Selecting High-Confidence Predictions from Ordinary Differential Equation Models of Biological Networks

by

Caitlin Anne Bever

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

at the
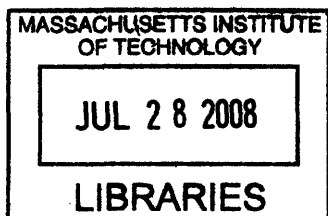
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

[June 2008]

May 2008

Author . . . .
Department of Biological Engineering
May 23, 2008
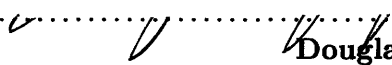
Certified by . . . .
Bruce Tidor
Professor of Biological Engineering and Computer Science
Thesis Supervisor

Accepted by . . . . / . . . // . . . . // . . . / . . . .
Alan Grodzinsky
Chairman, Department Committee on Graduate Students
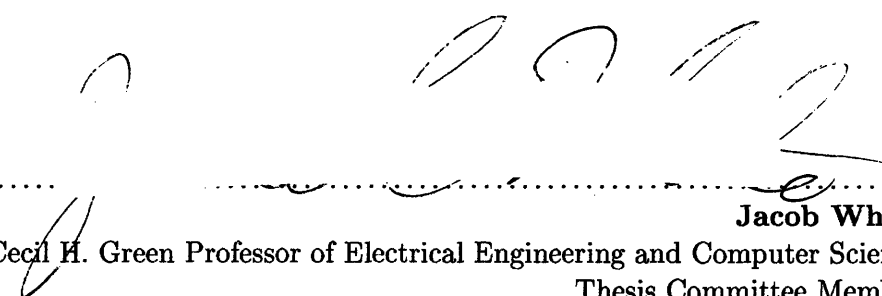
Thesis committee

Accepted by . . . .
Douglas Lauffenburger
Whitaker Professor of Biological Engineering, Chemical Engineering, and Biology
Chair of Thesis Committee

Accepted by . . . .
Bruce Tidor
Professor of Biological Engineering and Computer Science
Thesis Supervisor

Accepted by . . . .
Jacob White
Cecil H. Green Professor of Electrical Engineering and Computer Science
Thesis Committee Member

Accepted by . . . .
Paul Barton
Lammot du Pont Professor of Chemical Engineering
Thesis Committee Member

Selecting High-Confidence Predictions from Ordinary Differential Equation Models of Biological Networks

by

Caitlin Anne Bever

Submitted to the Department of Biological Engineering
on May 23, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biological Engineering

Abstract

Many cellular processes are governed by large and highly-complex networks of chemical interactions and are therefore difficult to intuit. Computational modeling provides a means of encapsulating information about these interactions and can serve as a platform for gaining understanding of the biology and making predictions about cellular response to perturbation. In particular, there has been considerable interest in ordinary differential equation (ODE) models, which have several attractive features: ODEs can describe molecular interactions with mechanistic detail, it is relatively straightforward to implement perturbations, and, in theory, they can predict the concentration and activity of every species as a function of time. However, both the topology and parameters in such models are subject to considerable uncertainty. We explore the ramifications of these sources of uncertainty for making accurate predictions and develop methods of selecting high confidence predictions from uncertain models. In particular, we promote a shift in emphasis from model selection to prediction selection, and use consensus among model ensembles to identify the predictions most likely to be accurate. By constructing decision trees, this consensus can also be used to partition the space of potential perturbations into regions of high and low confidence. We apply our methods to the Fas signaling pathway in apoptosis to satisfy two goals: first, to design a therapeutic cocktail to reduce cell death in the presence of high levels of stimulus, and second, to design experiments that may lead to a better understanding of the biological network.

Thesis Supervisor: Bruce Tidor
Title: Professor of Biological Engineering and Computer Science

Acknowledgments

I consider myself extraordinarily fortunate to have been a student at MIT, and, in particular, a member of the Biological Engineering Department. I have been exposed to so many diverse opportunities—both academic and extra-curricular—and so many amazing people, many of whom I am honoured to now call my friends. I only have great things to say about my department, so adeptly run by Doug Lauffenburger, that fosters a real sense of community amongst its members.

I am grateful to my advisor, Bruce Tidor, in general for his time and effort, but also for a number of specific things. For one, I admire his devotion to scientific integrity and his commitment to instilling the same values in his students. I also learned from Bruce not to get lost in the details, to always think of the day-to-day problems as pieces of a much larger puzzle. In addition, Bruce places a lot of emphasis on scientific communication and I benefited greatly from our weekly group meetings (once I conquered my initial fear). I would also like to thank the members of my thesis committee, Doug Lauffenburger, Jacob White, and Paul Barton, for their helpful comments along the way.

I feel so lucky to have been surrounded by such outstanding peers in my classes and in the Tidor Lab. I am sure that every member of the lab has contributed in some way to my thesis, either through insight, technical assistance, or encouragement. I am especially thankful to Josh Apgar and Katharina Wilkins for their regular and generous contributions to my work and my morale. I would also like to acknowledge Philip Kim and Bambang Adiwijaya for being the Systems Biology pioneers in the lab and paving the way for me to follow suit, and Aurore Zyto, for joining the lab simply because I asked nicely.

My family supported my interest in science from early on and provided me with all the resources I ever needed to reach my goals. Thanks, Mom, for being the sounding board for my successes and failures for all these years. Thanks, Dad, for teaching me algebra on napkins at the Dairy Queen. And thanks, Alea and Braden, for not disowning me, your absentee sister. I miss you!

I have also been blessed with wonderful roommates, Ale, Rebecca, and Heather, who made our home such a happy place to come to at the end of a long day. Ale very unselfishly offered his time and experimental expertise to help me untangle the Fas signaling network, and I am so grateful to Rebecca and Heather for feeding me when I was unable to accomplish that seemingly trivial task on my own.

The MIT Cycling Team provided not only a refuge from work, but innumerable fond memories and lifelong friendships. I am immensely proud of what the team has become from humble beginnings six years ago, a transformation that grew out of the dedication and love from so many. On that note, I thank Jason, my oldest and most faithful teammate, for being inspiring in his quiet and unassuming way, and for his unwavering confidence that we, and I, would find victory in the end.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Systems biology is a field that aims to study the function of the cell as a unit, rather than strictly as a sum of its parts [1,2], as is typical in traditional, reductionist, biology. While the general systems approach is not new, having found applications for many years in fields such as operations research and chemical engineering, it took several technological innovations for biologists to follow suit. These included genome sequencing methodologies (culminating in the success of the Human Genome Project [3,4]) along with several "high-throughput" measurement techniques that allow scientists to monitor the state of a cell in great detail: microarrays perform simultaneous genome-wide measurements of gene expression levels [5]; yeast-two hybrid screens [6] and protein arrays generate information about protein–protein interactions [7–9]; and improvements in mass spectrometry allow the quantification of protein concentrations and activation levels (for example, detection of phosphorylation and ubiquitination) [10,11]. At about the same time, there were complementary advancements in the ability to perform broad spectrum perturbation experiments: gene insertion and deletion technology improved [12] and RNA-mediated interference (RNAi) was discovered as a means of post-transcriptional control [13,14]. Taken together, the turn of the century saw a revolution in our ability to comprehensively measure a cell's response to stimulus.

Analysis methods lagged behind experimental progress, however, and we are still learning how best to take advantage of the massive amount of information being collected. Although it is possible to build on the work from other systems-oriented fields, systems biologists are presented with many new challenges unique to the study of life. In the human cell, for instance, there are 20,000-25,000 genes [15], each coding for a protein, many of

which have multiple splice variants [16] or states of activation [17]. In addition, cells contain a number of other small molecules that are neither nucleic acid nor protein, and each species can interact with a subset of all the cellular components. Collectively referred to as the interactome [7], the network of all interactions encapsulates the incredible complexity of the cell [18].

Noise is another issue that adds to our difficulty in understanding biology. The specific contents of each cell are unique but many biological measurement techniques are applied to populations of cells; it is important to recognize that the average dynamics of the group can hide the true behavior of the individual [19]. Even with the ability to measure properties of single cells (using flow cytometry, for example [20]), a cell's response to stimulus is not strictly a deterministic function of its state at the time of input. Heterogeneities in the spatial distribution of species and the stochasticity associated with chemical reactions among rare molecules combine to make cells intrinsically noisy entities [21–23]. Experimental noise adds yet another source of uncertainty [24, 25].

The long-term goal in systems biology is to understand how this large and complex network leads to observed cellular behavior. How does a cell process the information it receives from its environment and make decisions accordingly? How are cells able to transmit signals to each other? In a developing organism, how do cell types become differentiated: how does a cell know that it is a hepatocyte rather than a neuron? Once we achieve this forward level of understanding, it is hoped that we will be able to work backwards, to learn how to manipulate a cell or its environment to change its behavior. There are powerful clinical implications for such a high level of understanding and there is considerable interest in systems biology from the perspective of drug development [26–29].

Due to the size and complexity of the cell, the path linking noisy data to biological understanding and therapeutic utility often comes in the form of computational models. Whereas logic and human reasoning can come to terms with the results of a small number of experiments, it is impossible to intuit cellular function from thousands of measurements taken from distant nodes in the network. Algorithms and mathematics then step in to assimilate and analyze where we cannot.

1.1 Modeling approaches in systems biology

Modeling philosophies and approaches are diverse, in part because there are many ways of dealing with the hierarchical nature of the system and the different sources of complexity. Is it necessary, for instance, to consider both the mRNA and protein products of a particular gene or can the whole story be told by one or the other? Is it necessary to model the entire biological network to understand the function of one specific pathway and, if not, where do we draw the line of inclusion for relevant species? The nature of our models also depends highly on the detail of the predictions we hope for them to make. A phenomenological conclusion might be that a cancer cell will or will not die in response to drug treatment; alternatively, we might want to know how the concentration of a particular protein changes as a function of time following treatment. In broad terms, a modeler must decide how representative their model will be of the true underlying physics. Depending on the nature of the system, various details may be unimportant for capturing its behavior, just as it is frequently reliable to apply Newton's Laws without accounting for the quantum mechanics of the bodies involved. The difficulty in biology is that, unlike in physics, no one has been able to derive whether such limits of scale exist and whether the hierarchies of detail can be disentangled from each other.

Published models of biology range from the purely statistical to highly detailed and mechanistic [30] (Figure 1-1). At the least detailed end of the spectrum, statistical models are built upon correlations between components in the system and have no mechanistic underpinnings. Data mining techniques such as hierarchical clustering [31] or support vector machines [32, 33] have often been applied to microarray data to group together similarly expressed genes. If these gene groups share common function, for example, the clustering may provide the basis for a functional assignment rule. Correlation does not imply causality and statistical models as such cannot make statements as to the functional relationship between components.

Building in complexity, graph theoretic approaches such as Bayes nets [34, 35] generate a probabilistic description of the influence of each species on each other. With this type of model, conclusions can be drawn about the impact of removing a node in the network so

17
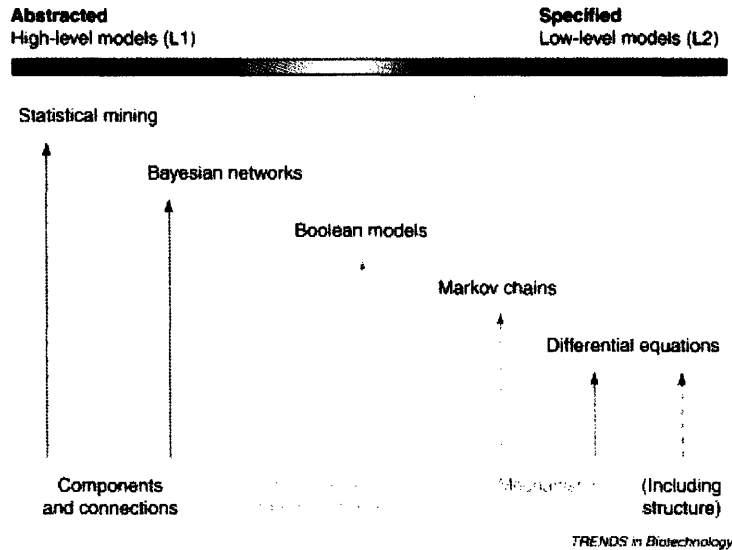
Figure 1-1: The scales of modeling in systems biology and the corresponding mathematical approaches. From Ideker and Lauffenburger, 2003 [30].

that some level of prediction about biological modification is possible. One shortcoming of Bayesian networks is their acyclic nature: if node A influences node B, it is not allowed for B to in turn influence A. As feedback loops are thought to be a key feature of biological networks(e.g. [36–38]), this is not a trivial issue. Nevertheless, Bayesian networks may serve as an excellent starting point on which to build more realistic models.

Differential equation models lie at the detailed end of the model spectrum, although among themselves they can capture very different levels of complexity. Most differential equation models are based on kinetic descriptions of gene–protein and protein–protein interactions, and are mechanistic as such. Models of this type are defined by their topology, which describes which species interact with which others and how, as well as their parameters, which specify the rates of reactions and processes in which these species participate. Ordinary differential equations are limited to systems of time-derivatives only and are deterministic given precise specification of initial conditions [39]. Partial differential equations are able to explicitly model dimensions of space in addition to that of time. Stochastic differential equations [40] aim to incorporate the effects of innate noise on top of time and space dynamics.

Thanks to their mechanistic underpinnings, differential equation models have great potential for extrapolating system behavior away from the training data. It is possible to

frame questions about changes to rate constants, protein concentrations, and network connectivity in terms of the mathematics contained within the equations. In terms of drug development, this allows scientists to explore the molecular mechanism of disease and devise relevant intervention strategies. With the increased power of extrapolation, however, comes an increased need for detail in specifying the models. Each distinct mechanism is accompanied by parameters, all of which must be specified either through real biological experimentation or through fitting of the model to data. Therefore, a model that includes many mechanisms will have many parameters and will likely require many data points for proper specification.

1.2 The traditional way to build (and use) differential equation models of biology

While different researchers may not use exactly the same specific techniques to building differential equation models, there is a typical iterative approach [1,2,41,42] that is outlined here.

1.2.1 Problem Definition

The process begins with the specification of the problem at hand and the identification of the pathway of interest, usually based on prior knowledge about the key elements associated with a disease or biological process. This stage of model building requires many choices to be made so that the model best takes advantage of the current state of knowledge without (a) becoming computationally unwieldy (though this is increasingly less of a concern), or (b) attempting to extract more information than the existing data provide. (Unrealistic ambition in model construction leads to "over-fitting" and reduces a model's predictive power.) In network terms, the problem definition stage is where primary decisions are made regarding the topology of the system.

Many choices are made about the number of components to consider and how many interactions to include. If two molecules are thought to have redundant functionality, it may or may not be necessary to include both individual species—rather, in some cases, one

lumped species may be used to represent both. If the interaction between two molecules is thought to be weak, it may be reasonable to omit the interaction from the model, depending on where the reaction occurs in the network. If an important part of the pathway lies downstream of the interaction, its inclusion may be critical for the model's utility, despite seemingly insignificant kinetics.

Many proteins undergo modifications in the cytoplasm, and different forms of the same protein may react with its chemical partners at different rates. Each chemical form can be thought of as a separate species acting independently of its alternative states. Doing so may come at a high computational cost that is not necessarily compensated for by increased biological insight. In the yeast pheromone pathway, for instance, Ste5 acts as a scaffold, facilitating activation between its multiple binding partners, the family of mitogen-activated protein kinases (MAPK, MAPKK, and MAPKKK), through localization. Because scaffold proteins have multiple docking sites, they exist in a combinatorial number of states, e.g. bound to MAPK only, bound to protein MAPK and MAPKK, bound to protein MAPK and MAPKKK, etc. Accounting for each state as an individual species can thus have an explosive effect on the size of any model to which it belongs. Clever mathematical tricks, however, have shown that the model does not have to explicitly account for every species to obtain accurate predictions about system behavior [43]. Another example of multiplicative species is the many phosphoylation states available for some proteins. Several of the proteins implicated in the epidermal growth factor receptor (EGFR) pathway belong to this class, and Wolf-Yadlin et al. [44] showed that different phosphorylation states of HER2 display different dynamics. It is also well-known that some proteins exhibit very different enzymatic activity depending on their level of phosphorylation [45]. Attempts to incorporate every unique phosphorylation state in a model, however, result in exponential increases in complexity. It is a difficult balancing act to incorporate enough detail while maintaining model usability.

Other choices about the depth of complexity relate to the mathematical detail used to describe the molecular interactions. The researcher must decide whether simplifying assumptions are to be made with regards to the reactions, such as whether to describe enzymatic activation using full mass-action kinetics or with the Michaelis–Menten approx-

imation [46,47]. The Michaelis–Menten approximation assumes that the rate at which an enzyme binds to its substrate is much faster than the rate of product formation, and that the concentration of the reaction intermediate is therefore at steady state. For a single enzymatic reaction, Michaelis–Menten kinetics require the specification of two rate constants, K_m and k_{cat}; in comparison, the equivalent mass-action reactions require three parameters. When incorporated into a model with 100 reactions, the savings in terms of parameter specification acquire real relevance. Furthermore, experimental measurements of enzymatic reactions often report K_m and k_{cat}, making it easier for these results to be incorporated into any model. If the assumption of fast binding is not correct, the application of Michaelis–Menten kinetics can lead to misleading dynamics for not only the species involved in that reaction, but for all species downstream.

1.2.2 Model Optimization and Selection

Models of biological networks are rife with uncertainties: uncertainties about the values of the parameters, about the scope and detail they include, as discussed above, and about the underlying biology itself. More often than not, standard practice in systems biology is to work with a single model at a time, defined by a unique topology and parameter set. Ideally, this model is chosen so as to maximize information while minimizing uncertainty, though in practice it is difficult to judge when this criterion is met, and other factors, such as computational expense, also weigh in. While many choices in model definition are *ad hoc*, it is sometimes difficult to eliminate candidate models, in which case a more rigorous model selection procedure is applied.

Traditional methods for model selection are borrowed from standard statistical practice and may be linked to the problem of parameter estimation [48]. While some model parameters may be known from prior experimentation, it is common for other parameters to be fit so as to optimize the relationship between the model and the data. The maximum likelihood estimate (MLE) of the parameters is the set of values that maximizes the probability of getting the data given the model. If the noise is assumed to be Gaussian, the MLE corresponds to the parameter that minimizes the sum of squares difference between the model and the data. In a Bayesian framework, the parameters are chosen to maximize

the posterior—the probability of the model given the data—and it is therefore referred to as the maximum a posteriori (MAP) estimate. In practice, an optimization is performed on each topologically unique model with the chosen statistical metric as the objective function. There are many algorithms for performing parameter estimation, some of which are described and evaluated in [49] and [50].

For model selection, both the MLE and MAP are typically used in a static way: the data have already been acquired, (for better or worse), each candidate model is parametrized to generate the best possible fit to the data, and a model may be selected as the most favorable if it produces the best MLE or MAP score. More sophisticated, but still static, methods exist for incorporating model size and/or complexity and these include the information criteria (Akaike, Bayesian, and Takeuchi [51,52]), minimum description length (MDL) [53] and information complexity (ICOMP) [54]. An alternative static approach is that of cross-validation [55, 56] whereby the models are trained on subsets of the data and evaluated on their ability to reproduce the remaining, or testing, data. None of these methods were developed specifically for use with differential equations models.

Another approach to model discrimination is that of experimental design followed by further data collection [57–60]. The idea is to choose experiments that maximize the expected differences between model outcomes; then by virtue of this choice, not all models will be consistent with the new data and these disparate models can be eliminated from the pool of reasonable candidates. More recently, interactive algorithms that are specifically designed for selecting among multiple differential equation models have emerged in which data acquisition and model evaluation are coupled and combined in a single procedure [61]. The major drawback of experimental design for model selection is, of course, the need for more experiments, which may be both time-consuming and costly.

1.2.3 Model Analysis

Once a model has been developed, there are numerous modes of analysis for determining various system properties. The particular property of interest may depend on whether the study is more pedagogical or practical, whether it is being used to further our understanding of the way biology works or if it was constructed with extrapolation in mind. There is

inevitably some overlap between the two.

Many biological networks demonstrate robustness to changes in parameter values and some to changes in topology as well [62–71]. The effect of stochastic reactions on cellular dynamics is roughly analogous to changing the values of kinetic rate constants in a model. As cells are clearly functional under noise, it is expected that ODE models of biological networks should display a robustness to changes in parameters. Genetic permutation incurred under evolutionary pressures may also cause effective changes to system parameters and has the potential to affect topology as well. Small genetic modifications, such as the mutation of a single amino acid, might generate a difference in that protein's enzymatic activation strength. More significant alterations might lead to complete loss of function for that protein, and the removal of the corresponding node from the biological network model. Sensitivity analysis, and variants thereof, are commonly used to assess a model's robustness [72]. This provides a measure of the expected change to a system output given a small perturbation to the parameter of interest. Similar information can be acquired through Monte Carlo simulations in which the model parameters are sampled many times from a distribution and the ODE integrated for each sample [56].

Bistability is a phenomenon in which a system can achieve one of two possible steady states; the state achieved by the system depends on the initial concentrations of the species involved. There are many examples of bistable systems in biology, of cells having to choose between two fates. Apoptosis is a key example of this behavior, where a cell exposed to a stimulus must decide whether to die or to proceed on its regular schedule [73]. Models of biological processes known to be bistable, as in apoptosis, are analyzed to determine whether they can recapitulate the expected bistability and to assess which part of the network assumes responsibility for it [74–77]. Non-linear dynamics techniques to address bistability largely center around fixed point analysis [78]. The standard extension to this type of analysis is to determine whether the bistable system can be made into a monostable one by adjusting the parameters or topology.

It is of note that many of the methods of model analysis are used in reverse to aid in model development. If it is assumed that good models of biological networks are robust, then this property can factor into the model selection procedure [79].

1.2.4 Model Validation and Refinement

Models of biology are never exact representations of the true physics and all models are incomplete and incorrect and many ways. Because of this, no model will be able to accurately predict the system behavior under all circumstances and it is always possible for new experimental data to be incongruous with the currently used model. When a model can no longer agree with all of the available data, it must be modified to generate a better match. If, on the other hand, the model made a prediction, new data were acquired, and there was good agreement between the two, some consider this to be validation of the model. This is a somewhat dangerous position to take given that no model will be correct in all cases, but it does indicate the breadth of its domain of utility.

1.3 Intended impact of this work

The work presented here examines the way that differential equation models are currently applied to biological networks. Given that models in biology are always uncertain in many ways, we investigated whether it was still possible for them to be useful for gaining insight into the biology and for making accurate predictions about behavior in response to perturbation. Furthermore, we wanted to develop ways to evaluate predictions without the requirement that further experiments be performed.

Computational biology is taking an increasingly large role at pharmaceutical companies that are having difficulty both with keeping the drug pipeline fed with new candidate therapies and with having to remove drugs already in the pipeline for lack of effectiveness or unintended side-effects. It is hoped that mathematical models will lead to improved target identification, drug safety analysis, and clinical trial design. In each of these tasks, computational biology has the potential to serve as a financial risk assessor, were there a way to evaluate the confidence in model predictions. Time is also a critical factor in drug development, and it is unrealistic to spend years performing experiments to get one (still imperfect) model.

In this thesis we have addressed several issues relating to these problems and suggest a number of possible solutions. In Chapter 2, we examined the role of uncertainty in parame-

ters and topology using simple toy systems and simulation. We found that model selection can sometimes be an unreliable method for identifying the highest accuracy predictions from non-linear differential equation models. We further highlighted that quality of model fit is poorly correlated with quality of prediction when models are perturbed away from the training data. And, though uncertainty in parameters should be addressed when making predictions, methods for assessing this source of error are incapable of capturing the variation from uncertainties in topology. The distance in phase space from the training data to a prediction is a weak indicator of prediction accuracy but the consensus among an ensemble of well-fitting models is a strong classifier of prediction error.

In Chapter 3, we further developed the use of consensus as a means of identifying high and low confidence predictions. Our first application was to apply model averaging with the method of propagated statistical uncertainty to generate time-dependent probability distributions from predictions using model ensembles. The distributions were able to successfully capture the true variation in predictions that resulted from both parametric and topological sources. We developed a consensus score based on this method of model averaging as a classifier of the overall, time-integrated, validity of model predictions. In order to define regions of parameter space that we would expect to lead to high or low consensus, we subjected ensembles to Monte Carlo perturbation experiments and built decision trees using the resulting ensemble variance. We demonstrated how these trees can then be used to identify high-confidence, biologically-interesting perturbations, as well as perturbations for which the models of the ensemble disagree. The method therefore combines selection of good predictions with experimental design for refinement of the ensemble.

Chapter 4 describes how we used our ensemble methods on the Fas signaling network in apoptosis. This was a considerably larger system of study, model uncertainties represented real differences between existing published models, and models were fit to actual experimental data. We constructed an ensemble of 64 models by considering the combinatorial set of six independent additions to a base network. We used our ensemble to explore treatment options for sepsis in the form of cocktails of caspase inhibitors. We subjected our ensemble to perturbations representing different treatments and built a decision tree to classify the consensus of the ensemble under each condition. We found that with two irreversible cas-

pase inhibitors, one each for caspases-6 and -8, the models agree poorly on the treatment response and we therefore suggested that this experiment be performed to refine the ensemble. With a cocktail of irreversible inhibitors for caspases-3, -8, and -9, all models agree that cells will experience a significant reduction in caspase-3 activation, despite exposure to high levels of stimulus. We also performed a secondary analysis of the impact of flux balance on the model fits and predictions and recommend further experimentation to identify the relative contributions to caspase-3 activation by caspase-8 and -9.

Chapter 2

Uncertainty in ordinary differential equation models of biological networks: Implications for the assessment of predictions and a role for model ensembles

Abstract

Biological modeling places a strong emphasis on building a model that accurately reproduces existing data. One anticipated benefit of a highly accurate model is that it can make accurate predictions. Here we investigate other possibilities for identifying reliable predictions that may be applicable in the absence of sufficient experimental data to uniquely determine even the topology of the model. Using synthetic data and small genetic networks, we explore the use of statistical uncertainty propagated from model fitting to prediction as a metric of predictions accuracy, as well as various metrics relating the training data to the predicted trajectory. The former is problematic because statistical uncertainty is difficult to accurately assess in the absence of a correct topology, and the latter shows some promise but tends to identify perturbations that remain close to the training data. Interestingly, we find consensus among an ensemble of topologically diverse models to be an excellent indicator of high-confidence predictions.

2.1 Introduction

Computational modeling of biological systems provides a means to increase our understanding of complex cellular and higher-order processes, as well as to make predictions of behavior resulting from altered conditions or pharmaceutical interventions [27,28,62,80–83]. Cellular processes are regulated by a complex network of physical and chemical interactions taking place among thousands of different species, including proteins, nucleic acids, and small molecules. Past advances in biology have enabled detailed studies to elucidate the role of individual molecules and reactions; systems biology aims to expand the field of focus to simultaneously probe large numbers of interrelated components and processes. By integrating system-wide information one aims simultaneously for a functional understanding from the systems perspective and the ability to make quantitative predictions about cellular behavior in response to a wide variety of perturbations, across a wide range of conditions. Because of the inherent complexity, computational modeling is a necessary tool in accounting for the large amount of system detail.

In particular, an emphasis has been placed on ordinary differential equation models that describe the kinetics of chemical reactions in the cell [30]. Such models are attractive for their mechanistic nature and power of extrapolation. A sufficiently well-defined ODE model can accurately predict system behavior under very different conditions than those under which the model was trained. The power of prediction does not come for free, however, and greater mechanistic detail imposes greater requirements on system specification.

Differential equation models in biology are typically developed in an iterative fashion [30, 82, 84]. Typically, the topology (or connectivity) of the network is developed by searching the literature for known interactions or applying a reverse engineering algorithm to existing data [35, 66, 85–89]. A mathematical description of the network is written according to that topology and values of any unknown parameters are fit by minimizing the difference between the model and experimental data. The model is then mined for biological insight and key predictions are tested experimentally. Eventually, new experiments show that the model requires refinement, which might involved tweaking the topology and then refitting the parameters, wherein the process is repeated. Models of cellular processes are

therefore characterized by their topology and their parameters, both of which are subject to uncertainty.

Currently, the standard approach is to work with one model at a time, described uniquely by its topology and parameters. If there are multiple topologies under consideration for the network, there are several ways to proceed. Often, a single topology will be selected for *ad hoc* reasons related to preference, clarity, or computational feasibility. If no experiments are to be done before continuing further, a more sophisticated alternative is to choose a set of biologically reasonable candidate models , and then parametrize each through optimization. The objective function for the optimization—the sum of squares error between measured and computed trajectories, for instance—is evaluated at the minimum, and this value may form the basis for a statistical test to select the "best" model. These tests range from simple to complex and represent Bayesian and frequentist philophies [90, 91]. The chosen model is then used to make all predictions about the system's response to perturbation. Proceeding in this fashion amounts to implicitly assuming that the "best" model, as judged by a statistical test, will make the best predictions. It should be noted that many of the standard selection metrics were not designed for use with differential equation models.

If there is an opportunity for further data acquisition, experimental design procedures can be implemented for model discrimination, reducing the pool of reasonable candidates [58–61, 92]. This may not always be an option, as time and cost can influence decisions. In these cases, it is important to consider the extent to which a model is useful, given its uncertainties.

In this regard, often little attention is paid to the evaluation of predictions from ODE models, outside of validation through further experimentation. Several notable exceptions include the efforts by a number of groups to assess the uncertainty contribution from parameters [93–97]. Biology itself is stochastic [65] and biological measurement techniques provide additional noise to any data. When fitting a specific topology to such noisy data, there will necessarily be uncertainty in the parameters. (There is some evidence that, relating to robustness properties of biological networks, the parameters would be ill-defined even for noiseless data [70]. In this way, biological models may be unidentifiable.) The uncertainty in parameters at the fit can be propagated onto any post-perturbation prediction using the

Fisher information [93, 98] or the method of bootstrapping [56, 96].

Accounting for the impact of topology on prediction uncertainty has received even less attention in systems biology modeling. Other fields have, however, highlighted its importance [72, 99, 100] and in particular have noted that model predictions tend to be more sensitive to differences in topology than they are to differences in parameters [101]. Recently, Kuepfer et al. [102] investigated the ability of 19 topologically different models to predict the behavior of the mTor pathway. Using an ensemble of models in this way can be compared to a topological bootstrap, whereby one samples the range of predictions due to differences in structure.

In this paper, we present an investigation into several aspects of ODE modeling of biological networks. The first aim was to assess the standard practice of model selection and its impact on our ability to make accurate predictions. The second goal was to test several heuristics for prediction evaluation including: (1) the distance between training data and a prediction (as a measure of extrapolation); (2) propagated parametric uncertainty; and (3) consensus among an ensemble of topologically different models. All evaluations were performed in a simulation framework in which the correct answer was available for each data set and prediction (see Figure 2-1, with more details included in the Methods section).

A major conclusion of this study is that it is valuable to examine the predictions from the ensemble of all reasonable models, rather than to pick one model and label it all-purpose. Furthermore, the consensus among the ensemble of models is a strong indicator of prediction accuracy. In Chapter 3, we will further address the concept of ensemble consensus and its utility in various forms for selecting high-confidence predictions.

2.2   Methods

The overall approach pursued here is outlined in Figure 2-1. An actual system was selected as representative of the true biology from which synthetic noisy data was produced through model simulation and treated as experimental data. A variety of trial network topologies was parameterized using this data, resulting in an ensemble of fit models. The fit models
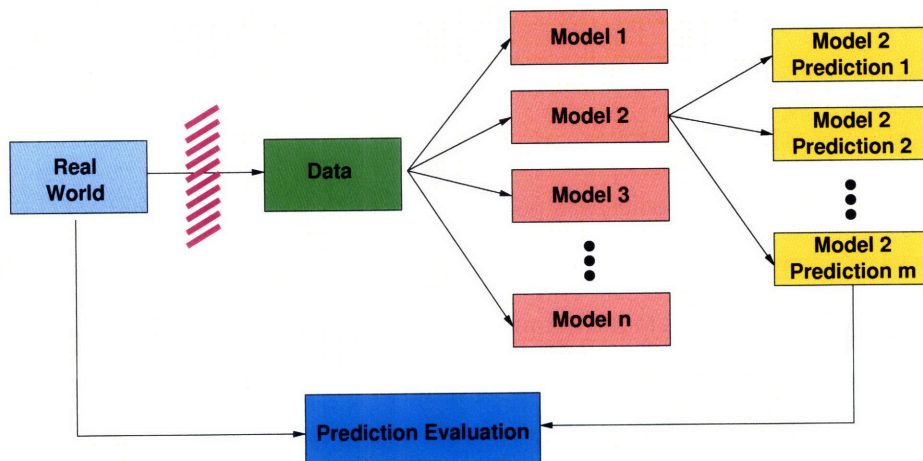
Figure 2-1: Schematic of the algorithm used in this study. We began by choosing an actual system ("Real World") and simulating it with ordinary differential equations. The integrated model was sampled and noise was added to produce simulated data ("Data"). From the point of data generation onwards, the methods were blinded with regards to the actual system. Multiple models were fit to each noisy data set to create an ensemble of models ("Model i"); these models differed in their topologies, or their parameter sets, or both. Rate parameters and/or initial concentrations of the models were then perturbed and the models were re-integrated under these new conditions to generate predictions ("Model i, Prediction j"). Properties of the ensemble under perturbation were computed and then evaluated against the predicted behavior of the actual system ("Prediction Evaluation").

were subjected to perturbations in the form of parameter changes, and both the fit and perturbed models were evaluated by comparing the results to those generated by the actual system. The procedure was repeated for different choices of actual system, different levels of noise in the data, and different samplings for each level of noise. The process can be divided into several steps: model selection, simulation, model fitting, modification and prediction, and analysis. Methods for each of these steps are outlined below, along with descriptions of the tools used for analysis.

2.2.1   Initial Topology Selection

We selected a set of small, simple genetic regulatory networks (GRNs) with which to develop the methods. The selected GRNs consist of three or four nodes with a single input, as shown in Figure 2-2A. The six network topologies all follow the same logical rules with respect to proteins A, B, and C, as described by the truth table in Figure 2-2B. That is, if the concentration of protein A goes up, the concentrations of proteins B and C go down, and

vice versa. It was assumed that the concentration of protein D was not measured and therefore did not figure into the truth table. Such would be the case if the role of protein D were only a hypothesized one.

The regulation between species was positive or negative transcriptional control and measurements were taken to be the concentrations of the proteins. Transcription and translation were lumped together in our models (similar to [103] or [46]), and genetic regulation of one species by another, or by itself was described using Hill coefficients. For protein i under promotion by protein j and j alone, the dynamics were written as

$$\frac{dx_i^{m_s}}{dt} = \gamma_i + \frac{\alpha_{ij}(x_j^{m_s})^{\beta_{ij}}}{(\xi_{ij})^{\beta_{ij}} + (x_j^{m_s})^{\beta_{ij}}} - k_i x_i^{m_s} \tag{2.1}$$

where $x_i^m(t)$ is the time-dependent concentration of protein i, as described by model m_s; γ_i is the constitutive production rate of protein i in the absence of any other proteins; α_{ij} is the maximum promotion rate of protein i due to regulation by protein j; ξ_{ij} is the concentration of protein j leading to the half-maximal rate of promotion; β_{ij} is the Hill coefficient of regulation; and k_i is the degradation rate constant. Repression of protein i by protein j was alternatively modeled using equations of the form

$$\frac{dx_i^{m_s}}{dt} = \frac{\gamma_i}{1 + (x_j^{m_s}/\kappa_{ij})^{\beta_{ij}}} - k_i x_i^{m_s} \tag{2.2}$$

where κ_{ij} is the concentration of protein j when, due to repression, the rate of production of protein i is reduced to half the constitutive rate. Full equations for all six topologies are included in Appendix A. Topologies 1 and 4 were parameterized (as specified in Appendix A) to give Models 1 and 4, and each in turn was used as the actual system. (The use of the term *Model* will be restricted to a topology and associated parameterization. In the absence of specific parameters, it will be referred to as a Topology.)

2.2.2 Data Simulation

Either Model 1 or Model 4 was pre-selected as the actual system and was then simulated, using the Jacobian numerical software package (Release 3.0A, Numerica Technology LLC,

A
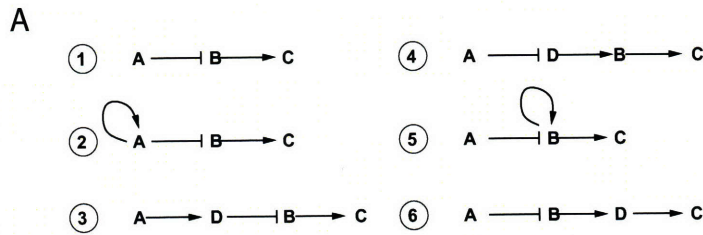
B

A	B	C	
1	1	1	0
1	1	0	0
1	0	1	0
1	0	0	1
0	1	1	1
0	1	0	0
0	0	1	0
0	0	0	0

Figure 2-2: The six small genetic regulatory networks chosen for study are shown in (A). The gene A has a basal level of transcription and its corresponding protein either upregulates (arrow) or downregulates (dashpot) the next gene in the sequence. All six of these networks show qualitatively similar responses in B and C when we increase or decrease the rate of transcription of A. The logic rule followed by all six models corresponds to AND(XOR(A,B),XNOR(B,C)), which is easily understood in terms of a truth table (B). The gene/protein D is considered invisible; that is, we can hypothesize its presence and how its presence affects the overall topology but direct measurements of its concentration are not available.

Cambridge, MA) for integration of the differential equations. The resulting trajectories of time-dependent concentrations were sampled at eleven equally-spaced time points (from $t = 0$ to 10, scaled time units). Gaussian noise was added to generate data values $y_{i,t} = x_i^{m_s}(t) + \epsilon$ where $\epsilon \sim N(x_i^{m_s}(t), \sigma^2 x_i^{m_s}(t))$. (The m_s label is removed from y here to indicate that once a data set was created, we assumed no knowledge of the actual system used to generate those data.) Each full data set consisting of the concentrations of proteins A, B, and C at all eleven time points is denoted with the matrix Y, with rows as species, and columns as time points. All analyses were carried out for both $\sigma^2 = 0.2$ and 0.3 (corresponding to 20% and 30% Gaussian noise).

While noise in chemical reactions is best captured by detailed stochastic simulations that track the evolution of species over time, such methods come at a high computational expense [21, 22]. We performed trial studies on the impact of using stochastic simulations with Langevin dynamics rather than additive noise (data not shown) and found sufficient similarity in the properties of the trajectories to continue with the more efficient deterministic approach. (It is not generally true, however, that deterministic diffential equations can necessarily capture the dynamics of noisy systems [104].)

2.2.3 Model fitting

The two primary descriptors of our models are the topology (which for each model is one of the topologies previously described) and the parameter values, $\boldsymbol{\theta}$. While some parameters in the models were assumed to be known, (hypothetically, from previous experimental results), others were fit. Parameters that were fit included a subset of the initial protein concentrations and rate constants, a summary of which is given in Table A.2. Optimizations were performed separately for each of the six different topologies—no adjustments were made to the topology during the fit.

We used Jacobian to find the vector of parameters $\boldsymbol{\theta}^*$ that yielded the best fit of each topology to a particular noisy data set Y. The objective function for the optimization was the log likelihood, equivalent to weighted least squares if the noise is assumed to be normally distributed, and the value of this function at the optimum is denoted $\log(L(\boldsymbol{\theta}^*))$. Jacobian uses a BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm to perform local op-

34

timization [105]; the bounds on each parameter were set to be sufficiently wide such that
the optima rarely included a value at the bound. (If optimizations to different data sets
repeatedly led to a parameter reaching a bound, that bound was stretched and the fits were
recomputed.) Optimizations were seeded with initial parameter guesses equal to the values
used in the actual system, i.e. from Table A.2.

To attempt to account for the variable number of parameters fit in the different models,
we applied the Akaike and Bayesian Information Criteria [51, 52]. Both the AIC and the
BIC balance the likelihood of the fit with terms accounting for model degrees of freedom.
Specifically,

$$AIC = -2\log(L(\boldsymbol{\theta}^*)) + 2K \tag{2.3}$$

where $L(\boldsymbol{\theta}^*)$ is the maximum likelihood valued at the best fit parameters and K is the
number of estimated parameters. The Bayesian information criterion is

$$BIC = -2\log(L(\boldsymbol{\theta}^*)) + K\log(n) \tag{2.4}$$

where the additional parameter n is the sample size of the data set used to fit the model.

2.2.4   Model modification and predictions

Optimized models with best fit parameters were subject to perturbations of random di-
rections and magnitudes. For this study, a perturbation consisted of altering one or more
parameters. The pool of potentially variable parameters included the initial concentrations
of proteins A, B, and C, $\{x_A(0), x_B(0), x_C(0)\}$, the degradation rate constants for these
same proteins, $\{k_A, k_B, k_C\}$, and the constitutive production rate of protein A, γ_A. These
are relevant parameters in all six topologies. One, two, or three parameters, randomly
selected from this set, were chosen to be varied at a time. The limit of three was chosen
to avoid obscuring the effects of uncertainty in the fit parameters. Among the set of seven
parameters available for perturbation, four were fit during optimization: $x_A(0), k_A, k_B$, and
k_C. If one of these parameters was selected for perturbation, the fit value of that parameter
was replaced with the perturbed value. Similarly, parameters that were initially fixed in the

optimizations were reassigned perturbed values. Values were selected from a logarithmic uniform distribution ranging two orders of magnitude above and below the values set in the original model.

