important role in determining the trajectory of reactant C whose value can be predicted with good accuracy. Reflecting the non-identifiability of the system and the presence of experimental error, the prediction has a manifold of uncertainty, depicted in Figure A-2D) as 60% and 90% confidence limits (in the case of this prediction, we trained the model on measurement of species A and B only). The inability of estimation to recover the parameter values used to generate synthetic data is not due to problems with the computational procedures. Instead, it represents a fundamental limit on our ability to understand biochemical systems based on time-course data alone.

Bibliography

[1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[2] H. Akaike. *On Entropy Maximization Principle.* Applications of Statistics. North-Holland, Amsterdam, 1977.

[3] H. Akaike. Information measures and model selection. *Bulletin of the International Statistical Institute*, 50(1):277–291, 1983.

[4] J. G. Albeck, J. M. Burke, B. B. Aldridge, M. Zhang, D. A. Lauffenburger, and P. K. Sorger. Quantitative analysis of pathways controlling extrinsic apoptosis in single cells. *Mol Cell*, 30(1):11–25, 2008.

[5] J. G. Albeck, J. M. Burke, S. L. Spencer, D. A. Lauffenburger, and P. K. Sorger. Modeling a snap-action, variable-delay switch controlling extrinsic cell death. *PLoS Biol*, 6(12):2831–52, 2008.

[6] B. B. Aldridge, J. M. Burke, D. A. Lauffenburger, and P. K. Sorger. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol*, 8(11):1195–203, 2006.

[7] B. B. Aldridge, S. Gaudet, D. A. Lauffenburger, and P. K. Sorger. Lyapunov exponents and phase diagrams reveal multi-factorial control over trail-induced apoptosis. *Mol Syst Biol*, 7:553, 2011.

[8] B. B. Aldridge, J. Saez-Rodriguez, J. L. Muhlich, P. K. Sorger, and D. A. Lauffenburger. Fuzzy logic analysis of kinase pathway crosstalk in tnf/egf/insulin-induced signaling. *PLoS computational biology*, 5(4):e1000340, 2009.

[9] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids.* Clarendon Press ;Oxford University Press, Oxford [England] :New York, 1987.

[10] A. R. Anderson and V. Quaranta. Integrative mathematical oncology. *Nature Reviews Cancer*, 8(3):227–234, 2008.

[11] E. Z. Bagci, Y. Vodovotz, T. R. Billiar, G. B. Ermentrout, and I. Bahar. Bistability in apoptosis: roles of bax, bcl-2, and mitochondrial permeability transition pores. *Biophysical journal*, 90(5):1546–59, 2006.

[12] S. Bandara, J. P. Schloder, R. Eils, H. G. Bock, and T. Meyer. Optimal experimental design for parameter estimation of a cell signaling model. *PLoS Comput Biol*, 5(11):e1000558, 2009.

[13] D. Battogtokh, D. Asch, M. Case, J. Arnold, and H. Schuttler. An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of neurospora crassa. *Proceedings of the National Academy of Sciences*, 99(26):16904, 2002.

[14] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 2nd edition, 1999.

[15] L. P. Billen, C. L. Kokoski, J. F. Lovell, B. Leber, and D. W. Andrews. Bcl-xl inhibits membrane permeabilization by competing with bax. *PLoS biology*, 6(6):e147, 2008.

[16] A. H. Briggs. A bayesian approach to stochastic cost-effectiveness analysis. an illustration and application to blood pressure control in type 2 diabetes. *Int J Technol Assess Health Care*, 17(1):69–82, 2001.

[17] L. J. Broadbelt and J. Pfaendtner. Lexicography of kinetic modeling of complex reaction networks. *AIChE journal*, 51(8):2112–2121, 2005.

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

[19] K. S. Brown, C. C. Hill, G. A. Calero, C. R. Myers, K. H. Lee, J. P. Sethna, and R. A. Cerione. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys Biol*, 1(3-4):184–95, 2004.

[20] K. S. Brown, C. C. Hill, G. A. Calero, C. R. Myers, K. H. Lee, J. P. Sethna, and R. A. Cerione. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Physical Biology*, 1(3):184, 2004.

[21] K. S. Brown and J. P. Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Phys Rev E Stat Nonlin Soft Matter Phys*, 68(2 Pt 1):021904, 2003.

[22] B. Calderhead and M. Girolami. Estimating bayes factor via thermodynamic integration and population mcmc. *Computational Statistics and Data Analysis*, 53:4028–4045, 2009.

