Mechanistic Hypothesis Exploration of Signaling Network Processes via Bayesian Inference Methods

By

Michael A. Kochen

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Informatics

January 31, 2020

Nashville, Tennessee

Approved:

Carlos Lopez, PhD

Daniel Fabbri, PhD

Vito Quaranta, MD

Jacob Hughey, PhD

Mark Ellingham, PhD

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my adviser, Dr. Carlos Lopez, for welcoming me to his lab midway through my PhD training. Guidance from Dr. Lopez, as well as a great deal of freedom to explore new and interesting ideas, resulted in the work laid out in this dissertation. I would also like to thank my other committee members, Dr. Vito Quaranta, Dr. Jacob Hughey, Dr. Mark Ellingham, and Dr. Daniel Fabbri, for keeping me on track and helping me avoid pitfalls that could have substantially set back my progress.

I would also like to express my gratitude to the National Library of Medicine as this work was, in part, made possible by the NLM training grant T15 LM007450 through the Department of Biomedical Informatics at Vanderbilt University. The DBMI faculty and staff who design and implement the departments graduate program are dedicated and professional. I cannot praise them highly enough.

TABLE OF CONTENTS

Page
ACKNOWLEDGMENTS ii
LIST OF TABLES vi
LIST OF FIGURES vii
DEFINITION OF TERMS ix

1 Introduction 1
1.1 Background 1
1.2 Mathematical Modeling of Biological Systems 3
1.3 Calibration of Kinetic Models 4
1.4 Bayesian Machinery 5
1.4.1 Bayesian Evidence 5
1.4.2 Nested Sampling 6
1.4.3 Point Selection Algorithms 7
1.5 Purpose and General Methodology 8
2 Model and Methods 9
2.1 Extrinsic Apoptosis Model 9
2.1.1 Summary 9
2.1.2 Description of the Extrinsic Apoptosis Signaling Network 11
2.1.3 Parameter Ranges and Initial Conditions 12
2.1.4 Calibration of the Base Model to Experimental Data 13
2.2 Methods 17
2.2.1 Summary 17
2.2.2 Modeling and Simulation 17
2.2.3 Bayesian Evidence Estimation 17
2.2.4 Nested Sampling Software 18
2.2.5 Objective Functions 19
2.2.5.1 Multimodel Inference Method 19
2.2.5.2 Pathway Targeted Method 19
2.2.6 Evidence Ratios 20
2.2.7 Computational Resources 21
3 Bayesian Evidence Based Analysis of Network Dynamics 22
3.1 Summary 22
3.2 Challenges in the Analysis of Dynamics of Physicochemical Networks 23
3.3 Overview of Bayesian Inference-Based Analysis of Network Dynamics 23
3.4 Results 24
3.4.1 General Strategy and Workflow 24
3.4.2 Differential Downregulation of Extrinsic Apoptosis Subnetworks by XIAP 29
3.4.2.1 XIAP Control of the Type I/II Phenotype 30

3.4.2.2 Mechanistic Interpretation of the Extrinsic Apoptosis Subnetworks 33
3.4.3 Apoptosis Signal Strength Determines the Signal Route Through the Network.....37
3.4.4 Evidence Ratios and XIAP Influence on Type I/II Apoptosis Phenotype.....41
3.4.5 Precision vs Computational Cost.....45
4 Bayesian Evidence-Based Analysis of the Extrinsic Apoptosis Signaling Network.....51
4.1 Summary.....51
4.2 Introduction.....52
4.3 Results.....56
4.3.1 XIAP/Caspase-3 Concentration Ratios and Type I/II Phenotype Outcome 57
4.3.2 Effect of Receptor Count and DISC Formation on Apoptosis Subnetworks.....63
4.3.3 Effect of Receptor Count and DISC Formation on Apoptosis Signal Flux.....69
4.3.4 Incomplete Apoptosis Recovery from MOMP Sensitizers using Common Priors...71
4.3.5 Near Full Apoptosis Recovery from MOMP Sensitizers using Adjusted Priors.....77
5 HypBuilder: Automatic Generation of Ensembles of Physicochemical Models.....80
5.1 Summary.....80
5.2 Introduction.....80
5.3 Software and Usage 82
5.3.1 General Workflow 82
5.3.2 Input File.....83
5.3.3 Run File.....83
5.3.4 Output 84
5.3.4.1 Monomers 84
5.3.4.2 Parameters.....85
5.3.4.3 Observables.....86
5.3.4.4 Rules 86
5.3.4.5 Initials 87
5.3.5 Molecular Interaction Library.....87
5.3.6 Features.....89
5.3.6.1 Initial Value Ranges.....90
5.3.6.2 Optional Reactions.....90
5.3.6.3 Grouped Reactions.....92
5.3.6.4 Data Nodes.....93
5.3.6.5 Don’t Write Don’t Care 94
5.3.6.6 Binding Sites.....95
5.3.6.7 Reaction Sequences 98
5.3.6.8 Initial Binding 99
5.3.6.9 Text 100
5.4 Examples.....101
5.4.1 Extrinsic Apoptosis Reaction Model 101
5.4.2 Direct, Indirect, and Embedded MOMP Models 104
6 Discussion and Future Directions 106
6.1 Discussion.....106

6.1.1 Bayesian Evidence-Based Exploration of Network Dynamics 106
6.1.2 Bayesian Evidence-Based Analysis of Extrinsic Apoptosis Regulatory Axes.....108
6.1.3 Large Scale Construction of Physicochemical Models with HypBuilder 110
6.2 Future Directions 110
6.2.1 Expansion of Use Cases for the Probabilistic Analysis of Network Dynamics 110
6.2.2 Exploration of Generated Biological Hypotheses.....111
6.2.3 HypBuilder Development 112
BIBLIOGRAPHY 113

APPENDICES 126

LIST OF TABLES

Table Page
2.1 Initial values for the baseline extrinsic apoptosis model 13
3.1 Averages for the standard error, clock time, and number of evaluations before termination for different nested sampling population sizes 47

Data tables containing results of each of the in-silico experiments carried out for this dissertation can be found in the repository https://github.com/LoLab-VU/BIND

LIST OF FIGURES

Figure Page
2.1 Schematic of apoptotic signal flow through the extrinsic apoptosis network 10
2.2 Trajectories of cleaved Bid, exported Smac, and cleaved PARP for the calibrated EARM model along with the data used for the calibration 14
2.3 Bid trajectory along with estimated error bars for the associated FRET data 15
2.4 PARP trajectory along with estimated error bars for the associated FRET data 16
3.1 Workflow for Bayesian evidence-based analysis of network dynamics and signaling regulation 27
3.2 Extrinsic apoptosis subnetworks 28
3.3 Comparison of the Bayesian evidence for achieving apoptosis via the isolated caspase pathway and the complete network and for increasing concentration of XIAP 31
3.4 Bayesian evidence for achieving apoptosis through all six subnetworks of the extrinsic apoptosis network and for increasing concentration of XIAP 34
3.5 Log-evidence version of plots in Figure 3.4 with estimated errors generated by MultiNest..36
3.6 Expected PARP cleavage for the caspase pathway and complete network under both low and high DISC conditions 38
3.7 Signal flux through the caspase and mitochondrial pathways as well as the estimated total flux through the system.....40
3.8 Evidence ratio calculations with no MOMP inhibition at increasing concentrations of XIAP 42
3.9 Evidence ratio calculations with MOMP inhibited by Bcl-2 at 328,000 molecules per cell and at increasing concentrations of XIAP.....44
3.10 Precision vs. computational cost.....46
3.11 Evidence and evidence ratio plots over a range of XIAP concentrations at increasing population levels for the nested sampling algorithm 48

4.1 Schematic of the extrinsic apoptosis model and axes of regulation 54
4.2 XIAP vs Caspase-3 (C3) evidence values for the caspase pathway over a range of XIAP and Caspase-3 59
4.3 XIAP vs Caspase-3 (C3) evidence values for the complete network over a range of XIAP and Caspase-3 60
4.4 Evidence ratios (complete/caspase) from the values in Figures 4.2 and 4.3 for XIAP vs Caspase-3 (C3).....62
4.5 Receptor and Fadd/Procaspase-8 level effects on the complete network and caspase pathway evidence for achieving apoptosis (expected values of PARP cleavage).....65
4.6 Statistics for the average evidence difference and the range of evidence values over increasing receptor concentrations for sequential/increasing Fadd/Caspase-8 levels.....66
4.7 Evidence ratio values over the range of receptor concentrations at each level of Fadd/ProCaspase-8 67
4.8 Associated log-evidence values with standard error bars for the evidence values in Figure 4.5 68
4.9 Signal flux through the caspase and mitochondrial pathways over increasing receptor concentration and at increasing levels of Fadd/Procaspase-8.....70
4.10 Bayesian evidence characteristics of Bcl-2 and Bad MOMP regulation.....73
4.11 Statistics for the recovery of apoptosis when inhibiting Bcl-2 with Bad 74
4.12 Bayesian evidence characteristics of a complex MOMP regulatory network 77
4.13 Bayesian evidence characteristics of MOMP regulation with binding rate parameters adjusted to more accurately reflect experimentally derived binding affinities.....78
4.14 Statistics for the recovery of apoptosis when binding affinities are closer to experimentally defined K_D values 79
5.1 HypBuilder workflow and processing 82
5.2 All possible networks created for the Data node example.....94
5.3 Comparison of manually and automatically constructed EARM models.....103
5.3 MOMP regulatory networks for the Direct, Indirect, and Embedded models.....105

DEFINITION OF TERMS

Chapter 1 Terms

Physicochemical model: A mathematical representation of a complex system that captures the specific biochemical interactions between system components. They are typically formulated stochastically or as a system of ordinary differential equations (ODE) using an established chemical kinetics formalism. Physicochemical models produce quantitative predictions for variations of the models inputs [22].

Rule-based model: An abstracted model in which reactions between individual sites or domains on a protein are written down as rules. The network of all possible species configurations and the underlying representative system of equations is then automatically generated [26].

Model calibration: Fitting model outputs to experimental data via adjustment of the models parameters. When Bayesian methods are used, posterior distributions of the parameters can be obtained [36].

Bayesian evidence: The normalizing constant for the derivation of the posterior distribution using Bayes theorem. Also called the marginal likelihood, it is the integral of the likelihood function over the prior distribution. In this work, likelihood functions are replaced with objective functions that represent quantities of interest and the prior distribution represents the current knowledge of model parameters. The evidence calculation then becomes the expected value of the quantity of interest over the given prior distribution of parameters.

Nested sampling: An efficient method for the computation of the Bayesian evidence. It takes the multi-dimension integral over the parameter space and converts it to a one-dimensional integral over the prior likelihood volume. The method depends on an efficient method to continually narrow the prior search area to high likelihood regions [38].

Expected value of a continuous function: The integral of the function over the given probability density function. In this work the (objective) function represents a quantity of interest and the probability density function represents the normalized parameter distribution.

Chapter 2 Terms

Apoptosis: Apoptosis, or programmed cell death, is a sequence of cellular events that allows cells in multi-cellular organisms to be removed in an orderly controlled fashion [51].

Extrinsic apoptosis reaction model (EARM): A physicochemical model of receptor mediated apoptosis [31]. The model includes two pathways, a direct caspase activation pathway, and a more complex mitochondrial pathway.

Death inducing signaling complex (DISC): The DISC is a protein complex consisting of a death inducing member of the tumor necrosis factor family of receptors, a death domain containing protein such as FADD, and an initiator caspase like Caspase-8. Formation of the DISC immediately follows ligand binding to the receptor and propagates the apoptosis signal. [53, 54]

Mitochondrial outer membrane permeabilization (MOMP): The process by which the mitochondria release the proapoptotic factors Cytochrome c and Smac/DIABLO. Regulation of MOMP, including its activation, inhibition, and execution, is controlled by the Bcl-2 family of proteins [60, 61].