A prediction consisted of the concentrations of proteins A, B, and C as a function of time generated by integrating the fit model with the perturbed parameter set. In the following text, $\mathbf{y}_t^{m_s,\Delta_r}$ is the vector of predictions for model s as a result of perturbation r at time t. Each element of the vector, $y_{i,t}^{m_s,\Delta_r}$, corresponds to the concentration of protein i. The ensemble prediction consisted of the union of the $\mathbf{y}_t^{m_s,\Delta_r}$ for all six models.

2.2.5   Heuristic measures of prediction confidence

Propagation of statistical error

All model parameters in real systems are subject to some degree of uncertainty whether they are optimized or determined independently through experimentation; sources of uncertainty include disagreement between the fit model and the data as well as experimental measurement error. When an ODE model with uncertain parameters is used to make predictions, the deterministic trajectories are also subject to uncertainty. The covariance matrix of a prediction under perturbation r, $\mathbf{V}^{m_s,\Delta_r}$ can be approximated as [50]

$$\mathbf{V}_t^{m_s,\Delta_r} \approx \left(\frac{\partial \mathbf{y}_t^{m_s,\Delta_r}}{\partial \boldsymbol{\theta}} \right) \mathbf{V}_\theta \left(\frac{\partial \mathbf{y}_t^{m_s,\Delta_r}}{\partial \boldsymbol{\theta}} \right)^T + \mathbf{V}_\epsilon \qquad (2.5)$$

where $\mathbf{V}_\theta$ is the covariance matrix of the parameters, and the partial derivatives are sensitivities of the model predictions to the various parameters. The $\mathbf{V}_\epsilon$ term can be used to account for additive measurement noise. In the absence of the correct model, the matrix $\mathbf{V}_\theta$ is estimated using

$$\mathbf{V}_\theta \approx -(\mathbf{H}^*)^{-1} = - \left(\frac{\partial^2 \log L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \right)^{-1}_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \qquad (2.6)$$

H^* is the Hessian matrix valued at the best fit and $L(\boldsymbol{\theta})$ is the likelihood of the data given the model with parameters $\boldsymbol{\theta}$, as in section 2.2.3.

$\mathbf{V}_t^{m_s,\Delta_r}$ was computed for all predictions r. As a measure of the total uncertainty of a prediction due to parameters alone, we calculated a score $U_{PE}^{m_s,\Delta_r}$ for the propagated error

36

by summing the variances at unit time points from $t = 0$ to 10 across all species.

$$U_{PE}^{m_s, \Delta_r} = \sum_{t=0}^{T} \text{Tr}(V_t^{m_s, \Delta_r}) \tag{2.7}$$

Distance from training data to prediction

We measured the distance between the training data and a prediction in three ways. The simple sum of squares distance computes the difference between data value $y_{i,t}$ for species i at time t and the predicted value under perturbation for the same species, at the same time, $y_{i,t}^{m_s, \Delta_r}$ and integrates over species and time:

$$D_{SSS}^{m_s, \Delta_r} = \sum_{i=1}^{N} \sum_{t=0}^{T} (y_{i,t} - y_{i,t}^{m_s, \Delta_r})^2 \tag{2.8}$$

We also computed the nearest point sum of squares which adds together the distances between each time point on the predicted trajectory and the nearest datum to which the model was fit:

$$D_{NPSS}^{m_s, \Delta_r} = \sum_{i=1}^{N} \sum_{t=0}^{T} \left(\min_{\tau \in [0,T]} (y_{i,\tau} - y_{i,t}^{m_s, \Delta_r}) \right)^2 \tag{2.9}$$

A third metric of distance between the training data and a prediction was formulated in which a box was drawn around the data in phase space. For each prediction time point lying outside the box, the distance was computed from the center of the box; points inside the box do not contribute to the metric:

$$D_{\text{Box}}^{m_s, \Delta_r} = \sum_{i=1}^{N} \sum_{t=0}^{T} (y_i^{BC} - y_{i,t}^{m_s, \Delta_r})^2, \quad \text{for all points outside the box} \tag{2.10}$$

where y_i^{BC} is the center of the box in phase space, i.e.

$$y_i^{BC} = \frac{1}{2} (\max_{t \in [0,T]} y_{i,t} - \min_{t \in [0,T]} y_{i,t}) \tag{2.11}$$

Consensus among models in an ensemble

We developed and tested a number of similarity scores to express the level of consensus among our ensemble in making a given prediction. The combinatorial consensus score, C_{CC}

sums up the differences between all combinations of models in the ensemble, over evenly spaced time points and across all species:

$$C_{CC}^{\Delta_r} = \sum_{i=1}^{N} \sum_{t=0}^{T} \sum_{s=1}^{M} \sum_{j \neq s}^{M} (y_{i,t}^{m_s,\Delta_r} - y_{i,t}^{m_j,\Delta_r})^2 \qquad (2.12)$$

where M is the number of models in the ensemble, N is the number of proteins that are measured and T is the last measured time point. Similarly, the weighted combinatorial consensus score C_{WCC} weights the differences according to the mean value of each model pair:

$$C_{WCC}^{\Delta_r} = \sum_{i=1}^{N} \sum_{t=0}^{T} \sum_{s=1}^{M} \sum_{j \neq s}^{M} \frac{(y_{i,t}^{m_s,\Delta_r} - y_{i,t}^{m_j,\Delta_r})^2}{((y_{i,t}^{m_s,\Delta_r} + y_{i,t}^{m_j,\Delta_r})/2)^2} \qquad (2.13)$$

The mean consensus score C_{MC} calculates the mean prediction and then sums the differences between each model in the ensemble and this mean:

$$C_{MC}^{\Delta_r} = \sum_{s=1}^{M} \sum_{i=1}^{N} \sum_{t=0}^{T} (y_{i,t}^{m_s,\Delta_r} - \overline{y}_{i,t}^{\Delta_r})^2 = \sum_{s=1}^{M} C_{MC}^{m_s,\Delta_r} \qquad (2.14)$$

The weighted mean consensus score C_{WMC} weights the differences according to the mean prediction:

$$C_{WMC}^{\Delta_r} = \sum_{s=1}^{M} \sum_{i=1}^{N} \sum_{t=0}^{T} \frac{(y_{i,t}^{m_s,\Delta_r} - \overline{y}_{i,t}^{\Delta_r})^2}{(\overline{y}_{i,t}^{\Delta_r})^2} = \sum_{s=1}^{M} C_{WMC}^{m_s,\Delta_r} \qquad (2.15)$$

C^{m_s} is a measure of the distance from model s to the mean ensemble prediction. Summing over the C^{m_s} therefore gives the overall variance of an ensemble prediction. All consensus measures have the property that they are low (close to zero) for a prediction for which all models predict a similar trajectory and increase with increasing disagreement among the models.

2.3 Results and Discussion

Our aim was to develop an understanding for how the quality of predictions from ODE models varies as a function of different features of the data, the perturbations, and the models themselves. Our overall strategy in working towards this goal was to generate a large

database of simulations across variant models from which to draw statistics that relate these features to prediction accuracy. As one tangible outcome, we sought methods to choose high confidence model predictions. A key feature of the work involved investigating the potential role of ensembles of models to improve upon predictions that are traditionally made using a single model. A computational framework for addressing both facets of the project was created and is described in Figure 2-1. Actual systems were chosen and simulated with additive noise; the noisy data sets were then fit to models with different topologies to create an ensemble of models; and, finally, numerous perturbations were applied to each member of the ensemble and the simulation outcome was recorded. The entire process was repeated for multiple actual systems, different levels of noise, and different subsets of the ensemble to yield millions of predictions for study.

The six models of our ensemble represent simple genetic regulatory networks that share the same Boolean logic with respect to the concentrations of the proteins involved (Figure 2-2). That is, if we represent high and low expression of a protein as on and off states, we would expect the six models to behave the same way, under all conditions. In choosing the ensemble this way, it is reasonable to assume that every topology has a good chance of fitting data generated from any of the other models.

We selected Model 1 and then Model 4 as our actual system (see Figure 1), each defined by assigning a single set of rate constants and initial conditions to Topology 1 and Topology 4. We then used the models to simulate deterministic time courses of protein concentration. To each simulation we added proportional Gaussian noise in which the variance of the distribution at each time point was a fixed percentage of the calculated deterministic value (simulating measurement noise). We generated fifty different data sets for each actual system at each of two levels of noise, 20% and 30%, for a total of 200 simulated initial data runs.

Optimization was performed to individually fit Topology 1 through Topology 6 to each of the 200 data sets. Fit models were generated by selecting parameters that minimized the sum of squares error between the model and data, subject to constraints (see Methods). Each of the 1200 fit models was uniquely specified by its topology and parameters; even for fit models where the topology matched that of the actual system, parameter values differed

39

from the actual system due to the noise added to produce the data.

Quality of fit cannot always select the model with the correct topology

We rank-ordered the fits to each data set according to the maximum likelihood (ML) and found that, according to that metric, fit models with Topology 1 were rarely selected as the best model when Model 1 was used as the actual system. Figure 2-3 (top row) shows that in the case of single data sets with 30% noise, in only one case out of 50 was a fit model with Topology 1 identified as the best model. Due to over-fitting, Topologies 3, 4, and 6, which all have an additional fit parameter compared to Topologies 1, 2, and 5, were preferred. The Akaike Information and Bayesian Information Criteria (AIC and BIC) are both metrics that attempt to improve the model selection procedure by balancing the quality of fit with the number of degrees of freedom. (The criteria differ in that they are derived from either a frequentist or Bayesian perspective.) After applying these metrics, Topology 1 was correctly selected as the best in about half of all trials. Similar results were obtained with 20% noise (see Figure A-1 in Appendix A).

When the actual system was instead Model 4, and the data had 30% noise, the maximum likelihood metric correctly chose Topology 4 in about one fifth of the trials. It was outselected by Topologies 3 and 6 which are equivalent in size. In this case, the AIC and BIC overpenalized the larger models, including the correct topology, always preferring a simpler one (Figure 2-3, bottom row). It was in fact impossible to find a linear correction term for the degrees of freedom that could consistently select the correct topology from data derived from Model 1 and Model 4. (We tested the Takeuchi Information Criterion (TIC) [52] and the Information Complexity (ICOMP) [54] as well (data not shown) and observed that these were also incapable of consistently selecting the correct topology from models of different complexity.) These results demonstrate that when the topologies under consideration are expected to behave similarly, and can all recapitulate the data well, the quality of fit between model and data is not generally capable of selecting the correct topology, even when supplemented by information criteria (specifically, the AIC and BIC).

The concept of statistical indistinguishability encapsulates the premise that when models are very similar near the data on which they are trained, there may not be enough
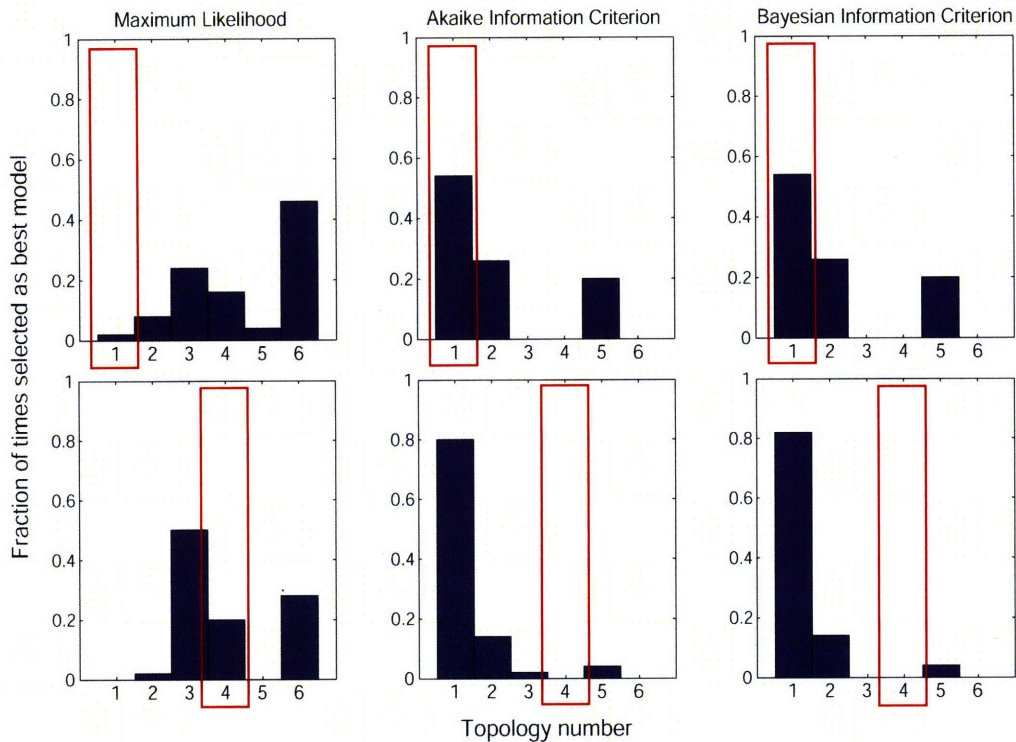
Figure 2-3: Histograms for the fraction of times each topology was selected as the best model by Maximum Likelihood, AIC, or BIC. The top row corresponds to data generated from Model 1, whereas the bottom row is for Model 4 (see Figure 2-2 for topologies). By ML, fit models with Topology 1 were rarely selected as the best model when Model 1 was used to generated the data because this topology has fewer parameters than Topologies 3, 4, and 6. While accounting for this difference with AIC or BIC improves the ability to select the correct topology in this particular case, these model selection metrics over-penalize fit models with the correct topology when the data was generated from Model 4.

information to tease them apart. The same may be true if the data are simply poor. It can be difficult to establish whether the ML, AIC, and BIC values associated with biological models imply distinguishability. Differences in the likelihoods of two candidate models i and j are typically assessed for significance using the likelihood ratio test (LRT) [90], in which the test statistic $\Lambda = -2(\log L_i - \log L_j)$ is compared to a χ^2 distribution with degrees of freedom equal to $(\mathrm{df}_i - \mathrm{df}_j)$. For models of the same size that have different underlying topologies, the procedure is not straightforward. Moreover, when the models are based in differential equations, Λ may not follow a χ^2 distribution.

To establish distinguishability using the AIC or BIC, the usual method consists of simi-

41

larly computing the differences associated with the candidate models, $\delta_i = \text{AIC}_i - \text{AIC}_{\text{min}}$.[1]
Models can be accepted or rejected based on the δ_i, although the thresholds for these decisions are somewhat empirical in nature [52].

We do not wish to imply that the statistical methods commonly applied to non-differential equations models will never be applicable to ODE models such as we've described here. Rather, we wish to highlight that there are cases where these methods will not generate the expected result and, lacking access to the correct answer as we had here, it is impossible to know *a priori* whether one's own particular test case falls in the category of applicable or not.

A discussion of identifiability is also warranted here, which we consider to be the ability to select the correct topology given perfect data. (There is an alternate notion of identifiability, that being whether multiple parameterizations of the same topology can yield identical fits to the data. We will use the first definition.) Noiseless data were generated using Model 1 or Model 4 and each topology was optimized to this data. The likelihoods for the fit models, normalized to the best fit, are plotted in Figure A-2. Model 1 is not identifiable as an actual system because through parameter adjustment, Topology 4 can be made to generate as good as fit as Topology 1. In contrast, Model 4 is identifiable as no other model topology can fit the data as well as Topology 4. There are also greater discrepancies between the fits of candidate topologies to Model 4 data, and this had an impact on other results that we discuss below.

Quality of fit is uncorrelated with quality of prediction for models that fit the data well

One thousand perturbations of random magnitudes and directions were applied to each of the 1200 fit models and over a million predictions were simulated (see Methods). The predictions were evaluated by comparing them to simulations of the actual system (without added noise) for the same perturbation. Figure 2-4 shows the sum of squares error of prediction versus the negative log likelihood valued at the fit for two subsets of the perturbations.

---

[1]The standard notation for these differences is Δ though we have used δ here to avoid confusion with the model perturbations, Δ_r as described in the Methods section.

In these subsets, the models were all parametrized to data from Model 1 (top) or Model 4 (bottom) with 30% noise. The stripes in the plot arise from the fact that a single fit model is associated with a particular likelihood score and that each such model was used to generate multiple predictions.

There is no correlation between quality of fit and quality of prediction in these cases ($\rho^2 = 0.024$ and 0.060 for Model 1 and Model 4 as actual system, respectively). Even the best fitting models can make very poor predictions (those points in the upper left) while less well-fitting models can sometimes make good predictions (lower right). It is also not true that for any particular data set the best fitting model generates the highest quality predictions. With Model 1 as the actual system, the best fitting model also made the best prediction for 10.2% of perturbations. With Model 4, this was a still modest 18.4%. (Figure A-3 shows the specific breakdown of these results by topology. In Figure A-4 these results have been normalized for the number of times that each model was selected as the best fit to the data.)

The disconnect between quality of fit and quality of prediction in this case is partly due to the fact that all of the models provide a reasonable fit to the data. If a model fits the data very poorly, it would be unlikely to ever make accurate predictions. In a setting in which we are interested in using such models to draw biological insight, however, candidate topologies that cannot match the data would in general not be used in any further analysis. The emphasis here is that while models may be indistinguishable at the level of fitting, they most certainly may vary in the quality of prediction when the actual system is perturbed.

Models with the correct topology do not necessarily make the best predictions

It is not true that the model with the correct topology generates the best predictions because fitting to noisy data sets results in imprecise parameter values. Our experiment with Model 1 as the actual system showed that models with Topology 1 made the best prediction for fewer than 40% of perturbations (Figure 2-5A). When Model 4 was instead the actual system, models with Topology 4 made better predictions than any other model nearly 60% of the time (Figure 2-5B).
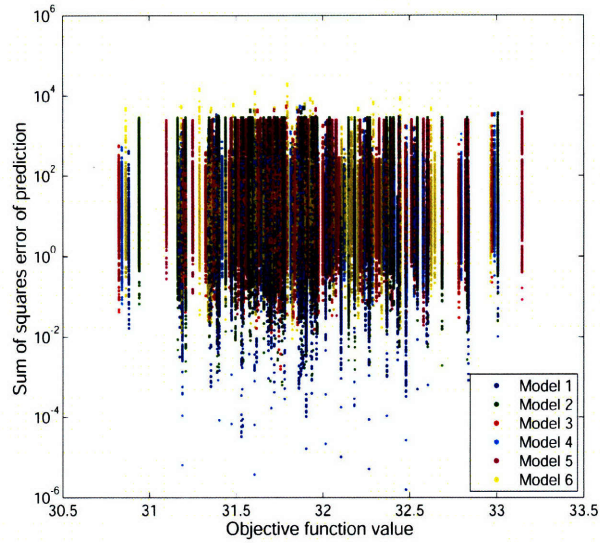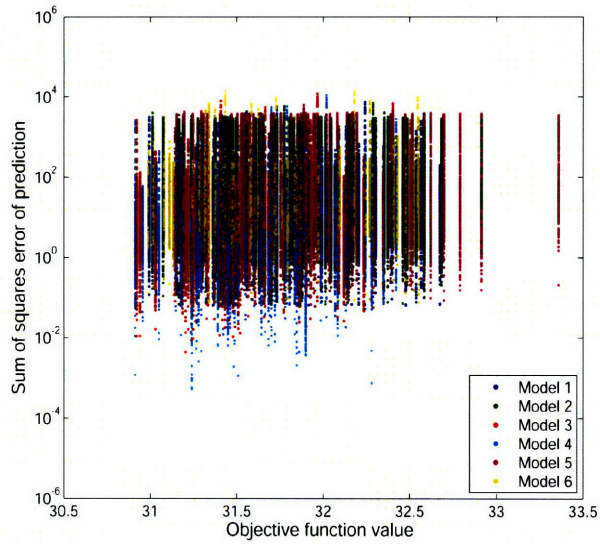
Figure 2-4: Plotting the error in prediction against the maximum likelihood at the fit demonstrates that there is little or no correlation between quality of fit and quality of prediction. Colors are for the different model topologies. (A) Data from Model 1 as the actual system. (B) Data from Model 4.

Unfortunately, the relationship between the data and the model cannot distinguish between the first case and the second—that would require detailed knowledge of the actual system, which we do not currently have when studying biology. It should be noted that sometimes, even when one particular topology makes better predictions than all others in the ensemble, the absolute error of prediction is still intolerably large.
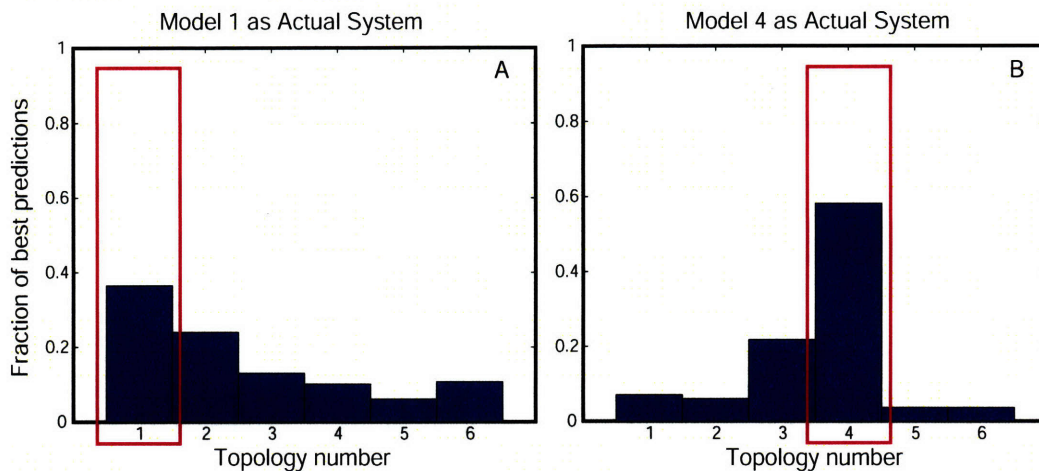
Figure 2-5: Histograms for the fraction of times each topology makes the best prediction, i.e. has the lowest sum of squares error compared to the other ensemble models (same data as Figure 2-4): (A) data generated using Model 1 as the actual system, (B) data generated with Model 4 as the actual system. When Model 1 is used to generate the data, models with that topology are only rated best for about 40% of the perturbations. When data is generated from Model 4, the model with the correct topology makes the best prediction for about 60% of all perturbations. That in either case the correct topology doesn't make the best prediction 100% of the time is due to uncertainty in the parameters.

Quantifying parametric uncertainty is of greatest utility when the correct topology is known

Many of the prior efforts to account for model uncertainty have focused on understanding errors in the parameter values [93, 95, 98]. We investigated the utility of parametric uncertainty to evaluate predictions by computing the estimated variance of predictions and comparing it to the actual error of those predictions over our database of perturbations. When only the correct topology was considered, the error of prediction was consistent in U_{PE}. In Figure 2-6A and B, data was generated using Model 1 and all fit models had Topology 1. For perturbations of these models, the error appeared to be linearly bounded

by the parametric uncertainty[2]. That is, given the estimated variance of a prediction, a reasonable limit could be placed on its associated error. When other topologies were introduced, unsurprisingly the estimated variance no longer provided this bound (Figure 2-6C and D); this follows from the fact that $V_t^{m_s, \Delta_r}$ (from Equation 2.5) only describes the width of the distribution of solutions while the mean is a function of both parameters and topology. Although there are other methods to estimate the uncertainty of predictions based on parameter errors, and we have not tested those here, it is reasonable to conclude that these are also incapable of capturing the variance due to differences in topology. We have merely illustrated the issue that both parameters and topology must be considered in order to fully understand the range of predictions that could be expected under a particular perturbation.

Distance from prediction to training data is weakly correlated with prediction quality

We were interested in learning whether the distance between the training data and a prediction was related to prediction accuracy. The thought was that predictions lying close to the data are in some way smaller extrapolations than predictions farther away and may therefore be less prone to error. We calculated the distance between the data and a prediction in three ways, as described in the Methods section. The distances for a subset of the simulations in which the data came from Model 1 with 30% noise are plotted against the error of prediction in Figure 2-7. The correlation coefficients for Figure 2-7, corresponding to the time space (A), phase space (B), and phase box (C) distances are 0.41, 0.44, and 0.35, respectively. While the correlation is weak in all three cases, it is clearly evident that the largest errors correspond to predictions that lie far away from the training data.

The red squares in Figure 2-7 are taken from models that were fit to the data but left otherwise unperturbed. With all three distance metrics, there are a number of post-perturbation predictions (blue dots) that are just as close to the data as the fit models, but exhibit errors up to an order of magnitude greater. Therefore, one cannot expect a

---

[2]For a given value of U_{PE}, the expectation is for the density of the error to drop off exponentially and the apparent bound arises due to limited sampling and plotting on a log scale.
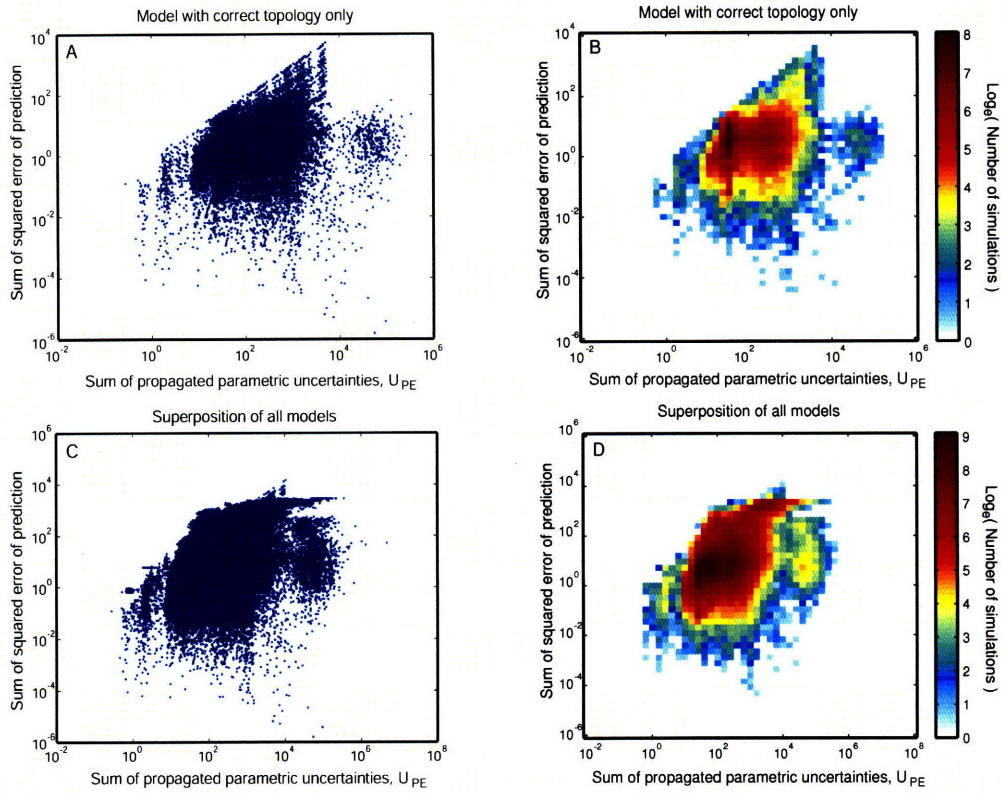
Figure 2-6: The propagated statistical uncertainty, U_{PE}, was computed for all models under all perturbations, and plotted versus the error of prediction. When the model under study has the correct topology, calculations with the estimated parametric uncertainty can provide a roughly linear bound on the actual error of prediction ((A) and (B)). When topology is uncertain, this linear relationship no longer exists and the prediction errors are generally larger: subplots (C) and (D) show superimposed results for models from all six topologies, fit to data from Model 1. The right column shows two-dimensional histograms of the data on the left.

prediction to be accurate simply because it lies close to the data; however, if the distance from a prediction to the data is large, one might conservatively be wary of its quality.

Consensus among model ensembles is a strong classifier of prediction error

For each of the perturbations in our database, we computed the ensemble distance from consensus (see Methods). Figure 2-8 shows the relationship between the ensemble variance, C_{MC} and the error of the ensemble mean prediction. There is an increasing trend in the error of prediction as the ensemble gets further away from consensus, though the correlation itself is not particularly strong ($\rho = 0.61$ for Model 1 as actual system (top row)). Notably, however, when the ensemble shows a high level of agreement, it is rare that the average
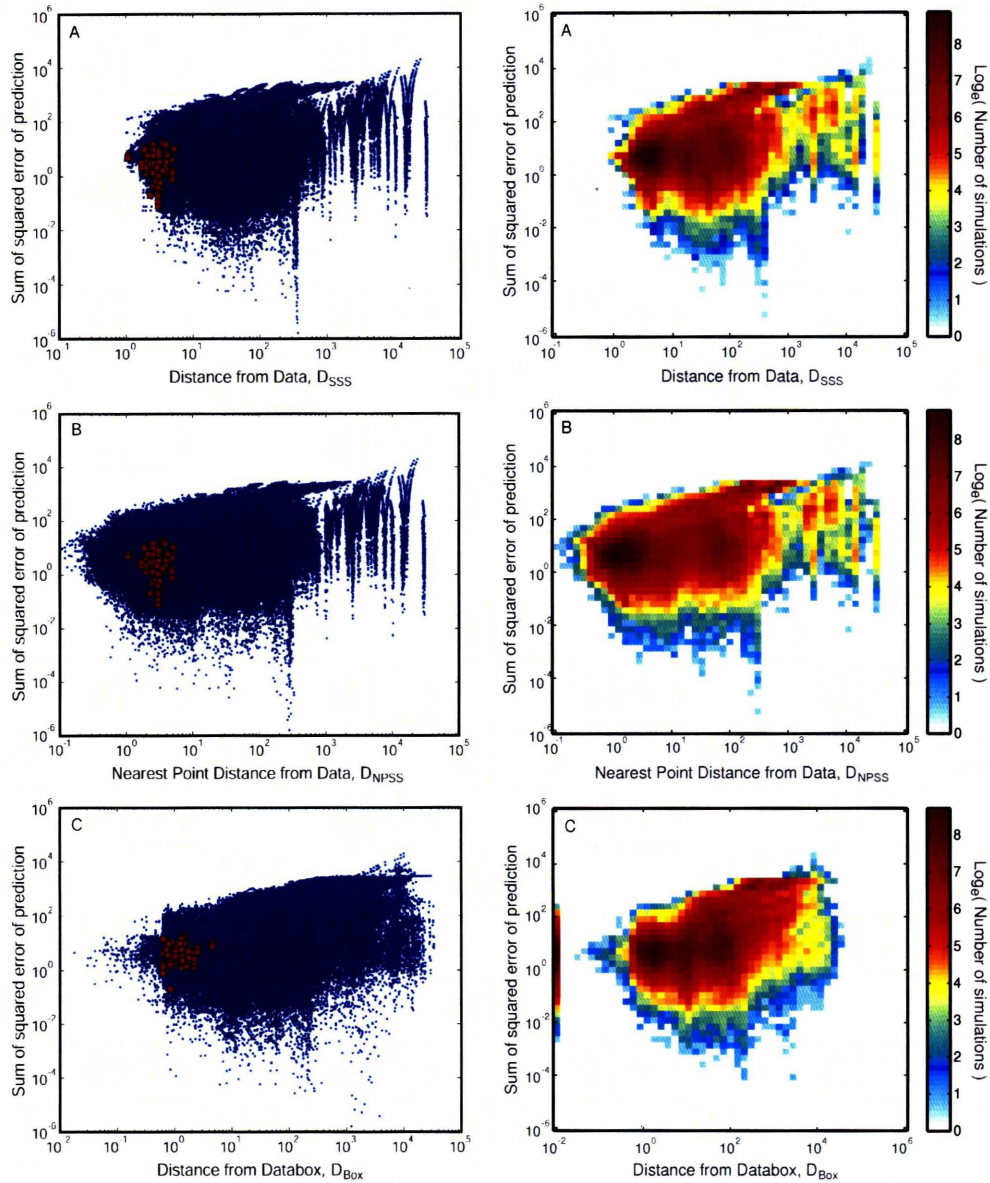
Figure 2-7: Distance from data to prediction measured three ways and plotted versus error of prediction: (A) Simple sum of square distance from data to prediction, (B) distance from data to prediction using the nearest point in phase space, (C) distance between data and prediction measured by drawing a box around the data in phase space (see Methods for details.) The right column contains two-dimensional histograms of the plots in the left column. Red squares on the left-hand plots show the distance and error for fits of all six model topologies to each of 50 noisy data sets.

prediction is poor. As in Figure 2-7, the red squares show the consensus and error for models that were fit but not perturbed. Once a perturbation has been applied, it is uncommon to have a similar consensus score to the unperturbed case and yet make a worse prediction. We concluded therefore that if the ensemble is perturbed and still shows a high level of consensus, that the mean prediction is likely to be accurate[3].

To determine the power of consensus as a classifier of prediction error, we analyzed the data in Figure 2-8 to produce a receiver operating characteristic (ROC) curve for such a classification test. The real designation of a prediction was assigned according to the ensemble mean prediction error: if it was smaller for the perturbation than it was at the fit, it was judged as being positive, and vice versa. The measure of quality was taken to be the ensemble variance: predictions for which the ensemble variance was below a chosen threshold were labelled positive and others, negative. Then a true positive by this test was a prediction for which the ensemble variance was below the threshold and the mean ensemble error was less than at the fit. To generate the curve, the stringency of this threshold was varied from the minimum to the maximum value of the computed variances across all perturbations and the rates of true and false positives were computed. The same test was performed separately for each of the 50 data sets built from Model 1 or Model 4 with 30% noise (see Figure A-6). The results were then averaged to produce Figure 2-9.

The ROC curve emphasizes the strength of the ensemble consensus as a classifier of prediction error as it is highly distinct from the curve expected for a random classifier (shown as the dotted line). The curves are very similar for Model 1 and Model 4 as actual systems, implying that, at least for this particular set of topologies, the ability of consensus to classify is independent of which model was used to generate the data. The squares in Figure 2-9 indicate the test where the consensus threshold was the ensemble variance of the fit. Using this threshold for classification is very conservative. The sensitivity is quite low (0.31 and 0.42 for Model 1 and Model 4 as actual systems, respectively) but the specificity (equal to one minus the false positive rate, 0.971 and 0.957) is very favorable. The use of the consensus at the fit as the threshold of discrimination is a good choice if one is mainly

---

[3]Similar conclusions were drawn using the combinatorial consensus score, C_{CCC}, and weighted consensus scores (data not shown).
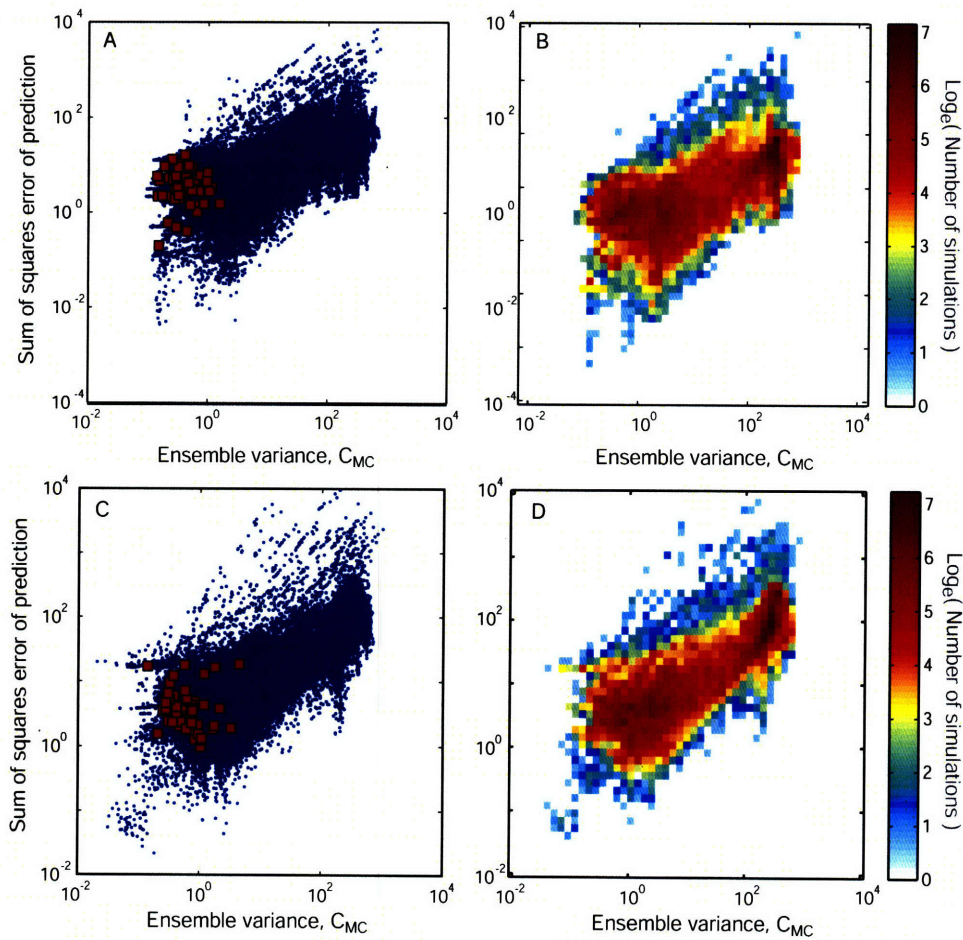
Figure 2-8: Distance between the true prediction and ensemble average prediction as a function of distance from consensus, C_{MC}. When models in the ensemble agree, i.e. the C_{MC} is small, predictions are accurate. The right hand column consists of two-dimensional histograms of the data in the left column. Red squares show the distance and error for the unperturbed fits of all six model topologies to each of 50 noisy data sets. (A,B) Model 1 as the actual system. (C,D) Model 4 as the actual system.
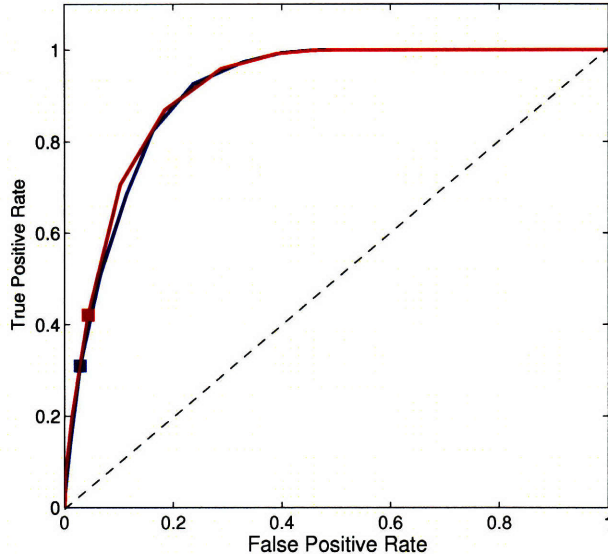
Figure 2-9: Averaged ROC curves for consensus as a classifier of prediction error. The data was drawn from either Model 1 (blue) or Model 4 (red) with 30% noise. The squares indicate the test in which the threshold of discrimination between positive and negative predictions is the ensemble variance valued at the fit. While this choice of consensus is conservative, it does result in the selection of highly accurate predictions.

interested in selecting a small number of predictions that are likely to be accurate. This choice also has no system-dependency which might lead to bias.

2.4   Conclusion

Here we report an investigation into the use of collections of mathematical models to represent incomplete knowledge of biological networks. The study used a topologically diverse ensemble of ODE models and focused on identifying high-confidence predictions from limited and noisy training data. Six model topologies were fit to simulated data, and predictions under a variety of perturbations were made from the fit models and compared to "actual" results from the same model that produced the original simulated data. The results provide interesting insights into procedures for model selection in biological network and heuristics for identifying high-confidence predictions.

When multiple topologies fit the available data reasonably well, the quality of fit is not a good indicator of topological correctness or prediction quality. Model selection, however,

has an important role in eliminating topologies that agree poorly with the available data.

Heuristics for evaluating the quality of model predictions were also evaluated. Parametric uncertainty, which certainly contributes to prediction error in biological modeling, is difficult to propagate through to prediction outcomes accurately in the face of topological uncertainty. Quantifying the distance between training data and testing data has some potential for identifying conservative predictions likely to be of high confidence. However, the most interesting predictions may be those relatively distant from the training data, which could limit the usefulness of this approach.

The most striking result from this study is that consensus across multiple models was useful for selecting high-confidence predictions. It will be important to validate these result with larger and more complex biological networks. We anticipate that approaches based on this methodology will be useful in applications where making a small number of high-confidence predictions are important, such as the design of clinical trials.

A prominent issue with ensemble techniques is of course how to choose the models in the first place. In this example, we hand-picked a number of models expected to behave similarly. In other applications, it might be reasonable for an expert to mine the biological literature for uncertainties in network structure, and to construct alternative topologies based on these differences. There also exist methods to automatically generate model ensembles [106–109]. A related concern might be the sensitivity of these methods to particular choices for the ensemble. For instance, how would the results change if the correct model were omitted? In this example using the six genetic regulatory networks, leaving out the actual system bore little effect on the general conclusions (see Figure A-5); however, one might imagine a pathological case in which the ensemble was very sensitive to the removal of a particular model.