[23] C. Chen, J. Cui, H. Lu, R. Wang, S. Zhang, and P. Shen. Modeling of the role of a bax-activation switch in the mitochondrial apoptosis decision. *Biophysical journal*, 92(12):4304–15, 2007.

[24] C. Chen, J. Cui, W. Zhang, and P. Shen. Robustness analysis identifies the plausible model of the bcl-2 apoptotic switch. *FEBS Lett*, 581(26):5143–50, 2007.

[25] T. Chen, H. L. He, G. M. Church, et al. Modeling gene expression with differential equations. In *Pacific symposium on biocomputing*, volume 4, page 4, 1999.

[26] W. W. Chen, M. Niepel, and P. K. Sorger. Classic and contemporary approaches to modeling biochemical reactions. *Genes development*, 24(17):1861–75, 2010.

[27] W. W. Chen, B. Schoeberl, P. J. Jasper, M. Niepel, U. B. Nielsen, D. A. Lauffenburger, and P. K. Sorger. Input-output behavior of erbb signaling pathways as revealed by a mass action model trained against dynamic data. *Molecular systems biology*, 5:239, 2009.

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

[29] J. E. Chipuk, T. Moldoveanu, F. Llambi, M. J. Parsons, and D. R. Green. The bcl-2 family reunion. *Mol Cell*, 37(3):299–310, 2010.

[30] M. Christie, J. Glimm, J. Grove, D. Higdon, D. Sharp, and M. Wood-Schultz. Error analysis and simulations of complex phenomena. *Los Alamos Science*, 29(6), 2005.

[31] G. Cohen. Caspases: the executioners of apoptosis. *Biochemical Journal*, 326(Pt 1):1, 1997.

[32] J. Cui, C. Chen, H. Lu, T. Sun, and P. Shen. Two independent positive feedbacks and bistability in the bcl-2 apoptotic switch. *PLoS One*, 3(1):e1469, 2008.

[33] B. C. Daniels, Y. J. Chen, J. P. Sethna, R. N. Gutenkunst, and C. R. Myers. Sloppiness, robustness, and evolvability in systems biology. *Current opinion in biotechnology*, 19(4):389–95, 2008.

[34] V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine. Rule-based modeling of cellular signaling concurrency theory. *Lecture Notes in Computer Science*, 4703:17–41, 2007.

[35] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag. Modular cell biology: retroactivity and insulation. *Molecular systems biology*, 4(1), 2008.

[36] M. A. Duran and B. S. White. Bayesian estimation applied to effective heat transfer coefficients in a packed bed. *Chemical Engineering Science*, 50(3):495–510, 1995.

[37] P. Flaherty, M. Radhakrishnan, T. Dinh, R. Rebres, T. Roach, M. Jordan, and A. Arkin. A dual receptor crosstalk model of g-protein-coupled signal transduction. *PLoS Comput Biol*, 4(9):e1000185, 2008.

[38] S. L. Frederiksen, K. W. Jacobsen, K. S. Brown, and J. P. Sethna. Bayesian ensemble approach to error estimation of interatomic potentials. *Phys Rev Lett*, 93(16):165501, 2004.

[39] N. Friel and A. N. Pettitt. Marginal likelihood estimation via power posteriors. *Journal of Royal Statical Society*, 70(3):589–607, 2008.

[40] P. Fuentes-Prior and G. S. Salvesen. The protein structures that shape caspase activity, specificity, activation and inhibition. *Biochemical Journal*, 384(Pt 2):201, 2004.

[41] M. Fussenegger, J. E. Bailey, and J. Varner. A mathematical model of caspase function in apoptosis. *Nature biotechnology*, 18(7):768–74, 2000.

[42] S. Gaudet, S. L. Spencer, W. W. Chen, and P. K. Sorger. Exploring the contextual sensitivity of factors that determine cell-to-cell variability in receptor-mediated apoptosis. *PLoS Comput Biol*, 8(4):e1002482, 2012.

[43] A. Gelman. *Bayesian data analysis*. Chapman Hall/CRC, Boca Raton, Fla., 2nd edition, 2004.

[44] A. Gelman and X.-L. Meng. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, 13(2):163–185, 1998.

[45] A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient metropolis jumping rules. *Bayesian Statistics*, 5:599–607, 1996.

[46] A. Gelman and D. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

[47] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman Hall, London, 1996.

[48] D. T. Gillespie. Exact stochastic simulation of coupled chemical-reactions. *Journal of Physical Chemistry*, 81(25):22, 1977.