Parameter range: The range, in Log_{10} space, for a model parameter. This is cast as a uniform prior distribution in the evidence calculation.

Initial value: The starting value for any model species that changes over time as a simulation proceeds.

Multimodel inference method: In the context of this work, the multimodel inference method breaks the model into relevant sub-models and calculates the expected values of a quantity of interest for each under various regulatory conditions. Inferences regarding regulatory control of cellular fate are then made through examination of the differences in signal throughput. These are effectively in silico knockout experiments.

Pathway targeted method: In the pathway targeted method the full model is retained and the targeted quantities of interest are the signal flux through different pathways. These flux values are estimated over various regulatory conditions and inferences are made regarding how the signal is directed through the network. Other targeted processes could also be analyzed with this method.

Signal flux: The amount of signal carried through a particular pathway.

Evidence ratios: Evidence ratios are equivalent to Bayes factors. In the context of this work the evidence ratios are used to infer key points in regulatory control of phenotypic shifts.

Chapters 3 and 4 Terms

Emergent behavior: Emergent behaviors are nonintuitive behavioral properties of complex networks that stem from interactions between networks components and crosstalk between pathways. Such behaviors are only observed in the context of the complete system. [1, 2]

Phenotype: In the context of this work phenotype is the set of characteristics, both behavioral and in the concentrations of proteins, displayed by a particular cell type.

Type I apoptosis: A cellular phenotype in which apoptosis can be successfully executed independent from mitochondrial involvement [93].

Type II apoptosis: A cellular phenotype in which successful execution of apoptosis depends on mitochondrial involvement [93].

Regulatory axes: In this work regulatory axes are those key points in the network that are hypothesized to control the regulation of signal throughput. They may take the form of an effector-target pair of proteins like XIAP and Caspase-3, a complex like the death inducing signaling complex (DISC), or a network module like the mitochondria.

Precision: In the context of this work precision is the estimated error on the estimates for the evidence calculations.

Resolution: In the context of this work resolution is the (relative) density of evidence (expected value) calculations over a range of values for a given regulator. We have assumed here that evidence estimates are made at evenly space intervals throughout a given range.

Chapter 1

Introduction

1.1 Background

Elucidation of the regulatory mechanisms that govern signaling dynamics through large complex biochemical networks is a challenging task. The interactions between the various pathways and the combined effects of regulatory elements can give rise to unexpected behaviors that are often difficult to explain [1, 2, 3]. Nevertheless, characterization of network dynamics and the underlying mechanisms that control their regulation is essential if we are to understand the complexities of signaling networks, the diseases their dysregulation causes, and effective treatments for those conditions [4, 5]. Cancer is a prime example [6, 7]. Crosstalk between pathways can result in multiple deleterious effects from a single driver mutation. For example, constitutively active Ras, and its interaction with PI3K, can increase both growth and survival signals via the MAPK/Erk and PI3K/Akt pathways respectively [8]. Signal dysregulation extends to the intercellular signaling network as well. Tumors shape the tumor microenvironment via a complex system of signals that remodel the surrounding extracellular matrix, promote angiogenesis, induce immune evasion, and promote many other growth and survival functions [9]. The fact that cancers typically obtain multiple driver mutations before becoming tumorigenic makes the analysis of the altered signaling dynamics enormously complex.

Deficiencies in knowledge regarding the complex interactions that govern the regulation of signaling dynamics can lead to difficulties in treating the diseases that result when signals become aberrant [10]. Target validation and target-based screening typically work on a “one-drug-one-gene paradigm” [11]. This can render effective treatments for diseases with complex etiologies out of reach. Drug development in such cases may benefit from combination treatments that attack multiple pathways and/or points of regulation [12, 13]. Drugs for complex diseases that pass the initial screening often fail in Phase II and III

clinical trials [12, 15, 16]. They often fail due to a lack of efficacy stemming from an incomplete understanding of disease pathogenesis or from the advent of unanticipated side-effects [14, 15]. The former is common in tumors with a heterogenous architecture, a feature that is implicated in treatment resistance after encouraging initial results [12, 16, 17, 18]. Side-effects can result from drug interactions with unexpected targets or unexpected cross-talk between pathways [10, 11, 14]. Overall, the unknown complexities of disease systems along with the most prevalent methodologies used in drug discovery results in missed treatment opportunities, lower success rates, higher costs, and worse outcomes [19].

A number of experimental techniques from molecular biology and proteomics can, in general, be used to help examine the dynamics of complex networks. For example, gene knockouts, and conversely protein overexpression, can be used to inhibit or activate proposed regulatory proteins or targeted pathways in order to examine the effects they elicit and attempt to infer their role in the overall dynamics of the system. Other methods like Förster resonance energy transfer (FRET) can be used to measure protein-protein interactions in vivo at different time points and locations to characterize the dynamics of known pathways. Unfortunately, conclusive establishment of mechanistic explanations for observed behaviors in large complex systems using these techniques would require their application in a combinatorial fashion to examine the interplay between all pathways and regulatory elements. The requisite number of experiments would thus be prohibitively costly and time consuming.

Fortunately, computational modeling of biochemical networks has advanced greatly in recent years and has become an established adjunct to experimental methods. In silico experiments can be designed to account for network complexities like cross-talk between pathways and multi-factor regulation [20, 21]. Hypotheses regarding emergent regulatory behavior can then be made that will guide experimentation to the most promising cost-effective targets.

1.2 Mathematical Modeling of Biological Systems

The use of mathematical modeling of biological systems alongside experimental methods typically follows an iterative approach [22, 23]. In short, a model that currently represents the best knowledge of a biological system of interest is used to make predictions of system behavior under various in silico experimental conditions. Those predictions are then used to guide wet bench experiments that ideally lead to novel insights. The additional knowledge gained from the bench experiments, whether new discoveries are made are not, are subsequently used to improve the computational model, beginning the cycle once again.

A wide range of well-established modeling techniques are available to systems modelers, the choice of which depends on the size, type, and level of detail of the system [24, 25]. Qualitative methods can be used to get a broad snapshot of the interactions or coarse-grained dynamics of a system. Conversely, quantitative methods are used to make analyses of system dynamics with much more precision, so long as enough mechanistic detail of the system is available. Quantitative physicochemical modeling requires knowledge of the important components of the system, the specific interactions between those components (i.e. inhibitory binding, enzymatic cleavage, etc.), and the rates at which those interactions take place. With that knowledge in hand, a mathematical representation of the system can be constructed to simulate network dynamics and formulate new hypotheses. Quantitative models are typically simulated as either stochastic or deterministic systems of equations and deterministic systems often take the form of ordinary differential equations (ODEs) [22, 23, 24, 25]. Although one could opt to write down such systems directly the number of binding interactions and molecular modifications in realistic representations of biological systems results in a combinatorial explosion of new molecular species and a concomitant explosion of equations [26]. Rule-based modeling systems are aimed at simplifying construction of these models.

Rule-based systems such as BioNetGen and the Kappa language are designed to more easily construct models that incorporate detailed biochemical interactions of the system components [27, 28, 29, 30]. These formalisms are designed specifically to address the

problem of the combinatorial explosion of unique protein states that arises from protein modification and complex formation. The solution is to write rules for the individual sites or domains on a protein rather than for every possible protein complex configuration. The software then automatically generates a network of interactions as well as all the underlying equations. In general, these systems allow for ODE or stochastic based simulation of mechanistic models and form the foundation of more sophisticated systems like PySB [31].

1.3 Calibration of Kinetic Models

When reaction rate parameters for kinetic models are unknown, and they often are, they must be calibrated. The goal is to find the parameter set that best explains a set of experimental data. Any calibration scheme for finding a suitable parameter set starts with an objective function that provides a measure of 'fit' to the data for any proposed parameter set. A typical objective function is the residual sum of squares.

Next, a method to search the given parameter space for an optimal parameter set is employed, and many methods have been devised for the task. Local methods can be applied in a multi-start fashion to find global minima or used in combination with global methods to form hybrid approaches [32]. Deterministic methods will find a minimum that is guaranteed to be globally optimal within a defined level of certainty but at a high computational cost [32]. Stochastic techniques that give no assurances of optimality can often efficiently find a minimization that is, in practice, often close to optimal. Many stochastic methods are inspired by natural processes such as simulated annealing (cooling metal), particle swarm optimization, (group dynamics such as a flock of birds), and the class of evolutionary computation algorithms (improved fitness through selection).

Another approach to the parameter calibration problem is through Markov chain Monte Carlo methods [36, 37]. These methods allow for both calibration and estimation of parameter posterior distributions that can be used to analyze their uncertainty given the available data. This is a key point in traditional computational modeling. Estimated outcomes that depend on parameters with broad uncertainty are suspect at best. Parameter distribution

estimates, in turn, depend on the available experimental data, which may be sparse or specific to a particular biological question. This can potentially result in a need to generate new data for every proposed in silico experiment, a costly and time-consuming endeavor. Methods to guide experimentation to productive targets without the need for calibrated parameters would close a gap in the cycle of experimentation and computational modeling.

1.4 Bayesian Machinery

1.4.1 Bayesian Evidence

Given a prior distribution $P(\theta|M) = \pi(\theta)$ and likelihood function $P(D|\theta, M) \equiv L(\theta)$ there are two unknowns in a Bayesian calculation (1.1), the posterior distribution $P(\theta|D, M)$ and the evidence $P(D|M) \equiv Z$ where θ represents a parameter set, D represent the data, and M represents a model [38, 39].

$$P(\theta|D, M) = \frac{P(D|\theta, M)\, P(\theta|M)}{P(D|M)} \tag{1.1}$$

The normalizing evidence term is typically ignored for parameter estimation purposes as inferences can be made directly from the unnormalized posterior generated using MCMC methods. However, when comparisons of different models are desired the evidence provides a means to do so. Calculating the evidence is equivalent to calculating the expected value of the likelihood function over the prior distribution, i.e., finding the average of the function over the prior. Typically, the likelihood, or objective function, calculates a fit to experimental data. As will be explained in Chapter 3, the objective functions used in this work represent quantities of interest and the evidence calculation estimates the expected values of those quantities.

To compute the Bayesian evidence for a model, also called the marginal likelihood, one integrates the likelihood function over the model's prior distribution (1.2).

$$P(D|M) = \int P(D|\theta, M)\, P(\theta|M)\, d\theta \tag{1.2}$$

This is done numerically when no analytical solution exists, as is often the case, and is typically a computationally intensive task, particularly for high parameter models. Unfortunately, simpler methods that use the maximum likelihood, like the Bayesian information criterion (BIC) and related algorithms [40], are inappropriate when the models to be compared can obtain equally good results for a given likelihood function. This is often the case for highly parameterized biological models. Thermodynamic integration is an often-used option to estimate evidence values [41]. In that method a series of unnormalized posterior averages are estimated using MCMC methods with a power (*T*) on the likelihood function that increases from 0, representing the prior, to 1, representing the true posterior. These averages are then integrated over *T* to produce an estimate of the log-evidence. This is a very computationally intensive exercise. Fortunately, a much more efficient algorithm called *nested sampling* has been developed [38].