We further investigated the promising avenue of consensus as a measure of prediction quality through the use of existing and novel mathematical techniques applied to the same six-model ensemble presented here. The results of that study are discussed in Chapter 3.

Chapter 3

Consensus methods for differentiating high- and low-confidence predictions from ODE models in biology

Abstract

Computational models are increasingly relevant for understanding biology at the systems level. Although the cell is both noisy and complex, improved computational power and high-throughput experimental techniques are making it easier to build detailed pathway models with differential equations. At this stage, however, these models are still subject to considerable uncertainty, in both topology and parameters. Ensembles of models that differ in their topologies, parameters, or both, may provide assistance in assessing the overall uncertainty in model predictions. Here, we have investigated the use of consensus among models of such an ensemble for evaluating predictions and for identifying perturbations that lead to either high or low consensus. We use model averaging, combined with propagated parametric uncertainty, to compute time-dependent probability distributions for protein concentrations in simple genetic regulatory networks. With the same test system, we apply decision trees to partition parameter space into regions of high or low consensus, as measured by an ensemble variance. The decision tree can also be used to design experiments for refining the ensemble. Finally, we introduce the concept of a consensus sensitivity, a continuous metric for evaluating how the ensemble variance changes in response to perturbation. The decision trees and sensitivities both have the ability to identify the relevant parameters for distinguishing the models of the ensemble. Overall, we demonstrate how consensus can be used to encapsulate multiple sources of error in predictions from differential equation models and to analyze the space of possible perturbations.

3.1 Introduction

Systems biology, and computational modeling of biological systems in particular, has great potential for building our understanding of complex biochemical networks. Models encapsulate and integrate information from diverse sources to assist our ability to predict cellular behavior both under growth conditions or in response to stimulus [1, 110, 111]. Mechanistic differential equation models lie at the detailed end of the spectrum of computational methodologies [30] and provide opportunities to probe the impact of specific chemical perturbations to the cell. These perturbations may represent changes to a cell that occur in a diseased state, as a result of therapeutic intervention, or both, and as such, these models hold much promise for the field of drug development [26–29, 112, 113].

Detailed models of the chemical kinetics inside a cell require large numbers of parameters to be specified, from rate constants to initial protein concentrations. Typically, many of these parameter values are unknown. Some of these values can be drawn from the literature, some can be measured specifically for building the model of interest, and the rest can be set to nominal values or fit to existing data through optimization [49, 84, 114, 115].

These models are also defined by a topology—or network structure—that provides an additional source of uncertainty. Not only is it difficult to choose where to delimit the network in constructing a model of a given cellular function, but often there are many biological uncertainties about the interaction partners of each protein and the specific mechanism by which these interactions occur.

Can models with so many sources of uncertainty still be useful for identifying drug targets or proposing their impact on biological function? Probably. At the very least, uncertain models have a good chance of making accurate predictions near the training data, provided they can recapitulate the data well. If a model is perturbed strongly away from the data, its validity must be questioned. The model *may* still make good predictions, but there is a chance it may not. Consider the role of model predictions in a drug development setting: if a pharmaceutical company were to assign development projects based on the predictions from a model, their importance (or funding) would ideally be ranked as a function of the potential impact balanced by the likelihood of success. In this way, uncertainty analysis for

a model's predictions may present significant value.

Several important papers have described methods for assessing and accounting for uncertainty in model predictions due to poorly determined parameters [93–97]. In biological modeling, the focus regarding topological uncertainty has been on solving problems of model selection [116–119] and experimental design for model discrimination [58–61, 92]. Rather few efforts have been made to explicitly account for topology as a second source of uncertainty in model predictions [102, 109, 120], despite the fact that predictions may in fact be more sensitive to changes in structure than to parameters [72, 100, 101, 121]. In previous studies, (see Chapter 2), we demonstrated a method of accounting for structural indeterminancy using model ensembles and showed that where the ensemble collectively agreed on a prediction, the prediction was more likely to be accurate.

The notion of consensus among models has been used in other disciplines for either improving or validating model predictions. The field of statistics, in particular, has been a proponent of consensus models into which structural uncertainty is explicitly incorporated [72, 100, 122–126]. While the study of econometrics has also employed these methods, they have additionally made use of implicit techniques, where forecasts from individual models are linearly weighted to generate an "average" prediction [127–130]. Both approaches have been adopted for use in diverse other domains, such as weather and climate forecasting [131–133], phylogenetics [134], and ecology [135, 136]. Several common classification algorithms in articial intelligence, such as boosting [137] and bagging [138], also take advantage of consensus among an ensemble of possible models.

There are a number of differences between the application of consensus in biological modeling and in other fields. Many other consensus approaches have been developed for use with classification problems in which a prediction consists of a vector of quantities that can take on finite number of discrete values. Here, a prediction consists of a time-course of continuous protein concentrations. Many classification algorithms do not as yet have continuous analogs.

Models in other disciplines may also not necessarily have the save level of explicit mechanistic interpretation. In statistics and econometrics, models are often built around observed correlations rather than causal descriptions of physical interactions. With correlation mod-

els, quality is assessed in terms of ability to accurately predict future measurements on which the model was not trained. In models of biology, there is the additional constraint of physical reality. There exists no model of signal transduction or genetic regulation that completely describes this physical reality, but it is certainly desirable to eliminate interactions that contradict known biological facts.

Our goal here was to determine the feasibility of different consensus methods both for identifying high-confidence predictions and for determining which types of system perturbations would lead to those accurate predictions. We tested a model averaging approach as a means for assessing the uncertainty of model predictions due to parameters and topologies. The result is a method that yields a time-dependent probability distribution for protein concentrations. We then demonstrated how regression trees could be used to partition perturbations into regions of parameter space that lead to either high or low consensus among the ensemble of models under consideration. In doing so, we derived a method for simultaneously selecting high-confidence predictions and suggesting experiments for refining the ensemble. Finally, we developed the consensus sensitivities, a novel method for analyzing how parameter choices lead to high or low ensemble consensus.

3.2 Methods

3.2.1 Models

The models used for the testing of methods presented here are parametrizations of the six simple genetic regulatory networks described in Chapter 2, as shown in Figure 2-2. All topologies in this ensemble agree at the Boolean level: increasing the concentration of protein A leads to a decrease in the expression of proteins B and C. Decreasing the concentration of protein A has the opposite affect. Equations for the models are provided in full detail in Appendix A.

3.2.2 Parameter estimation

Simulated data were generated either from Model 1 or Model 4, each of which is a specific parametrization of Topology 1 or 4. (The parameter values are given in Appendix A.) The

deterministic results of the integration of each model were sampled at 11 evenly spaced time points (0 through 10) and proportional Gaussian noise was added in the amount of either 20 or 30%. (The reader is referred to Chapter 2 for more details.)

Degradation constants and initial concentrations of all species for all six topologies were estimated by unconstrained local optimization with a maximum likelihood objective function. All parameter fitting was performed using the Jacobian numerical software package (Numerica Technology LLC, Cambridge, MA).

3.2.3 Monte Carlo simulation of perturbations

Fit models were subject to perturbations in the form of parameter changes. The set of variable parameters included the initial concentrations and degradation rates of proteins A, B, and C, and the constitutive production rate of protein A. These parameters are defined and have the same physical interpretation in all six topologies. For each perturbation, one to three parameters were selected at random from the set and assigned new values. These values were drawn from a uniform distribution that extended four logarithmic orders of magnitude (in base 10). A prediction in response to perturbation consisted of a 3×11 matrix of protein concentrations, with each row representing the time course of concentration of protein A, B, or C.

3.2.4 Parametric uncertainty

The Fisher information approach was used to estimate the uncertainty of a prediction associated with a given topology, due to parameters alone. For each perturbation, and for each model at each time point, we computed the covariance matrix of the predictions using [50]

$$
\mathbf{V}_{m,t} = \left(\frac{\partial \mathbf{y}_{m,t}}{\partial \boldsymbol{\theta}} \right) \mathbf{V}_{\theta} \left(\frac{\partial \mathbf{y}_{m,t}}{\partial \boldsymbol{\theta}} \right)^{T} \tag{3.1}
$$

where $\mathbf{V}_{\theta}$ is the covariance matrix of the parameters, $\mathbf{y}_{m,t}$ is the vector of concentrations of each protein species, as predicted by model m at time t, and the partial derivatives of this vector produce the sensitivities of the model predictions to the parameters, $\boldsymbol{\theta}$. $\mathbf{V}_{\theta}$ was estimated using the inverse of the Fisher information matrix, evaluated at the optimum of

the fit, as calculated by the Jacobian software.

3.2.5 Ensemble variance

We define the ensemble variance as the variance of predictions of the time-dependent protein concentrations across all models of the ensemble, summed over species and time:

$$V_E = \frac{1}{M} \sum_{t=0}^{T} \sum_{i=1}^{N} \sum_{m=1}^{M} \left(y_{i,m,t} - \frac{1}{M} \sum_{k=1}^{M} y_{i,k,t} \right)^2 \tag{3.2}$$

where T is the total time of the experiment, N is the number of measured species, M is the number of models in the ensemble, and $y_{i,m,t}$ is the concentration of species i, according to model m, at time t. The predictions were sampled ten times more finely than the data (101 time points). Subsequent results based on the calculation of V_E were shown to be insensitive to the specific sampling rate (data not shown).

3.2.6 Model averaging

We assigned each model the role of expert for the sake of combining their associated probabilities regarding predictions. Expert probabilities are typically combined linearly, using the linear opinion pool,

$$p^{\text{linear}} = \sum_{m=1}^{M} \lambda_m p_m \tag{3.3}$$

or exponentially, using the logarithmic opinion pool,

$$p^{\log} = \prod_{m=1}^{M} (p_m)^{\lambda_m} \tag{3.4}$$

where M is the number of experts, p_m is the probability associated with expert m, and λ_m is the weight given to this expert's opinion.

We assumed Gaussian distributions for each model with means equal to the deterministic prediction obtained by integrating the system of differential equations and variances equal to the propagated parametric uncertainty, as shown in Section 3.2.4. The probability that

at any time t the concentration of species j will be y according to model m is then given by

$$p_{i,m,t}(y) = \frac{1}{\sqrt{2\pi(\sigma_{i,m,t})^2}} \exp\left(-\frac{(y - y_{i,m,t})^2}{2(\sigma_{i,m,t})^2}\right)$$ (3.5)

Using the exponential opinion pool, the probability that any particular protein has concentration y at time t can be written as (see Supplemental Material)

$$p_{i,t}^{\log}(y) = \exp\left(-\left[\frac{1}{2}\log 2\pi + \sum_{m=1}^{M}\lambda_m \log \sigma_{i,m,t} + \frac{1}{2}\sum_{m=1}^{M}\frac{\lambda_m(y - y_{i,m,t})^2}{(\sigma_{i,m,t})^2}\right]\right)$$ (3.6)

Then for any protein species at any time, the concentration of maximum probability is

$$y_{i,t}^{MP} = \frac{\displaystyle\sum_{m=1}^{M} w_m y_{i,m,t}}{\displaystyle\sum_{m=1}^{M} w_m}$$ (3.7)

where

$$w_m = \frac{\lambda_m}{\sigma_{i,m,t}^2}$$ (3.8)

To evaluate the overall quality of a prediction, we computed a consensus probability score

$$P_C = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\log p_{i,t}^{MP}$$ (3.9)

where p^{MP} is the probability associated with the maximum probability prediction, y^{MP}. The overall procedure for generating the probability density distribution for a perturbation is illustrated in Figure 3-1.

Weighting schemes

The weighting for each model according to Equation 3.8 is directly proportional to our *a priori* confidence in the topology and inversely proportional to the uncertainty in the parameters. If the λ_{m_s} are the Bayesian priors, $P(m_s)$, this scheme is equivalent to Bayesian model averaging [123]. Because calculating these priors requires integrating over the accessible parameter space, we found that the computational cost was prohibitive for more
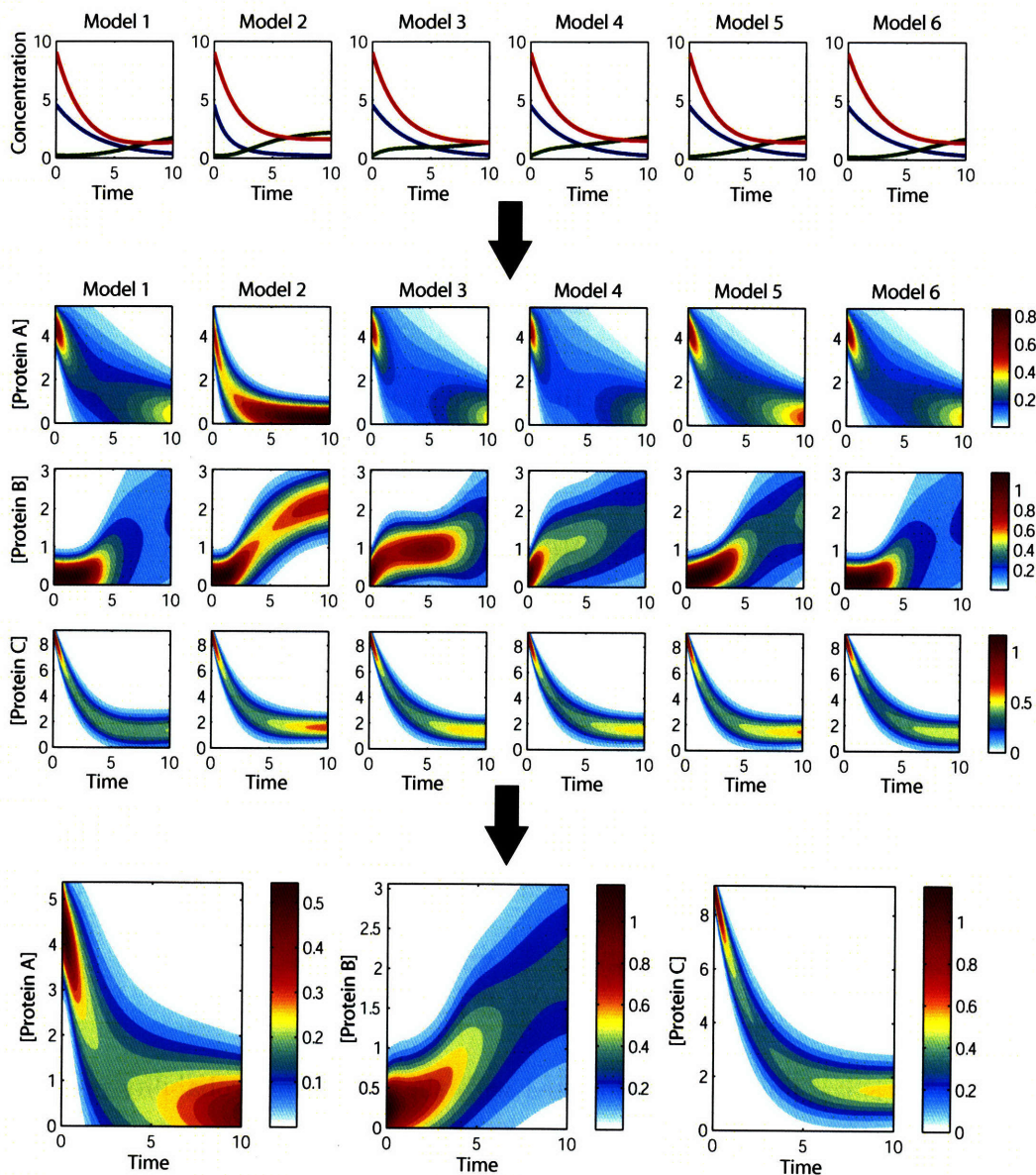
Figure 3-1: Procedure for generating probability distributions of protein concentration as a function of time. The six fit models of the ensemble are perturbed and then integrated to yield deterministic time courses (top row; blue, protein A; green, protein B; red, protein C). Each model is then assigned its own probability distribution by considering the contribution of parametric uncertainty to the accuracy of prediction (middle section). Distributions from individual models are then combined according to a logarithmic weighting (bottom row). The resulting probability distributions encapsulate both topological and parametric sources of uncertainty.

than a few fit parameters. We therefore tested two alternative weighting schemes for the ensemble: equal weighting ($\lambda_{m_s} = \frac{1}{M}$) and weighting according to the distance of each model from consensus:

$$\lambda_m = \frac{1/(y_m - \bar{y})^2}{\sum_{k=1}^{M} 1/(y_k - \bar{y})^2} \tag{3.10}$$

where the mean is taken over the models.

3.2.7   Decision trees

Decision trees were used to partition the space of perturbations into regions defined by either high or low ensemble consensus. The classification and regression tree algorithm (CART) [139] divides data into subsets that share similar values of a response variable; the division is performed by making recursive binary splits in the predictor variables. We defined the predictor variables as the directions and magnitudes of perturbations, expressed as log-fold changes to model parameters; the ensemble variance, V_E, was the response. Each split of the tree was chosen to yield the global optimum improvement to the tree's cost, given by

$$\text{cost} = \sum_{\ell \in \mathcal{L}} \sum_{r \in \ell} (V_E^{\Delta_r} - \overline{V}_E^{\ell})^2 \tag{3.11}$$

where $\mathcal{L}$ is the set of all terminal nodes (or leaves), $V_E^{\Delta_r}$ is the ensemble variance for perturbation r, and $\overline{V}_E^{\ell}$ is the mean ensemble variance of perturbations assigned to node ℓ. The overall cost of the tree can thus be described as the sum of the variances (of the ensemble variances) of all terminal nodes.

The decision tree analysis was performed in MATLAB (Version 2007a, The Mathworks, Natick, MA) with trees initially grown until they met the minimum split criterion of ten observations. Ten-fold cross-validation was performed to identify the optimal level of pruning, which was set to be the smallest tree within one standard error of the minimum cost tree.

3.2.8 Consensus sensitivity

We set C equal to the ensemble variance, V_E, as described above and calculated the sensitivities of this quantity to the vector of variable parameters, $\boldsymbol{\theta}_V$. (Recall that the variable parameters are shared among the models.)

$$\frac{\partial C}{\partial \boldsymbol{\theta}_V} = \frac{2}{M} \sum_{t=0}^{T} \sum_{i=1}^{N} \sum_{m=1}^{M} \left(y_{i,m,t} - \sum_{k=1}^{M} y_{i,k,t} \right) \left(\frac{\partial y_{i,m,t}}{\partial \boldsymbol{\theta}_V} - \sum_{k=1}^{M} \frac{\partial y_{i,k,t}}{\partial \boldsymbol{\theta}_V} \right) \qquad (3.12)$$

This can be re-written in terms of individual models and means:

$$\frac{\partial C}{\partial \boldsymbol{\theta}_V} = \frac{2}{M} \sum_{t=0}^{T} \sum_{i=1}^{N} \left[\sum_{m=1}^{M} \left(y_{i,m,t} \frac{\partial y_{i,m,t}}{\partial \boldsymbol{\theta}_V} \right) - \frac{1}{M} \left(\sum_{k=1}^{M} y_{i,k,t} \right) \left(\sum_{k=1}^{M} \frac{\partial y_{i,k,t}}{\partial \boldsymbol{\theta}_V} \right) \right] \qquad (3.13)$$

or decomposed into self- and mixture-terms:

$$\frac{\partial C}{\partial \boldsymbol{\theta}_V} = \frac{2}{M} \sum_{t=0}^{T} \sum_{i=1}^{N} \sum_{m=1}^{M} y_{i,m,t} \left[\frac{M-1}{M} \frac{\partial y_{i,m,t}}{\partial \boldsymbol{\theta}_V} - \frac{1}{M} \sum_{k=1,k\neq m}^{M} \frac{\partial y_{i,k,t}}{\partial \boldsymbol{\theta}_V} \right] \qquad (3.14)$$

We normalize the consensus sensitivities to the parameter values at which they are computed:

$$\frac{\partial C}{\partial \log \boldsymbol{\theta}_V} = \boldsymbol{\theta}_V \frac{\partial C}{\partial \boldsymbol{\theta}_V} \qquad (3.15)$$

3.3 Results

3.3.1 Model averaging provides time-dependent confidence information

Using our simulation framework, as described in the Methods, the actual system was selected as Model 1 or Model 4. The model was integrated then sampled and noise added to create data, and each of six topologies was parametrized to the data to generate an ensemble of fit models. This was repeated for 50 different data sets for each choice of actual system. The ensembles were perturbed in a Monte Carlo fashion and we developed and tested several metrics for identifying how consensus among the perturbed ensembles could be used to estimate the confidence of predictions. The first of these methods was the use of model averaging, combined with analysis of parametric uncertainty, as a means of computing

time-dependent confidence information.

Model averaging is a technique for combining the probabilities from a pool of experts into a unified probability distribution that accounts for the opinions of all the experts. In our case, each model of the ensemble was assigned the role of expert. The opinions of the experts can be weighted evenly in the case of no prior favoritism, or can be weighted based on past performance—models that have had prior success in assessing predictions may weigh in more heavily—or they may be weighted in other (possibly heuristic) ways. The procedure is similar in concept to a vote among models, though not completely analogous, as the "opinions" expressed are continuous rather than discrete.

Here we have assumed that the probability distribution for an individual model is Gaussian in protein concentration for any given time t. The mean of this Gaussian is the solution of the deterministic ordinary differential equation associated with the model. The variance is computed as the statistical uncertainty in parameters due to the fitting to noisy data, propagated to the point of perturbation by multiplication with the sensitivities. The Gaussian probability distributions are generated at each relevant time t for each model in the ensemble. The individual distributions are then combined according to the method of weighted logarithms to generate a single probability density plot for the ensemble, showing likelihood of a particular protein concentration as a function of time. This procedure is outlined in Figure 3-1; the results illustrate probability trajectories that demonstrate uncertainty due to incomplete knowledge of parameter values and topology.

To assess the ability of model averaging to capture the real uncertainty in predictions from the ensemble, we compared the resultant probability distributions to the variability that would have arisen either from parametric uncertainty within a particular topology, or from topological uncertainty across the models in the ensemble. In Figure 3-2, the leftmost column displays the probability distributions generated through averaging models 1 through 6 with equal weighting. The color at each point on the plot indicates the probability density that the protein of interest is present in the cell at that concentration, at that time. The solid and dotted lines correspond to the correct answer—the prediction from the actual system—and the maximum probability (MP) prediction according to the averaging procedure, respectively.

The second column shows the superposition of simulations from 50 ensembles of six models that were each fit to a different noisy data set from Model 1 as the actual system. The predictions from topologically distinct models are each plotted in a different color. Together, the simulations encapsulate the uncertainty from intra-model (parametric) and inter-model (topological) sources. The third column consists of two-dimensional histograms of the data in the second column to demonstrate the density of simulations. Each row represents a different specific perturbation from the set of Monte Carlo experiments. These perturbations are representative of the pool of simulations and were not selected because they were the most or least successful implementations of the model averaging; rather, they each demonstrate a different possible result for the averaging procedure. The top row corresponds to a perturbation of $\Delta \log_{10} C_o = 1.8331$ and $\Delta \log_{10} \gamma_A = 0.5327$ and represents a case where the models agree best with each other at intermediate time values but are less consenting at early and late times. The second row ($\Delta \log_{10} B_o = 1.7803$, $\Delta \log_{10} C_o = 1.8724$ and $\Delta \log_{10} k_C = -0.6456$) is a case where the models agree well with each at all times examined, and the MP trajectory is highly consistent with the actual system. The third row ($\Delta \log_{10} A_o = 0.9582$, $\Delta \log_{10} B_o = 0.3301$ and $\Delta \log_{10} C_o = 0.9562$) represents a case where the MP trajectory as predicted by model averaging is not particularly accurate in terms of capturing the concentration from the actual system, but predicts high confidence in the protein concentration at early and late times.

The goal of the model averaging is to capture in a probability distribution the variability expected due to the intra- and inter-model uncertainty and therefore we would expect that the model averaging in column 1 should reproduce the intra- and inter-model uncertainty shown in columns 2 and 3. For our particular test cases, the model averaging was successful in meeting this goal. The MP trajectories do not necessarily agree with the corresponding trajectories as predicted by the actual system. Nevertheless, in comparison to column 3, the method has accurately assessed the overall uncertainty and performs well in terms of identifying where the models are consistent in their predictions: high-confidence regions in the probability density plots correspond well with the regions of high-occurence in the multiple-data set simulations. It is also noteworthy that where the MP prediction is most accurate, i.e. in greatest agreement with the actual system, the assessed probability is

higher.

We computed an overall consensus score for predictions by integrating the probability density of the MP trajectories over time, summed over protein species A, B, and C. The consensus score serves as an internal check of the MP predictions. It is a self-assessment of the accuracy of the MP predictions, under the assumption that high levels of consensus equate to high quality predictions. Consensus scores are highest when the parametric uncertainty of individual models is low and the models also agree with one another. A low score can imply one of three scenarios: either the individual models agree but have low confidence in their predictive ability, or the models are individually confident but disagree, or, in the worst case, models are both dissonant and lacking confidence.

We subjected the 50 fit ensembles (each resulting from a different noisy data set) to 1000 perturbations in the form of parameter value changes. We calculated the consensus scores, P_C, and errors of MP prediction for each and found a negative correlation (Figure 3-3). Of particular note, predictions for which the consensus score is high are very unlikely to be poor: when the score assigns high-confidence to the MP prediction, that prediction is likely to agree well with that from the actual system. It is not necessarily true that an MP prediction with a low consensus score is inaccurate, similar to the way that a sample from a wide distribution still has some probability of lying very near the mean. Although the results shown in Figure 3-3 are a superposition of 1000 perturbations, the anti-correlation is not a result of the superposition. The same perturbations to a single ensemble of fit models yields a very similar distribution (Supplementary Figure B-1).

We tested the impact of two different weighting choices on the consensus scores: the first was equal weighting (Figure 3-3A) and the second was weighting by distance from consensus (Figure 3-3B), which gives models an extra boost in the vote if their prediction was very near the overall mean prediction of the ensemble. The weighting choice bears little effect on the relationship between consensus score and MP prediction error. The correlation coefficient for the errors of the two weightings is 0.9292 (Supplementary Figure B-2).

3.3.2 Decision trees for partitioning the space of perturbations into regions of high and low confidence

For random perturbations to the ensembles of fit models, we computed the ensemble variance to assess the degree to which there was consistency among model predictions. Perturbations consisted of changes to the values of one to three parameters chosen from a possible set of seven (see Methods). As shown in Figure 2-8 and described in Chapter 2, it was previously found that when models in an ensemble have low variance, it is rare for the average prediction to be inaccurate. With this result in mind, we built decision trees in which the predictor variables were the logfold parameter changes used to generate the perturbations, and the response was the ensemble variance. The resulting trees were used to predict consensus of an ensemble given the parameter values corresponding to a perturbation.

Figure 3-4 shows the decision tree built on a single data set derived from Model 1 as the actual system, at the 30% noise level. The color of each node, which varies from bright yellow to dark red, indicates the average value of the ensemble variance for all simulations belonging to that node. (This value is also printed inside each node, along with the node number.) The colors of the branching arrows indicate the dimension of parameter space in which the cut at that level was made. For instance, the first cut is in $\Delta \log_{10} k_B$ and all other cyan arrows represent cuts related to this parameter.

With node color as a visual guide, it is clear that different branches of the tree correspond to perturbations with differing average levels of consensus. The leftmost leaves in the tree [1] correspond to regions of parameter space in which the models agree poorly with each other. Simulations belonging to leaves on the rightmost branch of the tree have low ensemble variance. Because the simulations associated with these nodes have high consensus, they are more likely to be accurate.

In Figure 3-5A, we have tabulated the ensemble variance of all simulations, and, separately, only of those simulations in the best or worst leaves of the tree. For this particular example, there were 1000 perturbations in all, 561 of which were included in Node 64 (and for which the mean ensemble variance was 2.621), and 7 of which were members of Node 8

---

[1]Terminal nodes are referred to as *leaves*.

(with a mean ensemble variance of 585.41). The range of variances seen in all simulations is over three orders of magnitude wide, from nearly zero to over 600. The variances of simulations in the best node range up to 42.063, and the simulations in the worst node all have variances over 550. The variance for the models at the fit (prior to perturbation) was 0.593.

An important question in regards to the simulations in the best node is whether they have high consensus simply because these represent weak (i.e. small, insignificant, or trivial) perturbations. It is known that the six models agree well in their predictions at the fit to the data, so if the models are subjected to small perturbations, or large perturbations in insensitive directions, and the models do not stray far from the fit, it is likely that the predictions will still be accurate. That this might be true of some simulations does not negate the quality of these predictions, but more interesting biological insight might come from highly accurate predictions that are substantially different from the training data.

To examine whether there were interesting predictions in the best tree node, we generated a histogram of the squared Euclidean distance between post-perturbation predictions and the data to which the models were fit (Figure 3-5B). Although there are many predictions belonging to the best node for which the distance to the data is small, the overall distribution of distances is quite similar to that of the entire pool of simulations across the tree. As a reference, the average distance from the data to the mean ensemble fit model for this same experiment was about 3, whose base 10 logarithm is 0.477 (data not shown). In particular, there are a number of predictions for which there is both high consensus and a very large distance between the average ensemble prediction and the data.

Two such simulations are shown in Figure 3-6. The top row corresponds to a perturbation of $\Delta \log_{10} C_o = 1.8331$ and $\Delta \log_{10} \gamma_A = 0.5327$; the bottom to a perturbation of $\Delta \log_{10} B_o = 1.7803$, $\Delta \log_{10} C_o = 1.8724$, and $\Delta \log_{10} k_C = -0.6456$. (These are in fact the same perturbations shown in the top two rows of Figure 3-2.) The feature that high consensus/large distance-from-data simulations share is a positive perturbation to the initial concentration of protein C. This is unsurprising because the six genetic regulatory networks under consideration are linear, with protein C at the bottom of the reaction chain. Changes to the initial concentration of protein C bear no effect on the concentrations of proteins A

and B and provided the degradation constants k_C for each of the fit models are not too different, there should be good agreement between the resulting time-dependent trajectories from all models.

The mean ensemble predictions in the worst node of the tree are neither very far, nor very close, to the data. When the ensemble variance is high, however, it may not be meaningful to consider the mean ensemble prediction because the models may have very different opinions about the trajectories. If there is no good reason to favor one model over another *a priori*, it may not be possible to evaluate which model (if any) is correct. If an experiment were to be performed corresponding to the perturbations suggested by these high variance nodes, it would not be possible for all of the ensemble models to agree with the data. (If they cannot agree with each other, they cannot possibly all agree on the same data.) The parameter choices corresponding to these perturbations then make excellent candidates for further experimentation, as they will allow models to be discarded.

To demonstrate this concept, we took the six fit models that were used as the basis for building the decision tree, and we subjected them to a perturbation that would be suggested by the first two cuts in the tree: $\Delta \log_{10} k_A = -0.125$ and $\Delta \log_{10} k_B = -0.600$. The results of this simulated experiment are shown in Figure 3-7. The top row shows the fits of all six models along with the data to which they were optimized, while the bottom row shows the post-perturbation predictions from each model. In particular, while all other models predict a decrease in the concentration of protein A over the experimental time, Model 2 predicts upregulation. To adjust for the feedback in Topology 2, the corresponding fit estimated a very different value for the degradation rate constant of protein A compared to the other models. In resetting the value of this constant, Model 2 is no longer able to accurately predict the system behavior.

Regarding the predicted trajectories of protein B, the perturbation experiment leads to a strong disagreement between Model 2 and Model 5, while Models 1, 3, 4, and 6 would probably not be distinguished due to the differences in predictions being less than the experimental error. An accurate time course experiment measuring the concentration of protein C might be able to eliminate Models 2, 5, and 6 from the pool of candidate models. However, when we simulated the actual collection of data with 30% noise (equal to the level

of noise in the simulated data used for training the models), the differences in predictions were insufficiently large compared to the noise (Figure 3-7, bottom right).

3.3.3 Consensus sensitivity: A novel metric for identifying the separatrices of consensus space

Using the decision tree, we were able to identify regions of parameter space where the ensemble predictions were both in good agreement among themselves but differed greatly from the training data. The implication was that although individual models might be sensitive to a particular perturbation, the consensus of the ensemble may be insensitive to this same change. We defined a new metric called the consensus sensitivity in which we compute the sensitivity of the ensemble variance to parameter changes. For the set of 1000 perturbations to a single ensemble of fit models (Model 1 as the actual system, 30% noise), we calculated the consensus sensitivities to each of the variable parameters (Figure 3-8).

The magnitudes of the sensitivities make it clear that some parameters have a much greater impact on the consensus than others. The initial concentration of protein A has the ability to separate models when it is large. This is intuitive because the A protein is at the head of the linear genetic regulatory network for all six topologies considered. Both the initial concentration of protein B and its degradation rate constant may also have a significant effect on the ensemble consensus. This again is unsurprising considering that many of the differences among the candidate topologies relate to the position of protein B in the network or whether or not it is autoregulated through positive feedback. The consensus sensitivities to the degradation rate constant for protein C, k_C, are particularly small because protein C is the output of the system and no other species lay downstream to be affected by changes to its expression.

The most striking features of these plots are the minima for the sensitivities to k_A and k_B. Interestingly, these minima coincide very well with the values of $\Delta \log_{10} k_B = -0.45$ and $\Delta \log_{10} k_A = -0.04/-0.10$, the first two levels of cuts chosen by the decision tree built on the same simulations. The consensus sensitivities also provide a means for identifying these separatrices. We constructed a tree and performed consensus sensitivity analysis with Model 4 as the actual system rather than Model 1, and observed the same type of pairing

(see Supplementary Figures B-5 and B-6). The decision tree is built according to a greedy algorithm: each cut is made to generate the best improvement in the tree cost, without consideration of the impact to successive cuts. The first few cuts to the tree define the major axes of separation between regions of parameter space where the consensus is high and those where it is low.

3.4 Discussion

Using consensus as the basis for our approaches, we tailored existing means and developed new ones for analyzing the quality of predictions from ensembles of ODE models. Our test case consisted of six very simple genetic regulatory networks that follow the same Boolean logic. The data for training and testing models was simulated so that we were able to evaluate the performance of the different metrics under consideration.

The first use of consensus we described here was as the underlying basis for model averaging. Each model of the ensemble acted as an expert, providing an opinion about the time-dependent protein concentrations. These opinions were framed mathematically as Gaussian probability densities with means set to the deterministic values predicted by each model, and variances equal to the propagated parametric uncertainty. Probabilities from individual models were combined logarithmically and the maximum probability (MP) prediction was identified, along with its associated consensus score.

For our simple test case, model averaging performed very well in capturing the true variation in predictions resulting from uncertainty in both parameters and topology. In previous work, (see Chapter 2), we found that parametric uncertainty was insufficient for evaluating prediction error when topology was also undetermined. Model averaging as we have done then presents a method for combining both sources of error and capturing that information in a single probability density function.

We deliberately chose to use simulated—rather than real—data and very simple test systems to demonstrate the utility of the method. There are many possible scenarios, however, where adjustments to the method might have to be made to deal with the complexities of more realistic systems. For instance, the Fisher information method of estimating pa-

rameter uncertainty, as used here, may not accurately assess the contribution of the fitting procedure to prediction error within a single model topology. First, the Fisher information used in the calculation is only an estimation of the true uncertainty matrix, which is too unwieldy to compute, and, even if we could derive the actual Fisher information, it is only a lower bound on the true variance [90]. Secondly, the sensitivities used in the calculation are only valid locally and some evaluation of the smoothness near a prediction may be required to establish the appropriateness of the assumptions. Thirdly, our assumption of normality for prediction uncertainty is also likely to break down in larger, more non-linear systems.

Rather than make so many assumptions—about the quality of the Fisher information estimate, about the smoothness, and about Gaussian distributions—one could try to reproduce column 3 of Figure 3-2 using a bootstrap method [56, 96]. The simulations in column 2 were derived from ensembles fit to 50 data sets that differed in their noise alone. A non-parametric bootstrap approach might try to recreate these curves by sampling a single data set (with replacement) many times. An ensemble would be generated for each sampled data set and predictions would analyzed for overlap among all ensembles. A major drawback to this approach would be the computational expense in increasing the number of model optimizations. Whereas the model averaging approach introduced here requires only a single optimization for each topology in the ensemble, the non-parametric bootstrap would require optimization for each topology, for each sample of the data.

A parametric bootstrap might also be considered with the re-introduction of some assumptions. Once models are fit to a single data set, and a Fisher information matrix (FIM) is estimated, multiple simulations of each topology can be performed by sampling the parameters from Gaussian distributions described by the FIM. This method again comes at a computational cost due to the increased number of integrations of the differential equations. In general, this cost is significantly less than that related to the optimization.

There are a number of different choices in the implementation of model averaging that have not been fully explored. We chose to use the logarithmic opinion pool as the method of combining the probabilities from individual models. An alternate formulation is that of the linear opinion pool in which probabilities are added directly, without logarithmic normalization. The logarithmic pool was selected primarily because of the algebraic simplifications

to the maximum probability prediction function. We did not perform a full analysis of the differences between linear and logarithmic pools for capturing the true prediction variability, although individual distributions from the two methods were qualitatively very similar (see Supplementary Figure B-3 for an example comparison).

We tested two different weighting methods, equal and distance from consensus, which were similar in terms of their performance. There are, of course, many other reasonable weighting choices. If the weights are set to the true model priors, $P(M)$, then our method is equivalent to Bayesian Model Averaging [123,125]. The difficulty here is in the calculation of these priors, which requires integration over all accessible parameter space:

$$P(M) = \int P(M|\boldsymbol{\theta})P(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{3.16}$$

Even the approximation of this integral through Monte Carlo simulation is time-consuming and the computation is exponential in the number of fit parameters. Signal transduction pathway models, as an example, may have on the order of 50 or more unknown parameters that are optimized to the data [74,140].

Another statistically-based choice would be that of Akaike weights [52]. The Akaike Information Criterion (AIC) [51] (see Chapter 2) is computed for each topology, and the lowest AIC of the candidate models is subtracted off the total for each:

$$D_i = \text{AIC}_i - \text{AIC}_{\text{min}} \tag{3.17}$$

The weight associated with any model i is taken as

$$\lambda_i = \frac{\exp(-\frac{1}{2}D_i)}{\sum_{m=1}^{M} \exp(-\frac{1}{2}D_m)} \tag{3.18}$$

The MP trajectory of protein concentration resulting from the model averaging may not necessarily capture the true dynamics better than any single model; furthermore, it will not always be the case that high consensus among models is indicative of predictive accuracy. In any situation, there exists a possibility that all of the models will be making the same wrong prediction. That said, the model averaging procedure is nonetheless useful

for describing the variation of predictions that result from the parametric and topological uncertainties, as they have been assessed, and for providing information about where that variability is large or small.

We developed a consensus score that integrates the probability associated with the MP trajectory which is helpful in identifying those predictions where there is good agreement among models and a low level of parametric uncertainty. One shortcoming is that the score is relative rather than absolute. There is no threshold above which one can say for certain that a prediction is accurate. In a drug research setting, this may nevertheless be sufficient for ranking predictions in terms of their confidence given the current amount of information.

Our second application of consensus was the use of decision trees to partition a space of perturbations into regions or high or low ensemble variance. The trees additionally provide information about the parameters with the most impact on consensus, those being the first few cuts of the tree. Once parameter space has been partitioned, it becomes dually useful. Regions corresponding to low ensemble variance can be used to identify higher-confidence predictions. Regions of high variance suggest further experiments for testing and refining the models of the ensemble. In the process of ensemble refinement, it is important to recognize that elimination of a model from the ensemble does not necessarily imply that the topology of that model is incorrect. A further investigation of feasible parameterizations should be considered before assuming that a mismatch to new data has real biological implications.

A possible concern with decision trees is the robustness of the greedy algorithm to noise in the data. In our test case, this was not a problem, as a second tree, built on a data set with different noise, looked very similar to the original one (Supplementary Figure B-4). In practice, the random forests technique [141] could be applied to build tree-based classifiers that are less sensitive to noise.

In constructing one's own trees, consideration must be given to the particulars of the Monte Carlo perturbations used to generate the ensemble variances. If experimental design is to be extracted from the tree, then perturbations should only consist of changes that can be applied through real experimental manipulation. The sampling rate in each dimension may depend on the smoothness of the consensus surface. Our third tool, the consensus sensitivities, can provide assistance in this respect.

The consensus sentivities are a new metric for analyzing the dependence of the ensemble variance on parameters. The sensitivities at any specific point may not be particularly informative but when measured across parameter space, they map out the relative importance of the parameters when they take on different values, and define separatrices for regions of high and low consensus.

A key feature of all three methods discussed here is that there is no requirement for the correct model—that which represents the actual system—to be in the ensemble. They simply state whether the existing members of the ensemble agree or not, and attempt to evaluate the uncertainty of any predictions using the information contained in this agreement.