[49] F. Gonzalvez and A. Ashkenazi. New insights into apoptosis signaling by apo2l/trail. *Oncogene*, 29(34):4752–65, 2010.

[50] D. R. Green. Apoptotic pathways: paper wraps stone blunts scissors. *Cell*, 102(1):1, 2000.

[51] D. R. Green, G. I. Evan, et al. A matter of life and death. *Cancer cell*, 1(1):19–30, 2002.

[52] S. Gupta, S. S. Bisht, R. Kukreti, S. Jain, and S. K. Brahmachari. Boolean network analysis of a neurotransmitter signaling pathway. *Journal of theoretical biology*, 244(3):463–469, 2007.

[53] R. Gutenkunst, J. Waterfall, F. Casey, K. Brown, C. Myers, and J. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*, 3(10):1871–1878, 2007.

[54] A. Hilfinger and J. Paulsson. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci U S A*, 108(29):12167–72, 2011.

[55] K. L. Ho and H. A. Harrington. Bistability in apoptosis by receptor clustering. *PLoS computational biology*, 6(10):e1000956, 2010.

[56] F. Hua, M. G. Cornejo, M. H. Cardone, C. L. Stokes, and D. A. Lauffenburger. Effects of bcl-2 levels on fas signaling-induced caspase-3 activation: molecular genetic tests of computational model predictions. *J Immunol*, 175(2):985–995, 2005.

[57] S. Huang and D. E. Ingber. Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Experimental cell research*, 261(1):91–103, 2000.

[58] Z. Huang and J. Hahn. Fuzzy modeling of signal transduction networks. *Chemical Engineering Science*, 64(9):2044–2056, 2009.

[59] I. A. Hubner, E. J. Deeds, and E. I. Shakhnovich. High-resolution protein folding with a transferable potential. *Proceedings of the National Academy of Sciences of the United States of America*, 102(52):18914–18919, 2005.

[60] N. T. Ingolia. Topology and robustness in the drosophila segment polarity network. *Plos Biology*, 2(6):805–815, 2004.

[61] K. Jaqaman and G. Danuser. Linking data to models: data regression. *Nat Rev Mol Cell Biol*, 7(11):813–9, 2006.

[62] F. J. M. Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46:68–78, 1951.

[63] R. E. Kass and A. E. Raftery. Bayes factors and model uncertainty. Report, University of Washington, March 1993 1993.

[64] R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

[65] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of theoretical biology*, 22(3):437–467, 1969.

[66] M. Kaufman, F. Andris, and O. Leo. A logical analysis of t cell activation and anergy. *Proceedings of the National Academy of Sciences*, 96(7):3894–3899, 1999.

[67] S. H. Kaufmann and W. C. Earnshaw. Induction of apoptosis by cancer chemotherapy. *Exp Cell Res*, 256(1):42–9, 2000.

[68] J. D. Kearns and A. Hoffmann. Integrating computational and biochemical studies to explore mechanisms in nf-kappab signaling. *J Biol Chem*, 284(9):5439–43, 2009.

[69] H. A. Kestler, C. Wawra, B. Kracher, and M. Kühl. Network modeling of signal transduction: establishing the global view. *Bioessays*, 30(11-12):1110–1125, 2008.

[70] L. B. Kleiman, T. Maiwald, H. Conzelmann, D. A. Lauffenburger, and P. K. Sorger. Rapid phospho-turnover by receptor tyrosine kinases impacts downstream signaling and drug binding. *Mol Cell*, 43(5):723–37, 2011.

[71] D. Klinke. An empirical bayesian approach for model-based inference of cellular signaling networks. *BMC bioinformatics*, 10(1):371, 2009.

[72] L. Kuepfer, M. Peter, U. Sauer, and J. Stelling. Ensemble modeling for analysis of cell signaling dynamics. *Nature biotechnology*, 25(9):1001–1006, 2007.

[73] M. Laakso et al. Data integration methods to interpret genome-scale data from cancers. 2012.

[74] N. Lartillot and H. Philippe. Computing bayes factors using thermodynamic integration. *Systematic Biology*, 55(2):195–207, 2006.

[75] I. N. Lavrik, R. Eils, N. Fricker, C. Pforr, and P. H. Krammer. Understanding apoptosis by systems biology approaches. *Mol Biosyst*, 5(10):1105–11, 2009.

[76] B. Leber, J. Lin, and D. W. Andrews. Still embedded together binding to membranes regulates bcl-2 protein interactions. *Oncogene*, 29(38):5221–30, 2010.