1.4.2 Nested Sampling

Nested sampling takes the multi-dimensional integral over θ and converts it to a one-dimensional integral over the prior volume X, where $dX = \pi(\theta)d\theta$. Then the accrued volume of the prior for $L(\theta) > \lambda$ is given by the monotonically decreasing function

$$X(\lambda) = \int_{L(\theta)>\lambda} \pi(\theta)d\theta. \tag{1.3}$$

This function's inverse, $L(X) \equiv \lambda$, is similarly monotonically decreasing from 0 to 1 and we write the evidence calculation as

$$Z = \int_0^1 L(X)dX. \tag{1.4}$$

The error estimate is given for $\log(Z)$ as $\pm\sqrt{H/N}$ where *N* is the number of live points in the nested sampling algorithm and *H* is the relative *information* contained in the posterior vs the prior. It is a measure of concentration of the posterior mass within the prior and is expressed as

$$H = \int log\left(\frac{dP}{dX}\right) dP \tag{1.5}$$

where P is the prior and X is the posterior mass. With this foundation, a scheme can be used to continually narrow the sampling of the prior distribution to areas of higher likelihood. Note that the notation used above is drawn directly from [38]. An overview of the numerical computation of the evidence is in Chapter 3 and a more detailed treatment for both the evidence and the error estimate can be found in [38, 39]. We also note that the varied complexity of the models is automatically accounted for by the integration process. The better fit to data afforded by the higher parameter models is countered by the increase in the number of dimensions that must be integrated.

1.4.3 Point Selection Algorithms

After the nested sampling method was introduced by Skilling, much effort has gone into finding efficient algorithms to search the prior for high likelihood regions. The original work suggests an MCMC walk from an existing point to find a new point with likelihood higher than the lowest likelihood point in the current population (L_{low}). The new point is then added to the population and L_{low} is added to the sum that estimates equation 1.2 (see Section 2.2 and [38]). A number of more sophisticated and efficient algorithms have since been developed. MultiNest, for example, clusters the points in the existing population and forms hard ellipsoid bounds around the clusters at each iteration [42, 43]. The next point is then searched for within those bounds. The PolyChord algorithm uses slice sampling, where, given a density function f(x) and current point x_0, a new point x_1 within the density is found by randomly choosing a level y between 0 and $f(x_0)$, and then randomly choosing a new point within a horizontal interval at that level [44, 45]. Points are chosen until one is found within the density and those outside put new bounds on the interval. The program DNest4 is an implementation of a method called diffusive nested sampling in which a mixture of successive distributions is sampled to better handle complications like multimodal

posteriors [46, 47]. Other methods such as importance sampling and dynamic sampling have often been incorporated into existing software packages like those above [48, 49, 50].

1.5 Purpose and General Methodology

The overarching purpose of this work is to provide a methodology for the exploration of mechanistic biological networks with limited or even no knowledge of the reaction rate parameters and/or the data needed to calibrate them. We take a probabilistic approach toward accomplishing this task. If one cannot produce a value for an outcome of interest due to the uncertainty in the parameters, one may still produce an *expected* value based on all topological and parameter information that *is* known. Initial hypotheses based on in silico experiments can then suggest targets for experimentation and data collection. Expected values are calculated using the Bayesian machinery described in Section 1.4, as Bayesian evidence is by definition an expected value (1.2). We can leverage this by writing objective functions that represent quantities of interest. In the chapters below we use two such objective functions. The first is the proportion of the protein PARP that is cleaved, which is a proxy for apoptosis, and the second is the flux of the apoptosis signal through various pathways. Comparison of these outputs under different in silico experiments can then lead to exploitable insights and guided experimentation. To summarize the idea: if we cannot confidently calibrate, simulate, and produce values for quantities of interest, an alternative is to find prior distributions, integrate, and produce *expected* values for those quantities. In chapters 4 and 5 below we take the worst possible case, i.e., generic priors that represent reasonable uniform (in log space) parameter ranges and test this methodology against a model of extrinsic apoptosis. We show that even in the worst case, good knowledge of the reaction topology and generic knowledge of reaction rates, the qualitative results strongly agree with existing experimental evidence.

Chapter 2

Model and Methods

2.1 Extrinsic Apoptosis Model

2.1.1 Summary

Extrinsic apoptosis is a receptor mediated process for programmed cell death. It is essential for normal development and homeostasis and is therefore a well-studied system, making it ideal for demonstrating the efficacy of new computational approaches. In Chapters 3 and 4 below we focus heavily on the two known extrinsic apoptosis phenotypes: Type II, in which the apoptosis signal is dependent on amplification via the mitochondrial pathway, and Type I, in which it is independent. The base model used in those chapters is a modified version of the Extrinsic Apoptosis Reaction Model (EARM) from Lopez et al. [31]. Here we describe the signaling network that is represented by the modified model (Figure 2.1) and show that the model is calibratable to existing experimental data. We also define the literature derived initial values and the parameter ranges used for the nested sampling-based evidence/expected value estimation.

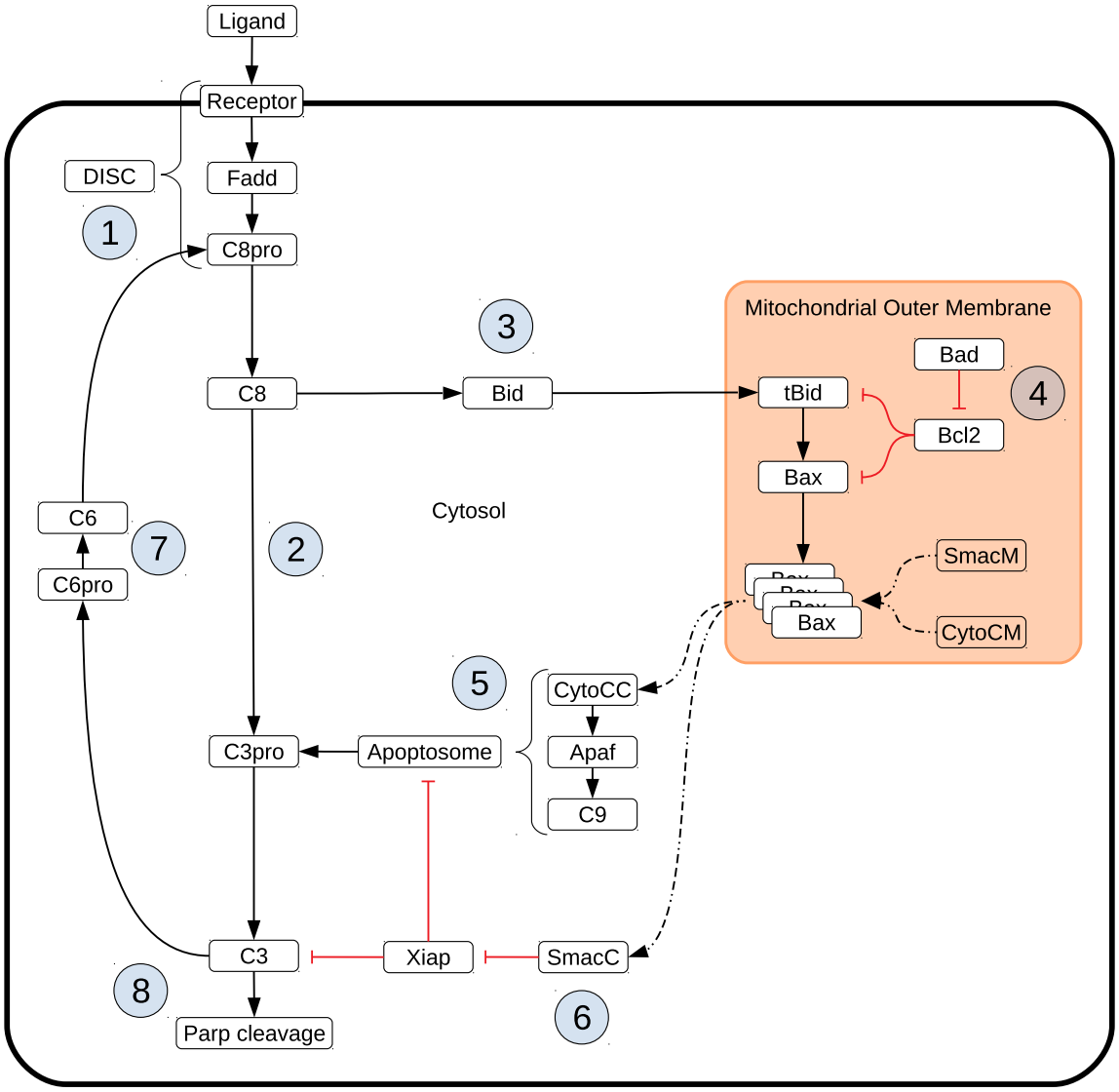


Figure 2.1. Schematic of apoptotic signal flow through the extrinsic apoptosis network.

2.1.2 Description of the Extrinsic Apoptosis Signaling Network

The apoptosis signal is initiated when a death inducing member of the tumor necrosis factor (TNF) superfamily of receptors (FasR, TNFR1, etc.) is bound by its respective ligand (FasL, TNF-α, etc.), setting off a sequence biochemical events that result in the orderly deconstruction of the cell [51]. The first stage of this sequence is the assembly of the DISC at the cell membrane ① and the subsequent activation of Caspase-8. Upon ligand binding and oligomerization of a receptor such as FasR or TRAIL, an adapter protein, like FADD (Fas-associated protein with death domain), is recruited to the receptors cytoplasmic tail [52, 53, 54]. FADD, in turn, recruits Caspase-8 via their respective death effector domains (DEDs), thus completing DISC formation [53, 54]. Other DISC components could also be included here, such as the regulator cFlip [55]. Once recruited, proximal Procaspase-8 monomers dimerize, inducing their autoproteolytic activity and producing active Caspase-8 [56, 57, 58].

After Caspase-8 activation the apoptotic signal can progress down two distinct pathways that both lead to the activation of Caspase-3 and the ensuing proteolysis of downstream targets. One pathway consists of a caspase cascade in which active Caspase-8 directly cleaves and activates Caspase-3 ② [59], while another more complex pathway is routed through the mitochondria. In the mitochondrial pathway Caspase-8 cleaves the proapoptotic Bcl-2 family protein Bid in the cytosol, which then migrates to the mitochondria ③ where it initiates mitochondrial outer membrane permeabilization (MOMP) and the release of proapoptotic factors that lead to Caspase-3 activation [60, 61].

MOMP has its own set of regulators that govern the strength of apoptotic signaling through the mitochondria ④. After Caspase-8 activated Bid, (tBid), migrates to the mitochondria it activates proteins in the outer mitochondrial membrane, such as Bax, that subsequently self-aggregate into membrane pores and allow exportation of Cytochrome c and Smac/DIABLO to the cytosol [62]. Bid and Bax are examples of proapoptotic proteins from the Bcl-2 family, all of which share BH domain homology [63]. Other members of this family act as MOMP regulators; the antiapoptotic Bcl-2, for example, binds and

inhibits both Bid and Bax while the proapoptotic Bad similarly binds and inhibits its target, Bcl-2 [64, 65, 66, 67]. Many other pro- and antiapoptotic members of the Bcl-2 family have been discovered and together regulate MOMP [68].

Regardless of which pathway is chosen, the intermediate results are Caspase-3 activation and subsequent cleavage of PARP ⑧, a proxy for cell death in the analyses here [69, 70]. XIAP (X-linked inhibitor of apoptosis protein) is an inhibitor of Caspase-3 and has been proposed to be a key regulator in determining the apoptotic phenotype of a cell (Type I/II cells are, respectively, independent/dependent on the mitochondrial pathway) [71]. XIAP sequesters Caspase-3 but also contains a ubiquitin ligase domain that directly targets Caspase-3 for degradation. The inhibitor also sequesters and inhibits the Caspase-3 activating Caspase-9 residing within the apoptosome complex [72, 73, 74]. Apoptosome formation is initiated by Cytochrome c exported from the mitochondria during MOMP ⑤. Cytochrome c induces the protein APAF-1 to oligomerize and subsequently recruit and activate Caspase-9, thus forming the complex [75]. Another MOMP export, the protein Smac/DIABLO ⑥, binds and inhibits XIAP, working in tandem with Cytochrome c to oppose XIAP and carry out the apoptosis inducing activity of the Type II pathway [76]. Finally, Procaspase/Caspase-6 constitutes a feed forward loop between Caspase-3 and Caspase-8 ⑦ [77].