3.5 Conclusions

The work presented here provides a proof-of-concept of three different approaches, all related to consensus, for investigating prediction quality from differential equation models. With model averaging we can either obtain time-dependent information about our confidence in the prediction of protein concentration, or compute a cumulative score for the weighted average prediction as a whole. Decision trees can be built to relate parameter changes to ensemble variance and delimit parameter space into regions of high or low consensus. Low consensus regions can provide experimental leads for improving our knowledge of reasonable models; high consensus regions suggest high-confidence predictions. Consensus sensitivities can provide a continuous means to analyze what types of perturbations lead to high or low ensemble variance.

There are many issues that will likely need to be studied further before these methods can be practically used on larger, more realistic systems. We did not address the topic of initial topology selection which may not be trivial. For now, the topologies can be built manually through careful mining of the biological literature to gain an understanding of which aspects of a given network are well-established, and which may be a source of mechanistic uncertainty. Alternatively, reverse engineering methods that try to build topology from scratch given biological data [35, 86] often provide a ranked list of probable topologies,

which would serve as a starting point for the ensemble.

Another topic to be examined is the role of the choice of selected outputs. With our simple test case, we measured proteins A, B, and C, thereby providing dense information for fitting the models. In a large system, the network is typically probed very sparsely. As a result, models in the ensemble might be overfit and likely to diverge when the system is subject to perturbation. In particular, in branched pathways, where an output may be receiving flux from diverse sources, it may be difficult to establish the balance across these branches. Differently balanced models may be able to fit the output equally well, but will make very different predictions about system response to perturbation of a particular branch. Therefore, it may be helpful to do a consensus analysis of the fluxes, in addition to the outputs.

Although a substantial amount of development remains, applications of consensus to understanding the predictions from biological models hold much promise. Our future work will apply these techniques to the Fas signaling network in apoptosis, discussed in detail in Chapter 4.
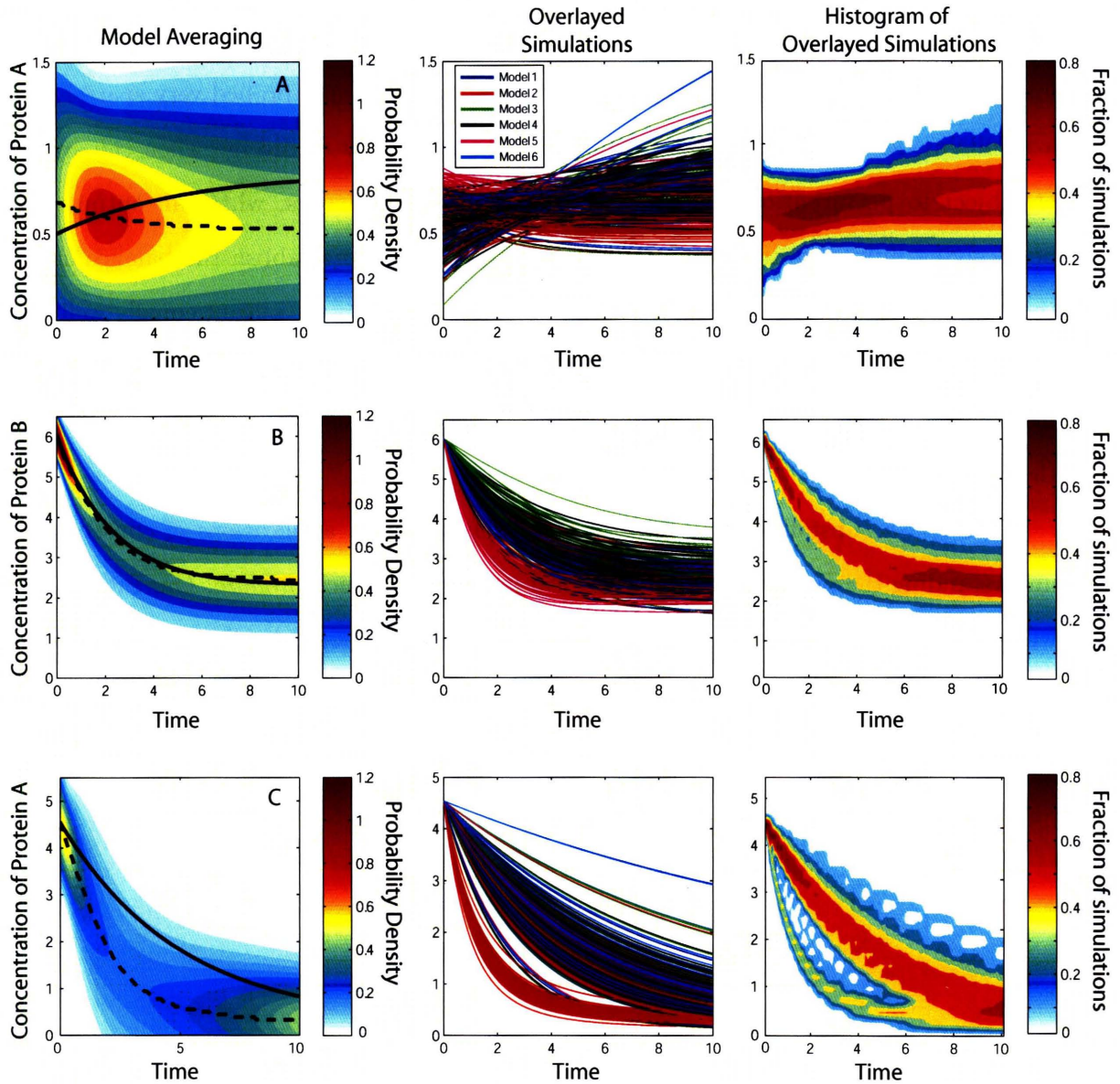
Figure 3-2: Comparison of model averaging to the actual prediction variability. To assess the performance of model averaging for capturing real uncertainty, we simulated 50 noisy data sets, generated an ensemble of fit modesl for each, the applied the same perturbation to all models. Each row represents one such perturbation: (A) $\Delta \log_{10} C_o = 1.8331$ and $\Delta \log_{10} \gamma_A = 0.5327$, (B) $\Delta \log_{10} B_o = 1.7803$, $\Delta \log_{10} C_o = 1.8724$, and (C) $\Delta \log_{10} B_o = 0.3301$ and $\Delta \log_{10} C_o = 0.9562$. The first column shows the probability distributions generated through the logarithmic opinion pool. The second column shows the time courses generated by all 300 perturbed models (50 data sets × 6 models) and the third column shows the corresponding density of simulations.
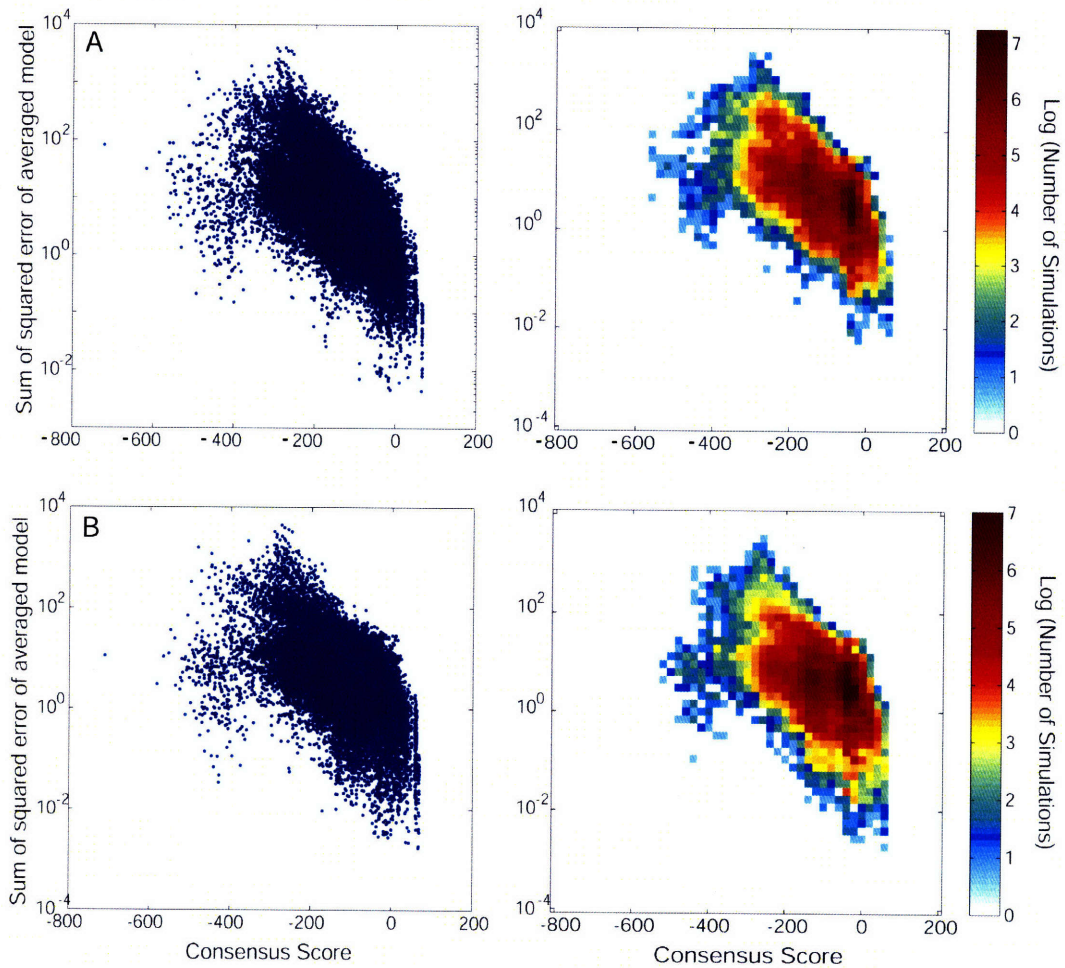
Figure 3-3: Error of the maximum probability time course vs. cumulative consensus scores, P_C, from model averaging. Consensus scores were computed by integrating over time the value of the probability density at the maximum probability concentration. Scores were additionally summed over species A, B, and C. High consensus scores are rarely associated with large errors though highly accurate predictions may have a low score. Two different model weighting schemes were tested: equal (top), and distance from consensus (bottom) in which models were proportionally assigned more weight if they were closer to the ensemble average. The right column consists of two-dimensional histograms of the data on the left.
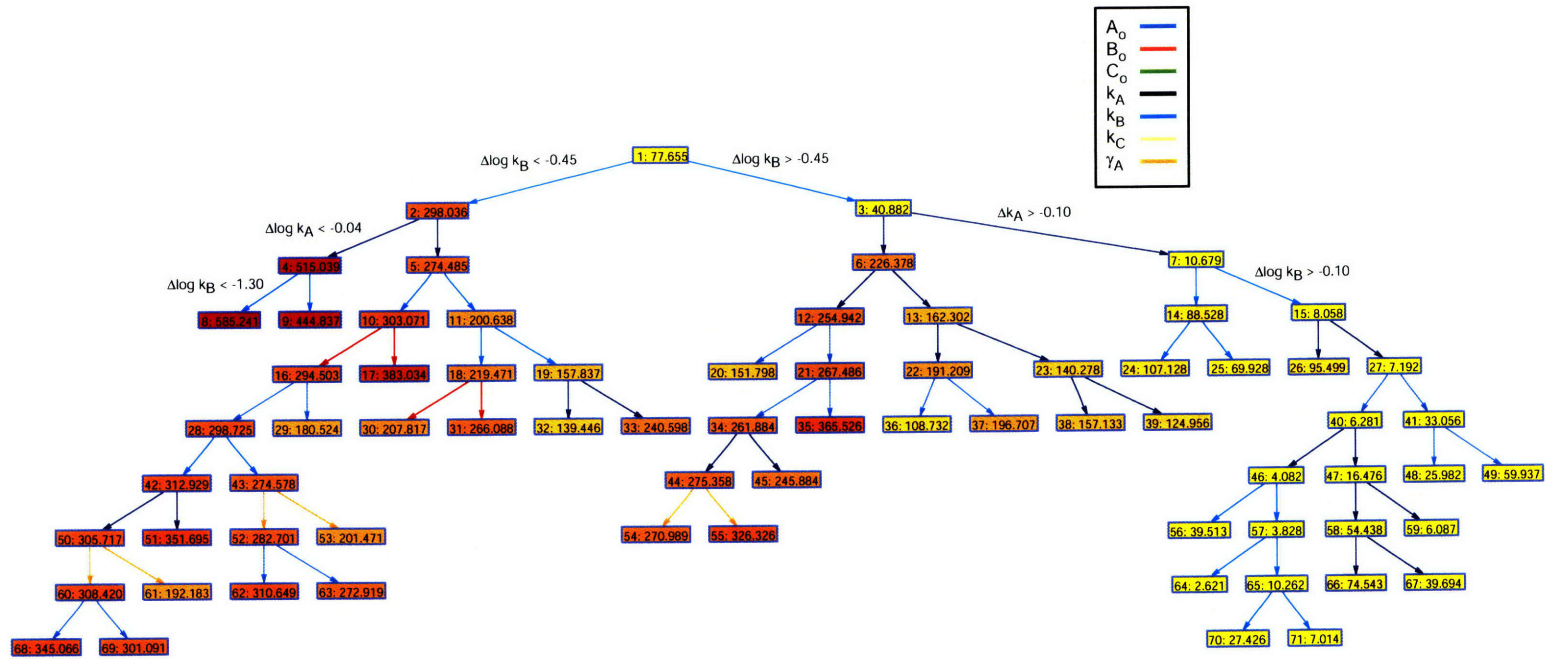
Figure 3-4: Regression tree built to partition prediction space into regions of high and low consensus. This tree was built using the set of 1000 parameter perturbations as predictor variables and the resulting ensemble variances as the response variable. Red nodes correspond to perturbations for which there is high ensemble variance and the model predictions are the most dissimilar; yellow nodes correspond to perturbations for which there is good consensus among the models in the ensemble. The colors of the arrows indicate the dimension of parameter space in which that particular cut was made (see legend for specific assignments). Nodes are labelled with an identity, e.g. Node 6, and the mean ensemble variance of simulations belonging to that node.
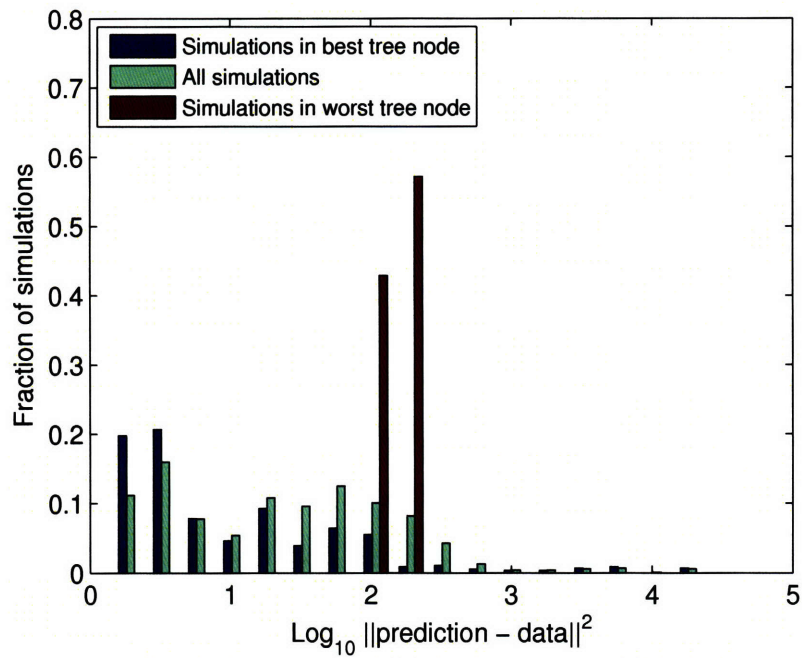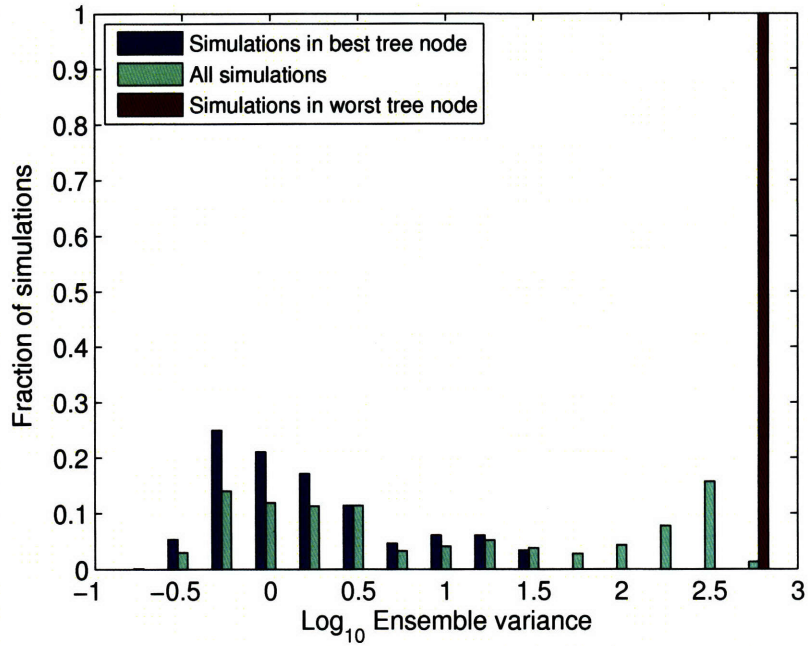
Figure 3-5: Comparison of the properties of simulations at opposite ends of the decision tree. The ensemble variance and the mean squared Euclidean distance of ensemble predictions to the data were computed for all 1000 simulations. The simulations were then sorted according to the decision tree in Figure 3-4. 561 simulations were assigned to the node of lowest ensemble variance while 7 were assigned to the node of highest variance, (the "best" and "worst" nodes, respectively). Simulations in the worst node invariably have very poor consensus while those belonging to the best node are on average of higher consensus than the entire pool of perturbations considered. Many of these high consensus predictions are trivial (in that they lay very close to the training data), although a select number of perturbations result in consistent model predictions very far from the data.
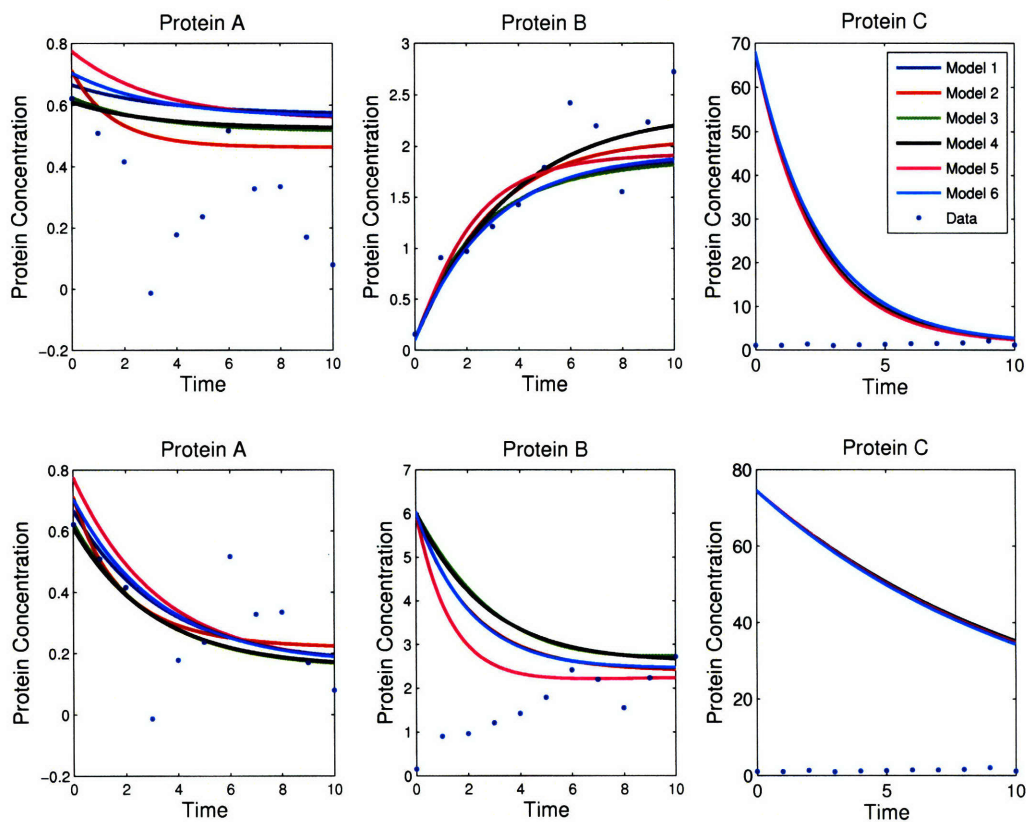
Figure 3-6: Protein time courses for two perturbations resulting in both high consensus and a large distance between the mean ensemble predictions and the training data. The perturbations applied here correspond to the same perturbations shown in rows A and B of Figure 3-2.
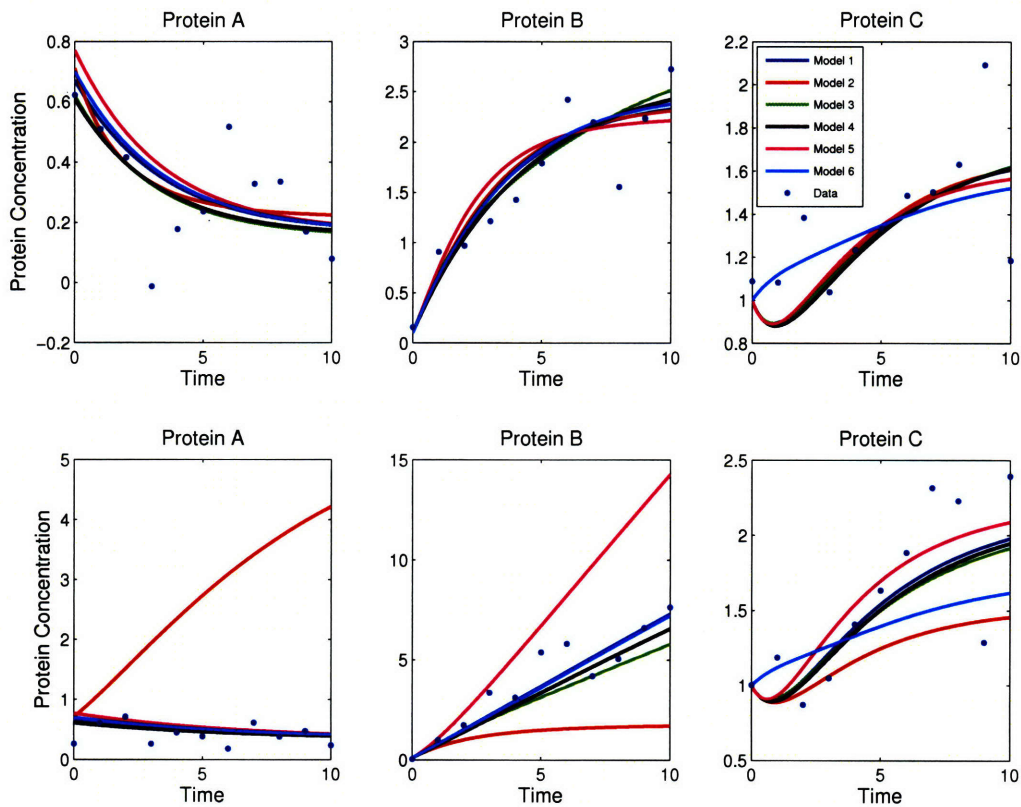
Figure 3-7: An example simulated experiment performed at the suggestion of the decision tree. All six topologies were fit to data generated by Model 1 as the actual system with 30% noise (top row). All fit models were then perturbed using $\Delta \log_{10} k_A = -0.125$ and $\Delta \log_{10} k_B = -0.600$. The time courses for the perturbed models show considerable disagreement and, in particular, models 2 and 5 make very different predictions than models 1, 3, 4, and 6 (bottom row). Data from a real experiment were simulated using the perturbed actual system with 30% noise and show that measurements of proteins A and B would be more informative for distinguishing between models than those of protein C.
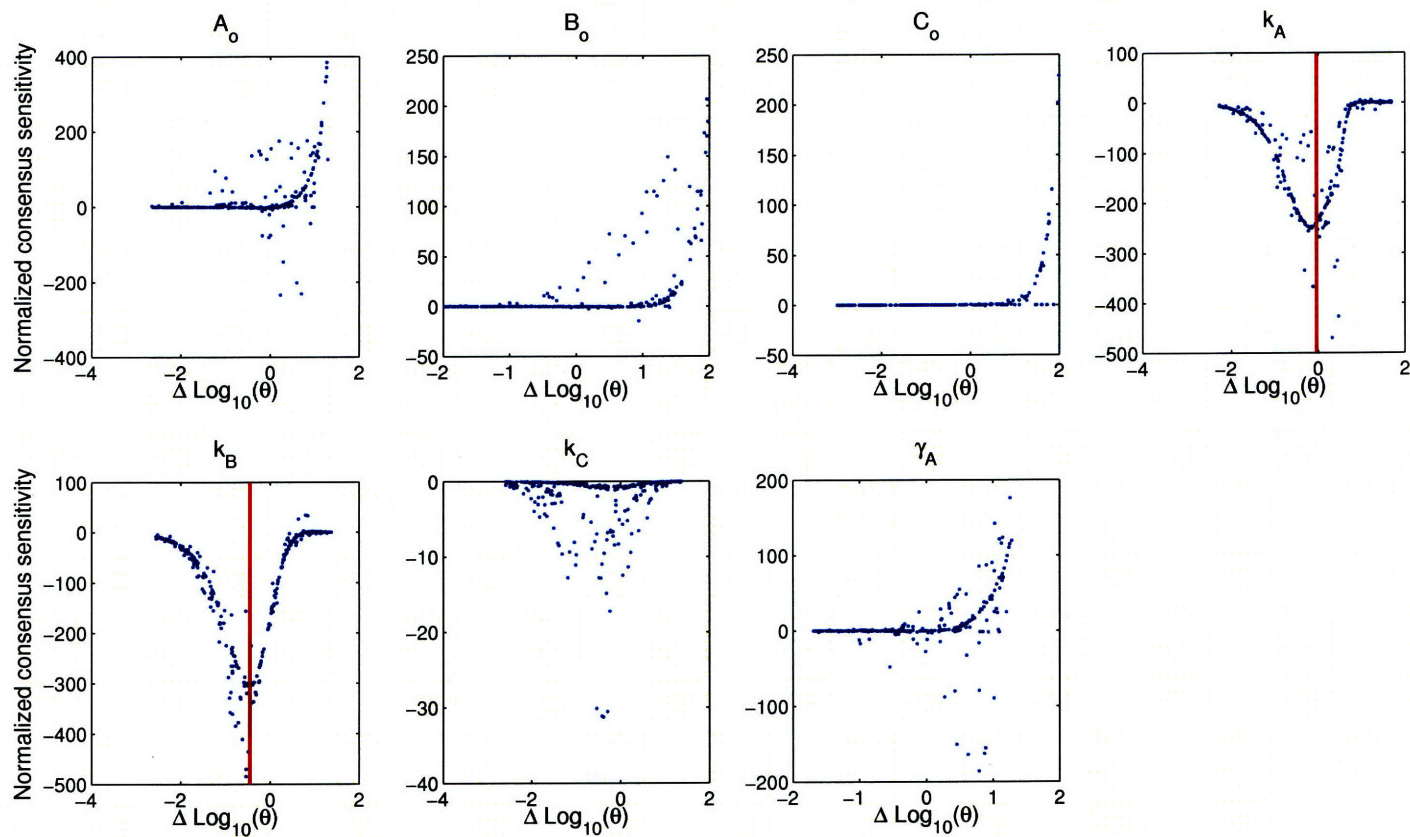
Figure 3-8: Normalized consensus sensitivities for the Monte Carlo perturbation experiment of the six genetic regulatory networks. The consensus sensitivities to the seven variable parameters were computed for each of the 1000 perturbed ensembles. The magnitudes of sensitivities vary significantly from parameter to parameter, with k_A and k_B as the overall most impactful determinants of consensus. The vertical red lines indicate the values of k_A and k_B corresponding to the first two splits in the decision tree built on the same set of perturbations (see Figure 3-4).

Chapter 4

An ensemble of Fas signaling network models to elucidate pathway function and design an optimal cocktail of caspase inhibitors

Abstract

Apoptosis, or programmed cell death, is a critical regulatory process that plays an important role in many organismal functions, from development to homeostatis. Dysregulation of apoptosis resulting in either too much or too little cell death is implicated in a wide range of diseases; a systems-level understanding of the underlying biochemical network is therefore highly sought-after. In this regard, there has been considerable interest in developing differential equation models of the Fas signaling network. While these models can provide detailed answers to mechanistic biological questions, they also require detailed specification at the level of topology and parameters, both of which are subject to uncertainty. Here, we examine the role of topological choice in predicting the response of overstimulated cells, such as would be observed in sepsis, to potential therapies. We created an ensemble of 64 models through combinatorial expansion of a base network and subjected this ensemble to perturbations in the form of simulated treatment with cocktails of caspase inhibitors. The ensemble variance at each point of perturbation was used to train a decision tree for classifying consensus based on treatment. Using the tree, we were also able to solve two design problems: (a) selection of the optimal inhibitor cocktail for reducing cell death due to apoptosis, and (b) selection of the optimal inhibitor cocktail for refining the model ensemble.

4.1 Introduction

Apoptosis is the process by which cells commit suicide in response to internal or external cues [73]. It is alternately known as programmed cell death because it is governed by a tightly-controlled network of interacting proteins. The primary players in this network are the cysteinyl aspartate-specific proteases (caspases [142]) and the Bcl-2 family of proteins [143, 144], the latter of which interact with the caspases and with each other at the mitochondrial membrane. Cell death regulated by apoptosis plays a positive, and indeed necessary, role in many physiological processes including development, homeostasis, and viral defense. Dysregulation of apoptosis, on the other hand, is the root cause of many diseases, either through absence or overabundance [145-148]. Cancer cells often lose their ability to undergo apoptosis, leading to uncontrolled proliferation and tumor growth [149]. Other afflictions, such as sepsis [150], stroke [151], and neurodegenerative disorders [152], are characterized by unconstrained death when cells become hypersensitive to apoptotic stimulus or when cells are presented with an excess of death signal.

A fundamental comprehension of the apoptotic pathway is critical to our ability to address the causes of these diseases and develop relevant treatments. Due to the complexity of the core protein network, mathematical modeling may be the best way to synthesize our knowledge of the individual components into a systems-level understanding. There are indeed several published differential equation models of the Fas signaling network [74-77, 153-161]. These models differ considerably in scope, detail, methodology, and focus; a summary is provided in Table 4.1.

The first major model of apoptosis by Fussenegger et al. [153] was developed as an exploratory tool for studying the network and investigating possible therapeutic interventions. Their model was parametrized with nominal values and their reaction kinetic equations incorporated several assumptions about rate-limiting steps. At the time of publication (2000), it was a fairly complete model capturing the dynamics from receptor–ligand binding to executioner caspase activation, through both caspase-8 and caspase-9. They did not include a direct path from caspase-8 to membrane permeabilization (which is included in more recent models through the mediator Bid) and rather caspase-3 activation was required to occur

before activation of caspase-9. They examined how over-expression or mutation of various proteins (either alone or in pairs) would affect apoptosis and concluded that it is more difficult to inhibit apoptosis than to promote it. All of the perturbations considered were related to proteins intrinsic to the network rather than exogenous additions to the system.

The work of Bentele [74], published four years later, took advantage of the greater state of knowledge about the apoptosis network and furthermore tackled several important technical aspects of using differential equations to describe biology. The authors separated known biochemical interactions from those lacking mechanistic detail and used mass-action kinetics to describe the former and black-box modules with thresholding to model the latter. Rather than assign nominal values to unknown rate constants and initial concentrations, they performed a considerable number of Western blot time-course experiments, the data from which were used to fit parameters through a novel optimization procedure. They subsequently used their fit model to make predictions about the roles of IAP and FLIP in providing the system with thresholding behavior. In particular, they proposed that the ability of a cell to withold from apoptosis in the presence of small, non-zero ligand concentrations depended critically on the nature of receptor complex inhibition by FLIP.

Three further studies [155, 157, 161] have constructed comprehensive models that encompass both the Type I and Type II pathways in detail. The first of two models by Hua et al. [155] very explicitly modeled formation of the DISC, follow-on activation of caspase-8, and inhibition by FLIP. Their model was parametrized using values drawn from the biological literature as well as values fit by hand to a set of experimental time-course data (Western blots of procaspase-3 and procaspase-8). They used their model to test three different hypotheses about the mechanisms by which Bcl-2 might inhibit the mitochondrial pathway and analyzed the impact of over-expressing or knocking-down its expression. The second model from Hua et al. [157] differed from the first in a number of ways. They reduced the overall model size through several simplifications (notably, removing the explicit description of receptor complex formation as well as the mechanism of mitochondrial permeabilization) but also added some new interactions, such as feedback from activated caspase-3 to caspase-8 through caspase-6. Also, parameters in this model were fit to data using optimization. They performed Monte Carlo simulations of the mechanistic model using different choices

of initial concentrations and applied decision trees to clusters of the outcomes. The tree approach was used to analyze how the initial concentrations of various species resulted in either slow or fast activation of caspase-3.

The work of Albeck et al. [161] stands out for its extensive integration of experiments and computation. The authors used three means of measurement (live-cell microscopy, flow cytometry, and immunoblotting) to probe apoptotic activation in response to external stimulus, both at the single cell and cell population level. Both their model and their single-cell experiments show that under baseline conditions PARP cleavage by caspase-3 occurs in an all-or-none fashion, but that stochastic differences between cells lead to variance in the time-to-activation following stimulus. They also showed (again, with both modeling and experiment) that a more graded response could be obtained in individual cells under perturbation, for example, by downregulating Smac or by upregulating Bcl-2 while downregulating XIAP.

Many other mathematical models of apoptosis have not attempted to include both the Type I and Type II pathways, preferring to concentrate of the role of a particular subnetwork [75–77,154,156,158–160]. A significant subset of these papers (as well as some of the larger models) have operated on the assumption that the apoptotic network is bistable—that cells exist in either dead or alive states—and have analyzed the network for sources of this bistability. Various parts of the network have been cited as generating bistability including the interactions of XIAP with caspases-3 and -9 [156,158], inhibition of the DISC by FLIP [74], feedback from caspase-3 to caspase-9 [158], sequestration of caspase-8 by BAR [75], interactions between Bcl-2 and Bax [77], cooperativity in formation of the apoptosome [76], and degradation of caspase-3 [154]. Other network properties that have been assumed or investigated are irreversibility (cells cannot die and then return to life) [158,159] and robustness (small changes to rate constants or initial concentrations should not destroy the bistability and/or irreversibility) [69,76,77,158–160].

While it is interesting to consider the properties of small parts of the network, independent of the system as a whole, it is not necessarily the case that two subsystems, each individually bistable or irreversible, will retain their dynamic properties when joined in a single model. While the larger models that aim to encompass the biology of the entire

network are less prone to such issues, these are also inevitably subject to differences and uncertainties in topology which may or may not impact their overall dynamic behavior.

Here we examine the role of topological choice on predictions from differential equation models of the Fas signaling network. In particular, we are interested in designing cocktails of caspase inhibitors for reducing cell death in the overstimulating conditions observed in sepsis. To this end, we built an ensemble of 64 models derived from a base network and the combinatorial enumeration of six possible add-ons. We subjected the ensemble to over one thousand possible treatment options involving caspase inhibitors and built a decision tree based on the consensus among models in response to perturbation. The decision tree allowed us to simultaneously solve two design issues: first, we were able to identify treatment options that would reliably lead to alleviation of apoptosis, and secondly, we were able to design an experiment with caspase inhibitors for refining the ensemble of reasonable models. In addition, we investigated the role of flux balance in this branched pathway for generating the different observed model behaviors.

4.2 Methods

4.2.1 The Base Model

There are several aspects of the Fas signaling network that are both well-accepted by the biological community and are consistently represented in computational models. The path from Fas ligand stimulus to caspase-3 activation can be sub-divided into three main parts which together constitute our base model, a diagram of which appears as Figure 4-1.

Part I consists of the initiation of the pathway in response to stimulus. The first step is binding of Fas ligand to its receptor, Fas, at the cell membrane, followed by recruitment of FADD and formation of the death-inducing signaling complex (DISC). The DISC recruits procaspase-8, leading to its subsequent cleavage and activation. Activation of caspase-8 can be tempered by the presence of FLIP which competitively binds to the DISC. Part II involves cleavage of procaspase-3 by caspase-8. The direct activation produced through the joint action of Parts I and II is sometimes known as the Type I or extrinsic pathway of apoptosis [162, 163]. Part III consists of a more circuitous route to caspase-3 cleavage

Author	Year	Species	Pathway	Parameters	Focus
Fussenegger	2000	18	I/II	Nominal values	Effect of combined therapies
Bentele	2004	41	I/II	Fit to data	Parameter estimation/Threshold mechanism for activation by ligand
Eissing	2004	8	I	Literature/Nominal values	BAR as a source of bistability
Stucki	2005	10	II	Nominal values	Interplay of SMAC, IAP, and Casp3
Hua	2005	85	I/II	Literature/Adjusted by hand	Effects of changing levels of Bcl-2
Rehm	2006	19	II	Literature/Supplementary experiments	Casp3 activation from mitochondrion and control by XIAP
Hua	2006	29	I/II	Fit to data/Adjusted by hand	Integration of model-driven and data-driven approaches
Legewie	2006	13	II	Literature	Role of XIAP is bistability of the mitochondrial activation pathway
Bagci	2006	31	II	Literature/Nominal values/Other models	Bistability of the mitochondrial pathway
Chen	2007a	11	II	Nominal values/Drawn from other models	Bax activation as a source of bistability
Chen	2007b	6/8	II	Literature/nominal values	Robustness analysis to compare direct and indirect mechanisms for MOMP
Cui	2008	up to 9	II	Nominal/Drawn from other models	Comparison of 4 models of the Bcl-2 apoptotic switch
Albeck	2008	58	I/II	Literature/Nominal values	Modeling the all-or-none response of Casp3 activation

Table 4.1: Prior published models of the Fas signaling pathway, broken down by year of publication, number of species included in the model, whether the Type I and/or Type II pathways were modeled, source of parameter values, and focus of the analysis.
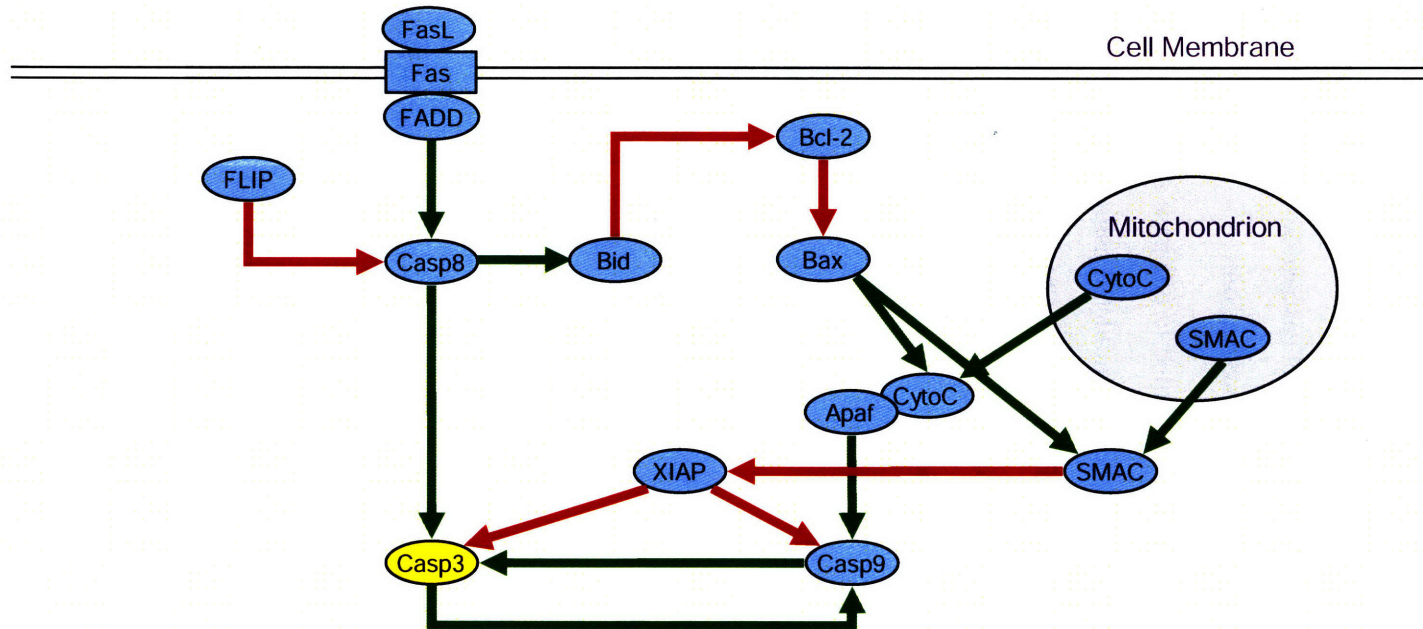
Figure 4-1: Base Model of the Fas signaling network. Green arrows represent an activating influence while red arrows represent inhibition.

following activation of caspase-8. This so-called Type II (or instrinsic) pathway of apoptosis is initiated when caspase-8 cleaves Bid, which then promotes formation of mitochondrial membrane pores by Bax, and leads to release of Smac and cytochrome c into the cytoplasm. cytochrome c subsequently assists formation of the apoptosome, a large multi-unit protein complex that facilitates activation of caspase-9. Activated caspase-9 can cleave procaspase-3, hence completing this path to activation.