[77] G. Lillacci and M. Khammash. Parameter estimation and model selection in computational biology. *PLoS Comput Biol*, 6(3):e1000696, 2010.

[78] F. Llambi, T. Moldoveanu, S. W. Tait, L. Bouchier-Hayes, J. Temirov, L. L. McCormick, C. P. Dillon, and D. R. Green. A unified model of mammalian bcl-2 protein family interactions at the mitochondria. *Molecular cell*, 44(4):517–531, 2011.

[79] C. F. Lopez, J. L. Muhlich, J. A. Bachman, and P. K. Sorger. Programming biological models in python using pysb. *Molecular systems biology*, 9(1), 2013.

[80] J. F. Lovell, L. P. Billen, S. Bindner, A. Shamas-Din, C. Fradin, B. Leber, and D. W. Andrews. Membrane binding by tbid initiates an ordered series of events culminating in membrane permeabilization by bax. *Cell*, 135(6):1074–84, 2008.

[81] D. J. C. MacKay. *Information theory, inference, and learning algorithms.* Cambridge University Press, Cambridge, UK ; New York, reprinted with corrections. edition.

[82] Z. Mai and H. Liu. Boolean network-based analysis of the apoptosis network: Irreversible apoptosis and stable surviving. *Journal of theoretical biology*, 259(4):760–769, 2009.

[83] P. Mendes and D. Kell. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14(10):869, 1998.

[84] M. Middendorf, E. Ziv, and C. H. Wiggins. Inferring network mechanisms: the drosophila melanogaster protein interaction network. *Proceedings of the National Academy of Sciences of the United States of America*, 102(9):3192–3197, 2005.

[85] M. K. Morris, J. Saez-Rodriguez, P. K. Sorger, and D. A. Lauffenburger. Logic-based models for the analysis of cell signaling networks. *Biochemistry*, 49(15):3216–3224, 2010.

[86] J. M. Murphy, D. M. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430(7001):768–772, 2004.

[87] G. Nelson, L. Paraoan, D. G. Spiller, G. J. C. Wilde, M. A. Browne, P. K. Djali, J. F. Unitt, E. Sullivan, E. Floettmann, and M. R. H. White. Multi-parameter analysis of the kinetics of nf-œb signalling and transcription in single living cells. *Journal of Cell Science*, 115(6):1137–1148, 2002.

[88] L. Neumann, C. Pforr, J. Beaudouin, A. Pappa, N. Fricker, P. H. Krammer, I. N. Lavrik, and R. Eils. Dynamics within the cd95 death-inducing signaling complex decide life and death of cells. *Mol Syst Biol*, 6:352, 2010.

[89] S. H. Northrup and H. P. Erickson. Kinetics of protein-protein association explained by brownian dynamics computer simulation. *Proc Natl Acad Sci U S A*, 89(8):3338–42, 1992.

[90] E. Pieroni, S. de la Fuente van Bentem, G. Mancosu, E. Capobianco, H. Hirt, and A. de la Fuente. Protein networking: insights into global functional organization of proteomes. *Proteomics*, 8(4):799–816, 2008.

[91] W. H. Press. *Numerical recipes in C : the art of scientific computing.* Cambridge University Press, Cambridge ; New York, 2nd edition, 1995. MITb10741026 (OCoLC)33066174 William H. Press ... [et al.]. ill. ; 25 cm. "First edition originally published 1988; second edition originally published 1992. This reprinting is corrected to software version 2.06"–T.p. verso. Includes bibliographical references (p. 926-929) and index.

[92] M. Rehm, H. Dussmann, and J. H. Prehn. Real-time single cell analysis of smac/diablo release during apoptosis. *The Journal of cell biology*, 162(6):1031–43, 2003.

[93] M. Rehm, H. J. Huber, H. Dussmann, and J. H. Prehn. Systems analysis of effector caspase activation and its control by x-linked inhibitor of apoptosis protein. *The EMBO journal*, 25(18):4338–49, 2006.

[94] M. Rehm, H. J. Huber, C. T. Hellwig, S. Anguissola, H. Dussmann, and J. H. Prehn. Dynamics of outer mitochondrial membrane permeabilization during apoptosis. *Cell death and differentiation*, 16(4):613–23, 2009.

[95] W. A. Reiners, S. Liu, K. G. Gerow, M. Keller, and D. S. Schimel. Historical and future land use effects on n2o and no emissions using an ensemble modeling approach: Costa rica's caribbean lowlands as an example. *Global Biogeochemical Cycles*, 16(4):1068, 2002.