2.1.3 Parameter Ranges and Initial Conditions

The prior distribution takes the form of a set of parameter ranges, one for each reaction rate parameter. The chosen ranges span four orders of magnitude around generic reaction rates deemed plausible [22] and are specific to the type of reaction taking place. The ranges of reaction rate parameters, in Log_{10} space, are 1^{st} order forward: [-4.0, 0.0], 2^{nd} order forward: [-8.0, -4.0], 1^{st} order reverse: [-4.0, 0.0], and catalysis: [-1.0, 3.0]. These ranges were also used in calibration of the base model. Initial conditions, where possible, were either gleaned from the literature [78, 79] or taken from a previous model of extrinsic apoptosis

[31]. Because the baseline model was designed to concur with Type II cells, literature derived initial conditions were based on Type II Jurkat or Hela cell lines (Table 2.1).

Table 2.1. Initial values for the baseline extrinsic apoptosis model.

Monomer	**Initial Value**	**Monomer**	**Initial Value**
Ligand	1000	C3pro	21000
ParpU	1000000	CytoCM	500000
C8A	0	CytoCC	0
SmacM	100000	BaxA	0
BaxM	40000	ApafI	100000
Apop	0	BidU	171000
Fadd	130000	BidT	0
SmacC	0	C3A	0
ParpC	0	Bad	53000
Xiap	42000	ApafA	0
C9	100000	BidM	0
C3ub	0	Receptor	100
C8pro	130000	C6A	0
Bcl2	328000	C6pro	100

2.1.4 Calibration of the Base Model to Experimental Data

The modified model recapitulates extrinsic apoptosis execution to experimental data [80] upon calibration to time-dependent trajectories of Bid, Smac, and PARP (Figures 2.2-2.4). Data can be found at https://github.com/clopezx/earm.

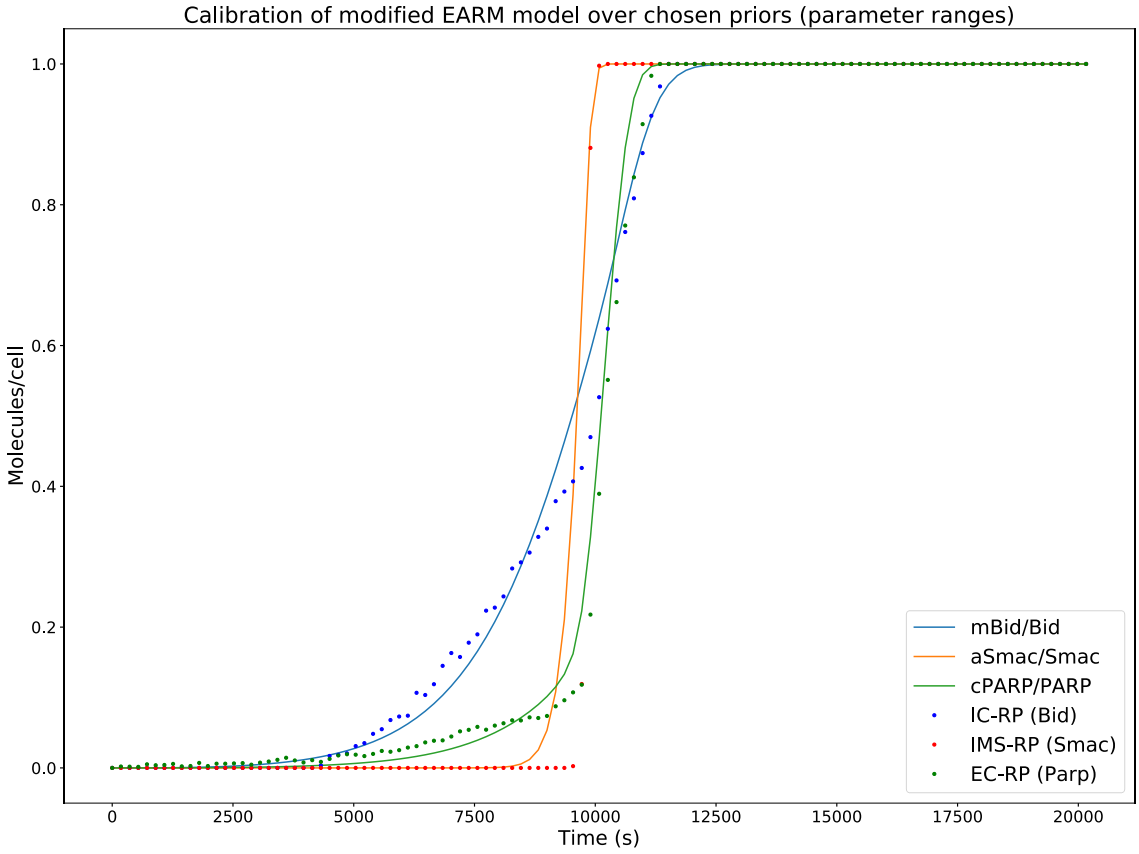


Figure 2.2. Trajectories of cleaved Bid, exported Smac, and cleaved PARP (solid lines) for the calibrated EARM model along with the data used for the calibration (points).

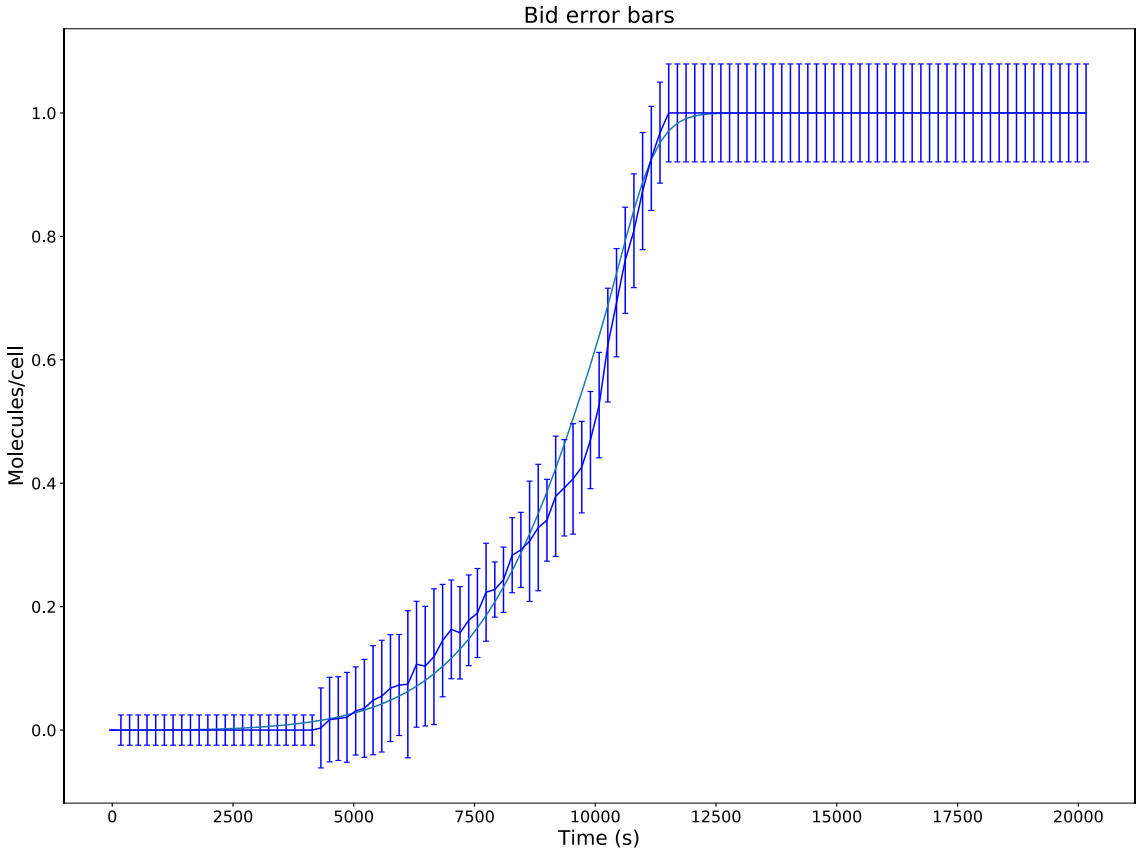


Figure 2.3. Bid trajectory along with estimated error bars for the associated FRET data.

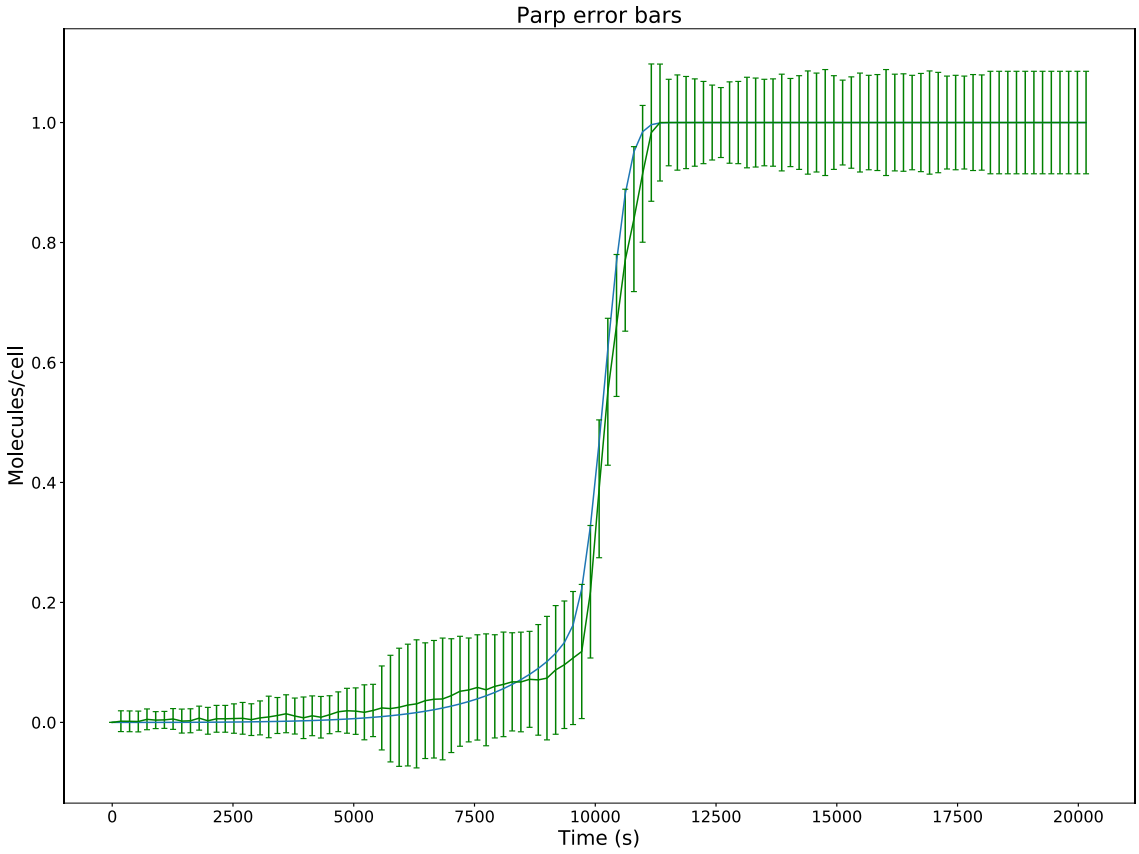


Figure 2.4. PARP trajectory along with estimated error bars for the associated FRET data.