The specific mechanisms included in our base model are described in more detail below. All reactions herein described were modeled using mass-action kinetics except for the release of Smac and cytochrome c from the mitochondria. For these reactions we assumed that the rate of release was proportional to the concentration of Bax pores, but that no complex was formed during the interaction.

DISC formation

Fas remains inactive until bound by Fas ligand (FasL) at which point they bind to form the Fas complex (FasC):

$$\text{Fas} + \text{FasL} \quad \rightleftharpoons \quad \text{FasC} \tag{4.1}$$

The Fas complex can then bind reversibly to intracellular FADD to form the Death-Inducing Signaling Complex (DISC) [164, 165]:

$$\text{FasC} + \text{FADD} \quad \rightleftharpoons \quad \text{DISC} \tag{4.2}$$

Reactions localized to the DISC

Binding of procaspase-8 to the DISC can lead to cleavage to the activated form, caspase-8 [166].

$$\text{DISC} + \text{Pro8} \quad \rightleftharpoons \quad \text{DISC:Pro8} \tag{4.3}$$

$$\text{DISC:Pro8} \quad \rightarrow \quad \text{DISC} + \text{Casp8} \tag{4.4}$$

90
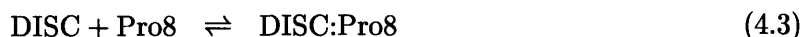
FLIP competitively binds to the DISC, reducing the activation of caspase-8 [167, 168].

$$DISC + FLIP \rightleftharpoons DISC:FLIP \tag{4.5}$$

Activation of effector caspase-3

Once initiator caspase-8 is activated, it proceeds to cleave and activate the effector caspase-3 [169, 170] which itself goes on to induce the apoptotic disassembly machinery.

$$Casp8 + Pro3 \quad \rightleftharpoons \quad Casp8:Pro3 \tag{4.6}$$

$$Casp8:Pro3 \quad \rightarrow \quad Casp8 + Casp3 \tag{4.7}$$
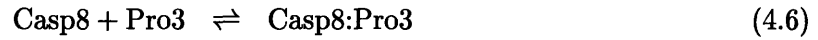
Activation of Bid by caspase-8

Activated caspase-8 also cleaves the BH3-only member of the Bcl-2 family, Bid [171]. Activated Bid is referred to as truncated Bid (tBid).

$$Casp8 + Bid \quad \rightleftharpoons \quad Casp8:Bid \tag{4.8}$$

$$Casp8:Bid \quad \rightarrow \quad Casp8 + tBid \tag{4.9}$$
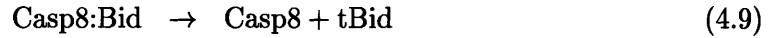
Interactions between Bcl-2 family members at the mitochondrion

Truncated Bid facilitates the dimerization of Bax, a key step leading to permeabilization of the mitochondrial membrane [171–174].

$$tBid + Bax \quad \rightleftharpoons \quad tBid:Bax \tag{4.10}$$

$$tBid:Bax + Bax \quad \rightleftharpoons \quad tBid:Bax_2 \tag{4.11}$$

$$tBid:Bax_2 \quad \rightleftharpoons \quad tBid + Bax_2 \tag{4.12}$$
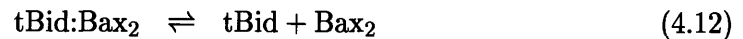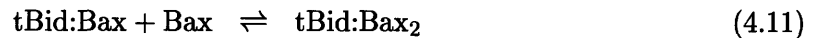
In the base model we assume that Bax dimers can spontaneously separate into monomers but that monomers cannot autoassemble into the dimer.

$$Bax_2 \rightarrow 2Bax \tag{4.13}$$

Bcl-2 can reversibly bind to both tBid and Bax [172, 175, 176], reducing the rate of Bax dimerization induced by truncated Bid.

$$\text{Bcl-2} + \text{tBid} \quad \rightleftharpoons \quad \text{Bcl-2:tBid} \qquad (4.14)$$

$$\text{Bcl-2} + \text{Bax} \quad \rightleftharpoons \quad \text{Bcl-2:Bax} \qquad (4.15)$$
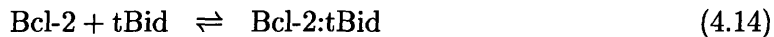
Release of Smac and cytochrome c from the mitochondrion

Bax dimers, Bax_2, are taken to signify formation of mitochondrial membrane pores, allowing the release of cytochrome c and Smac into the cytoplasm [174, 177–179].

$$\text{Bax}_2 + \text{cytoc}_{\text{mito}} \quad \rightarrow \quad \text{Bax}_2 + \text{cytoc} \qquad (4.16)$$

$$\text{Bax}_2 + \text{Smac}_{\text{mito}} \quad \rightarrow \quad \text{Bax}_2 + \text{Smac} \qquad (4.17)$$
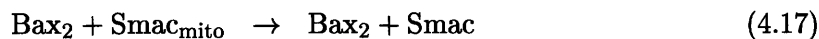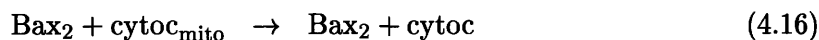
Construction of the apoptosome and activation of caspase-9

Formation of the apoptosome begins when cytochrome c is released from the mitochondria and binds to Apaf-1 [180]. Seven cytochrome c/Apaf dimers combine to form a single apoptosome, and it is assumed that this assembly is irreversible [181, 182].

$$\text{cytoc} + \text{Apaf} \quad \rightleftharpoons \quad \text{cytoc:Apaf} \qquad (4.18)$$

$$7\text{cytoc:Apaf} \quad \rightarrow \quad \text{Apop} \qquad (4.19)$$
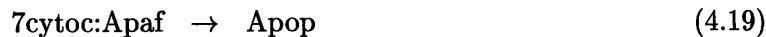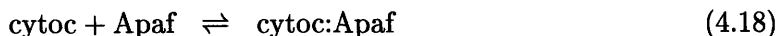
The apoptosome can bind two molecules of procaspase-9 and their co-localization leads to cleavage into the active form, caspase-9 [183]. It is assumed that the apoptosome is also

capable of directly binding to the active form.

$$\text{Apop} + \text{Pro9} \rightleftharpoons \text{Apop:Pro9} \tag{4.20}$$

$$\text{Apop:Pro9} + \text{Pro9} \rightleftharpoons \text{Apop:Pro9}_2 \tag{4.21}$$

$$\text{Apop:Pro9}_2 \rightarrow \text{Apop:Casp9}_2 \tag{4.22}$$

$$\text{Apop} + \text{Casp9} \rightleftharpoons \text{Apop:Casp9} \tag{4.23}$$

$$\text{Apop:Casp9} + \text{Casp9} \rightleftharpoons \text{Apop:Casp9}_2 \tag{4.24}$$
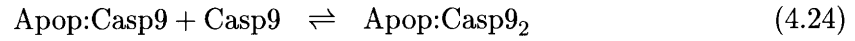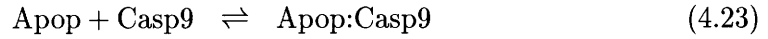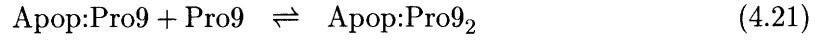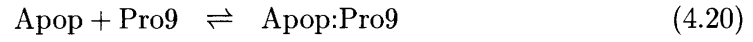
Activation of caspase-3 by caspase-9

Caspase-9 can cleave procaspase-3 into caspase-3 either as a free molecule in the cytoplasm [170] or bound to the apoptosome [184].

$$\text{Casp9} + \text{Pro3} \rightleftharpoons \text{Casp9:Pro3} \tag{4.25}$$

$$\text{Casp9:Pro3} \rightarrow \text{Casp9} + \text{Casp3} \tag{4.26}$$

$$\text{Apop:Casp9} + \text{Pro3} \rightleftharpoons \text{Apop:Casp9:Pro3} \tag{4.27}$$

$$\text{Apop:Casp9:Pro3} \rightarrow \text{Apop:Casp9} + \text{Casp3} \tag{4.28}$$

$$\text{Apop:Casp9}_2 + \text{Pro3} \rightleftharpoons \text{Apop:Casp9}_2\text{:Pro3} \tag{4.29}$$

$$\text{Apop:Casp9}_2\text{:Pro3} \rightarrow \text{Apop:Casp9}_2 + \text{Casp3} \tag{4.30}$$
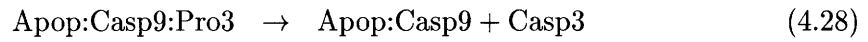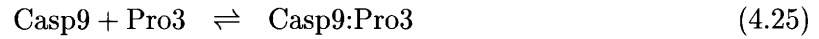
Activation of caspase-9 by caspase-3

In a positive feedback loop, caspase-9 can also be activated by its substrate caspase-3, either free or in complex with the apoptosome [185].

$$\text{Casp3} + \text{Pro9} \rightleftharpoons \text{Casp3:Pro9} \tag{4.31}$$

$$\text{Casp3:Pro9} \rightarrow \text{Casp3} + \text{Casp9} \tag{4.32}$$

$$\text{Casp3} + \text{Apop:Pro9} \rightleftharpoons \text{Casp3:Apop:Pro9} \tag{4.33}$$

$$\text{Casp3:Apop:Pro9} \rightarrow \text{Casp3} + \text{Apop:Casp9} \tag{4.34}$$
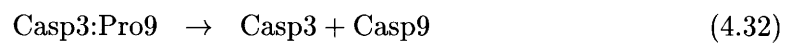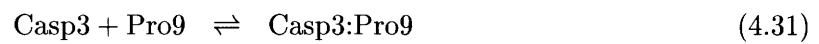
Inhibition by and of XIAP

XIAP functions to inhibit apoptosis by binding effector caspase-3, thereby preventing it from carrying out cellular disassembly [186, 187]:

$$\text{XIAP} + \text{Casp3} \ \rightleftharpoons \ \text{XIAP:Casp3} \tag{4.35}$$

XIAP can also bind full-length and cleaved caspase-9, either in the cytosol [188] or localized to the apoptosome [189, 190]. We assume that Apop:Pro9$_2$ is quickly converted to Apop:Casp9$_2$ so that it does not have the opportunity to bind XIAP.

$$\text{XIAP} + \text{Casp9} \ \rightleftharpoons \ \text{XIAP:Casp9} \tag{4.36}$$

$$\text{XIAP} + \text{Apop:Casp9} \ \rightleftharpoons \ \text{XIAP:Apop:Casp9} \tag{4.37}$$

$$\text{XIAP} + \text{Apop:Casp9}_2 \ \rightleftharpoons \ \text{XIAP:Apop:Casp9}_2 \tag{4.38}$$

$$\text{XIAP} + \text{Pro9} \ \rightleftharpoons \ \text{XIAP:Pro9} \tag{4.39}$$

$$\text{XIAP} + \text{Apop:Pro9} \ \rightleftharpoons \ \text{XIAP:Apop:Pro9} \tag{4.40}$$
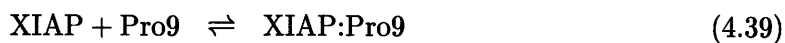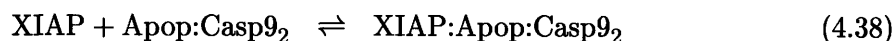
When Smac is released from the mitochondria, it binds to XIAP, relieving the inhibition of the caspases [191, 192].

$$\text{Smac} + \text{XIAP} \rightleftharpoons \text{Smac:XIAP} \tag{4.41}$$

4.2.2 Model add-ons to generate the ensemble

We chose six independent modifications to the base model, reactions that are differentially represented in the existing published models, and some of which are still uncertain from a biological perspective. By no means is this an exhaustive set of the differences among models, the biological uncertainties, or the possible mechanistic descriptions of the system. These were chosen, rather, as a representative sample of the types of choices that distinguish reasonable Fas models. They share the property that each may be included or not by switching rate constants to zero. An ensemble of models was created by constructing the combinatorial set of all possible combinations of inclusion or exclusion of the add-ons. The

ensemble therefore consisted of $2^6 = 64$ different topologies, each of which was parameterized to create a model.

The modifications are each described in detail here and are summarized in Figure 4-2. In the text, model topologies are referred to by a six-bit binary code where each bit represents inclusion (1) or exclusion (0) of each add-on, in order from 1 to 6 as described below. The Base Model is referred to as Model 000000 and the model that includes all possible add-ons, the Master Model, is Model 111111.

Add-on No.1: Inclusion of caspase-2

Though rarely included in models—see [74] as an exception—caspase-2 is known to be activated by caspase-3 [193] and there is further evidence that caspase-2, like caspase-8, has the ability to cleave Bid into tBid [194]. Inclusion of these reactions introduces a source of positive feedback for the Type II pathway and a source of crosstalk between the instrinsic and extrinsic pathways. The specific reactions for this model modification are as follows:

$$\text{Casp3} + \text{Pro2} \rightleftharpoons \text{Casp3:Pro2} \tag{4.42}$$

$$\text{Casp3:Pro2} \rightarrow \text{Casp3} + \text{Casp2} \tag{4.43}$$

$$\text{Casp2} + \text{Bid} \rightleftharpoons \text{Casp2:Bid} \tag{4.44}$$

$$\text{Casp2:Bid} \rightarrow \text{Casp2} + \text{tBid} \tag{4.45}$$
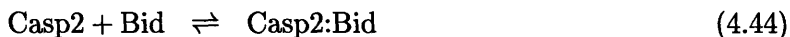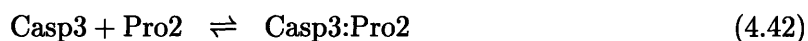
Add-on No.2: Inclusion of caspase-6

Caspase-6 is known to be a mediator of positive feedback through the Type I apoptotic pathway. Caspase-3 cleaves procaspase-6 into caspase-6 [170, 195] which then activates caspase-8 [196]. These reactions have been included explicitly in some models [74,157,161], while in others this feedback is modeled as occurring directly, without mediation of caspase-
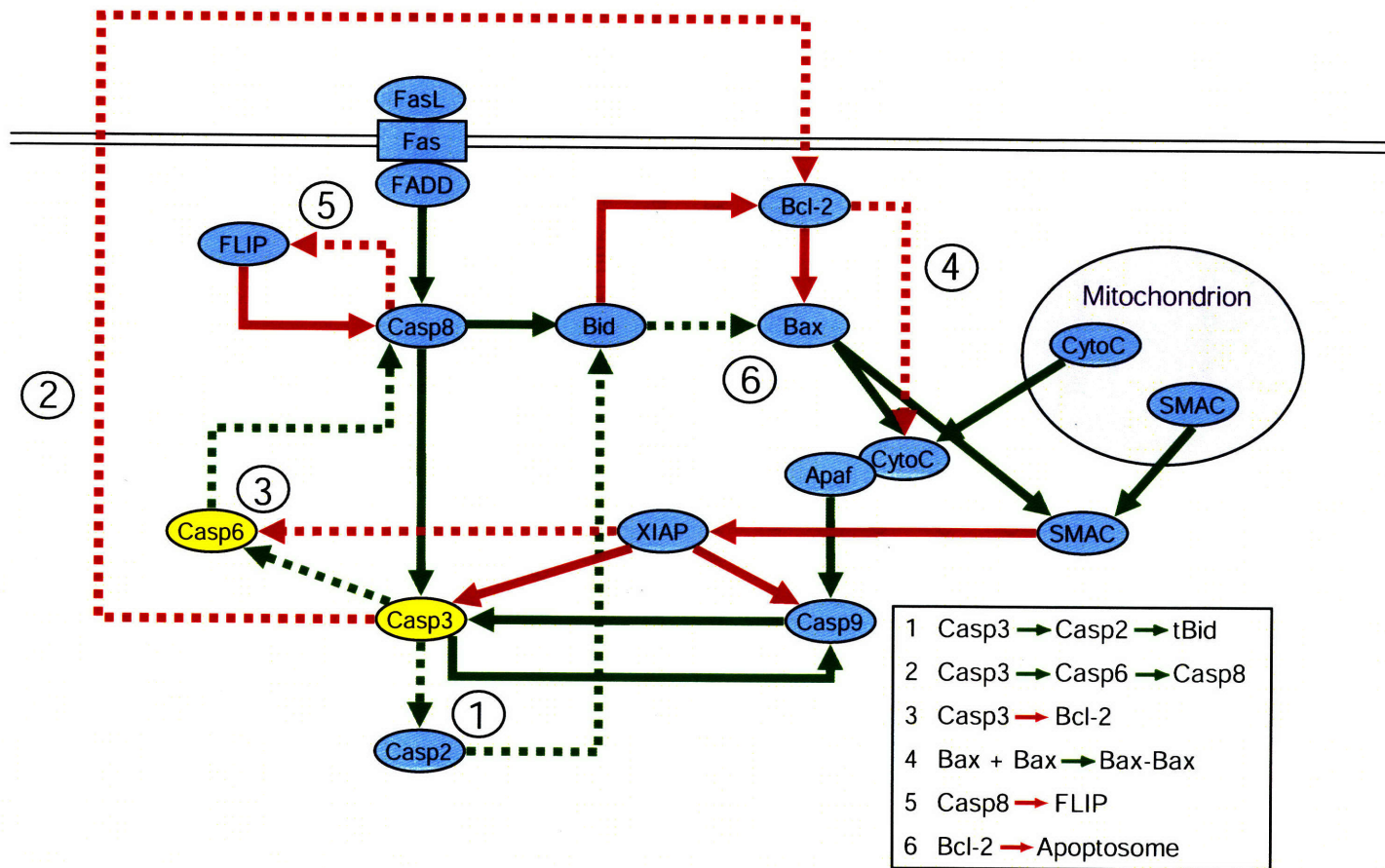
Figure 4-2: Master Model of the Fas signaling network, showing the six model add-ons used to build the library of topologies. Solid lines represent interactions included in the Base Model whereas dotted lines correspond to the various optional add-ons. As in Figure 4-1, green and red arrows indicate activating or inhibiting influence, respectively.

6 [75].

$$\text{Casp3} + \text{Pro6} \quad \rightleftharpoons \quad \text{Casp3:Pro6} \tag{4.46}$$

$$\text{Casp3:Pro6} \quad \rightarrow \quad \text{Casp3} + \text{Casp6} \tag{4.47}$$

$$\text{Casp6} + \text{Pro8} \quad \rightleftharpoons \quad \text{Casp6:Pro8} \tag{4.48}$$

$$\text{Casp6:Pro8} \quad \rightarrow \quad \text{Casp6} + \text{Casp8} \tag{4.49}$$
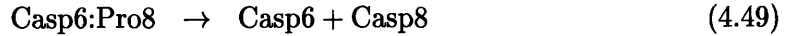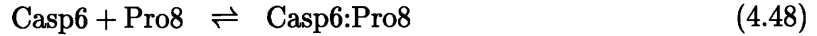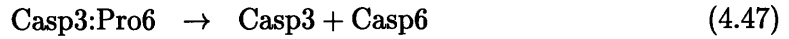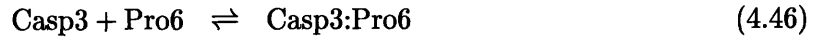
Add-on No.3: Interactions of caspase-3 with members of the Bcl-2 family of proteins

It has been proposed that there exists a positive feedback loop linking the Type I and Type II pathways, in the form of deactivation of Bcl-2 by caspase-3 [197–199]. If activation of caspase-3 were to occur quickly by direct action of caspase-8, these reactions could accelerate mitochondrial membrane permeabilization.

$$\text{Casp3} + \text{Bcl-2} \quad \rightleftharpoons \quad \text{Casp3:Bcl-2} \tag{4.50}$$

$$\text{Casp3:Bcl-2} \quad \rightarrow \quad \text{Casp3} + \text{Bcl-2}_{\text{inactive}} \tag{4.51}$$
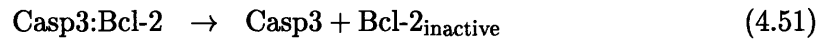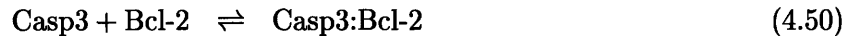
Add-on No.4: Autodimerization of Bax

The mechanism by which the Bcl-2 family of proteins interacts to control permeabilization of the mitochondrial membrane and subsequent release of cytochrome c and Smac is still contested [200,201]. Previously, it had been assumed that Bid functioned to directly activate Bax and that membrane permeabilization required interaction of these two molecules. More recently, it has been shown that the role of Bid may be, at least in part, an indirect one because cells with Bid mutated to prevent binding to Bax still die [202]. The hypothesis is that Bcl-2 acts as a stop for membrane permeabilization and that activated Bid operates by relieving this barrier. Several recent papers have addressed this uncertainty by investigating possible mechanisms of permeabilization, both direct and indirect, and concluded that a direct model was preferred for its robustness properties [77,159,160].

Our base model is one in which there is only direct activation by Bid: Bax dimerization

requires the assistance of tBid. As one of our model options, we allow autoactivation of Bax through spontaneous dimerization, a reaction which has been previously proposed [203]. In this model, if there is insufficient Bcl-2 in the cell to sequester Bax molecules and prevent them from binding to one another, the mitochondria become leaky, and release cytochrome c and Smac, even in the absence of a Fas ligand stimulus. This differs from the proposed models by Chen and Cui [77, 160] in which complete Bax activation requires at least one monomer to interact with a mediator molecule.

$$2\text{Bax} \rightleftharpoons \text{Bax}_2 \tag{4.52}$$

Add-on No.5: Inactivation of FLIP by caspase-8

In the base model, FLIP reduces the rate at which caspase-8 becomes activated by competing for the binding site of the DISC. This next model modification allows for a positive feedback in which activated caspase-8 can cleave FLIP, reducing its ability to bind the DISC, and increasing the rate at which caspase-8 is activated at the cellular membrane [204].

$$\text{Casp8} + \text{FLIP} \quad \rightleftharpoons \quad \text{Casp8:FLIP} \tag{4.53}$$

$$\text{Casp8:FLIP} \quad \rightarrow \quad \text{Casp8} + \text{FLIP}_{\text{inactive}} \tag{4.54}$$

Add-on No.6: Binding of Bcl-2 to the apoptosome

The last add-on is a putative mechanism by which Bcl-2 might act anti-apoptotically: by binding to the apoptosome, thereby preventing local activation of caspase-9. The mechanism of action may involve binding to either cytochrome c [205,206] and/or Apaf-1 [207–209] though the latter reaction has been challenged as occuring in C. elegans but not humans [210]. We modeled the interaction of Bcl-2 directly with the apoptosome rather than with cytochrome c or Apaf-1 individually, because in our model the apoptosomes are formed very quickly following mitochondrial membrane permeabilization. The overall effect of this reaction could be neutral or negative towards apoptosis considering that Bcl-2 at the apoptosome may displace activated as well as inactivated caspase-9, the consequence being

that more of the active form is present in the cytoplasm to activate free procaspase-3.

$$\text{Apop} + \text{Bcl-2} \quad \rightleftharpoons \quad \text{Apop:Bcl-2} \tag{4.55}$$
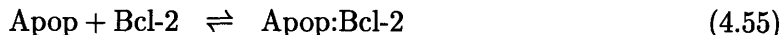
4.2.3 Parameter values

Where possible, values for rate constants were drawn from published experimental results. A summary of the rate constants and their sources is shown in Table 4.2. Little information is available on the initial concentrations of proteins, and there is likely strong variability across cell-types and even individual cells. As a result, we chose to fit the initial concentrations of all species that represent individual proteins, and to set the initial concentrations of all complexes to zero. At any given time, the concentrations of these complexes are not zero, and these values can be obtained by equilibrating the system of differential equations in the absence of FasL stimulus. However, it was found that equilibration occured on very fast time-scales (on the order of minutes, data not known) and did not affect the resulting stimulus responses. Due to the large number of models and simulations, as a time-saving measure we used zero concentrations for all complexes and equilibration was not performed.

4.2.4 Experimental Data

Data for fitting parameters that were not available in the literature were generously donated by Dr. Fei Hua, who used the same data in previously published work on Fas model simulation [155, 157]. The data can be divided into two subsets, though all experiments were performed on Jurkat.E6 human T cells. The first set consists of Western blot data for the fraction of uncleaved (pro)caspase-3 and -8, collected over an 8-hour time span following stimulation with 100 ng·ml^{-1} Fas ligand (FasL). The second set consists of two dose response experiments, one each in regular Jurkat cells and in Bcl-2 over-expressing Jurkat cells. Cells were stimulated with either 1, 10, 100, or 1000 ng·ml^{-1} FasL and the fraction of cells positive for cleaved caspase-3 was measured at times out to 8 hours using FACS analysis. The data used for fitting are shown in Figure 4-3. Further details on the nature of the experiments can be obtained from [155] and [157].

	R1	R2	P1	P2	Forward Rate Constant	Reverse Rate Constant	Forward Value $nM^{-1}\,s^{-1}$	Reverse Value s^{-1}
1	Fas	FasL	FasC		k1on	k1off	Fit	1.1 k1on [211]
2	FasC	FADD	DISC		k2on	k2off	Fit	Fit
3	DISC	Pro8	DISC:Pro8		k3on	k3off	Fit	Fit
4	DISC	Casp8	DISC: Pro8		k4on	k4off	0	Fit
5	DISC	FLIP	DISC:FLIP		k5on	k5off	Fit	Fit
6	Casp8	Pro3	Casp8:Pro3		k6on	k6off	8.7E-4(k6off+k7off)/k7off [212]	Fit
7	Casp8	Casp3	Casp8:Pro3		k7on	k7off	0	Fit
8	Casp8	Bid	Casp8:Bid		k8on	k8off	Fit	Fit
9	Casp8	tBid	Casp8:Bid		k9on	k9off	0	Fit
10	tBid	Bax	tBid:Bax		k10on	k10off	Fit	Fit
11	tBid:Bax	Bax	tBid:Bax$_2$		k11on	k11off	k10on	k10off
12	tBid	Bax$_2$	tBid:Bax$_2$		k12on	k12off	Fit	Fit
13	Bax	Bax	Bax$_2$		k13on	k13off	0	Fit
14	tBid	Bcl-2	tBid:Bcl-2		k14on	k14off	Fit	220 k14on [213]
15	Bax	Bcl-2	Bax:Bcl-2		k15on	k15off	Fit	Fit
16	Bax$_2$	cytoc$_{mito}$	Bax$_2$	cytoc	k16on	k16off	Fit	0
17	Bax$_2$	Smac$_{mito}$	Bax$_2$	Smac	k17on	k17off	k16on	0
18	cytoc	Apaf	cytoc:Apaf		k18on	k18off	Fit	Fit
19	cytoc:Apaf	—	Apop		k19on	k19off	Fit	0
20	Apop	Pro9	Apop:Pro9		k20on	k20off	Fit	Fit
21	Apop:Pro9	Pro9	Apop:Pro9$_2$		k21on	k21off	k20on	k20off
22	Apop:Pro9$_2$	—	Apop:Casp9$_2$		k22on	k22off	Fit	0
23	Apop	Casp9	Apop:Casp9		k23on	k23off	k20on	k20off
24	Apop:Casp9	Casp9	Apop:Casp9$_2$		k24on	k24off	k20on	k20off
25	Casp9	Pro3	Casp9:Pro3		k25on	k25off	(k25off+4.8)/(2.5E5) [214]	Fit
26	Casp9	Casp3	Casp9:Pro3		k26on	k26off	0	4.8 [214]
27	Apop:Casp9	Pro3	Apop:Casp9:Pro3		k27on	k27off	140/(k27off+k28off) [215]	Fit
28	Apop:Casp9	Casp3	Apop:Casp9:Pro3		k28on	k28off	0	Fit
29	Apop:Casp9$_2$	Pro3	Apop:Casp9$_2$:Pro3		k29on	k29off	k27on	k27off
30	Apop:Casp9$_2$	Casp3	Apop:Casp9$_2$:Pro3		k30on	k30off	0	k28off
31	Casp3	Pro9	Casp3:Pro9		k31on	k31off	Fit	Fit
32	Casp3	Casp9	Casp3:Pro9		k32on	k32off	0	Fit
33	Casp3	Apop:Pro9	Casp3:Apop:Pro9		k33on	k33off	k31on	k31off
34	Casp3	Apop:Casp9	Casp3:Apop:Pro9		k34on	k34off	0	k32off
35	Casp3	XIAP	Casp3:XIAP		k35on	k35off	2.5E-3 [187]	2.4E-3 [187]
36	Casp9	XIAP	Casp9:XIAP		k36on	k36off	k35on	k35off
37	Pro9	XIAP	Pro9:XIAP		k37on	k37off	k35on	k35off
38	Apop:Pro9	XIAP	Apop:Pro9:XIAP		k38on	k38off	k35on	k35off
39	Apop:Casp9	XIAP	Apop:Casp9:XIAP		k39on	k39off	k35on	k35off
40	Apop:Casp9$_2$	XIAP	Apop:Casp9$_2$:XIAP		k40on	k40off	k35on	k35off
41	Smac	XIAP	Smac:XIAP		k41on	k41off	7E-3 [216]	2.2E-3 [216]
42	Casp3	Pro2	Casp3:Pro2		k42on	k42off	Fit	Fit
43	Casp3	Casp2	Casp3:Pro2		k43on	k43off	0	Fit
44	Casp2	Bid	Casp2:Bid		k44on	k44off	Fit	Fit
45	Casp2	tBid	Casp2:Bid		k45on	k45off	0	Fit
46	Casp3	Pro6	Casp3:Pro6		k46on	k46off	Fit	Fit
47	Casp3	Casp6	Casp3:Pro6		k47on	k47off	0	Fit
48	Casp6	Pro8	Casp6:Pro8		k48on	k48off	Fit	Fit
49	Casp6	Casp8	Casp6:Pro8		k49on	k49off	0	Fit
50	Casp3	Bcl-2	Casp3:Bcl-2		k51on	k51off	Fit	Fit
51	Casp3	Bcl-2$_{inactive}$	Casp3:Bcl-2		k52on	k52off	0	Fit
52	Casp8	FLIP	Casp8:FLIP		k53on	k53off	Fit	Fit
53	Casp8	FLIP$_{inactive}$	Casp8:FLIP		k54on	k54off	0	Fit
54	Apop	Bcl-2	Apop:Bcl-2		k55on	k55off	Fit	Fit
55	I$_{Casp2}$	Casp2	I$_{Casp2}$:Casp2		k56on	k56off	Assigned	Assigned
56	I$_{Casp3}$	Casp3	I$_{Casp3}$:Casp3		k57on	k57off	Assigned	Assigned
57	I$_{Casp6}$	Casp6	I$_{Casp6}$:Casp6		k58on	k58off	Assigned	Assigned
58	I$_{Casp8}$	Casp8	I$_{Casp8}$:Casp8		k59on	k59off	Assigned	Assigned
59	I$_{Casp9}$	Casp9	I$_{Casp9}$:Casp9		k60on	k60off	Assigned	Assigned

Table 4.2: Reactions and corresponding rate constants for the Fas signaling network.
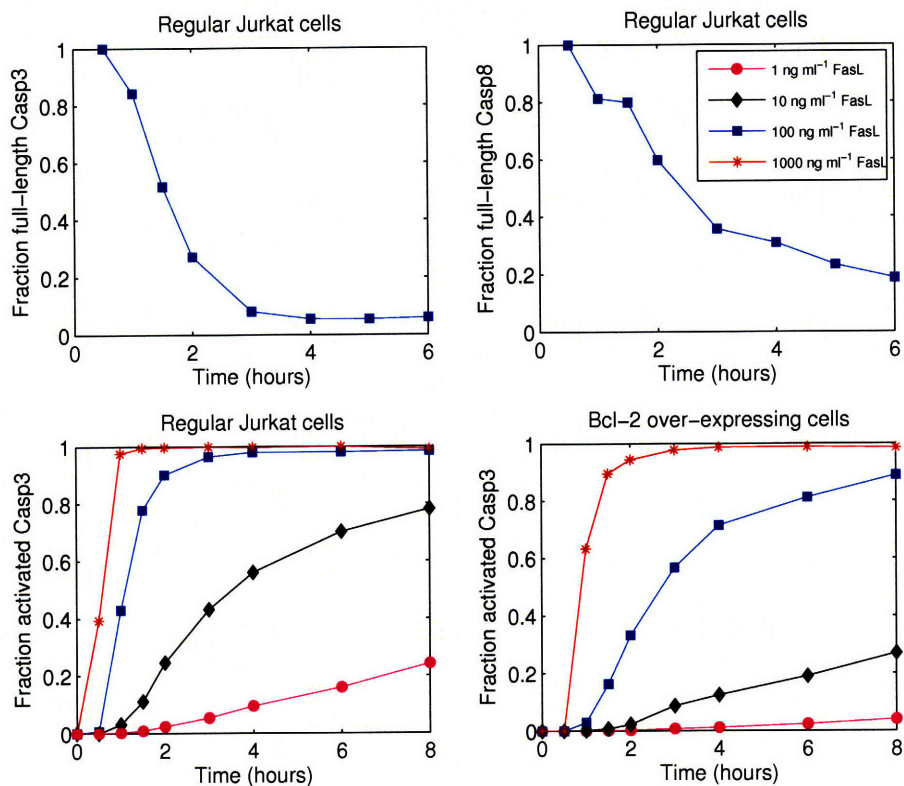
Figure 4-3: Data used for fitting the Fas models.

4.2.5 Model optimization

Parameters for which no value was available in the literature were fit to the data. Optimization was performed in Jacobian (Numerica Technology, Inc., Cambridge, MA) with the weighted least squares objective function. Optimization seeds for the non-zero initial concentrations are shown in Table 4.3. It was previously found by Hua et al. that a FasL concentration of 100 ng ml^{-1} corresponded to an effective FasL concentration of 2 nM for modeling [155]. Thus, the initial concentration of FasL in the model was set to 0.02, 0.2, 2, or 20 nM when fitting to experimental data from cells stimulated with 1, 10, 100, or 1000 ng ml^{-1} FasL, respectively.

4.2.6 Model perturbations and predictions

We implemented model perturbations that mechanistically represent binding of caspases by synthetic inhibitors in the presence of a strong Fas ligand signal such as might be seen in

Species	Initial value seeds (nM)
Apaf	100
Bax	75
Bcl-2	75
Bid	25
$cytoc_{mito}$	100
FADD	20
Fas	10
Pro2	20
Pro3	200
Pro6	20
Pro8	30
Pro9	20
$Smac_{mito}$	100
XIAP	30

Table 4.3: Optimization seeds for fitting initial concentrations in the Fas models.

sepsis [150]. There are a wide variety of caspase inhibitors available, that differ in a number of features:

1. **Specificity**: Caspase inhibitors come in a wide variety of specificities, from those that bind strongly to a single caspase, to broad spectrum inhibitors that bind with some strength to all caspases.

2. **Reversibility**: Inhibitors can either bind their targets irreversibly, removing them entirely from the apoptotic network, or they can bind reversibly, mimicking the natural behavior of XIAP. (Unlike XIAP, however, they are not bound by Smac.)

3. **Binding to active or inactive forms of caspase**: It is possible to design inhibitors that bind to procaspases as well as active ones, just as XIAP can bind caspase-9 and procaspase-9, though in practice most inhibitors bind to the active site of their targets, regions that are not exposed in the inactive forms.

We modeled the addition of cocktails of inhibitors [217] consisting of up to five specific caspase inhibitors, one each for caspases-2, -3, -6, -8, and -9. We assumed that all inhibitors act on the active form of each caspase, and are ineffective against caspases that are complexed

with other molecules, yielding 5 additional reactions and 10 new species.

$$I_{Casp2} + Casp2 \rightleftharpoons I_{Casp2}:Casp2 \tag{4.56}$$

$$I_{Casp3} + Casp3 \rightleftharpoons I_{Casp3}:Casp3 \tag{4.57}$$

$$I_{Casp6} + Casp6 \rightleftharpoons I_{Casp6}:Casp6 \tag{4.58}$$

$$I_{Casp8} + Casp8 \rightleftharpoons I_{Casp8}:Casp8 \tag{4.59}$$

$$I_{Casp9} + Casp9 \rightleftharpoons I_{Casp9}:Casp9 \tag{4.60}$$

For specific inhibitors, the k_{on} of each reaction was either 0 or 10^{-3} nM^{-1} s^{-1}. (Setting k_{on} to zero effectively removes that inhibitor from the cocktail.) If k_{on} was non-zero, then k_{off} for that reaction could be either 0, 10^{-2} or 10^{-1} s^{-1}. We used $k_{off} = 0$ to represent irreversible binding of an inhibitor. Reversible binding could occur with a K_D of 10 or 100 nM, representing strong or weak inhibition within the range seen among commerically available inhibitors[1]. The four possible settings for each of five inhibitors results in $4^5 = 1024$ perturbations. All inhibitors were assumed to be present in the cell at a concentration of 50 nM and the external Fas ligand concentration was set to 20 nM (equivalent to the highest level of FasL simulus in the experiments used for fitting the data).

4.3 Results

4.3.1 Poorly constrained models in a branched pathway have very different flux profiles

In Figure 4-4 we show the superposition of fits by the 64 topologies to the dose response data shown in Figure 4-3. Models were fit to the entire set of data simultaneously, and while they are able to recapitulate the activation curve of caspase-3 in the presence of high stimulus concentrations, in general, the models are poor fits to the time courses for lower concentrations of Fas ligand. There are several possible explanations for this result that highlight many of the real difficulties in building complex differential equation models of biological pathways.

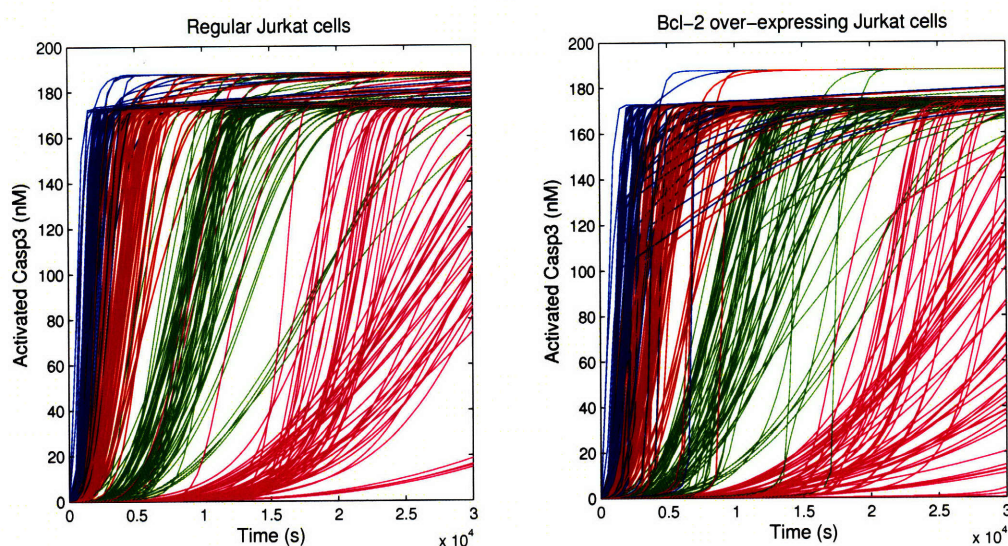[1]See http://www.calbiochem.com for a subset of available inhibitors.

Figure 4-4: Superposition of fits by all 64 Fas signaling pathway topologies to the data regular (left) and Bcl-2 over-expressing (right) Jurkat cells. The various colors correspond to FasL concentrations of 20 nM (blue), 2 nM (red), 0.2 nM (green), and 0.02 nM (magenta). The agreement between models is progressively poorer for lower stimulus levels.

One possibility is that parameter space was not explored sufficiently to find values that would reproduce the data well. The objective function space for these optimizations was plagued by many local minima, as different initial seeds for fit parameters invariably led to different solutions (data not shown). Due to time constraints (optimization of a single topology required on the order of 2 hours using Jacobian) and a very large search space (on the order to 50 fit parameters per topology) we did not do full multi-start optimization. Also related to optimization was the fact that, although we did allow initial protein concentrations to vary, the optimizer did not move along these directions, and final estimates of initial concentrations never changed more than 1% from the seed values provided to the optimizer. To alleviate this issue, we tried scaling these variables prior to optimization so that they were similar in magnitude to the rate constants. We observed no effect on the quality of the resulting fits (data not shown).

Another issue that may have affected the ability to reproduce the data is the inappropriate fixing of parameter values. As much as possible rate constants were assigned according to values derived from the biological literature. Also in accordance with the literature, we constrained several rate constants to equal each other, provided this was chemically rea-

sonable. All of these choices were made to improve the size of the optimization space in relation to the number of data points used in the fits. However, if these values were incorrect, then we may have inadvertently made the optimization problem more difficult. (We tried optimizing the models with no fixed parameters and found that they could reproduce the time course data from the regular Jurkat cells remarkably well; however, they were almost universally insensitive to Bcl-2 over-expression because they channeled very little flux through the mitochondrial pathway. These models were therefore judged as being overfit.)