[96] D. S. Reis Jr. and J. R. Stedinger. Bayesian mcmc flood frequency analysis with historical information. *Journal of Hydrology*, 313:97–116, 2005.

[97] D. R. Rhodes, S. A. Tomlins, S. Varambally, V. Mahavisno, T. Barrette, S. Kalyana-Sundaram, D. Ghosh, A. Pandey, and A. M. Chinnaiyan. Probabilistic model of the human protein-protein interaction network. *Nature biotechnology*, 23(8):951–959, 2005.

[98] H. H. Robertson. *The solution of a set of reaction rate equations*, volume IV.1, IV.10, pages 178–182. Academ. Press, 1966.

[99] M. Rodriguez-Fernandez, J. A. Egea, and J. R. Banga. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics*, 7:483, 2006.

[100] P. J. Roebber, D. M. Schultz, B. A. Colle, and D. J. Stensrud. Toward improved prediction: High-resolution and ensemble modeling systems in operations. *Weather and Forecasting*, 19(5):936–949, 2004.

[101] J. Saez-Rodriguez, L. G. Alexopoulos, J. Epperlein, R. Samaga, D. A. Lauffenburger, S. Klamt, and P. K. Sorger. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Molecular systems biology*, 5(1), 2009.

[102] C. Scaffidi, S. Fulda, A. Srinivasan, C. Friesen, F. Li, K. J. Tomaselli, K.-M. Debatin, P. H. Krammer, and M. E. Peter. Two cd95 (apo-1/fas) signaling pathways. *The EMBO Journal*, 17(6):1675–1687, 1998.

[103] S. Schuster, M. Marhl, and T. Hofer. Modelling of simple and complex calcium oscillations. from single-cell responses to intercellular signalling. *Eur J Biochem*, 269(5):1333–55, 2002.

[104] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

[105] A. Shamas-Din, J. Kale, B. Leber, and D. W. Andrews. Mechanisms of action of bcl-2 family proteins. *Cold Spring Harbor perspectives in biology*, 5(4), 2013.

[106] A. B. Singer, J. W. Taylor, P. I. Barton, and W. H. Green. Global dynamic optimization for parameter estimation in chemical kinetics. *The Journal of Physical Chemistry A*, 110(3):971–976, 2006.

[107] T. Smith, A. Ross, N. Maire, N. Chitnis, A. Studer, D. Hardy, A. Brooks, M. Penny, and M. Tanner. Ensemble modeling of the likely public health impact of a pre-erythrocytic malaria vaccine. *PLoS medicine*, 9(1):e1001157, 2012.

[108] S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, and P. K. Sorger. Non-genetic origins of cell-to-cell variability in trail-induced apoptosis. *Nature*, 459(7245):428–32, 2009.

[109] S. L. Spencer and P. K. Sorger. Measuring and modeling apoptosis in single cells. *Cell*, 144(6):926–39, 2011.

[110] L. M. Tran, M. L. Rizk, and J. C. Liao. Ensemble modeling of metabolic networks. *Biophysical journal*, 95(12):5606–5617, 2008.

[111] D. L. Vaux and S. J. Korsmeyer. Cell death in development. *Cell*, 96(2):245–254, 1999.

[112] G. Von Dassow, E. Meir, E. M. Munro, and G. M. Odell. The segment polarity network is a robust developmental module. *Nature*, 406(6792):188–192, 2000.

[113] J. Wang, L. Zheng, A. Lobito, F. Chan, J. Dale, M. Sneller, X. Yao, J. M. Puck, S. E. Straus, M. J. Lenardo, et al. Inherited human caspase 10 mutations underlie defective lymphocyte and dendritic cell apoptosis in autoimmune lymphoproliferative syndrome type ii. *Cell*, 98(1):47, 1999.

[114] L. J. White, N. D. Evans, T. J. Lam, Y. H. Schukken, G. F. Medley, K. R. Godfrey, and M. J. Chappell. The structural identifiability and parameter estimation of a multispecies model for the transmission of mastitis in dairy cows with postmilking teat disinfection. *Math Biosci*, 180:275–91, 2002.

[115] M. B. Yaffe. Signaling networks and mathematics. *Science Signalling*, 1(43):eg7, 2008.

[116] R. J. Zheng, Q. Comparison of deterministic and stochastic kinetics for nonlinear systems. *J Chem Phys*, 94(5):3644–3648, 1991.