2.2 Methods

2.2.1 Summary

Suppose we have a certain amount of knowledge with regard to the reaction topology for a biological signaling network and the associated reaction rates. Given that knowledge, we estimate expected values for quantities of interest under various in silico experimental conditions. Hypotheses are then generated based on differences in those values and how fast they change as regulatory conditions are varied. The methods needed to accomplish this are detailed in this section and applied to the extrinsic apoptosis model, or variants of it, in Chapters 3 and 4.

2.2.2 Modeling and Simulation

Models were written in the format of PySB, a rule-based system for constructing and simulating physicochemical models under mass-action kinetics [31]. Simulations were run, in the context of Bayesian evidence estimation, using the PySB software (http://pysb.org/). All representative models and software are distributed with open-source licensing and can be found in the GitHub repository https://github.com/LoLab-VU/BIND.

2.2.3 Bayesian Evidence Estimation

Bayesian evidence is expressed as

$$P(D|M) = \int L(D|\theta, M)\, P(\theta|M)\, d\theta \tag{1}$$

where M is the model under consideration, D is the data, θ is a particular set of parameter values, $L(D|\theta, M)$ is the likelihood function describing the fit of the data to the model under those parameter values, and $P(\theta|M)$ is the prior distribution of parameters. Note that this also represents the expected value of the likelihood function over the given prior distribution. All evidence estimates were made using nested sampling; introduced by Skilling in [38]. This method simplifies the evidence calculation by introducing a prior mass element

$dX = P(\theta|M)\, d\theta$ that is estimated by $(X_{i-i} - X_i)$ where $X_i = e^{-i/N}$, i is the current iteration of the algorithm, and N is the total number of live points. The evidence is then written as

$$Z = \int_0^1 L\, dX \approx \sum_{i=1} L_i (X_{i-1} - X_i) \tag{2}$$

Initialization of the algorithm is carried out by randomly selecting an initial population of parameter sets (points in parameter space) from the prior distribution, scoring each one with the likelihood function, and ranking them from L_{high} to L_{low}. At each iteration of the algorithm a new set of parameter values is selected and scored. If that score is higher than L_{low}, then it is added to the population, at the appropriate rank, and L_{low} is removed from the population and added to the evidence sum (2).

2.2.4 Nested Sampling Software

All evidence estimates in this work are calculated with MultiNest, a nested sampling-based algorithm designed for efficient evidence calculation on highly multimodel posterior distributions [42, 43]. MultiNest works by clustering the live points (population of parameter sets) and enclosing them in ellipsoids at each iteration. The enclosed space then constitutes a reduced space of admissible parameter sets. This lowers the probability of sampling from low likelihood areas and evaluating points that will only be discarded. The evidence estimate is accompanied by an estimate of the evidence error. The algorithm terminates when the presumed contribution of the highest likelihood member of the current set of live points, $L_{high}X_i$ is below a threshold. Here, we use a threshold of 0.0001 and a population size and 16,000 unless otherwise noted. See [42, 43], for more details on the MultiNest algorithm. We use MultiNest with the Python wrapper PyMultiNest [81], which facilitates the integration of PySB into the nested sampling pipeline.

2.2.5 Objective Functions

In this work we will present two methodologies with two different objective functions that use changes evidence values as regulatory conditions are varied to make inferences on changing network dynamics. We note that our objective functions are not true likelihood functions but instead represent values for quantities of interest.

2.2.5.1 Multimodel Inference Method

In the multimodel inference method we break down the network into various subnetworks and test each over increasing values of proposed regulatory elements for efficacy in achieving apoptosis. A proxy for apoptosis in this model, and the objective function for the nested sampling calculation, is the proportion of the protein PARP that has been cleaved by caspase-3 at the end of the in-silico experiment (20160 seconds). The function is thus

$$Obj_{multimodel} = \frac{cParp}{tParp}$$

where $cParp$ is the amount of PARP that has been cleaved and $tParp$ is the total amount of PARP in the system. When substituted into equation (1) in place of the likelihood function, the evidence calculation produces an expected value for the proportion of PARP that has been cleaved. It is an estimate of the average PARP cleavage over the chosen parameter ranges.

2.2.5.2 Pathway Targeted Method

In the pathway targeted method, we again vary proposed regulators but retain the full model while using targeted objective functions that represent the signal flux through different pathways in the network, similar to [83]. We consider the signal flux through the caspase pathway and the mitochondrial pathway. The objective function estimating signal flux through a pathway is

$$Obj_{pathway} = \sum_{t=0}^{T} \frac{\sum_0^t C3_{pathway}}{\sum_0^t C3_{total}} \times (cParp_t - cParp_{t-1})$$

where t represents time in seconds, $\sum_0^t C3_{pathway}$ is the amount of Caspase-3 activated via the target pathway up to time *t*, $\sum_0^t C3_{total}$ is the total Caspase-3 activated up to time *t*, and $\frac{\sum_0^t C3_{caspase}}{\sum_0^t C3_{total}}$ is the proportion of activated Caspase-3 that was produced via the target pathway up to time t. $(cParp_t - cParp_{t-1})$ is the total PARP that has been cleaved, and activated, by Caspase-3 from time $t-1$ to time t. Thus, at time *t* we know how much Caspase-3 has been activated via the target pathway and we have an estimate for how much PARP was activated (via Caspase-3) for the discrete time period $t-1$ to t. Multiplication of these two terms returns an estimate for the amount of PARP cleaved via the target pathway at time *t*. Summing over *T* then returns an estimate for the total apoptosis signal flowing through the target pathway. Like the PARP cleavage objective function, the signal flux objective substituted into equation (1) produces an expected value for this quantity on the given prior distributions.

2.2.6 Evidence Ratios

Evidence estimates are often used to select between two competing models by calculating the Bayes factor, or the ratio of their evidence values. This provides a measure of confidence for choosing one model over another. We can likewise use changes in evidence values to track changes in evidence ratios that provide additional insights into the dynamical relationship between pathways. To facilitate visualization of the changes in evidence ratios, with a continuous and symmetric range, the Bayes factors were calculated as

$$Bf = \begin{cases} -\frac{Z_2}{Z_1} + 1 \ \ if \ Z_1 < \ Z_2 \\ \frac{Z_1}{Z_2} - 1 \ \ if \ Z_1 > \ Z_2 \end{cases}$$

where Z_1 and Z_2 are the evidence estimates for two pathways under comparison.

2.2.7 Computational Resources

Because of the high computational workload necessary for this analysis, a wide range of computational resources were used. The bulk of the work was done on the ACCRE cluster at Vanderbilt University which has more than 600 compute nodes running Intel Xeon processors and a Linux OS. As many as 300 evidence calculations were run in parallel on this system. Additional resources included two local servers, also running Intel processors and a Linux OS, as well as a small local four node cluster running Linux and AMD Ryzen 1700 processors.

Chapter 3

Bayesian Evidence Based Analysis of Network Dynamics

3.1 Summary

Characterization of signal execution dynamics within complex biochemical networks is highly challenging but necessary to understand how cells process signals and commit to a biological phenotype. Mechanistic interpretation of experimental results can be inaccurate due to limited data or the need for an unrealistic number of measurements. Mathematical simulation of network dynamics has emerged as a complement to experimentation for the exploration of signal execution mechanisms. However, traditional computational methods require either detailed knowledge of model parameters or sufficient data to calibrate models to experiments, both of which can be difficult to obtain.

To address this challenge, we take a probabilistic approach to the analysis of network-driven biochemical processes using a Bayesian inference formalism to explore network dynamics when data is limited and identify the regulatory mechanisms of biochemical signaling. We applied the approach to the well-studied signal execution pathways of mammalian extrinsic apoptosis and produced results consistent with experimental evidence as well as additional (theoretical) hypotheses regarding the mechanisms of mitochondrial signal amplification.

This chapter, along with Chapter 2, was written as a stand-alone manuscript and is meant to both introduce the concept of probabilistic, objective function-based modeling and to give proof of concept examples of the methodology. A version of the manuscript can be found here: https://www.biorxiv.org/content/10.1101/732396v2.

3.2 Challenges in the Analysis of Dynamics of Physicochemical Networks

Emergent behaviors of complex biological networks are difficult to characterize [1, 2]. They arise from the interplay between various components and pathways that make up the larger system and because they appear only when those pieces are brought together, determining the role of any singular part can pose a significant challenge. To study the myriad of possible cellular regulatory conditions and accelerate the formation of predictive hypotheses, computational modeling is often used alongside experimental methods [3]. Physicochemical models representing the currently known biochemical interactions of a system are constructed from knowledge garnered from years or even decades of experimentation. Unfortunately, such models depend on reaction rate parameters that are typically unknown and must therefore be calibrated to experimental data – data that is often scarce [20]. An especially difficult behavior to study is the evolution of signal execution dynamics under shifting regulatory conditions, the analysis of which is essential to the identification of regulatory elements that govern signal transduction between pathways or those with the potential to elicit phenotype transitions. In such cases, calibration of a complex model to inadequate data can result in equally good fits for very different parameter sets [37], potentially leading to inconsistent conclusions regarding the signaling dynamics of the system. Methods that retain the extensive mechanistic knowledge gleaned from the literature without using explicit parameter values will allow for greater exploration of network dynamics and accelerate the generation of new biological hypotheses.

3.3 Overview of Bayesian Inference-Based Analysis of Network Dynamics

In this work, we have taken a probabilistic approach to the exploration of network dynamics over regulatory perturbations by utilizing methods from the field of model selection and multimodel inference [82, 83]. We use a Bayesian inference-based framework for the analysis of network dynamics across regulatory conditions. In particular, we use Bayesian model evidence as a metric for comparisons of signal flow through different pathways and

subnetworks. Two complementary approaches are used. The first employs a multimodel inference approach that deconstructs the test model into subnetworks, all of which can successfully execute the apoptotic signal. Bayesian evidence estimates, for an objective function representing a quantity of interest and over a realistic range of reaction rate parameter values, are computed over a range of regulatory conditions. This is equivalent to calculating the expected value of that quantity over the given parameter range for various in silico experiments. The second approach retains the full network but defines the objective functions to represent processes on target pathways, in this case the signal flux through those pathways. The expected flux values are then compared to provide a view of signal execution that inherently includes crosstalk between pathways. Because the evidence estimate takes the form of an expected value calculation these approaches are effectively probabilistic analogs to traditional modeling methods that calibrate and simulate to obtain explicit values of interest.

As a test case, we apply these methods to a modified version of the extrinsic apoptosis reaction model (EARM; see Section 2.1) [17]. Apoptosis is a well-studied system making it ideal for demonstrating the efficacy of computational approaches [17, 84-88]. There exist two extrinsic apoptosis phenotypes: Type I, which is independent of mitochondrial amplification of the apoptotic signal, and Type II, which is dependent on it. In the following we will use our method to examine the effects of changes in various regulatory axes on signaling dynamics and phenotypic outcome in the context of the established experimental literature.

3.4 Results

3.4.1 General Strategy and Workflow

To investigate the dynamics of apoptotic signaling, we take a Bayesian inference approach to examine how different network components contribute to signal execution. The goal is to build a composite description of system dynamics by observing variations in signal throughput between these subnetworks relative to changes in regulatory conditions.

This differs from traditional model selection and multimodel inference applications where models are typically ranked based on their fit to experimental data and high-ranking models may be averaged to obtain a composite model [82, 83, 86, 89, 90, 91]. In contrast, the objective functions used here represent quantities of interest and evidence calculations result in expected values of the quantities given the prior distributions of parameters. A higher evidence estimate thus indicates a more robust signal over the prior range of parameters. It should be noted that Bayesian evidence inherently incorporates model complexity as the objective functions are integrated over normalized prior distributions that consist of as many dimensions as parameters [86, 42, 92].