A third concern regards assumptions about differences between regular and Bcl-2 over-expressing cells. We assumed, as was done in [157], that only the concentration of Bcl-2 was different between the two cell types, and that the initial concentrations of all other proteins were unchanged. Note, however, that the observed reduction in sensitivity to Fas ligand in the over-expressing cells could similarly be caused by a simultaneous reduction in the number of Fas receptors (data not shown). With our assumptions, the optimizer had no possibility of finding this solution.

Lack of knowledge of the initial concentrations, and in particular differences in initial concentrations between the two cell types, could be alleviated somewhat if we had more information about the relative fluxes through the Type I and Type II pathways. The data that were used to fit the models were drawn from the two branch-points in the network: caspase-8 is where the two pathways diverge, and caspase-3 is where they reconverge. There is, therefore, very little information to constrain the relative amount of caspase-3 activation from caspase-8 and caspase-9.

To address this issue further, we calculated the time-dependent fluxes into activated caspase-3 from either caspase-8 or caspase-9 and clustered the 64 models according to their flux patterns (Figures 4-5 and 4-6). Clustering was performed using the k-means algorithm with $k = 8$. Figure 4-7 shows the cumulative contribution to caspase-3 activation by caspase-8 and -9 for four topologies, each drawn from a different flux cluster. We immediately observed that in many cases, the relative contribution of caspase-8 to caspase-3 activation was considerably more than that by caspase-9. In some topologies, up to 99% of the flux was through caspase-8, regardless of Bcl-2 level or concentration of Fas ligand. In other classes of models, caspase-9 contributed up to 99% of the caspase-3 activation, though only for

very low concentration of ligand. Other clear features were that the flux through caspase-8 was delayed and the maximum flux decreased for lower stimulus concentrations. Quite to the contrary, lower ligand concentrations often resulted in significantly more flux through caspase-9 (though, as with caspase-8, the activation did not occur for several hours following stimulus). As would be expected based on path length, activation by caspase-9 always occured after the initiation by caspase-8. We hypothesize that when large concentrations of caspase-8 become activated on a short time-scale, direct activation of caspase-3 quickly follows, so quickly in fact that by the time the mitochondrial pathway has activated caspase-9, there is little procaspase-3 left on which it can act. If caspase-8 is activated more slowly, as it is when there is little ligand present, the mitochondrial pathway has a chance to catch up and play a greater role in activation.

Figure 4-7 also serves to demonstrate how the time course of caspase-3 activation can be altered by over-expression of Bcl-2 and how this change is dependent on the particular flux balance of a model. A large surplus of Bcl-2 was observed to invariably repress the mitochondrial pathway. Caspase-8 was able to compensate for this loss to different degrees, depending on the specific parameters and topology of the model.

4.3.2 Graded response data can be reconciled with switch-like action by considering activation noise

Another major concern regarding our inability to fit the data well relates to the nature of the data itself and the known switch-like behavior of the apoptotic network. Several experimental studies using FRET reporters have shown caspase activation in individual cells to occur very rapidly once initiated, on the time scale of minutes [161, 218, 219]. The same studies also showed that variations between individual cells in a population result in differences in the time to activation. Averaging over a population of switch-like individuals can generate an apparent graded response, masking the all-or-none behavior. Our models were intended to predict the expected dynamic behavior occurring within a single cell, so it is unsurprising that at low ligand concentrations, the models do a poor job of recapitulating the population data. For high levels of FasL, this should not be as much of an issue as the switching times will vary proportionately less.
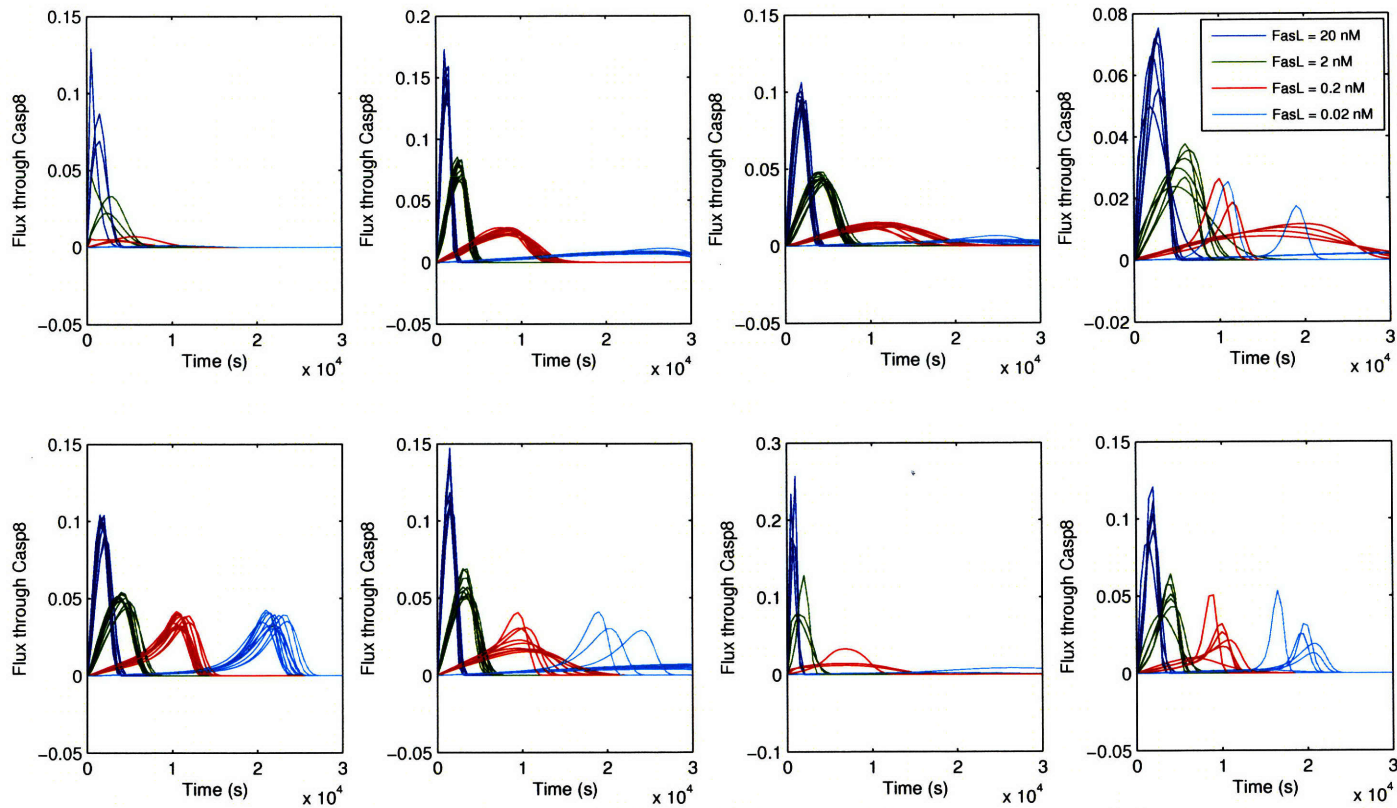
Figure 4-5: Clustering of the flux into activated caspase-3 by caspase-8 for the 64 fit models. Clustering was performed using k-means with a correlation distance metric and 100 replicates. As the concentration of ligand is decreased, the flux through caspase-8 is delayed and broadened.
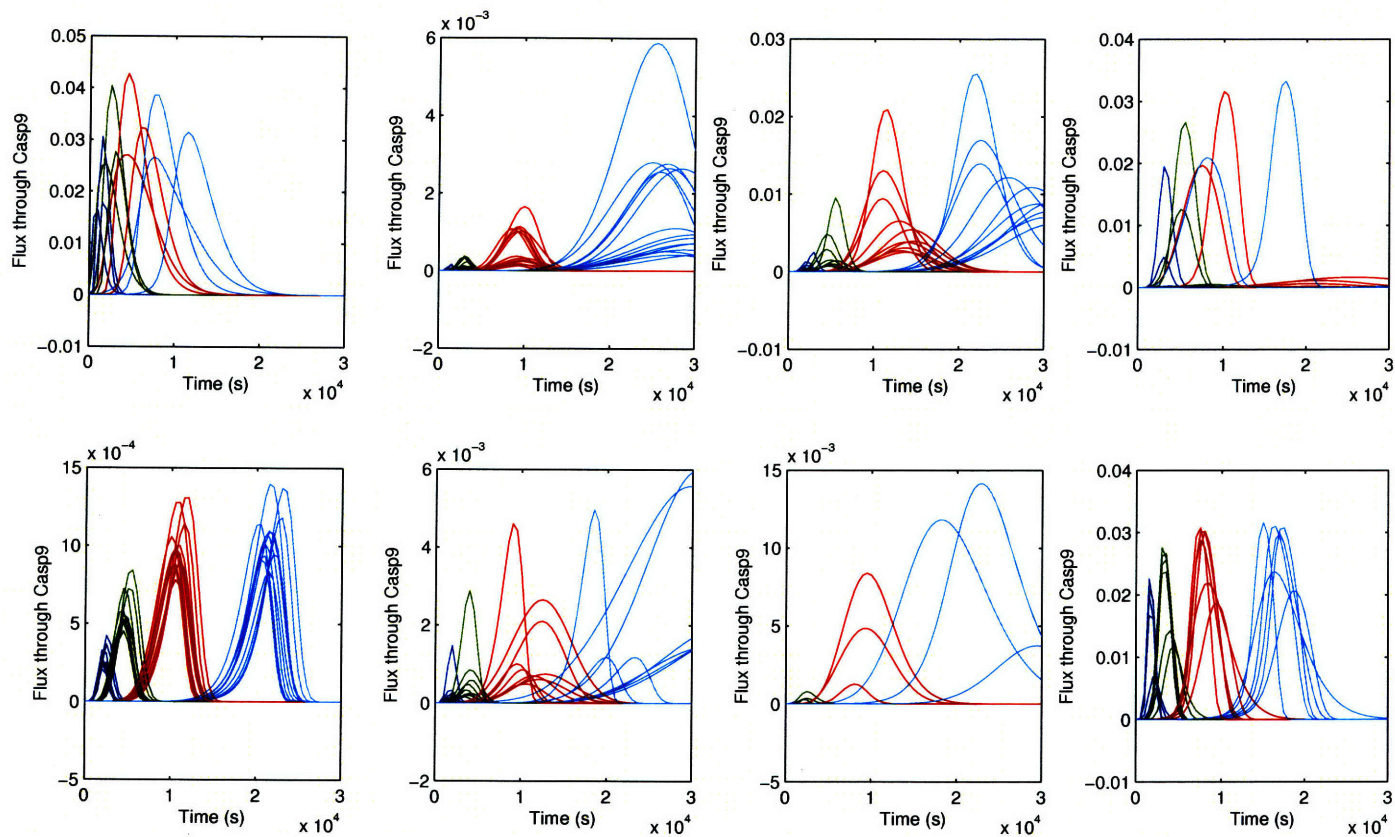
Figure 4-6: Clustering of the flux into activated caspase-3 by caspase-9 for the 64 fit models. As with caspase-8, smaller stimuli result in delayed and broadened fluxes. Overall, in these models, caspase-9 contributes relatively little to activation of caspase-3, which is evident through comparison of the flux magnitudes. However, interestingly, the proportional flux through caspase-9 is considerably more for low rather than high FasL. This relates to the delay between activation from the Type I and Type II pathways.
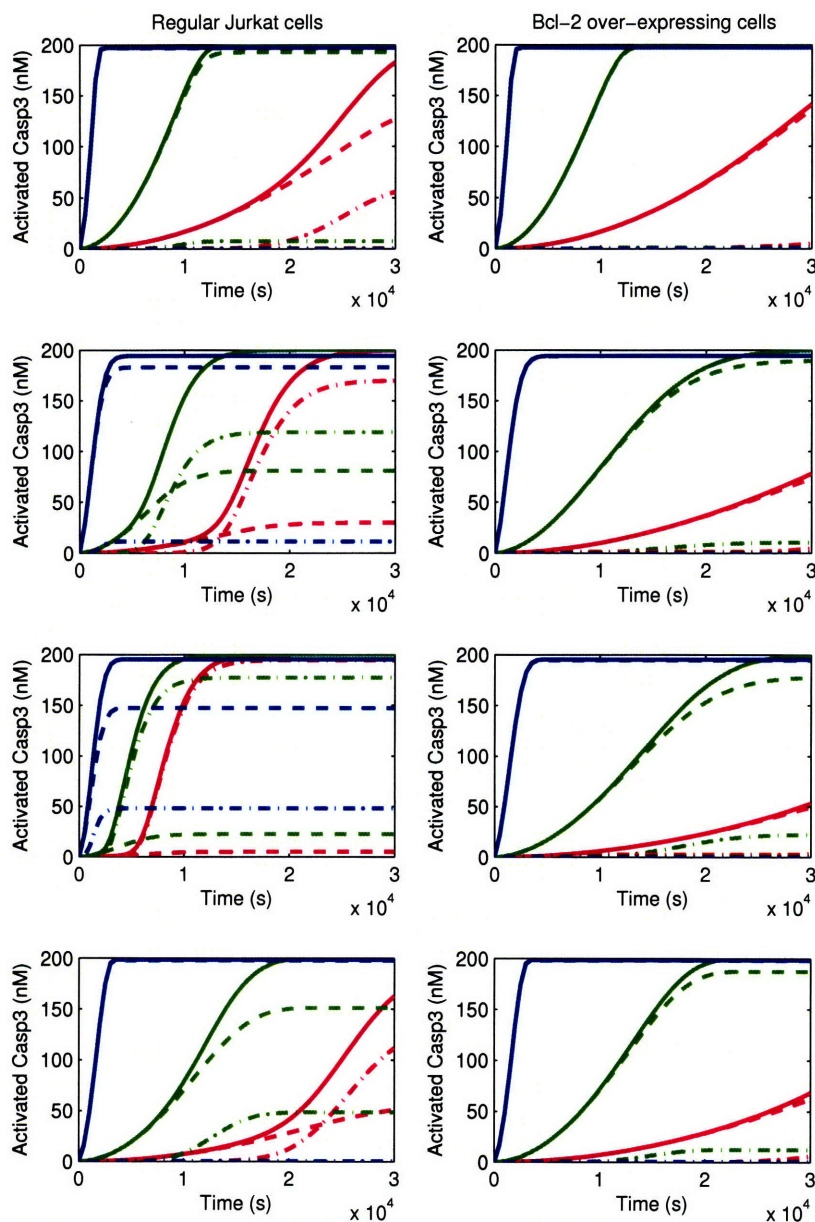
Figure 4-7: Contributions of caspase-8 (dash) and caspase-9 (dot-dash) to activation of caspase-3 (solid) under baseline values or Bcl-2 over-expression. The relative amount of caspase-3 activated by the Type I and Type II pathways varies considerably by model (Topology 000000, 000010, 000110, and 000111 from top to bottom), by ligand concentration (either 20 nM (blue), 0.2 nM (green) or 0.02 nM (magenta) within each plot), and by Bcl-2 concentration (left and right columns). The models shown belong to clusters 2, 1, 8, and 3 (in order, again, from top to bottom).
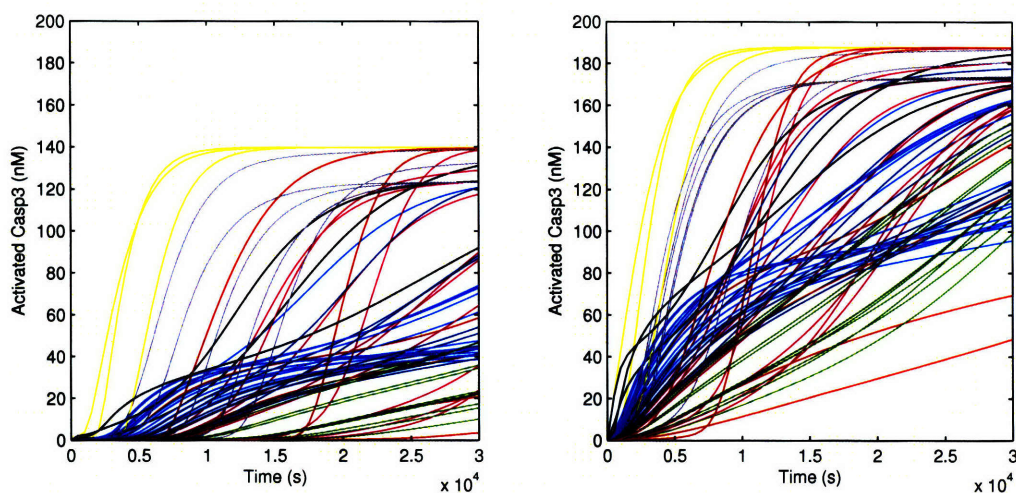
Figure 4-8: Comparison of behaviors among flux clusters under perturbation. The models of the ensemble were clustered according to their flux profiles into activated caspase-3 (as shown in Figures 4-5 and 4-6) and models were subject to perturbations in the form of treatment with irreversible inhibitors to caspases-3, -8, and -9 (left) or treatment with irreversible inhibitors of caspase-8 and caspase-9 only (right). Simulations are colored according to the cluster to which they belong, not according to the concentration of Fas ligand, which was constant at 20 nM.

It may or may not be the case that Bcl-2 over-expressing cells also display such switch-like activation profiles. Prior experiments in HeLa cells showed that Bcl-2 over-expression may cause more gradual switching, though the specifics depend on the concentrations of other species in the network, e.g. XIAP [161]. It is therefore unclear whether the data collected from Bcl-2 over-expressing cells show a graded response due to population-averaging or because the response within each cell is actually more gradual in this case. If mitochondrial membrane permeabilization is the major source of the all-or-none response in apoptosis [76, 161], then shutting down the Type II pathway may very well lead to the latter result.

We tested the ability our models with more switch-like activation profiles (and therefore generally poorer fits to the data) to generate a graded response when subjected to stimulus noise. As shown in Figure 4-9, we simulated activation of caspase-3 in model 110001 for 100 different concentrations of Fas ligand, randomly selected from a log-normal distribution with mean 0.02 nM (corresponding to the 1 ng ml^{-1} FasL experiment) and 5% or 30% proportional variance. These simulations reflect the stochastic nature of ligand molecules

binding to the cell surface, a possible source of switching-time variation across individual cells. The population mean trajectories, shown in red, are in stark contrast to the deterministic prediction with [FasL] = 0.02 nM (green). For the same model, receptor concentration would not similarly result in different switching times for the cells. The same proportional levels of noise in Fas receptor would be expected to bear little effect because it is at a saturating concentration with respect to the ligand for this particular choice of initial concentrations. We did not explore the impact of noise in other species on the time to caspase-3 activation, but it is very reasonable to expect that there may be other, intracellular, sources of variation.

4.3.3 The optimal caspase cocktail for differentiating between models

While the models agreed poorly on the activation profile of caspase-3 under conditions of low ligand concentration, they were unanimous in the quick onset of apoptosis for high concentrations. We therefore retained all 64 topologies in the ensemble as we were interested in the response of cells to different caspase inhibitor treatments in the presence of high stimulus concentrations, mimicking cellular conditions in a septic environment.

We subjected all models of the ensemble to 1024 different perturbations resulting from four different inhibitors choices for each of the five caspases in the model. For each perturbation representing a treatment condition, we computed the ensemble variance of the activation profiles of caspase-3, as described in Chapter 3. We used the inhibitor specifications for the cocktails as the independent variables for constructing a decision tree to classify the ensemble variance of the responses (Figure 4-10).

The nodes of the decision tree are color-coded to indicate the level of consensus of treatment choices belonging to that node with yellow representing high consensus, and red, low consensus. There is a distinct separation of high and low consensus perturbations, with only two major cuts required to separate the good from the bad. Namely, simulated addition of an irreversible inhibitor of caspase-8 and any choice of inhibitor *except* irreversible for caspase-9 results in large ensemble variance, regardless of whether there are inhibitors present at any strength for caspases-2, -3, and -6. All other choices involving caspase-8 and -9 lead to continued agreement among the models. In Figure 4-11B we have plotted the

simulated time courses for caspase-3 activation from all 64 models in the ensemble under treatment with an irreversible inhibitor of caspase-6 and caspase-8. This was the treatment condition corresponding to the largest ensemble variance of all 1024 tested treatment options. Referring back to Figure 4-8, we see that a large fraction of this variation is captured by the differences in the model fluxes at the fits. Therefore, we would expect that experimental testing of the cellular response to this inhibitor cocktail would lead to elimination of whole clusters of models that do not agree with the data.

4.3.4   Optimal inhibition of apoptosis involves a three-pronged attack

While experiments for refining the ensemble can be designed immediately by looking at the tree, the design of the optimal treatment plan for reducing apoptosis in the presence of high stimulus involves a second step. Once perturbations of high consensus have been identified by the tree, predictions from those treatment profiles need to be analyzed further to determine therapeutic effect. In Figure 4-12, we have plotted the ensemble variance against the mean concentration of active caspase-3 at 8 hours. From this plot we identified those perturbations corresponding to high ensemble variance and low activation. It turns out that, according to the models, in order to significantly reduce caspase-3 activation we require at a minimum three points of intervention: caspase-3, caspase-8, and caspase-9. (Choices involving caspase-2 and caspase-6 are redundant to the effects of the other three inhibitors.)

We also looked into the ensemble variance as the best metric for measuring the consensus of the ensemble. Another possible measure, which is even simpler, is the variance of the predicted caspase-3 concentration at a single time-point, say at 8 hours. In Figure 4-13, the value of the full ensemble variance is plotted against the variance at 8 hours for all treatment options. Importantly, the metrics agree on all counts when there is good consensus among the models. Using the 8-hour time points to generate the decision tree did not alter any of the first three levels of cuts in the tree (data not shown).

4.4 Discussion

4.4.1 Simplifications and uncertainties not accounted for in the models

In this work our goal was to investigate the diversity or similarity of predictions from an ensemble of topologically different models. It was not our goal to capture all of the biological or modeling uncertainty about Fas signaling in these choices, but rather to consider some of the disparities in the existing literature and examine their impact. There are many simplications and omissions in both our Base and Master models. As a first example, there are several species that we left out of our models that are purportedly redundant in function or behavior. Caspase-10 is thought to behave as an initiator caspase, in a similar manner to caspase-8, and in most models is omitted for this reason [74–76, 155, 157, 161]. Similarly, for the sake of modeling, caspase-7 is usually considered redundant to effector caspase-3, at least in terms of apoptotic initiation, and thus has typically been lumped with caspase-3 in models. Experiments have neverthless shown several differences between caspase-7 and caspase-3. For instance, XIAP binds to each with slightly different affinities [187, 216] and c-IAP-1 appears to be a stronger inhibitor of caspase-7 while the reverse is true of c-IAP-2 [220]. It is also known that, once activated, caspase-3 has a more dominant role in the execution of cellular disassembly, and that the two proteins may have distinct targets [221]. These omissions are potentially relevant to our therapeutic cocktail design, as the inhibitor intended to bind caspase-8 may not also act on caspase-10, leaving an open channel for propagation of the apoptotic signal.

The Bcl-2 proteins are those that interact at the mitochondrion to prevent or promote release of cytochrome c and Smac, both of which behave pro-apoptotically. The Bcl-2 family itself consists of both pro- and anti-apoptotic proteins: Bax, Bak, Bim, Bid, Bik, Bad, and Noxa belonging to the former, and Mcl-1, Bcl-2, Bcl-W, A1A, and Bcl-xL, the latter [144]. The pro-apoptotic group can further be divided into two sub-classes. Bax and Bak are directly involved in mitochondrial membrane pore formation while the others, known as BH3-only proteins for their basic structural unit, are facilitators of this activity. In our model, there is a single representative from each of the main classes: Bax represents both Bax and Bak; Bid represents itself plus Bim, Bik, Bad, and Noxa; and Bcl-2 is the sole

anti-apoptotic protein.

Knockout or knockdown experiments of the different Bcl-2 family members have demonstrated that there are certainly within-class differences. Deletion of Bax in mice led to male sterility and lymphoid hyperplasia [222] while Bak deletion mutants showed no side-effects [223]. Double mutants of both Bax and Bak were inviable [223] so, while the precise roles of the two proteins may be distinct, they each have some ability to compensate for lack of the other. Another difference between Bax and Bak lies in localization. Bax resides in the cytosol until apoptotic initiation, whereas Bak is thought to be localized to the mitochondrial membrane regardless of apoptotic stimulus [224].

We also modeled a single inhibitor of apoptosis species, which we labeled XIAP. There are in fact a number of similar proteins, namely c-IAP-1 and c-IAP-2. All three act by binding to caspases and preventing downstream activity, but they likely exist at different concentrations in the cell, have different affinities for the various caspases, and also have different non-caspase binding partners. Roy et al. [220] showed that the c-IAPs can bind TRAF proteins but XIAP does not, and that XIAP is a more potent inhibitor of both caspase-3 and caspase-7 than either c-IAP by 2–3 log-fold.

One major concern about the potential success of the caspase inhibitor cocktail is that previous research has shown that, in some circumstances, relief of apoptosis only leads to necrotic or autophagic cell death [225–227]. It is unclear whether this cellular "back-up plan" is cell type-dependent and how specific it is to the particular cell stimulus. If the bypass occurs at the receptor level, rather than in the cytosol, the same inhibitors may be effective against cell-death in response to Fas ligand but not to TRAIL, for instance. In general, there is a poor understanding of how the caspase network connects to and influences other cellular pathways (see [228] as an example) and, furthermore, we have not attempted to include these connections in our models.

There is also a question of whether cells that are "rescued" from apoptosis are viable after treatment. For example, our simulations suggest that the best reduction of caspase-3 activation with the inhibitor choices tested is still only about 85%. The presence of that much activated caspase may or may not result in cellular disassembly. In this regard, we did not include any degradation (or constitutive production) processes in the models. Other

researchers have found that active caspase-3 is targetted for ubiquin-mediated degradation on the order of hours, which is within the time scale of processes considered here [161]. This degradation may be sufficient to account for the last 15% of activation, though we cannot do more than speculate without directly incorporating this effect into the model.

4.4.2   Recommended follow-up experiments

We suggest a number of experiments as follow-up to the computational studies performed here. The first set of recommended experiments relates to improved model building. In an ideal world, one would perform flow cytometry experiments to measure the initial (pre-stimulus) concentrations of several important components of the network with special attention to any differences between cell lines for which data is to be collected.

The model fitting could further be greatly improved with data collected at the single cell level, both to eliminate the impact of averaging over a population of stochastically distinct cells and to make clear whether the all-or-none property exists only in baseline cells or whether this phenomenon is also exhibited in the presence of Bcl-2 over-expression.

The flux balance in this branched pathway needs to be established and it cannot be done by measuring only network components that are shared by both the Type I and Type II pathway. One way to establish this balance would be to repeat the dose response experiments and to measure the time-dependent activation of Bid and/or caspase-9, in addition to caspase-3 and -8.

To eliminate candidate models from our ensemble, we suggest an experiment in which cells are subjected to irreversible inhibitors specific to caspase-6 and caspase-8, respectively. Finally, we would want to test our prediction of a cocktail of inhibitors of caspases-3, -8, and -9 as a potential treatment for preventing apoptosis in the presence of high concentrations of Fas ligand.

4.5   Conclusions

In this work we built an ensemble of 64 models that differed in their topologies, both for understanding the apoptosis pathway and designing treatment options for preventing cell

death. The construction of the ensemble highlighted many of the difficulties relating to the modeling of real biological pathways with differential equations, and, specific to this network and dataset, found that underspecification of the fluxes can lead to a wide variety of predicted behaviors under certain conditions. Nevertheless, all models of the ensemble agreed strongly on the cellular response to high ligand concentration, and we used our unique ensemble approach to test the sensitivity of this activation to different treatments using caspase inhibitors. Each specific combination of inhibitors led either to continued agreement on a quick activation, continued agreement but with reduced activation, or disagreement between models. We hope to conduct our suggested experiments in order to validate our methods but also to test the viability of the proposed caspase inhibitor therapy.
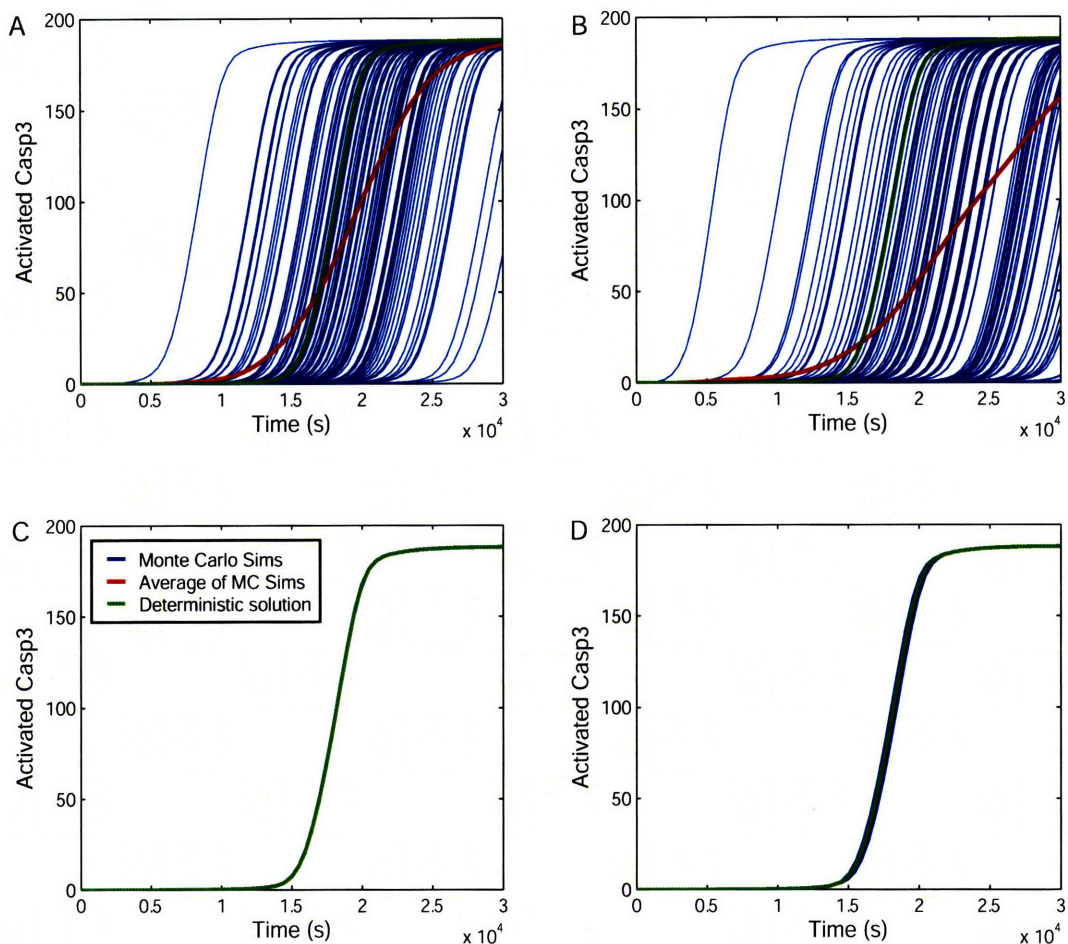
Figure 4-9: Simulations of the stochastic response of apoptotic activation to changes in Fas receptor and ligand concentrations. It is known that non-Bcl-2 over-expressing cells exhibit a switch-like response to apoptotic stimuli. Our data were from cell populations, but the models represent the dynamics of single cells. To rectify this difference, we verified that averaging over individual models could reproduce the data. Simulations in which the Fas ligand concentration was varied (at either the 5% (A) or 30% (B) level) produced vastly different switching times while the same variability in the concentration of Fas receptor (C and D) had little effect. All simulations shown here are for model 110001 with a mean Fas ligand concentration of 0.02 nM and a mean Fas receptor concentration of 10 nM.
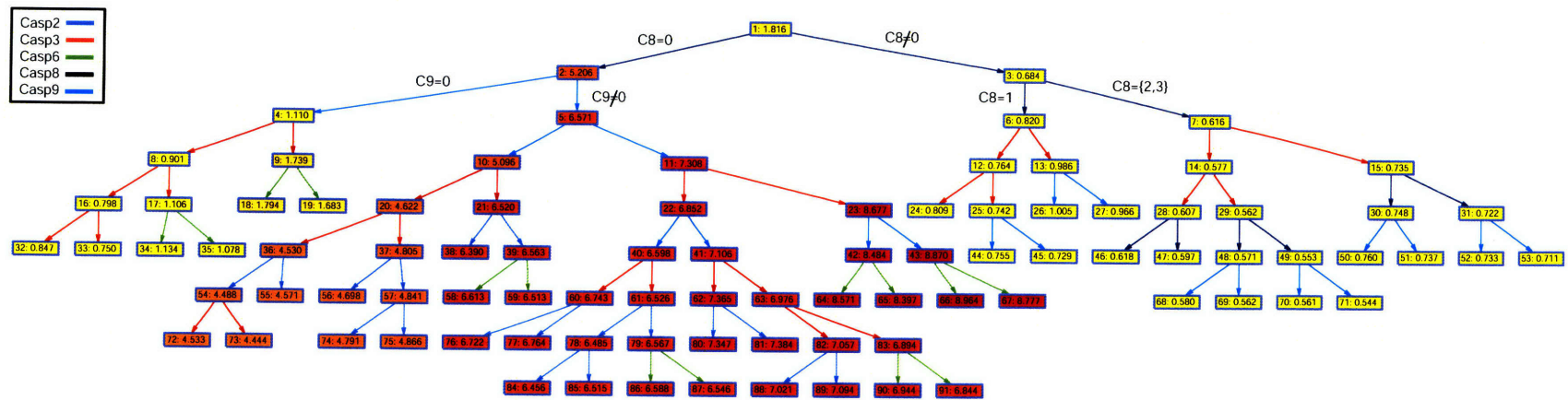
Figure 4-10: Decision tree for analyzing the prediction space of the perturbations by caspase inhibitor cocktail. Node color represents the ensemble variance ($\times 10^4$) for the predictions from the 64 models of the ensemble. Yellow corresponds to low variance while red corresponds to high variance. Cuts in the tree are made regarding inhibitor choices: 0 = irreversible inhibitor for that caspase, 1 = strong reversible inhibitor ($K_D = 10$ nM), 2 = weak reversible inhibitor ($K_D = 100$ nM), 3 = no inhibitor. There is high variance when the models are subjected to an irreversible inhibitor of caspase-8 and no inhibitor of caspase-9.
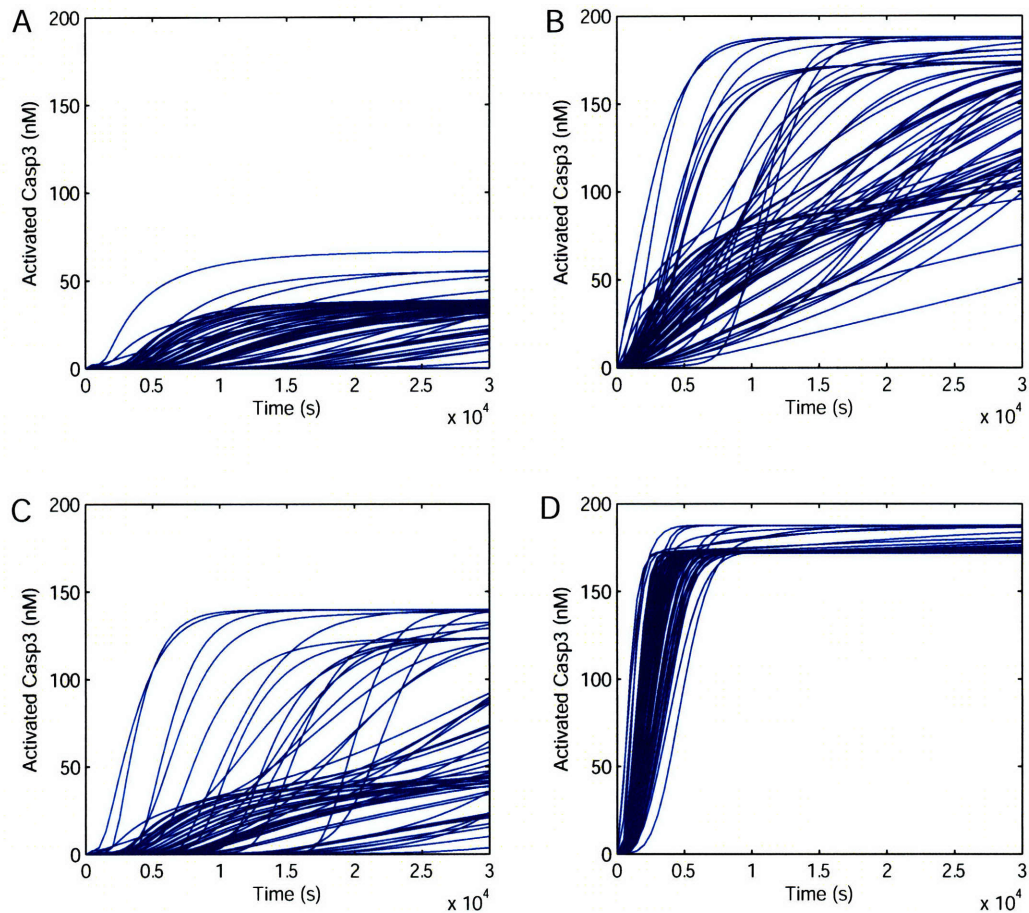
Figure 4-11: Demonstration of the predictions from the ensemble of models for different choices of caspase inhibitors. Models were subjected to a (A) irreversible inhibitors of caspases-3, -8, and -9; (B) irreversible inhibitors of caspases-6 and -8; (C) irreversible inhibitors of caspases-3, -6, and -8; and (D) irreversible inhibitors of caspases-2 and -6, a strong reversible inhibitor of caspase-8, and a weak reversible inhibitor of caspase-9. The first cocktail represents a strong decrease in caspase-3 activation while maintaining agreement among the models. Cocktails (B) and (C) represent the highest ensemble variance and highest steady-state variance of all the possible combinations: if cells were experimentally subjected to these combinations, it should be possible to eliminate some of the candidate models. Cocktail (D) demonstrates that given the right combinations, even four simultaneous inhibitors can have very little effect on caspase-3 activation.
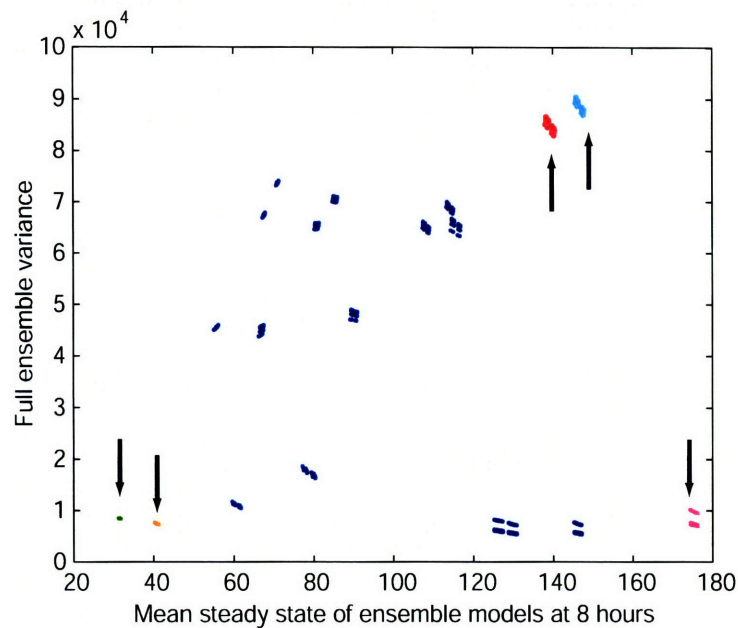
Figure 4-12: Ensemble variance vs. mean concentration of Casp3 at 8 hours for the 1024 caspase inhibitor cocktails. There are several clear groupings based on the particular composition of the inhibitor cocktail. Some notable cocktails are distinctly colored in the plot: irreversible inhibitors of caspase-3, -8, and -9 (green); irreversible inhibitors of caspase-8 and -9, and a strong reversible inhibitor of caspase-3 (orange); no inhibitor of caspase-3, irreversible inhibitor of caspase-8, weak reversible inhibitor of caspase-9 (red); no inhibitor of caspase-3 or -9, irreversible inhibitor of caspase-8 (cyan); no inhibitor of caspase-3 and any inhibitor of caspase-8 *except* irreversible (magenta).
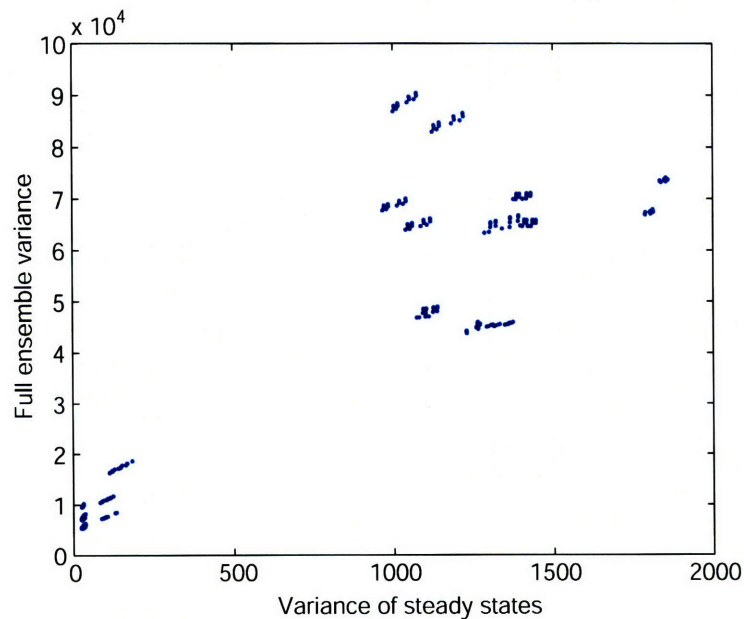

Figure 4-13: Comparison of the full ensemble variance to the variance of steady state values, computed across the 64 models of the ensemble for each of 1024 inhibitor perturbations.

Chapter 5

General Conclusions and Future Directions

In this thesis, we studied the current modeling paradigm in systems biology and proposed new tools for improving the way biological models are used, both for making predictions about response to perturbation and for learning about the biology. In Chapter 2, we used a simulation framework to explore the impact of parametric and topological sources of uncertainty on the ability of differential equation models to make accurate predictions. The simulation results highlighted several key points:

1. Model selection methods may not be able to identify the correct topology from a set of plausible candidates.

2. The best-fitting model, whether with the correct topology or otherwise, will not necessarily generate the most accurate predictions.

3. The model with the correct topology, whether the best-fitting model or otherwise, will not necessarily generate the most accurate predictions.

4. Uncertainty due to parameters cannot capture the full variation in predictions when topology is also unknown.

In Chapter 2 we also found that the consensus among a family of well-fitting models, as measured by an ensemble variance, was shown to be a very strong classifier of prediction

error. Together, these results emphasize the advantage of supplementing model selection with prediction selection when many models can capture the data well.

In Chapter 3, we used ensemble consensus as the basis for the development of methods designed for identifying high- and low-confidence predictions from differential equation models. One such method combined a model averaging approach with propagated statistical uncertainty, resulting in probability distributions of protein concentration as a function of time. Such distributions are useful in a number of ways. First, they can be used to make predictions that incorporate information about uncertainty in both topology and parameters. Second, they can highlight the species and time points for which we have the most or least forecasting information. Finally, they can be used to compute an overall score for a prediction that incorporates information from all species and time points, and this score can be used to determine whether the predictions under a particular perturbation are reliable or not.

Also in Chapter 3, we developed two methods for exploring the parameter space accessible by experimental perturbation, the goal being to identify the regions of this space corresponding to either high or low confidence predictions. In one method, we subjected model ensembles to Monte Carlo perturbations within the accessible parameter space, and computed the corresponding ensemble variance. We then used decision trees to map parameter space into regions of consistent consensus, and used opposing branches of the tree to (a) identify biologically interesting high-confidence predictions and (b) design experiments to refine the ensemble of plausible models. Our second method for mapping consensus in parameter space consisted of computing a new metric, the consensus sensitivities, which measures the sensitivity of the ensemble variance to changes in parameters.

We tested our method of mapping parameter space by applying it to a real biological system, as described in Chapter 4. We built 64 models of the Fas signaling network and used consensus among the ensemble to identify the best sepsis treatment possible with a cocktail of specific caspase inhibitors. In working with a larger system and real data, we encountered many of the difficulties associated with mechanistic modeling of biology. In particular, in the apoptosis network, more work needs to be done to understand the relative contributions of the Type I and Type II pathways, as model behavior in response to stimulus

was shown to be largely determined by the fluxes through these two parts of the network. Our experiences reinforced the need to be able to conduct systems biology iteratively, where experiments support modeling, and modeling promotes further experiments, as it is often unknown exactly what information will be needed to constrain a model properly. That said, our ensemble technique is one approach for maximizing the utility of the information at hand.

There are numerous avenues to be explored further based on the work presented here. Due to time and technical limitations, we did not test our model averaging technique on the apoptosis network, nor did we compute the consensus sensitivities. I recommend that both of these tests be performed to validate the techniques using a larger, more realistic system.

Another possible extension to this work is to implement consensus ideas in an optimization framework. Several experimental design algorithms for model discrimination already employ this basic idea, but rather than use them to seek out perturbations where models disagree, they could also be applied in the opposite way, to find perturbations for which consensus is reached. Instead of optimizing on the difference between model predictions, our objective function could consist of an ensemble variance term, balanced by the mean distance of the predictions from the fits:

$$f(\Delta\boldsymbol{\theta};\boldsymbol{\theta}) = \sum_{m=1}^{M} \left(x_m(\boldsymbol{\theta}+\Delta\boldsymbol{\theta}) - x_m(\boldsymbol{\theta})\right)^2 - \beta \sum_{m=1}^{M} \left(x_m(\boldsymbol{\theta}+\Delta\boldsymbol{\theta}) - \frac{1}{M}\sum_{k=1}^{M} x_k(\boldsymbol{\theta}+\Delta\boldsymbol{\theta})\right)^2$$

where $x_m(\boldsymbol{\theta})$ and $x_m(\boldsymbol{\theta}+\Delta\boldsymbol{\theta})$ are the values of the mth model at the fit and under perturbation, and the sums are over all models M. β is a tuning parameter to weight the relative importance of the ensemble variance compared to the distance from the fit. Optimizing such an objective function on $\Delta\boldsymbol{\theta}$ may lead to predictions that are both biologically interesting (because they are far from the data used to train the models) and of high-confidence (because the models agree on the predictions).

In the work presented here, the ensembles were constructed using single parameterizations of models with distinct topologies. It would not be difficult to expand the ensembles to include multiple parameterizations of each topology: the optimization functions of large

models with many unknown parameter values and limited data are often plagued by multiple local minima.

Yet another interesting avenue of further research might be to consider not only experimental perturbations, but also variation among patients or across cell types. For instance, one could imagine exploring the consensus among an ensemble of models built to represent different patients subjected to in silico drug treatments. The decision tree could then be used to identify treatments for which all patients were expected to respond similarly or treatments for which we would expect to see stratification. Ensembles could be built to include both biological uncertainties as well as patient differences and the methods could be used to analyse whether predicted patient differences were more or less significant than the dispersion seen across multiple biological representations.

Overall, the ensemble methods presented in this thesis lay the groundwork for many applications of biological modeling, and may be of particular interest in the field of drug development and clinical trial design. The depth of their utility, however, cannot be fully evaluated until they are integrated with experiments and it is hoped that future collaborations will enable this implementation.

Appendix A

Chapter 2 Supplementary Material

A.1   Model equations

Model 1

$$\frac{dx_A}{dt} = \gamma_A - k_A x_A \tag{A.1}$$

$$\frac{dx_B}{dt} = \frac{\gamma_B}{1 + (x_A/\kappa_{AB})^{\beta_{AB}}} - k_B x_B \tag{A.2}$$

$$\frac{dx_C}{dt} = \gamma_C + \frac{\alpha_{BC}(x_B)^{\beta_{BC}}}{(\xi_{BC})^{\beta_{BC}} + (x_B)^{\beta_{BC}}} - k_C x_C \tag{A.3}$$

Model 2

$$\frac{dx_A}{dt} = \gamma_A + \frac{\alpha_{AA}(x_A)^{\beta_{AA}}}{(\xi_{AA})^{\beta_{AA}} + (x_A)^{\beta_{AA}}} - k_A x_A \tag{A.4}$$

$$\frac{dx_B}{dt} = \frac{\gamma_B}{1 + (x_A/\kappa_{AB})^{\beta_{AB}}} - k_B x_B \tag{A.5}$$

$$\frac{dx_C}{dt} = \gamma_C + \frac{\alpha_{BC}(x_B)^{\beta_{BC}}}{(\xi_{BC})^{\beta_{BC}} + (x_B)^{\beta_{BC}}} - k_C x_C \tag{A.6}$$

Model 3

$$\frac{dx_A}{dt} = \gamma_A - k_A x_A \tag{A.7}$$

$$\frac{dx_B}{dt} = \frac{\gamma_B}{1 + (x_D/\kappa_{DB})^{\beta_{DB}}} - k_B x_B \tag{A.8}$$

$$\frac{dx_C}{dt} = \gamma_C + \frac{\alpha_{BC}(x_B)^{\beta_{BC}}}{(\xi_{BC})^{\beta_{BC}} + (x_B)^{\beta_{BC}}} - k_C x_C \tag{A.9}$$

$$\frac{dx_D}{dt} = \frac{\alpha_{AD}(x_D)^{\beta_{AD}}}{(\xi_{AD})^{\beta_{AD}} + (x_A)^{\beta_{AD}}} - k_D x_D \tag{A.10}$$

Model 4

$$\frac{dx_A}{dt} = \gamma_A - k_A x_A \tag{A.11}$$

$$\frac{dx_B}{dt} = \frac{\alpha_{DB}(x_D)^{\beta_{DB}}}{(\xi_{DB})^{\beta_{DB}} + (x_D)^{\beta_{DB}}} - k_B x_B \tag{A.12}$$

$$\frac{dx_C}{dt} = \gamma_C + \frac{\alpha_{BC}(x_B)^{\beta_{BC}}}{(\xi_{BC})^{\beta_{BC}} + (x_B)^{\beta_{BC}}} - k_C x_C \tag{A.13}$$

$$\frac{dx_D}{dt} = \frac{\gamma_D}{1 + (x_A/\kappa_{AD})^{\beta_{AD}}} - k_D x_D \tag{A.14}$$

Model 5

$$\frac{dx_A}{dt} = \gamma_A - k_A x_A \tag{A.15}$$

$$\frac{dx_B}{dt} = \frac{\gamma_B}{1 + (x_A/\kappa_{AB})^{\beta_{AB}}} + \frac{\alpha_{BB}(x_B)^{\beta_{BB}}}{(\xi_{BB})^{\beta_{BB}} + (x_B)^{\beta_{BB}}} - k_B x_B \tag{A.16}$$

$$\frac{dx_c}{dt} = \gamma_C + \frac{\alpha_{BC}(x_B)^{\beta_{BC}}}{(\xi_{BC})^{\beta_{BC}} + (x_B)^{\beta_{BC}}} - k_C x_C \tag{A.17}$$

Model 6

$$\frac{dx_A}{dt} = \gamma_A - k_A x_A \tag{A.18}$$

$$\frac{dx_B}{dt} = \frac{\gamma_B}{1 + (x_A/\kappa_{AB})^{\beta_{AB}}} - k_B x_B \tag{A.19}$$

$$\frac{dx_C}{dt} = \gamma_C + \frac{\alpha_{DC}(x_D)^{\beta_{DC}}}{(\xi_{DC})^{\beta_{DC}} + (x_D)^{\beta_{DC}}} - k_C x_C \tag{A.20}$$

$$\frac{dx_D}{dt} = \frac{\alpha_{BD}(x_B)^{\beta_{BD}}}{(\xi_{BD})^{\beta_{BD}} + (x_B)^{\beta_{BD}}} - k_D x_D \tag{A.21}$$

A.2 Parameters for Models 1 and 4 as actual systems

Table A.1: Parameters for Models 1 and 4 when they were used as actual systems to generate simulated data

Parameter Name	Biological Meaning	Units	Value in Actual System Model 1	Model 4
γ_A			0.05	0.05
γ_C	Constitutive production rate	[conc.]/[time]	0	0
γ_B			1	n/a
γ_D			n/a	1
κ_{AB}	Concentration at half-maximal	[conc.]	1	n/a
κ_{AD}	expression rate under repression		n/a	1
α_{BC}	Maximal expression rate under	[conc.]/[time]	1	1
α_{DB}	promotion		n/a	1
ξ_{BC}	Concentration at half-maximal	[conc.]	1	1
ξ_{DB}	expression rate under promotion		n/a	1
β_{AB}			2	n/a
β_{AD}	Hill coefficient of regulation	none	n/a	1
β_{BC}			1	1
β_{DB}			n/a	2
k_A			0.2	0.2
k_B	Degradation rate constant	1/[time]	0.4	0.4
k_C			0.4	0.4
k_D			n/a	0.4

A.3 Effect of reduced noise on model selection

The model selection results with 20% additive noise are very similar to those obtained with 30% noise. Maximum likelihood (ML) is unable to identify Topology 1 as the basis of the actual system although the AIC and BIC are more successful. When Model 4 was used as the actual system, Topology 4 gets selected with relative frequency using ML, though the AIC and BIC overpenalize the larger models. As with 30% noise then, there is an inconsistency with all three model selection procedures.
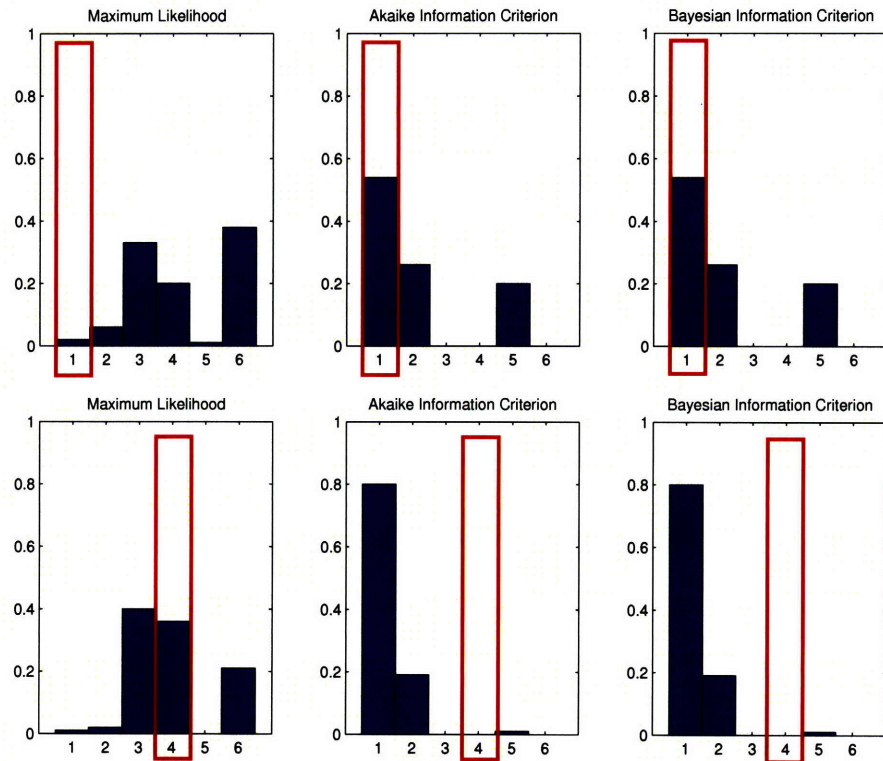
Figure A-1: Histograms for the number of times each model was ranked best with Model 1 (top) or Model 4 (bottom) as the actual system and 20% noise. Data represent optimizations of all six topologies to 100 datasets that differed only in the noise added to the deterministic simulation of the actual system.
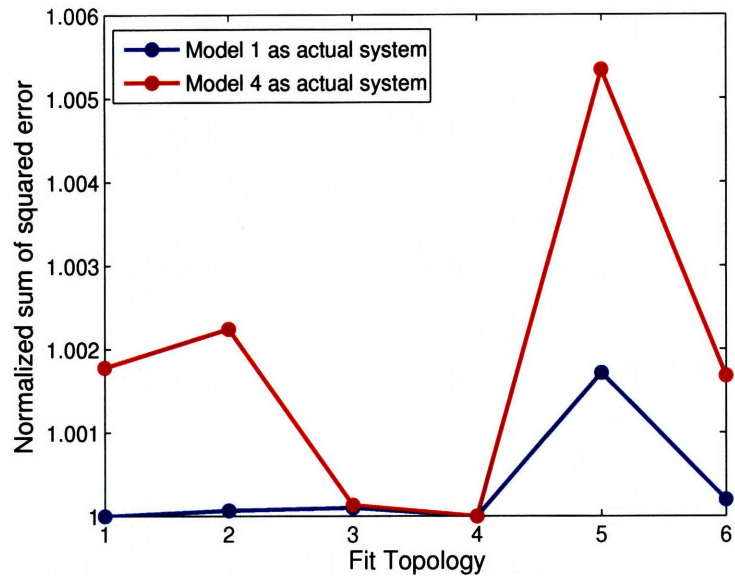
A.4   Model identifiability

Figure A-2: Identifiability of the different topologies with Model 1 or Model 4 as the actual system

Data were generated using Model 1 or Model 4 as the actual system, sampled at unit time points from $t = 0$ to $t = 10$, with no noise added. All six topologies were then fit to the noiseless data, producing the normalized negative log likelihood scores as shown above. In the case of Model 1 as the actual system, there is poor identifiability among the candidate topologies and, in particular, Topology 4 can be made to generate as good a fit as the correct topology, Topology 1. With Model 4 as the actual system, the identifiability is somewhat better, as demonstrated by the greater differences between the likelihood of the fit of the correct topology and those of the incorrect topologies. Topology 3 can, however, also generate high quality fits.

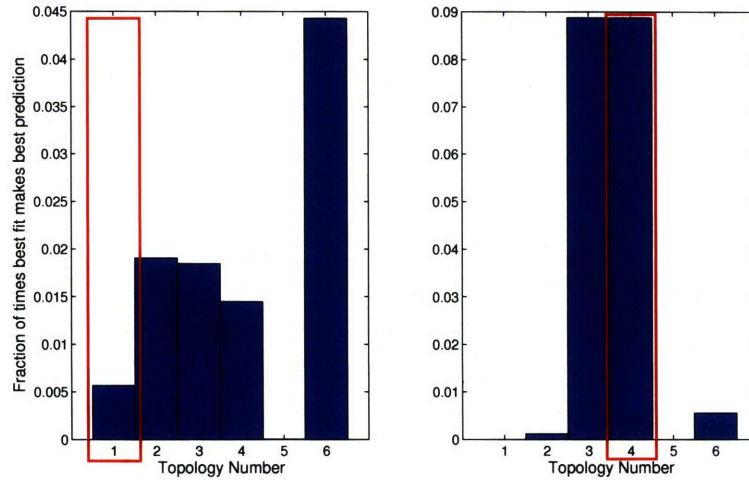# A.5 Probability that the best fitting model will make the best predictions

Figure A-3: Histograms for the fraction of times that each topology generates the best fit and also make the best predictions. The results are for Model 1 (A) and Model 4 (B) as the actual system, 50 datasets with 30% noise and 1000 perturbations.
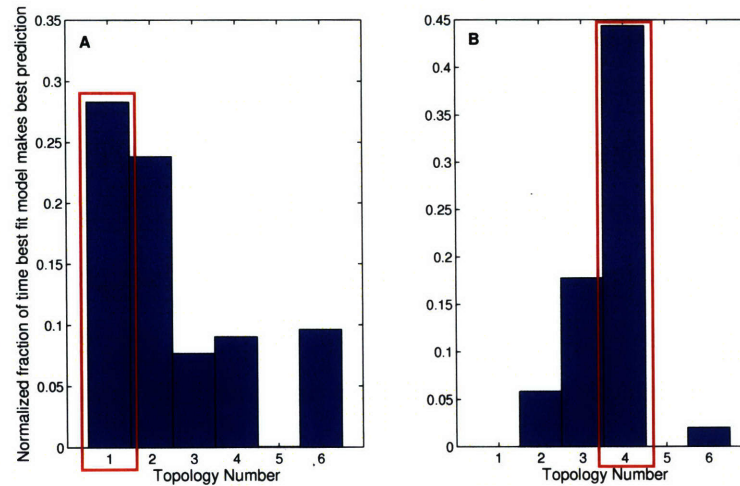

Figure A-4: Histograms for the number of times each model is the best fit and also makes the best prediction, normalized by the number of times each topology was selected as the best fit. The results are for Model 1 (A) and Model 4 (B) as the actual system, 50 datasets with 30% noise and 1000 perturbations.

A.6 Impact of ensemble choice on consensus

It is important to consider the impact of omitting the correct topology. In real biological examples, unlike the simulated systems tested here, it is certain that no model under consideration will be a complete and accurate description of the true underlying mechanism. We repeated the calculations of ensemble variance for the perturbation experiment, this time leaving out models with Topology 1. The conclusion remained the same in this case—that consensus among models implied higher prediction accuracy.
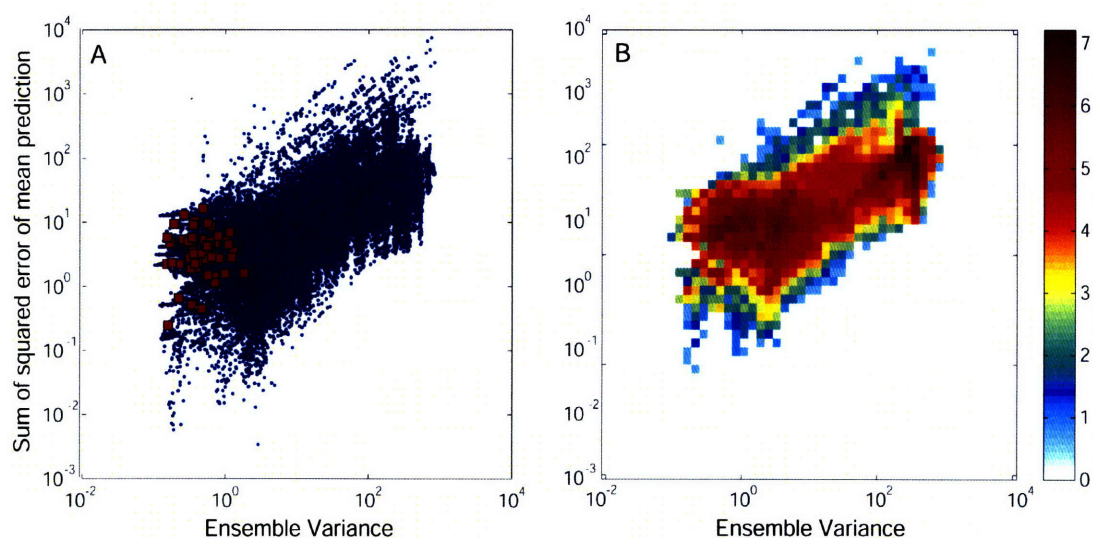

Figure A-5: Error of prediction vs. ensemble variance when the correct topology is omitted from the ensemble. The same data, fits and perturbations that were used to generate Figure 2-8 were again used here but predictions from models using Topology 1 were omitted from the calculations of the mean prediction error and the ensemble variance. Panel B is a two-dimensional histogram of the data in Panel A.

A.7 Receiver Operating Characteristic (ROC) curves for individual dataruns

The ROC curve shown in Figure 2-9 is an average over 50 trials with unique datasets. The ROC curves for individual datasets are noisier but also demonstrate the strength of consensus as a classifier. It is interesting that the position of the fit value of the consensus on these curves (shown as red stars) is quite variable: this indicates that the number of true positives inside the fit threshold is inconsistent. This might be the result of high sensitivity

to noise and suggests that either the Monte Carlo sampling rate should be increased or else the fit threshold might be too strict.
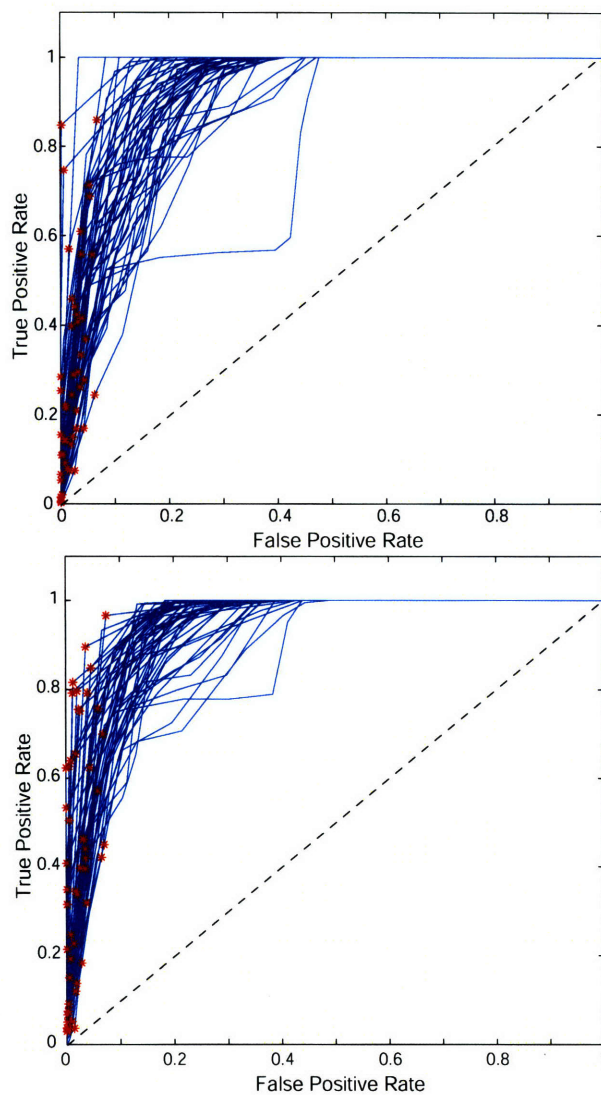

Figure A-6: ROC for 50 data runs with Model 1 (top) or Model 4 (bottom) as actual system. Red stars correspond to the classification scheme where the threshold of discrimination between good and poor predictions is the value of the ensemble variance at the fit.

Appendix B

Chapter 3 Supplementary Material

Derivation of probability under model averaging

$$
\begin{aligned}
\log p_{i,t}^{\log}(y) &= \log \prod_{m=1}^{M} (p_{i,m,t})^{\lambda_{i,m,t}} \\
&= \sum_{m=1}^{M} \lambda_{i,m,t} \log p_{i,m,t}(y) \\
&= \sum_{m=1}^{M} \lambda_{i,m,t} \log \left[\frac{1}{\sqrt{2\pi\sigma_{i,m,t}^2}} \exp\left(-\frac{(y - y_{i,m,t})^2}{2\sigma_{i,m,t}^2} \right) \right] \\
&= -\sum \lambda_{i,m,t} \left[\frac{1}{2}\log 2\pi + \log \sigma_{i,m,t} + \frac{(y - y_{i,m,t})^2}{2\sigma_{i,m,t}^2} \right] \\
&= -\left[\frac{1}{2}\log 2\pi + \sum_{m=1}^{M} \lambda_{i,m,t} \log \sigma_{i,m,t} + \frac{1}{2}\sum_{m=1}^{M} \frac{\lambda_{i,m,t}(y - y_{i,m,t})^2}{\sigma_{i,m,t}^2} \right]
\end{aligned}
$$

At the maximum probability:

$$
\begin{aligned}
\frac{d}{dy}\log p_{i,t}^{\log}\Big|_{y=y^{MP}} &= 0 \\
\Rightarrow \frac{1}{2}\sum_{m=1}^{M} 2\lambda_{i,m,t}\frac{(y^{MP} - y_{i,m,t})}{\sigma_{i,m,t}^2} &= 0 \\
\Rightarrow y^{MP} &= \sum_{m=1}^{M} \frac{\lambda_{i,m,t}y_{i,m,t}}{\sigma_{i,m,t}^2} \Bigg/ \sum_{m=1}^{M} \frac{\lambda_{i,m,t}}{\sigma_{i,m,t}^2}
\end{aligned}
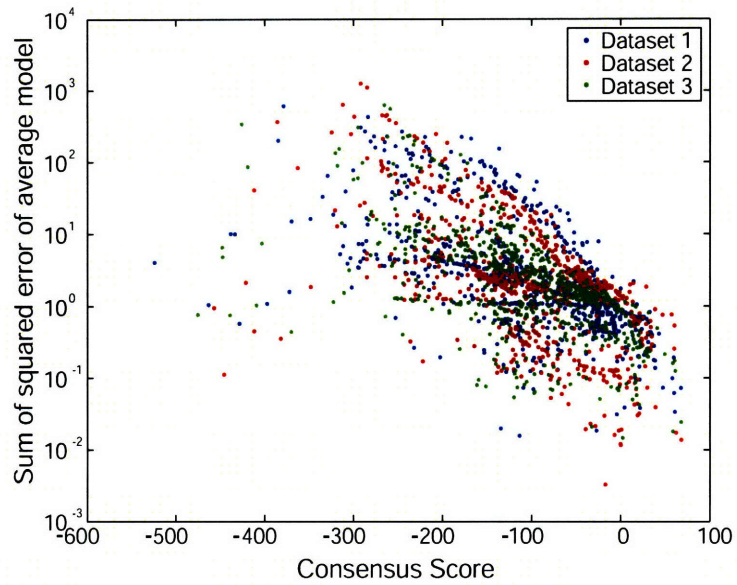\qquad (B.1)
$$

Additional figures

Figure B-1: Error of MP prediction vs. consensus score for individual data sets. Data sets were all generated from Model 1 as the actual system with 30% noise. The same perturbations were applied to all data sets. Noise affects the specific errors and scores but the overall correlation is independent.
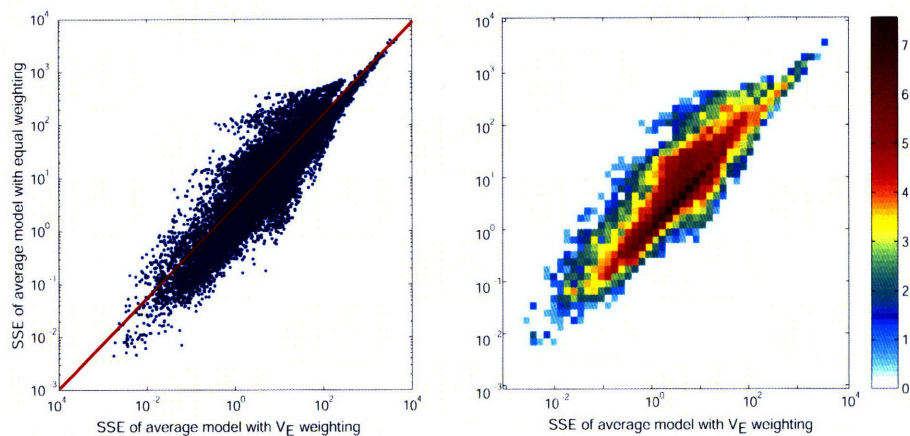
Figure B-2: Comparison of the errors of the MP predictions from equal or distance-from-consensus weighting. The correlation coefficient is 0.9292 and the distance-from-consensus measure performs slightly better on average than equal weighting. Importantly, for large errors, the two weighting methods are very nearly equivalent.
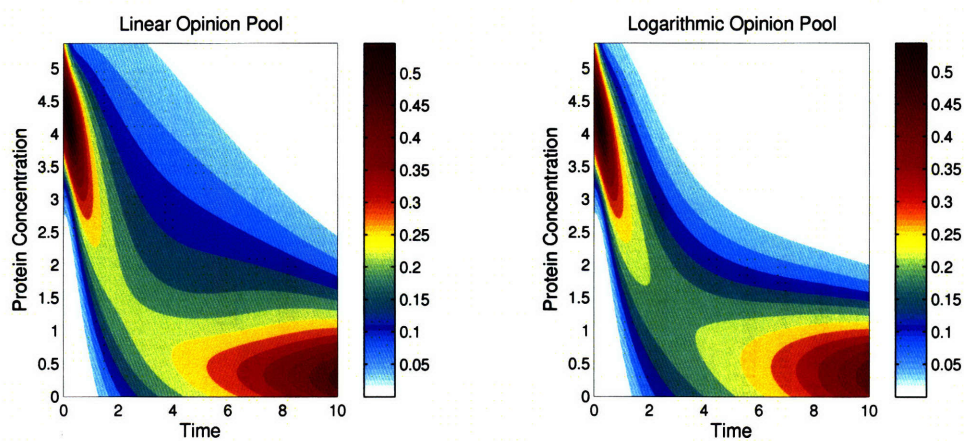
Figure B-3: Comparison of the linear and logarithmic opinion pools. For the same set of fit models under the same perturbation, the probability distributions associated with the linear (left) and logarithmic (right) opinion pools were computed. The maximum probability (MP) trajectory is very similar under the two methods, and the overall distributions share similar features.
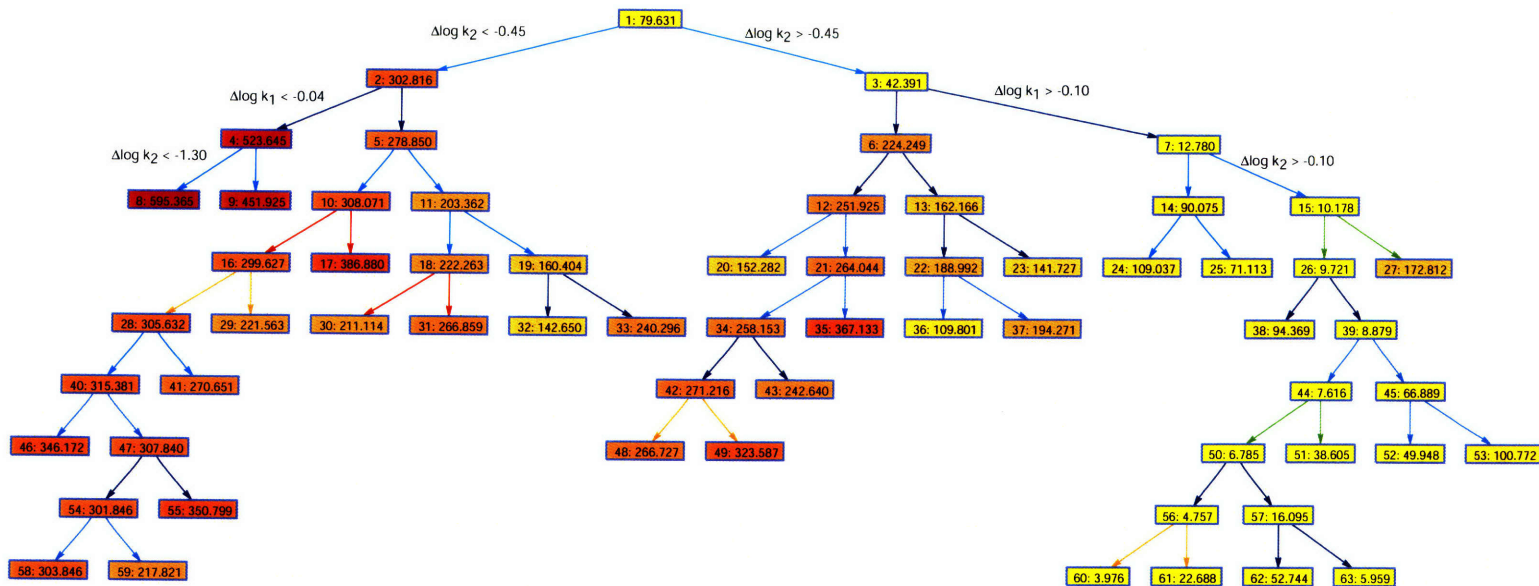
Figure B-4: The effect of a different noisy data set on the construction of the decision tree. An ensemble was built by fitting all six topologies to data from Model 1 as the actual system with 30% noise. The data differed only in the noise from that used to build the tree in Figure 3-4 and the same perturbations were applied. The first three levels of cuts are consistent between the trees and overall there is a high degree of similarity.
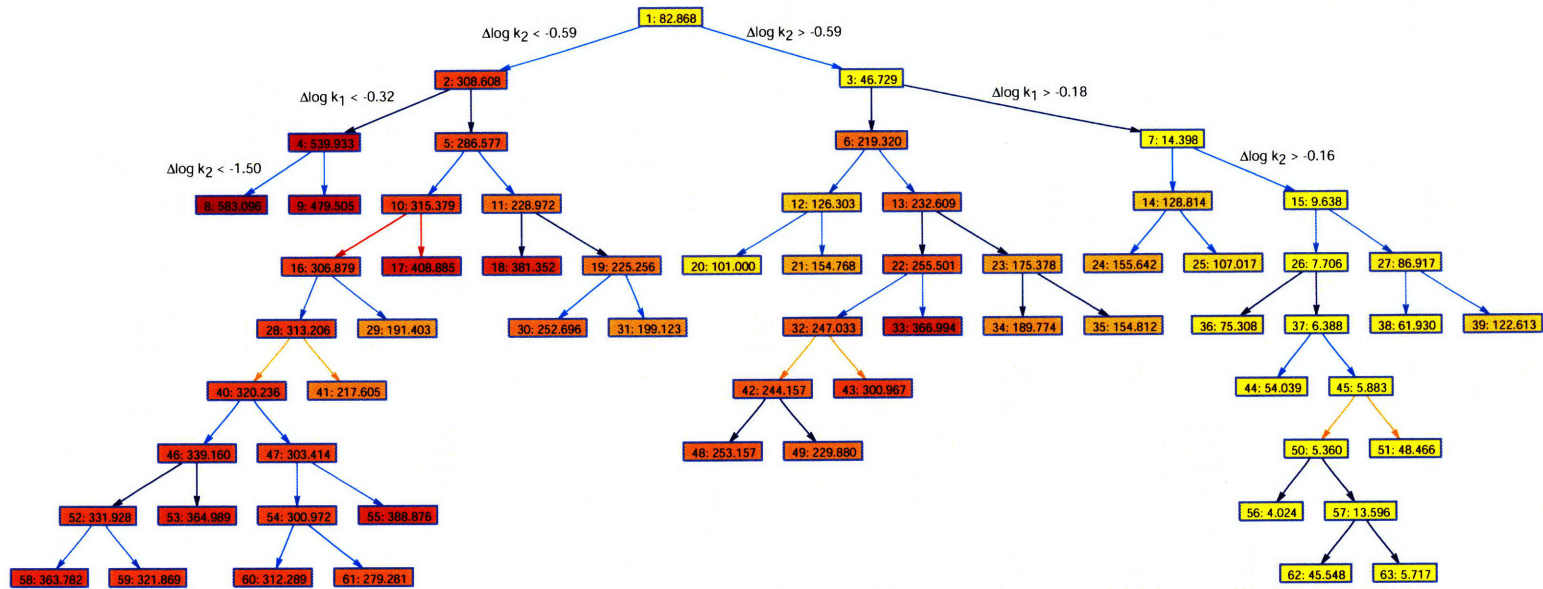
Figure B-5: Decision tree built on data from Model 4 as the actual system. The same procedure as was followed to build the tree shown in Figure 3-4 was applied using an ensemble fit to data from Model 4 as the actual system with 30% noise.
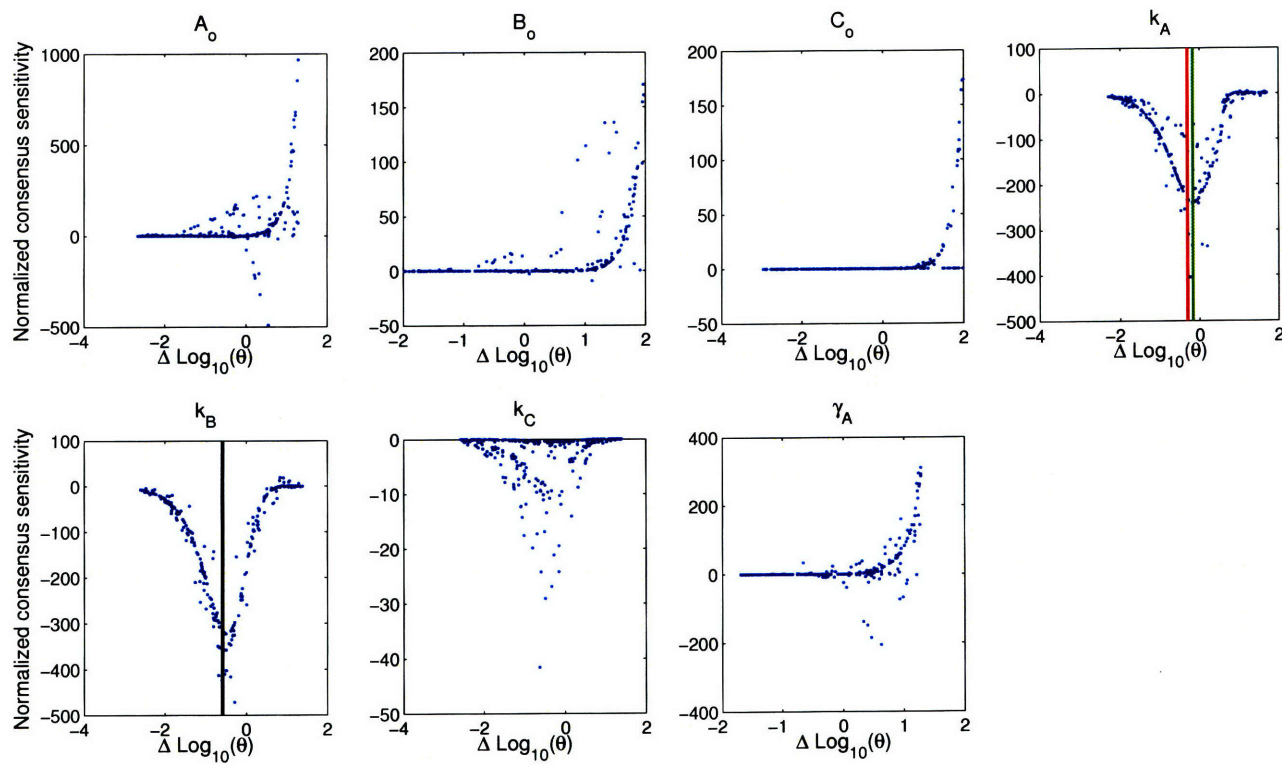
Figure B-6: Consensus sensitivities for an ensemble fit to data from Model 4 as the actual system. The black line shows the location of the first cut in the decision tree built on the same perturbation experiment (Figure B-5), and the red and green lines show the positions of the two second level cuts.