The general approach is shown schematically in Figure 3.1. Two complementary methods are used. In a multimodel inference approach the model is deconstructed into biologically relevant subnetworks and the probability of achieving apoptosis under various regulatory conditions is estimated via Bayesian evidence. If we tailor the objective function to represent signal execution strength, as measured by cleaved PARP at the end of the simulation run, then the evidence describes the likelihood that the signal is effectively transmitted through the network. As we will see, comparisons of changes in signal strength through relevant subnetworks subsequently allows inferences to be made on the effect of the perturbed network regulator as well as various network components on the overall dynamics of the system. We use these changes in Bayesian evidence to examine how variations in regulatory elements alter the dynamics of the extrinsic apoptosis model and gain insight into the mechanisms that commit the network to either Type I (mitochondria independent) or II (mitochondria dependent) execution modes.

The EARM reaction topology was deconstructed into several network variations (Figure 3.2A-3.2F). These include the full model, the caspase pathway, and the mitochondrial pathway including two subpathways that either directly transduce the apoptotic signal (via Caspase-3 cleavage) or inhibit XIAP, the inhibitor of activated Caspase-3 (and Caspase-9 in the Apoptosome). Also included are combinations of the caspase pathway with either of

the two mitochondrial subpathways and the mitochondrial signal transduction pathway in isolation. We focus primarily on the complete network and caspase pathway as these are the most relevant for the analysis of Type I/II phenotypic outcome. The second method, deemed the pathway targeted approach, retains the complete network and uses objective functions that measure apoptotic signal flux (see Section 2.2 for details) through the caspase and mitochondrial pathways, as well as the total flux through the network. We primarily consider the influence of the apoptosis inhibitor XIAP on regulatory dynamics and phenotypic fate but also look at the regulatory effect of the death inducing signaling complex (DISC).

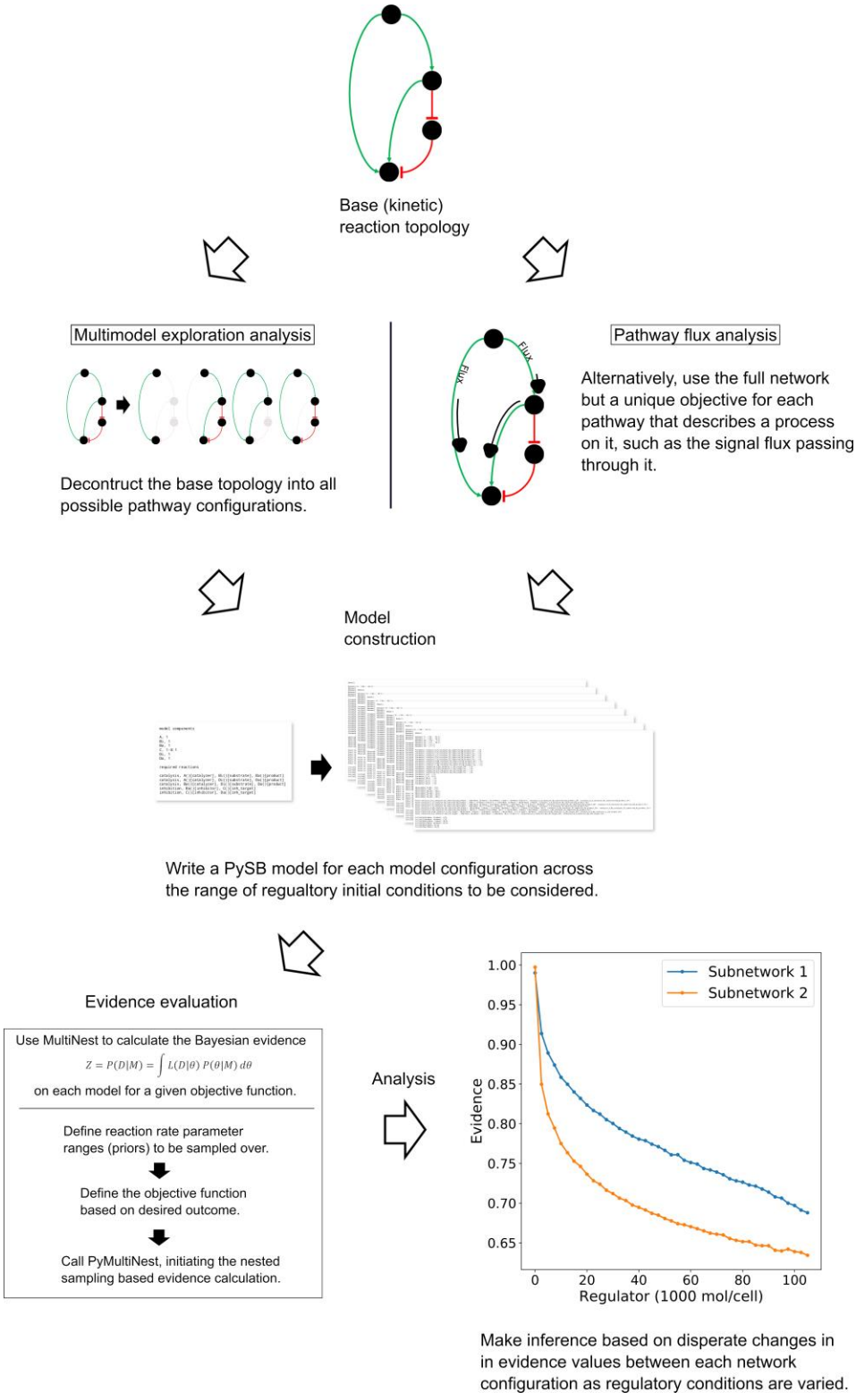
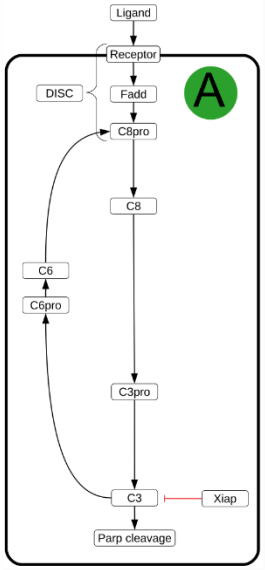
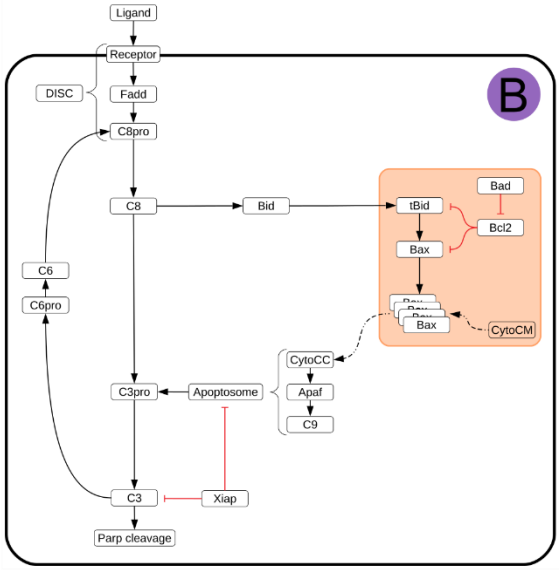
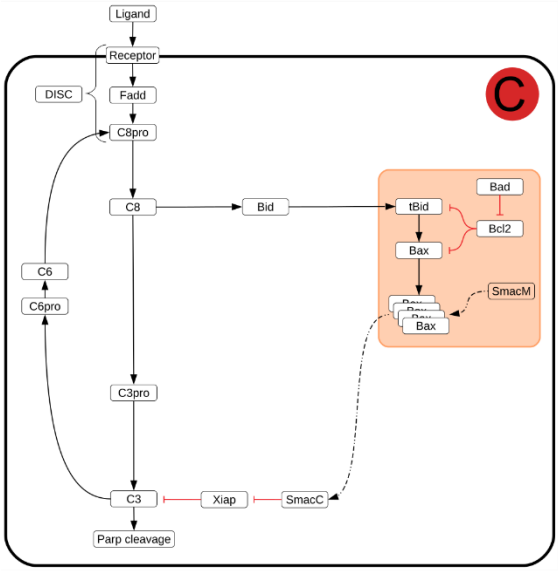
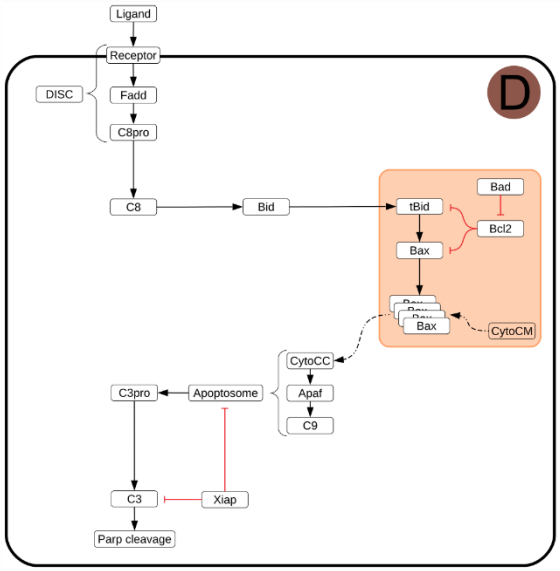


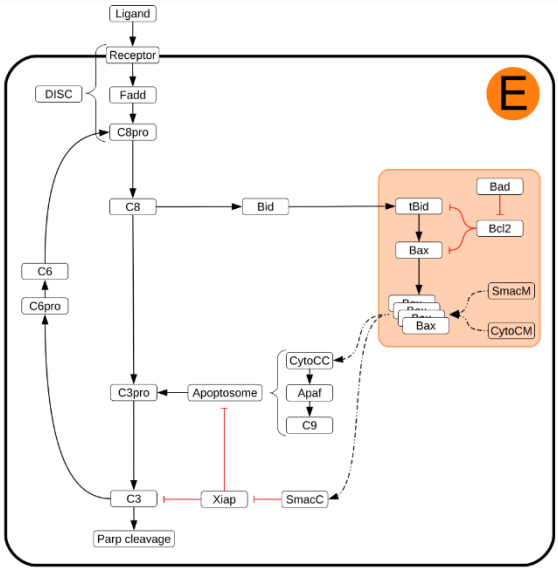
Figure 3.1. Workflow for Bayesian evidence-based analysis of network dynamics and signaling regulation.