Bibliography

[1] Ideker T, Galitski T, and Hood L. A new approach to decoding life: Systems biology. *Annu. Rev. Genom. Hum. G.*, 2:343–372, 2001.

[2] Kitano H. Systems biology: A brief overview. *Science*, 295:1662–1664, 2002.

[3] Lander ES, Linton LM, Birren B, Nusbaum C, and Zody MC. Initial sequencing and analysis of the human genome. *Nature*, 409:860–921, 2001.

[4] Venter JC, Adams MD, Myers EW, Li PW, and et al. Mural RJ. The sequence of the human genome. *Science*, 291:1304–1351, 2001.

[5] Schena M, Shalon D, Davis RW, and Brown PO. Quantitative monitoring of gene-expression patterns with a complementary-dna microarray. *Science*, 270:467–470, 1995.

[6] Durfee T, Becherer K, Chen PL, Yeh SH, Yang YZ, Kilburn AE, Lee WH, and Elledge SJ. The retinoblastoma protein associates with the protein phosphatase type-1 catalytic subunit. *Gene. Dev.*, 7:555–569, 1993.

[7] Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, and Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A.*, 98:4569–4574, 2001.

[8] MacBeath G and Schreiber SL. Printing proteins as microarrays for high-throughput function determination. *Science*, 289:1760–1763, 2000.

[9] Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, and et al. Knight JR. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, 403:623–627, 2000.

[10] Aebersold R and Mann M. Mass spectrometry-based proteomics. *Nature*, 422:198–207, 2003.

[11] Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, and Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.*, 17:994–999, 1999.

[12] Longtine MS, McKenzie A, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, and Pringle JR. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast*, 14:953–961, 1998.

[13] Fraser AG, Kamath RS, Zipperlan P, Martinez-Campos M, Sohrmann M, and Ahringer J. Functional genomic analysis of C-elegans chromosome I by systematic RNA interference. *Nature*, 408:325–330, 2000.

[14] Zamore PD, Tuschl T, Sharp PA, and Bartel DP. RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, 101:25–33, 2000.

[15] Collins FS, Lander ES, Rogers J, and Waterston RH. Finishing the euchromatic sequence of the human genome. *Nature*, 431:931–945, 2004.

[16] Brett D, Pospisil H, Valvarcel J, Reich J, and Bork P. Alternative splicing and genome complexity. *Nat. Genet.*, 30:29–30, 2001.

[17] Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, and Darnell J. *Molecular Cell Biology*. WH Freeman, United States, 1999.

[18] Somogyi R and Sniegoski C. Modeling complexity of genetic networks. *Complexity*, 1:45–63, 1996.

[19] Ferrell JE Jr and Machleder EM. The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. *Science*, 280:895–898, 1998.

[20] Darzynkiewixz Z, Bruno S, Delbino G, Gorczyca W, Hotz MA, Lassota P, and Traganos F. Features of apoptotic cells measured by flow-cytometry. *Cytometry*, 13:795–808, 1992.

[21] Gillespie DT. Exact stochastic simulation of coupled chemical-reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[22] Arkin A, Ross J, and McAdams HH. Stochastic kinetic analysis of developmental pathway bifurcation. *Genetics*, 149:1633–1648, 1998.

[23] Thattai M and van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.*, 98:8614–8619, 2001.

[24] Lee MLT, Kuo FC, Whitmore GA, and Sklar J. Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridations. *Proc. Natl. Acad. Sci. U. S. A.*, 97:9834–9839, 2000.

[25] Brown CS, Goodwin PC, and Sorger PK. Image metrics in the statistical analysis of DNA microarray data. *Proc. Natl. Acad. Sci. U. S. A.*, 98:8944–8949, 2001.

[26] Aksenov SV, Church B, Dhiman A, Georgieva A, Sarangapani R, Helmlinger G, and Khalil IG. An integrated approach for inference and mechanistic modeling for advancing drug development. *FEBS Lett.*, 579:1878–1883, 2005.

[27] Butcher EC, Berg EL, and Kunkel EJ. Systems biology and drug discovery. *Nat. Biotechnol.*, 22:1253–1259, 2004.

[28] Butcher EC. Can cell systems biology rescue drug discovery? *Nat. Rev. Drug Disc.*, 1:1–6, 2005.

[29] Kell DB. Systems biology, metabolic modelling and metabolomics in drug descovery and development. *Drug Discov. Today*, 11:1085–1092, 2006.

[30] Ideker T and Lauffenburger DA. Building with a scaffold: Emerging strategies for high- to low-level cellular modeling. *Trends Biotech.*, 21:255–262, 2003.

[31] Eisen MB, Spellman PT, Brown PO, and Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.*, 95:14863–14868, 1998.

[32] Brown MPS, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M, and Haussler D. Knoweldge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. U. S. A.*, 97:262–267, 2001.

[33] Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. U. S. A.*, 98:15149–15154, 2001.

[34] Sachs K, Gifford D, Jaakkola T, Sorger P, and Lauffenburger DA. Bayesian network approach to cell signaling pathway modeling. *Science STKE*, 20:PE38, 2002.

[35] Sachs K, Perez O, Pe'er D, Lauffenburger DA, and Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308:523–529, 2005.

[36] Lahav G, Rosenfeld N, Sigal A, Geva-Zatorsky N, Levine AJ, Elowitz MB, and Alon U. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat. Genet.*, 36:147–150, 2004.

[37] Yeger-Lotem E, Sattath S, Kashtan N, Itzkovitz S, Milo R, Pinter RY, Alon U, and Margalit H. Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proc. Natl. Acad. Sci. U. S. A.*, 101:5934–5939, 2004.

[38] Johnston RJ, Chang S, Etchberger JF, Ortiz CO, and Hobert O. MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc. Natl. Acad. Sci. U. S. A.*, 102:12449–12454, 2005.

[39] Boyce WE and DiPrima RC. *Elementary differential equations and boundary value problems*. John Wiley & Sons, Inc., New York, 1997.

[40] Gillespie DT. The chemical Langevin equation. *J. Chem. Phys.*, 113:297–306, 2000.

[41] Viswanathan GA, Seto J, Patil S, Nudelman G, and Sealfon SC. Getting started in biological pathway construction and analysis. *PLoS Comput. Biol.*, 4:e16, 2008.

[42] Orton RJ, Sturm OE, Vyshemirsky V, Calder M, Gilbert DR, and Kolch W. Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway. *Biochem. J.*, 392:249–261, 2005.

[43] Adiwijaya BS. *Simulation and optimization tools to study design principles in biological networks*. PhD thesis, Massachusetts Institute of Technology, 2006.

[44] Wolf-Yadlin A, Kumar N, Zhang Y, Hautaniemi S, Zaman M, Kim H.-D., Grantcharova V, Lauffenburger DA, and White FM. Effect of HER2 overexpression on cell signaling networks governing proliferation and migration. *Mol. Syst. Biol.*, 2:e54, 2006.

[45] Krebs EG and Beavo JA. Phosphorylation-dephosphorylation of enzymes. *Ann. Rev. Biochem.*, 48:923–959, 1979.

[46] Edelstein-Keshet L. *Mathematical Models in Biology*. Random House, New York, 1988.

[47] Lauffenburger DA and Linderman JL. *Receptors: Models for binding, trafficking, and signaling*. Oxford University Press, New York, 1993.

[48] Hill WJ, Hunter WG, and Wichern DW. A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics*, 10:145–160, 1968.

[49] Moles CG, Mendes P, and Banga JR. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.*, 13:2467–2474, 2003.

[50] Bard Y. *Nonlinear Parameter Estimation*. Academic Press, New York, 1974.

[51] Akaike H. Likelihood of a model and information criteria. *J. Economet.*, 16:3–14, 1981.

[52] Burnham KP and Anderson DR. *Model Selection and Multimodel Inference*. Springer, New York, 1998.

[53] Rissanen J. A universal prior for integers and estimation by minimum description length. *Ann. Stat.*, 11:416–431, 1983.

[54] Bozdogan H. Akaike's information criterion and recent developments in information complexity. *J. Math. Psych.*, 44:62–91, 2000.

[55] Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. *J. Am. Stat. Assoc.*, 78:316–331, 1983.

[56] Davison AC and Hinkley DV. *Boostrap methods and their application*. Cambridge University Press, New York, 1997.

[57] Box GEP and Hill WJ. Discrimination among mechanistic models. *Technometrics*, 9:57–71, 1967.

[58] Chen BH and Asprey AP. On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations. *Ind. Eng. Chem. Res.*, 42:1379–1390, 2003.

[59] Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, Bullinger E, Allgower F, and Gilles ED. A benchmark for methods in reverse engineering and model discrimination: Problem formulation and solutions. *Genome Research*, 14:1773–1785, 2004.

[60] Vatcheva I, de Jong H, Bernard O, and Mars NJI. Experiment selection for the discrimination of semi-quantitative model of dynamical systems. *Artif. Intell.*, 170:472–506, 2006.

[61] Apgar JF, Toettcher JE, Endy D, White FM, and Tidor B. Stimulus design for model selection and validation in cell signaling. *PLOS Comput. Biol.*, 4:e30, 2008.

[62] Barkai N and Leibler S. Robustness in simple biochemical networks. *Nature*, 387:913–917, 1997.

[63] Alon U, Surette MG, Barkai N, and Leibler S. Robustness in bacterial chemotaxis. *Nature*, 397:168–171, 1999.

[64] Yi TM, Huang Y, Simon MI, and Doyle J. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl. Acad. Sci. U. S. A.*, 97:46–49, 2000.

[65] Rao CV, Wolf DM, and Arkin AP. Control, exploitation and tolerance of intracellular noise. *Nature*, 420:231–237, 2002.

[66] Csete M and Doyle JC. Reverse engineering of biological complexity. *Science*, 285:1664–1669, 2002.

[67] Han JDJ, Bertin N, Hao T, Goldberg DS, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, 430:88–93, 2004.

[68] Stelling J, Sauer U, Szallasi Z, Doyle FJ, and Doyle J. Robustness of cellular functions. *Cell*, 118:675–685, 2004.

[69] Eissing T, Allgower F, and Bullinger E. Robustness properties of apoptosis models with respect to parameter variations and intrinsic noise. *IEE Proc.-Syst. Biol.*, 152:221–228, 2005.

[70] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, and Sethna JP. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comp. Biol.*, 3:1871–1878, 2007.

[71] Socolovsky M, Murrell M, Liu Y, Pop R, Porpiglia E, and Levchenko A. Negative autoregulation by FAS mediates robust fetal erythropoiesis. *PLoS Biol.*, 5:2296–2311, 2007.

[72] Draper D. Assessment and propagation of model uncertainty. *J. R. Statist. Soc. B*, 57:45–97, 1995.

[73] Nagata S. Apoptosis by death factor. *Cell*, 88:355–365, 1997.

[74] Bentele M, Lavrik I, Ulrich M, Stößer, Heerman DW, Kalthoff H, Krammer PH, and Eils R. Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis. *J. Cell. Biol.*, 166:839–851, 2004.

[75] Eissing T, Conzelmann H, Gilles ED, Allgower F, Bullinger E, and Scheurich P. Bistability analyses of a caspase activation model for receptor-induced apoptosis. *J. Biol. Chem.*, 279:36892–36897, 2004.

[76] Bagci EZ, Vodovotz Y, Billiar TR, Ermentrout GB, and Bahar I. Bistability in apoptosis: roles of Bax, Bcl-2 and mitochondrial permeability transition pores. *Biophys. J.*, 90:1546–1559, 2006.

[77] Cui J, Chen C, Lu H, Sun T, and Shen P. Two independent positive feedbacks and bistability in the Bcl-2 apoptotic switch. *PloS One*, 1:e1469, 2008.

[78] Strogatz SH. *Nonlinear Dynamics and Chaos*. Perseus Publishing, Cambridge, 1994.

[79] Kim BS, Tidor B, and White J. Unpublished results. 2008.

[80] Wiley SH, Shvartsman SY, and Lauffenburger DA. Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol.*, 13:43–50, 2003.

[81] Tyson JJ. Modeling the cell-division cycle Cdc2 and Cyclin interactions. *Proc. Natl. Acad. Sci. U. S. A.*, 88:7328–7332, 1991.

[82] Andrews SS and Arkin AP. Simulating cell biology. *Curr. Biol.*, 16:R523–R527, 2006.

[83] Brent R and Endy D. Modelling cellular behaviour. *Nature*, 409:391–395, 2001.

[84] Jaqaman K and Danuser G. Linking data to models: data regression. *Nat. Rev. Mol. Cell Biol.*, 7:813–819, 2006.

[85] Friedman N. Inferring cellular networks using probabilistic graphical models. *Science*, 303:799–805, 2004.

[86] Gardner TS, di Bernardo D, Lorenz D, and Collins JJ. Inferring genetic networks and identifying compound mode of action via expresion profiling. *Science*, 301:102–105, 2003.

[87] Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV, and Hoek JB. Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci. U. S. A.*, 99:12841–12846, 2002.

[88] Samoilov M, Arkin A, and Ross J. On the deduction of chemical reaction pathways from measurements of time series concentrations. *Chaos*, 11:108–114, 2001.

[89] Tegner J, Yeung MKS, Hasty J, and Collins JJ. Reverse engineering gene networks: Integrating genetic perturbations with dynamical modeling. *Proc. Natl. Acad. Sci. U. S. A.*, 100:5944–5949, 2003.

[90] Rice JA. *Mathematical Statistics and Data Analysis*. Thomson Books/Cole, United States, 2007.

[91] Buckland ST, Burnham KP, and Augustin NH. Model selection: an integral part of inference. *Biometrics*, 53:603–618, 1997.

[92] Gadkar KG, Gunawan R, and Doyle FJ. Iterative approach to model identification of biological networks. *BMC Bioinformatics*, 6:155, 2005.

[93] Brown KS and Sethna JP. Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E*, 68, 2003.

[94] Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, and Cerione RA. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys. Biol.*, 1:184–195, 2004.

[95] Battogtokh D, Asch DK, Case ME, Arnold J, and Schuttler H-B. An ensemble method for identifying regulatory circuits with special reference to the *qa* gene cluster of *Neurospora crassa. Proc. Natl. Acad. Sci. U. S. A.*, 99:16904–16909, 2002.

[96] Joshi M, Seidel-Morgenstern, and Kremling A. Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metab. Eng.*, 8:447–455, 2006.

[97] Liebermeister W and Klipp E. Biochemical networks with uncertain parameters. *IEE Proc.-Syst. Biol*, 152:97–107, 2005.

[98] Waterfall JJ, Casey FP, Gutenkunst RN, Brown KS, Myers CR, Brouwer PW, Elser V, and Sethna JP. Sloppy-model universality class and the vandermonde matrix. *Phys. Rev. Letters*, 97, 2006.

[99] Araujo RP, Petricoin EF, and Liotta LA. A mathematical model of combination therapy using the EGFR signaling network. *Biosystems*, 80:57–69, 2005.

[100] Chatfield C. Model uncertainty, data mining and statistical inference. *J. R. Statist. Soc. A*, 158:419–466, 1995.

[101] Wood SN and Thomas MB. Super-sensitivity to structure in biological models. *Proc. Roy. Soc. (B)*, 266:565–570, 1999.

[102] Kuepfer L, Peter M, Sauer U, and Stelling J. Ensemble modeling for analysis of cell signaling dynamics. *Nat. Biotechnol.*, 25, 2007.

[103] Gardner TS, Cantor CR, and Collins JJ. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403:339–342, 2000.

[104] Mettetal JT, Muzzey D, Pedraza JM, Ozbudak EM, and van Oudenaarden A. Predicting stochastic gene expression dynamics in single cells. *Proc. Natl. Acad. Sci. U. S. A.*, 103:7304–7309, 2006.

[105] Bertsekas DP. *Nonlinear Programming.* Athena Scientific, Cambridge, 1999.

[106] Yeang C-H, Mak HC, McCuine S, Workman C, Jaakkola T, and Ideker T. Validation and refinement of gene-regulatory pathways on a network of physical interactions. *Genome Biol.*, 6:R62, 2005.

[107] Sayyed-Ahmad A, Tuncay K, and Ortoleva PJ. Toward automated cell model development through information theory. *J. Phys. Chem. A*, 107:10554–10565, 2003.

[108] Haunschild MD, Freisleben B, Takors R, and Wiechert W. Investigating the dynamic behavior of biochemical networks using model families. *Bioinformatics*, 21:1617–1625, 2005.

[109] Wahl SA, Haunschild MD, Oldiges M, and Wiechert W. Unravelling the regulatory structure of biochemical networks using stimulus response experiments and large-scale model selection. *IEE Proc.-Syst. Biol.*, 153:275–285, 2006.

[110] Ma'ayan A, Blitzer RD, and Iyengar R. Toward predictive models of mammalian cells. *Annu. Rev. Biophys. Biomol. Struct.*, 34:319–349, 2005.

[111] Kremling A and Saez-Rodriguez J. Systems-biology—an engineering perspective. *J. Biotechnol.*, 8:447–455, 2007.

[112] Rajasethupathy P, Vayttaden SJ, and Bhalla US. Systems modeling: a pathway to drug discovery. *Cur. Opin. Chem. Biol.*, 9:400–406, 2005.

[113] Hood L and Perlmutter RM. The impact of systems approaches on biological problems in drub discovery. *Nat. Biotechnol.*, 22:1215–1217, 2004.

[114] Werner Horbelt, Thorsten Moller, Jens Timmer, Werner Melzer, and Karl Winkler. Analysis of nonlinear differential equations: Parameter estimation and model selection. In *ISMDA '00: Proceedings of the First International Symposium on Medical Data Analysis*, pages 152–158, London, UK, 2000. Springer-Verlag.

[115] Voss HU and Timmer J. Nonlinear dynamical system identification from uncertain and indirect measurements. *Int. J. Bifurcat. Chaos*, 14:1905–1933, 2003.

[116] Baker CTH, Bocharov GA, Paul CAH, and Rihan FA. Computational modelling with functional differential equations: Identification, selection, and sensitivity. *Appl. Num. Math.*, 53:107–129, 2004.

[117] Baker CTH, Bocharov GA, Ford JM, Lumb PM, Norton SJ, Paul CAH, Junt T, Krebs P, and Ludewig B. Computational approached to parameter estimation and model selection in immunology. *J. Comput. Appl. Math*, 184:50–76, 2005.

[118] Brause RW. Model selection and adaptation for biochemical pathways. In *ISBMDA '04: Proceedings of the 5th International Symposium on Biological and Medical Data Analysis*, pages 439–449, Berlin, 2004. Springer-Verlag.

[119] Gennemark P and Wedelin D. Efficient algorithms for ordinary differential equation model identification of biological systems. *IET Syst. Biol.*, 1:120–129, 2007.

[120] Steuer R, Nunes Nesi A, Fernie AR, Gross T, Blasius B, and Selbig J. From structure to dynamics of metabolic pathways: application to the plant mitochondrial TCA cycle. *Bioinformatics*, 23:1378–1385, 2007.

[121] Fussman GF and Blasius B. Community response to enrichment is highly sensitive to model structure. *Biol. Lett.*, 1:9–12, 2005.

[122] Bordley RF. The combination of forecasts: a Bayesian approach. *J. Op. Res. Soc.*, 33:171–174, 1982.

[123] Hoeting JA, Madigan D, Raftery AE, and Volinsky CT. Bayesian model averaging: A tutorial. *Stat. Science*, 14:383–401, 1999.

[124] Kass RE and Raftery AE. Bayes factors. *J. Amer. Stat. Assoc.*, 90:773–795, 1995.

[125] Raftery AE and Zheng YY. Discussion: Performance of Bayesian model averaging. *J. Am. Stat. Assoc.*, 98:931–938, 2003.

[126] Harrison PJ and Stevens SF. Bayesian forecasting. *J. R. Statist. Soc. B*, 38:205–247, 1976.

[127] McNees SK. The uses and abuses of 'consensus' forecasts. *J. Forecasting*, 11:703–710, 1992.

[128] Bates JM and Granger CWJ. The combination of forecasts. *Operations Research*, 20:451–468, 1969.

[129] Clemen RT and Winkler RL. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19:187–203, 1999.

[130] Genest C and Zidek JV. Combining probability distributions: a critique and annotated bibliography. *Stat. Sci.*, 1:114–148, 1986.

[131] Hansen JA. Accounting for model error in ensemble-based state estimation and forecasting. *Mon. Weather Rev.*, 130:2373–2391, 2002.

[132] Krishnamurti TN, Kishtawal CM, Zhang Z, LaRow T, Bachiochi D, and Williford E. Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, 13:4196–4216, 2000.

[133] Palmer TN. A nonlinear dynamicla perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Q. J. Roy. Met. Soc.*, 127:279–304, 2001.

[134] Posada D and Buckley TR. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst. Biol.*, 53:793–808, 2004.

[135] Araujo MB and New M. Ensemble forecasting of species distributions. *Trends Ecol. Evol.*, 22:42–47, 2007.

[136] Hoffmann JP. Simultaneous inductive and deductive modeling of ecological systems via evolutionary computation and information theory. *Simulation*, 82:439–450, 2006.

[137] Freund Y and Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55:119–139, 1995.

[138] Breiman L. Bagging predictors. *Mach. Learn.*, 24:123–140, 1996.

[139] Breiman L, Friedman JH, Olshen RA, and Stone CJ. *Classification and Regression Trees*. Chapman & Hall, London, 1984.

[140] Schoeberl B, Eichler-Jonsson C, Gilles ED, and Muller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.*, 20:370–375, 2002.

[141] Breiman L. Random forests. *Mach. Learn.*, 45:5–32, 2001.

[142] Thornberry NA and Lazebnik Y. Caspases: enemies within. *Science*, 281:1312–1316, 1998.

[143] Adams JM and Cory S. The Bcl-2 protein family: Arbiters of cell survival. *Science*, 281:1322–1326, 1998.

[144] Youle RJ and Strasser A. The Bcl-2 protein family: opposing activities that mediate cell death. *Nat. Rev. Mol. Cell. Biol.*, 9:47–59, 2008.

[145] Thompson CB. Apoptosis in the pathogenesis and treatment of disease. *Science*, 267:1456–1462, 1995.

[146] Fischer U and Schulze-Osthoff K. New approaches and therapeutics targeting apoptosis in disease. *Pharmacol. Rev.*, 57:187–215, 2005.

[147] Reed JC. Apoptosis-based therapies. *Nat. Rev. Drug Discov.*, 1:111–121, 2002.

[148] Philchenkov A. Caspases: potential targets for regulating cell death. *J. Cell. Mol. Med.*, 8:432–444, 2004.

[149] Kerr JFR, Winterford CM, and Harmon BV. Apoptosis—its significance in cancer and cancer-therapy. *Cancer*, 73:2013–2026, 1994.

[150] Oberholzer C, Oberholzer A, Clare-Salzler M, and Moldawer LL. Apoptosis in sepsis: a new target for therapeutic exploration. *FASEB J.*, 15:879–892, 2001.

[151] Dirnalg U, Iadecola C, and Moskowitz MA. Pathobiology of ischaemic stroke: an integrated view. *Trends Neurosci.*, 22:391–397, 1999.

[152] Cotman CW and Anderson AJ. A potential role for apoptosis in neurodegeneration and alzheimer's disease. *Mol. Neurobiol.*, 10:19–45, 1995.

[153] Fussenegger M, Bailey JE, and Varner J. A mathematical model of caspase function in apoptosis. *Nat. Biotechnol.*, 18:768–774, 2000.

[154] Stucki JW and Simon HU. Mathematical modeling of the regulation of caspase-3 activation and degradation. *J. Theor. Biol.*, 234:123–131, 2005.

[155] Hua F, Cornejo MG, Cardone MH, Stokes CL, and Lauffenburger DA. Effects of bcl-2 levels on Fas signaling-induced caspase-3 activation: molecular genetic tests of computational model predictions. *J. Immunol.*, 175:985–995, 2005.

[156] Rehm M, Huber HJ, Dussmann H, and Prehn JHM. Systems analysis of effector caspase activation and its control by x-linked inhibitor of apoptosis protein. *EMBO J.*, 25:4338–4349, 2006.

[157] Hua F, Hautaniemi S, Yokoo R, and Lauffenburger DA. Integrated mechanistic and data-driven modelling for multivariate anlysis of signalling pathways. *J. R. Soc. Interface*, 3:515–526, 2006.

[158] Legewie S, Blüthgen N, and Herzel H. Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feeback and bistability. *PLOS Comput. Biol.*, 2:1061–1073, 2006.

[159] Chen C, Cui J, Lu H, Wang R, Zhang S, and Shen P. Modeling of the role of a Bax-activation switch in the mitochondrial apoptosis decision. *Biophys. J.*, 92:4303–4315, 2007.

[160] Chen C, Cui J, Zhang W, and Shen P. Robustness analysis identifies the plausible model of the Bcl-2 apoptotic switch. *FEBS Lett.*, 581:5143–5150, 2007.

[161] Albeck JG, Burke JM, Lauffenburger DA, and Sorger PK. Modeling a snap-action, variable-delay switch controlling extrinsic cell death. *PLoS Biol.*, 2008.

[162] Scaffidi C, Fulda S, Srinivasan A, Friesen C, Li F, Tomaselli KJ, Debatin KM, Krammer PH, and Peter ME. Two CD95 (APO-1/Fas) signaling pathways. *EMBO J.*, 17:1675–1687, 1998.

[163] Barnhart BC, Alappat EC, and Peter ME. The CD95 type I/type II model. *Semin. Immunol.*, 15:185–193, 2003.

[164] Kischkel FC, Hellbardt S, Behrmann I, Germer M, Pawlita M, Krammer PH, and Peter ME. Cytotoxicity-dependent APO-1 (FAS/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor. *EMBO J.*, 14:5579–5588, 1995.

[165] Muzio M, Chinnaiyan AM, Kischkel FC, O'Rourke K, et al. FLICE a novel FADD-homolous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signalling complex. *Cell*, 85:817–827, 1996.

[166] Medema JP, Scaffidi C, Kischkel FC, Shevchenko A, Mann M, Krammer PH, and Peter ME. FLICE is activated by association with the CD95 death-inducing signaling complex (DISC). *EMBO J.*, 16:2794–2804, 1997.

[167] Irmler M, Thome M, Hahne M, Schneider P, et al. Inhibition of death receptor signals by cellular FLIP. *Nature*, 388:190–195, 1997.

[168] Krueger A, Baumann S, Krammer PH, and Kirchhoff S. FLICE-inhibitory proteins: regulators of death receptor-mediated apoptosis. *Mol. Cell. Biol.*, 21:8247–8254, 2001.

[169] Srinivasula SM, Ahmad M, Fernandes-Alnemri T, Litwack G, and Alnemri ES. Molecular ordering of the Fas-apoptotic pathway: The Fas/APO-1 protease Mch5 is a CrmA-inhibitable protease that activates multiple Ced-3/ICE-like cystein proteases. *Proc. Natl. Acad. Sci. U. S. A.*, 93:14486–14491, 1996.

[170] Slee EA, Harte MT, Kluck RM, Wolf BB, et al. Ordering the cytochrome c-initiated caspase cascade: Hierarchical activation of caspases-2, -3, -6, -7, -8, and -10 in a caspase-9-dependent manner. *J. Cell Biol.*, 144:281–292, 1999.

[171] Li HL, Zhu H, Xu CJ, and Yuan JY. Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell*, 94:491–501, 1998.

[172] Desagher S, Osen-Sand A, Nichols A, Eskes R, et al. Bid-induced conformational change of Bax is responsible for mitochondrial cytochrome c release during apoptosis. *J. Cell Biol.*, 144:891–901, 1999.

[173] Eskes R, Desagher S, Antonsson B, and Martinou JC. Bid induces the oligomerization and insertion of Bax into the outer mitochondrial membrane. *Mol. Cell. Biol.*, 20:929–935, 2000.

[174] Korsmeyer SJ, Wei MC, Saito M, Weller S, Oh KJ, and Schlesinger PH. Pro-apoptotic cascade activates BID, which oligomerizes BAK or BAX into pores that result in the release of cytochrome c. *Cell Death Differ.*, 7:1166–1173, 2000.

[175] Oltvai ZN, Milliman CL, and Korsmeyer SJ. Bcl-2 heterodimerizes in-vivo with a conserved homolog, BAX, that accelerates programed cell-death. *Cell*, 74:609–619, 1993.

[176] Cheng EHYA, Wei MC, Weiler S, Flavell RA, Mak TW, Lindsten T, and Korsmeyer SJ. Bcl-2, Bcl-xL sequester BH3 domain-only molecules preventing BAX- and BAK-mediated mitochondrial apoptosis. *Mol. Cell*, 8:705–711, 2001.

[177] Bossy-Wetzel E and Green DR. Caspases induce cytochrome c release from mitochondria by activating cytosolic factors. *J. Biol. Chem.*, 274:17484–17490, 1999.

[178] Wei MC, Lindsten T, Mootha VK, Weiler S, et al. tBID, a membrane-targeted death ligand, oligomerizes BAK to release cytochrome c. *Gene Dev.*, 14:2060–2071, 2000.

[179] Wei MC, Zong WX, Cheng EHY, Lindsten T, et al. Proapoptotic BAX or BAK: a requisite gateway to mitochondrial dysfunction and death. *Science*, 292:727–730, 2001.

[180] Saleh A, Srinivasula SM, Acharya S, Fishel R, and Alnemri ES. Cytochrome c an datp-mediated oligomerization of Apaf-1 is a prerequisite for procaspase-9 activation. *J. Biol. Chem.*, 274:17941–17945, 1999.

[181] Zou H, Li YC, Liu HS, and Wang XD. An Apaf-1·Cytochrome c multimeric complex is a functional apoptosome that activates procaspase-9. *J. Biol. Chem.*, 274:11549–11556, 1999.

[182] Acehan D, Jiang XJ, Morgan DG, Heuser JE, Wand XD, and Akey CW. Three-dimensional structure of the apoptosome: implications for assembly, procaspase-9 binding, and activation. *Mol. Cell*, 9:423–432, 2002.

[183] Srinivasula SM, Ahmad M, Fernandes-Alnemri T, and Alnemri ES. Autoactivation of procaspase-9 by Apaf-1-mediated oligomerization. *Mol. Cell.*, 1:949–957, 1998.

[184] Bratton SB, Walker G, Srinivasula SM, Sun XM, Butterworth M, Alnemri ES, and Cohen GM. Recruitment, activation and retention of caspases-9 and -3 by Apaf-1 apoptosome and associated XIAP complexes. *EMBO J.*, 20:998–1009, 2001.

[185] Zou H, Yang R, Hao J, Wang J, Sun C, et al. Regulation of the Apaf-1/caspase-9 apoptosome by caspase-3 and XIAP. *J. Biol. Chem.*, 278:8091–9098, 2003.

[186] Deveraux QL, Roy N, Stennicke HR, et al. IAPs block apoptotic events induced by caspase-8 and cytochrome c by direct inhibition of distinct caspases. *EMBO J.*, 17:2215–2223, 1998.

[187] Riedl SJ, Renatus M, Schwarzenbacher R, Zhou Q, et al. Structural basis for the inhibition of caspase-3 by XIAP. *Cell*, 104:791–900, 2001.

[188] Deveraux QL, Leo E, Stennicke HR, Welsh K, Salvesen GS, and Reed JC. Cleavage of human inhibitor of apoptosis protein XIAP results in fragments with distinct specificities for caspases. *EMBO J.*, 18:5242–5251, 1999.

[189] Srinivisula SM, Hegde R, Saleh A, et al. A conserved XIAP-interaction motif in caspase-9 and Smac-DIABLO regulates caspase activity and apoptosis. *Nature*, 410:112–116, 2001.

[190] Bratton SB, Lewis J, Butterworth M, Duckett C, and Cohen GM. XIAP inhibition of caspase-3 preserves its association with the Apaf-1 apoptosome and prevents CD95- and Bax-induced apoptosis. *Cell Death Differ.*, 9:881–892, 2002.

[191] Verhagen AM, Ekert PG, Pakusch M, Silke J, et al. Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing IAP proteins. *Cell*, 102:43–53, 2000.

[192] Du C, Fang M, Li Y, Li L, and Wang X. Smac, a mitochondrial protein that promotes sytochrom c-dependent caspase activation by eliminating IAP inhibition. *Cell*, 102:33–42, 2000.

[193] Paroni G, Henderson C, Schneider C, and Brancolini C. Caspase-2-induced apoptosis is dependent on caspase-9, but its processing during UV- or tumor necrosis factor-dependent cell death requires caspase-3. *J. Biol. Chem.*, 276:21907–21915, 2001.

[194] Guo Y, Srinivasula SM, Druilhe A, Fernandes-Alnemri T, and Alnemri ES. Caspase-2 induces apoptosis by releasing proapoptotic proteins from mitochondria. *J. Biol. Chem.*, 277:13430–13437, 2002.

[195] Srinivasula SM, Fernandes-Alnemri T, Zangrilli J, Robertson N, et al. The Ced-3/interleukin 1β converting enzyme-like homolog Mch6 and the lamin-cleaving enzyme Mch2α are substrates for the apoptotic CPP32. *J. Biol. Chem.*, 271:27099–27106, 1996.

[196] Cowling V and Downward J. Caspase-6 is the direct activator of caspase-8 in the cytochrome c-induced apoptosis pathway: absolute requirement for removal of caspase-6 prodomain. *Cell Death Differ.*, 9:1046–1056, 2002.

[197] Cheng EH, Kirsch DG, Clem RJ, Ravi R, Kastan MB, Bedi A, Ueno K, and Hardwick JM. Conversion of Bcl-2 to a Bax-like death effector by caspases. *Science*, 278:1966–1968, 1997.

[198] Kim YM, Kim TH, Seol DW, Talanian RV, and Billiar TR. Nitric oxide suppression of apoptosis occurs in association with an inhibition of Bcl-2 cleavage and cytochrome c release. *J. Biol. Chem.*, 273:31437–31441, 1998.

[199] Kirsch DG, Doseff A, Chau BN, Lim DS, et al. Caspase-3-dependent cleavage of Bcl-2 promotes release of cytochrome c. *J. Biol. Chem.*, 274:21155–21161, 1999.

[200] Meier M and Vousden KH. Lucifer's labyrinth—ten years of path finding in cell death. *Mol. Cell*, 28:746–754, 2007.

[201] Fletcher JI and Huang DCS. Controlling the cell death mediators Bax and Bak—Puzzles and conundrums. *Cell Cycle*, 7:39–44, 2008.

[202] Willis SN, Fletcher JI, Kaufmann T, et al. Apoptosis initiated when BH3 ligands engage multiple Bcl-2 homologs, not Bax or Bak. *Science*, 315:856–859, 2007.

[203] Mikhailov V, Mikhailova M, Pulkrabek DJ, Dong Z, Venkatachalam MA, and Saikumar P. Bcl-2 prevents Bax oligomerization in the mitochondrial outer membrane. *J. Biol. Chem.*, 276:18361–18374, 2001.

[204] Niikura Y, Nonaka T, and Imajoh-Ohmi S. Monitoring of caspase-8/FLICE processing and activation upon Fas stimulatioin with novel antibodies directed against a cleavage site for caspase-8 and its substrate FLICE-like inhibitory protein (FLIP). *J. Biochem.*, 132:53–62, 2002.

[205] Kharbanda S, Pandey P, Schofield L, Isreals S, et al. Role for Bcl-xL as an inhibitor of cytosolic cytochrome c accumulation in DNA damage-induced apoptosis. *Proc. Natl. Acad. Sci. U. S. A.*, 94:6939–6942, 1997.

[206] Chaloupka R, Petit PX, Israel N, and Sureau F. Over-expression of Bcl-2 does not protext cells from hypericin photo-induced mitochondrial membrane depolarization, but delays subsequent events in the apoptotic pathway. *FEBS Lett.*, 462:295–301, 1999.

[207] Hu UM, Benedict MA, Wu DY, Inohara N, and Nunez G. Bcl-xL interacts with Apaf-1 and inhibits Apaf-1-dependent caspase-9 activation. *Proc. Natl. Acad. Sci. U. S. A.*, 95:4386–4391, 1998.

[208] Huang DCS, Adams JM, and Cory S. The conserved N-terminal BH4 domain of Bcl-2 homologues is essential for inhibition of apoptosis and interaction with CED-4. *EMBO J.*, 17:1029–1039, 1998.

[209] Pan GH, O'Rourke K, and Dixit VM. Caspase-9, Bcl-X-L, and Apaf-1 form a ternary complex. *J. Biol. Chem.*, 273:5841–5845, 1998.

[210] Moriishi K, Huang DCS, Cory S, and Adams JM. Bcl-2 family members do not inhibitor apoptosis by binding the caspase activator Apaf-1. *Proc. Natl. Acad. Sci. U. S. A.*, 96:9683–9688, 1999.

[211] Pitti RM, Marster SA, Lawrence DA, Roy M, Kischkel FC, et al. Genomic amplification of a decoy receptor for fas ligang in lung and colon cancer. *Nature*, 396:699–703, 1998.

[212] Stennicke HR, Jürgensmeier JM, Shin W, Deveraux Q, Wolf BB, et al. Pro-caspase 3 is a major physiologic target of caspase-8. *J. Biol. Chem.*, 273:27084–27090, 1998.

[213] Letai A, Bassik M, Walensky L, Sorcinelli M, Weiler S, and Korsmeyer S. Distinct BH3 dominas either sensitize or activate mitochondrial apoptosis, serving as prototype cancen therpeutics. *Cancer Cell*, 2:183–192, 2002.

[214] Renatus M, Stennicke HR, Scott FL, Liddington RC, and Salvesen GS. Dimer formation drives the activation of the cell death protease caspase 9. *Proc. Natl. Acad. Sci. U. S. A.*, 98:14250–14255, 2001.

[215] Yin Q, Park HH, Chung JY, Lin SC, Lo YC, et al. Caspase-9 holoenzyme is a specific and optimal procaspase-3 processing machine. *Mol. Cell*, 22:259–268, 2006.

[216] Huang YH, Rick RL, Myszka DG, and Wu H. Requirement of both the second and third BIR domains for the relief of X-linked inhibitor of apoptosis protein (XIAP)-mediated caspase inhibition by Smac. *J. Biol. Chem.*, 278:49517–49522, 2003.

[217] Callus BA and Vaux DL. Caspase inhibitors: viral, cellular and chemical. *Cell Death Differ.*, 14:73–78, 2007.

[218] Luo KQ, Yu VC, Pu YM, and Chang DC. Application of the fluorescence resonance energy transfer method for studying the dynamics of caspase-3 activation during UV-induced apoptosis in living HeLa cells. *Biochem. Bioph. Res. Co.*, 283:1054–1060, 2001.

[219] Rehm M, Dussmann H, Janicke RU, Tavare JM, Kogel D, and Prehn JHM. Single-cell fluorescence resonance energy transfer analysis demonstrates that caspase activation during apoptosis is a rapid process – Role of caspase-3. *J. Biol. Chem.*, 277:24506–24514, 2002.

[220] Roy N, Deveraux QL, Takahashi R, Salvesen GS, and Reed JC. The c-IAP-1 and c-IAP-2 proteins are direct inhibitors of specific caspases. *EMBO J.*, 16:6914–6925, 1997.

[221] Slee EA, Adrain C, and Martin SJ. Executioner caspase-3,-6, and -7 perform distinct, non-redundant roles during the demolition phase of apoptosis. *J. Biol. Chem.*, 276:7320–7326, 2001.

[222] Knudson CM, Tung KS, Tourtellotte WG, Brown GA, and Korsmeyer SJ. Bax-deficient mice with lymphoid hyperplasia and male germ cell death. *Science*, 270:96–99, 1995.

[223] Lindsten T, Ross AJ, King A, Zong WX, et al. The combined functions of proapoptotic Bcl-2 family members Bak and Bax are essential for normal development of multiple tissues. *Mol. Cell*, 6:1389–1399, 2000.

[224] Wolter KG, Hsu YT, Smith CL, Nechushtan A, Xi XG, and Youle RJ. Movement of Bax from the cytosol to mitochondria during apoptosis. *J. Cell Biol.*, 139:1281–1292, 1997.

[225] Holler N, Zaru R, Michaeu O, Thome M, et al. Fas triggers an alternative caspase-8-independent cell death pathway using the kinase RIP as effector molecule. *Nat. Immunol.*, 1:489–495, 2000.

[226] Vandenabeele P, Vanden Berghe T, and Festjens N. Caspase inhibitors promote alternative cell death pathways. *Sci. STKE*, 358:pe44, 2006.

[227] Denmeade SR, Lin XS, Tombal B, and Isaacs JT. Inhibition of caspase activity does not prevent the signaling phase of apoptosis in prostate cancer cells. *Prostate*, 39:269–279, 1999.

[228] Kataoka T, Budd RC, Holler N, Thome M, et al. The caspase-8 inhibitor FLIP promotes activation of NF-κB and Erk signaling pathways. *Curr. Biol.*, 10:640–648, 2000.