A
Ligand
Receptor
DISC
Fadd
C8pro
C8
C6
C6pro
C3pro
C3
Xiap
Parp cleavage


B
Ligand
Receptor
DISC
Fadd
C8pro
C8
Bid
tBid
Bad
Bcl2
Bax
CytoCM
CytoCC
Apaf
C9
C6
C6pro
C3pro
Apoptosome
C3
Xiap
Parp cleavage


C
Ligand
Receptor
DISC
Fadd
C8pro
C8
Bid
tBid
Bad
Bcl2
Bax
SmacM
C6
C6pro
C3pro
C3
Xiap
SmacC
Parp cleavage


D
Ligand
Receptor
DISC
Fadd
C8pro
C8
Bid
tBid
Bad
Bcl2
Bax
CytoCM
CytoCC
Apaf
C9
C3pro
Apoptosome
C3
Xiap
Parp cleavage


E
Ligand
Receptor
DISC
Fadd
C8pro
C8
Bid
tBid
Bad
Bcl2
Bax
SmacM
CytoCM
C6
C6pro
CytoCC
Apaf
C9
C3pro
Apoptosome
C3
Xiap
SmacC
Parp cleavage

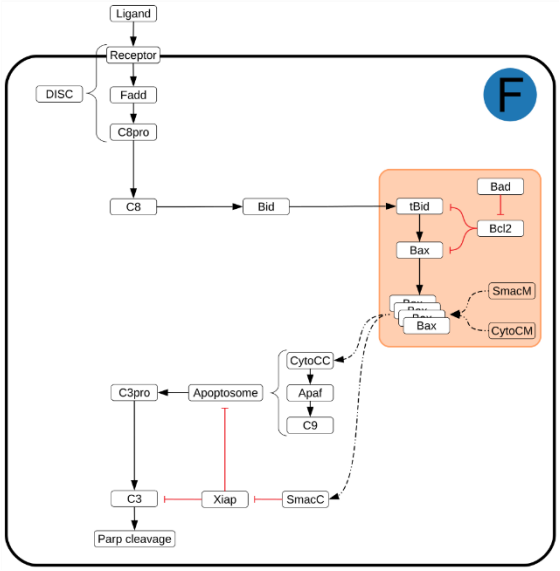

F
Ligand
Receptor
DISC
Fadd
C8pro
C8
Bid
tBid
Bad
Bcl2
Bax
SmacM
CytoCM
CytoCC
Apaf
C9
C3pro
Apoptosome
C3
Xiap
SmacC
Parp cleavage

Figure 3.2. Extrinsic apoptosis subnetworks. (A) The isolated caspase pathway. (B) The caspase pathway with the Caspase-3 activating component of the mitochondrial pathway. (C) The caspase pathway with the XIAP inhibiting component of the mitochondrial pathway. (D) The isolated Caspase-3 activating component of the mitochondrial pathway. (E) The complete network. (F) The isolated mitochondrial pathway.

3.4.2 Differential Downregulation of Extrinsic Apoptosis Subnetworks by XIAP

XIAP has been put forth as a critical regulator in the choice of apoptotic phenotype. In Jost et al. [71] they examined hepatocytes (Type II cells) and lymphocytes (Type I cells) under different conditions to examine the role XIAP plays in Type I/II determination and made several observations. They reported that Fas ligand (FasL) induced apoptosis resulted in increased levels of XIAP in hepatocytes but lowered levels in thymocytes. They then found that while XIAP deficient mice died earlier than wild-type when injected with hepatocyte targeted FasL or anti-Fas antibody, XIAP deficient thymocytes showed no increase in apoptosis. From this they concluded that XIAP must be a key regulator of apoptosis in hepatocytes. Lastly, they treated XIAP, Bid, and XIAP/Bid deficient mice, along with wild-type, with FasL or Fas-antibody. All but the Bid-only deficient mice showed hepatocyte effector caspase activation, implying that the loss of XIAP rendered previously apoptosis resistant Bid-only knockouts susceptible to apoptosis through the Type I pathway. Altogether, they concluded that XIAP is the key regulator that determines the choice of pathway.

The results in Jost et al. [71] imply that the cellular level of XIAP determines the preferred apoptotic pathway with higher levels specific to Type II cells and lower levels specific to Type I. To examine possible mechanistic explanations for this behavior, and to explore the general dynamics of the model, we computed the Bayesian evidence for the six apoptosis-inducing subnetworks described above at varying concentrations of XIAP.

3.4.2.1 XIAP Control of the Type I/II Phenotype

Here we compare two of those networks, the caspase only pathway and the complete model, to gauge the effect of XIAP on networks with and without a mitochondrial component (Figure 3.3). The effect of XIAP on all subnetworks are displayed in Figure 3.4. The log-evidence version of Figure 3.4 along with estimated errors generated by MultiNest are displayed in Figure 3.5. XIAP was varied from 0 to 200,000 molecules per cell in increments of 250 to explore how changes in XIAP affect the likelihood of apoptosis execution. For those networks that include components of the mitochondrial pathway Bcl-2 was excluded to ensure those components were fully active. All other initial values were fixed at the levels shown in Table 2.1. In the absence of XIAP both the caspase pathway and complete model have evidence estimates greater than 0.99, (Figure 3.3C) indicating that they both reach nearly full PARP cleavage, and by extension apoptosis, across the allowed range of parameters.

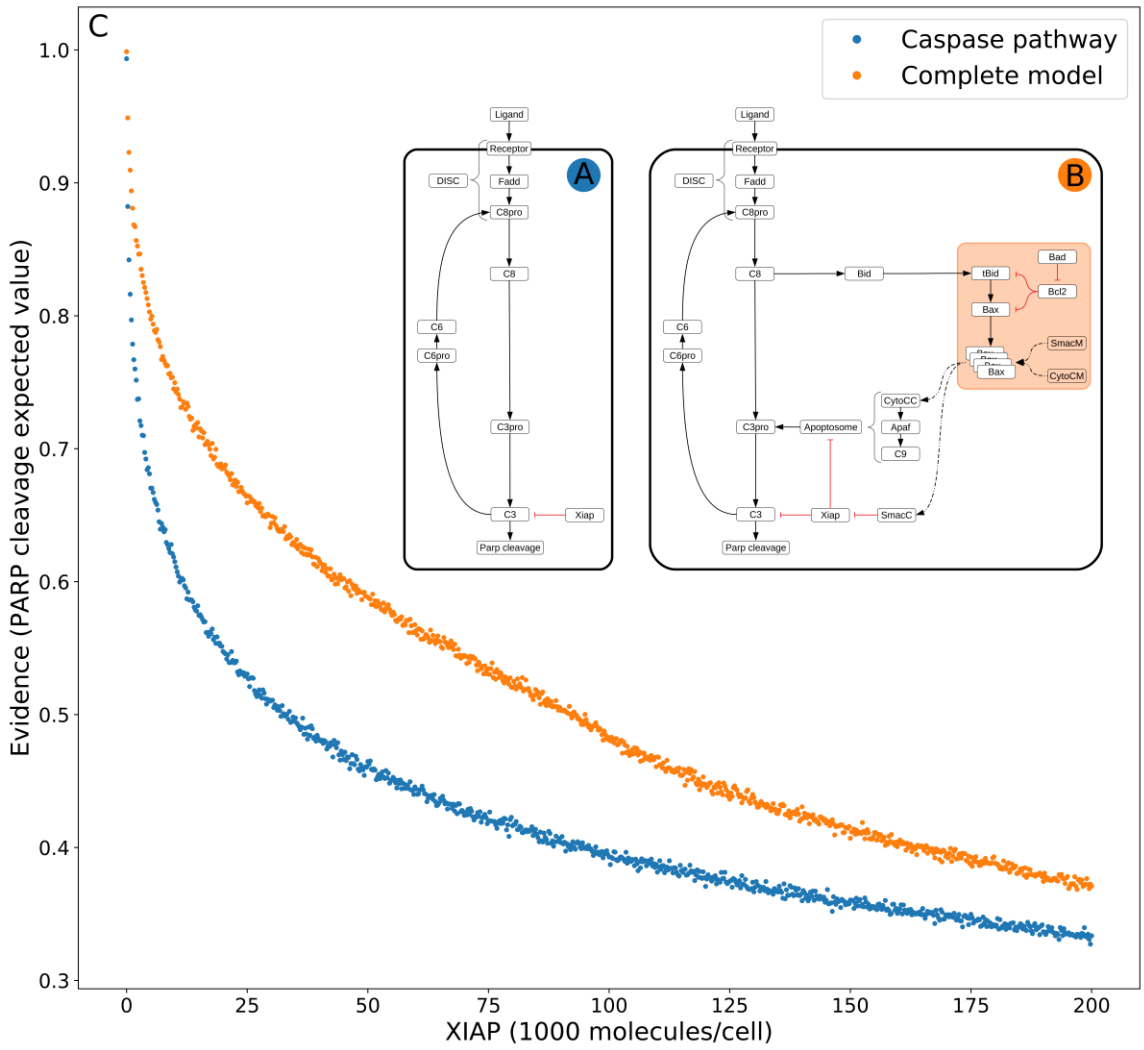


Figure 3.3. Comparison of the Bayesian evidence for achieving apoptosis via the isolated caspase pathway and the complete network and for increasing concentration of XIAP. (A) The isolated caspase pathway. (B) The complete network. (C) The trends in the expected values of cleaved PARP for each of the networks in (A)-(B) over a range of values of the apoptosis inhibitor XIAP and for an objective function that computes the proportion of PARP cleavage (a proxy for cell death) at the end of a simulation.

As XIAP levels increase we see differential effects on these subnetworks in the form of diverging evidence estimates indicating differences in the efficacy of XIAP induced apoptotic inhibition. The isolated caspase pathway (Figure 3.3C blue) shows the steepest decline which is most prominent for lower values of XIAP but diminishes as XIAP increases. The evidence values for the caspase pathway clearly diverges from those of the complete model. The expected value for the proportion of PARP cleavage, the average PARP cleavage over the provided parameter ranges, for the caspase pathway falls to 0.5 at an XIAP level of roughly 32,000. The complete network however, requires XIAP levels nearly 3x as high with the evidence value reaching 0.5 at around 92,000.

Because the caspase pathway is representative of the Type I phenotype, the disproportionate drop in its expected proportion of PARP cleavage as XIAP increases is consistent with experimental evidence showing XIAP induced transition to a Type II phenotype. The complete network, containing the full mitochondrial subnetwork, is also affected by XIAP but clearly shows a higher resistance to its antiapoptotic effects, particularly at moderate levels of the inhibitor. This suggests a growing dependence on mitochondrial involvement in apoptosis as XIAP increases from low to moderate levels. At higher levels of XIAP the evidence values for the caspase pathway level off and the gap between the two networks narrows. The disproportionate effect of XIAP inhibition of apoptosis on the caspase pathway suggests that the mechanism for XIAP induced transition to a Type II pathway is simply differential inhibition of the apoptotic signal through the isolated caspase pathway versus a network with mitochondrial involvement.

We note here that the small differences in evidence values between the two models should not be surprising since every subnetwork being considered is capable of transmitting the apoptotic signal. Thus, we should not expect differences of evidence that would rule out any of them under model selection criteria. In a classical model selection and multimodel inference scenario small differences in evidence estimates that would not allow for selection of a favored model might be used to construct a composite model, weighting the various components by the evidence values [82, 83]. Fortunately, we have no need to

choose a best model as the complete model already represents the biology as we understand it. The goal here is to use the differences in evidence to construct a composite picture of, not the structure of the model, but signaling dynamics. For that we consider relative changes in the evidence values as XIAP is increased.

3.4.2.2 Mechanistic Interpretation of the Extrinsic Apoptosis Subnetworks

Above the caspase pathway the next two highest trends in evidence values belong to the networks representing the caspase pathway with mitochondrial activation of Caspase-3 and mitochondrial activation of Caspase-3 alone (Figures 3.4 purple and 3.4 brown). For most of the range with XIAP below 100,000 these two trends have largely overlapping trajectories, despite the fact that the former has twice as many paths carrying the apoptotic signal. By the time XIAP reaches a level of 100,000 the two trends diverge as the decrease in evidence values for the mitochondrial activation only network accelerates. This could be explained by XIAP overwhelming the Apoptosome at these higher levels. The apoptosome is an apoptosis inducing complex (via Caspase-3 cleavage) consisting of Cytochrome c, APAF-1, and Caspase-9, and is an inhibitory target of XIAP. As XIAP increases past 125,000 the mitochondrial activation only evidence values fall below even the caspase only values, possibly due to the two-pronged inhibitory action of XIAP at both the Apoptosome and Caspase-3. An interesting observation here is that the addition of the caspase pathway to the mitochondrial activation pathway does not appear to increase the likelihood of achieving apoptosis for lower values of XIAP.

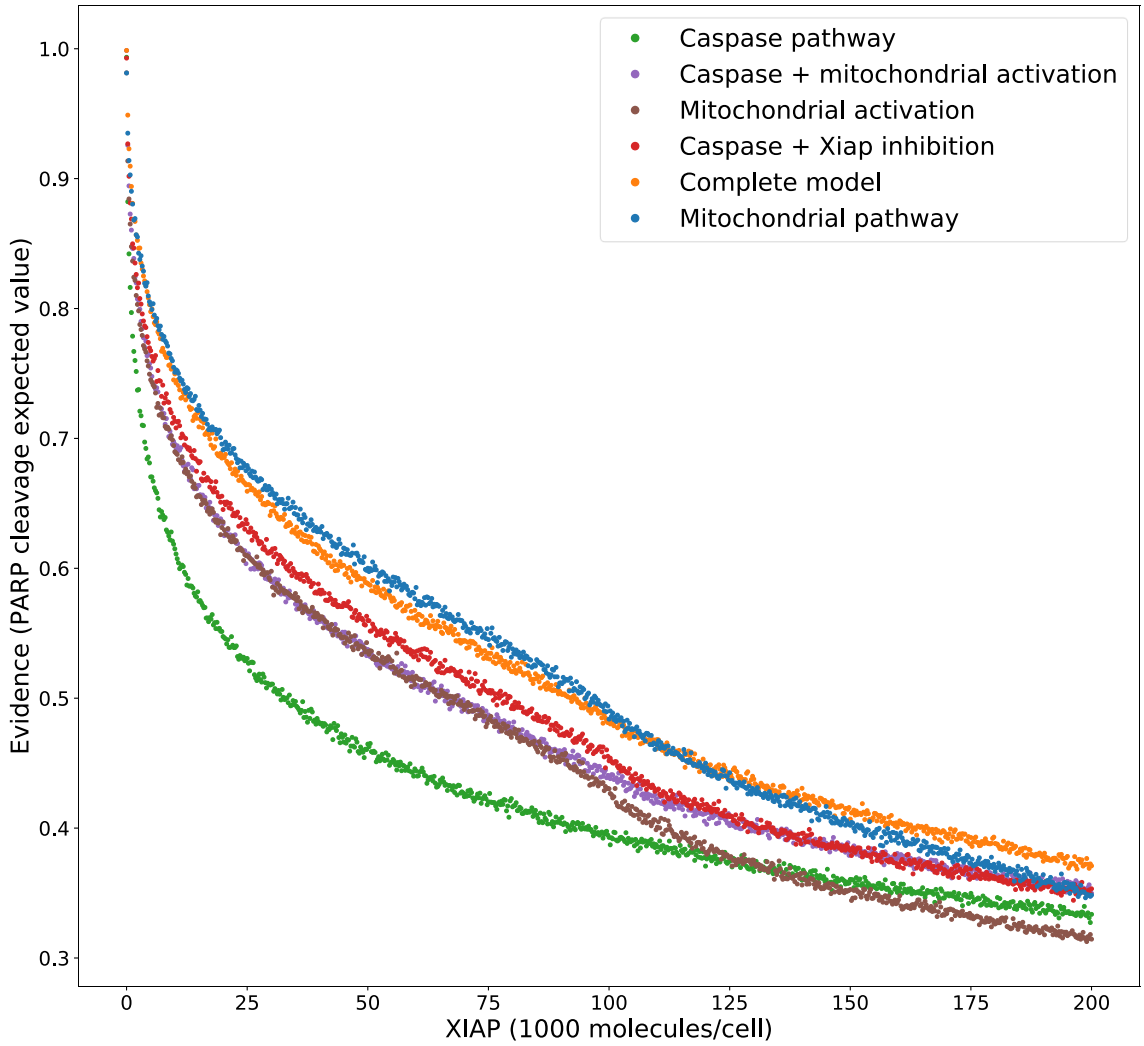


Figure 3.4. Bayesian evidence for achieving apoptosis through all six subnetworks of the extrinsic apoptosis network and for increasing concentrations of XIAP. (A) The isolated caspase pathway. (B) The caspase pathway with the Caspase-3 activating component of the mitochondrial pathway. (C) The caspase pathway with the XIAP inhibiting component of the mitochondrial pathway. (D) The isolated Caspase-3 activating component of the mitochondrial pathway. (E) The complete network. (F) The isolated mitochondrial pathway.

Above those two evidence trends is the trend for the network consisting of the caspase pathway and mitochondrial inhibition of XIAP (Figure 3.4 red). Below an XIAP level of 100,000 this trend is consistently above the evidence values for the network of the caspase pathway plus mitochondrial activation of Caspase-3. Note that while the caspase pathway does not appear to increase the likelihood of achieving apoptosis when added to the mitochondrial activation pathway (Figure 3.4 purple), the amplification of the caspase pathway via mitochondrial inhibition of XIAP leads to a higher likelihood than direct activation through the mitochondria. This suggests the possibility that the primary mechanism for mitochondrial apoptotic signal amplification, under some conditions, may be inhibition of XIAP, with direct signal transduction a secondary mechanism. Above an XIAP level of 100,000, the caspase with XIAP inhibition trend drops to levels roughly in line with the evidence values for the caspase pathway plus direct activation, possibly due to the fact that Smac, the mitochondrial export that inhibits XIAP, is also set to 100,000 molecules per cell. Both, however, remain more likely to attain apoptosis than the caspase only pathway.

The two subnetworks with the highest trends in evidence for apoptotic signal execution are the complete model and the isolated mitochondrial pathway (Figures 3.4 orange and 3.4 blue). As previously mentioned, both of these networks contain the full mitochondrial pathway implying that this pathway supports resistance to XIAP inhibition of apoptosis. Between XIAP levels of 0 to 100,000 the two trends track very closely, with the mitochondrial only pathway showing a slight but consistent advantage for apoptotic execution. The average difference between an XIAP level of 20,000 and 80,000 is roughly 0.014, meaning we expect the average PARP cleavage to favor the mitochondrial only pathway by about 1.4 percentage points, which may seem unremarkable. Context matters however, and the context here is that the complete network has potentially twice the bandwidth for the apoptotic signal, namely the addition of the more direct caspase pathway. Together, this raises the possibility that under some conditions the caspase pathway is not a pathway but a sink for the apoptotic signal. In such a scenario, the signal through the caspase pathway would get lost as Caspase-3 is degraded by XIAP. Not until the signal through the mitochondrial

pathway begins inhibiting XIAP could the signal proceed. Around the 100,000 level of XIAP the evidence trend for the mitochondrial pathway crosses below that for the complete network. This could be due to the parity with Smac, components of the Apoptosome, or a combination of the two.

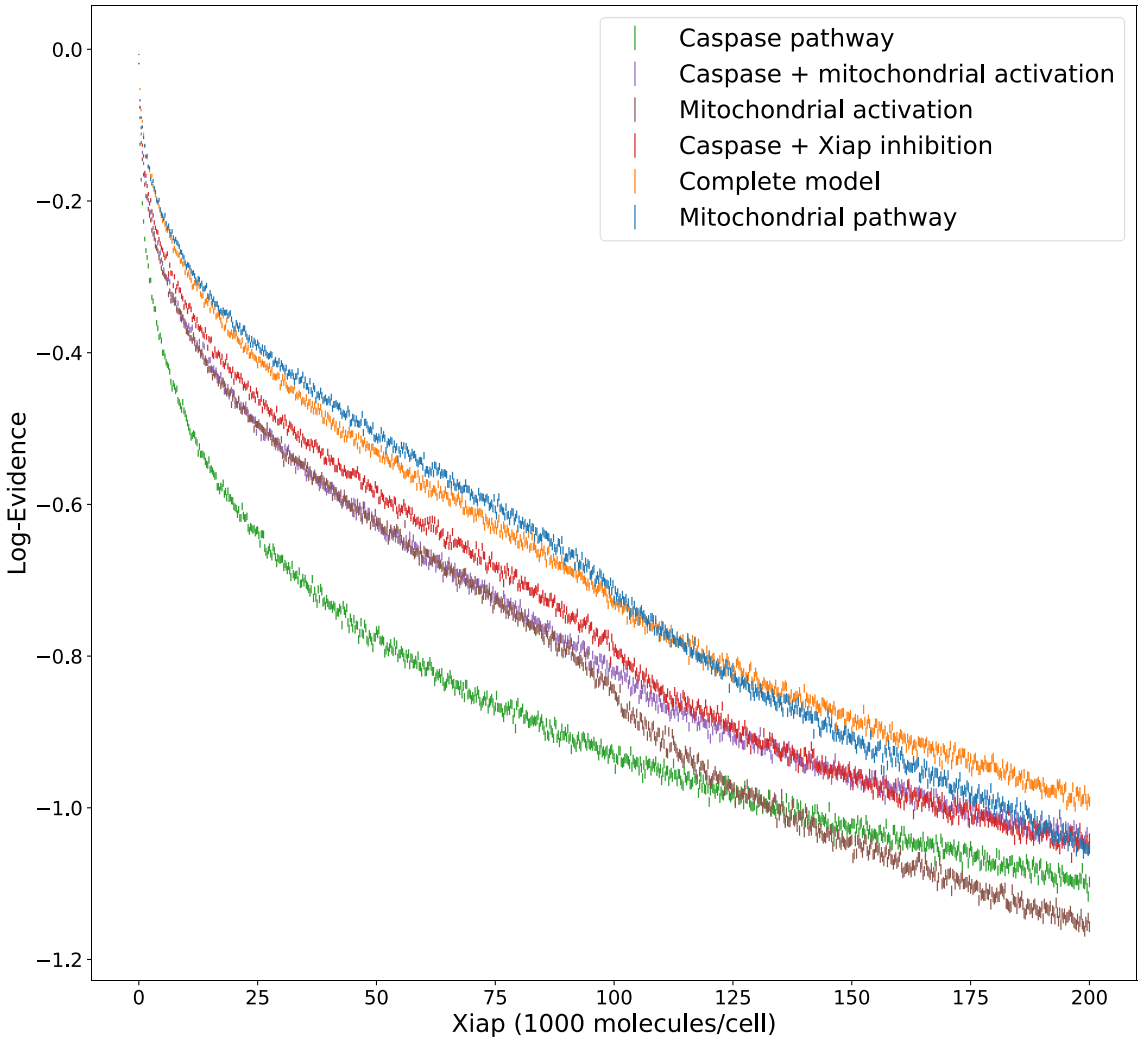


Figure 3.5. Log-evidence version of plots in Figure 3.4 with estimated errors generated by MultiNest.

3.4.3 Apoptosis Signal Strength Determines the Signal Route Through the Network

The Type I/II phenotypes for the extrinsic apoptosis system were first described by Scaffidi et al. [92]. In that work they examined several cell lines and classified them into those that required the mitochondrial pathway to achieve apoptosis (Type II) and those that don't (Type I). They made several interesting conclusions. They found that Type II cells had relatively weak DISC formation, that both phenotypes responded equally well to receptor mediated cell death, that there was a delay in caspase activation in Type II cells, and that caspase activation happened upstream of mitochondrial activation in Type I cells and downstream in Type II. Here we examine the effects of increasing XIAP, again from 0 to 200,000 in increments of 250, at a low level of DISC formation by lowering the initial values of both the scaffold protein FADD and the initiator Caspase-8, from 130,000 to 100 molecules per cell. In addition to the multimodel inference approach used above, we also use the pathway targeted approach, using the flux objective function (see Section 2.1), to get a holistic view of network dynamics that incorporates cross-talk from all pathways.

Figure 3.6 displays the PARP cleavage expected values from Figure 3.3 along with their low DISC counterparts. Two things are immediately apparent. The expected PARP cleavage for the caspase pathway under low DISC conditions is much lower across the entire range of XIAP. The complete network on the other hand shows almost no difference in low DISC conditions at lower values of XIAP and was even slightly higher at higher values of XIAP. This clearly supports the hypothesis that that mitochondrial involvement is necessary to overcome weak DISC formation and that this constitutes a Type II trait.

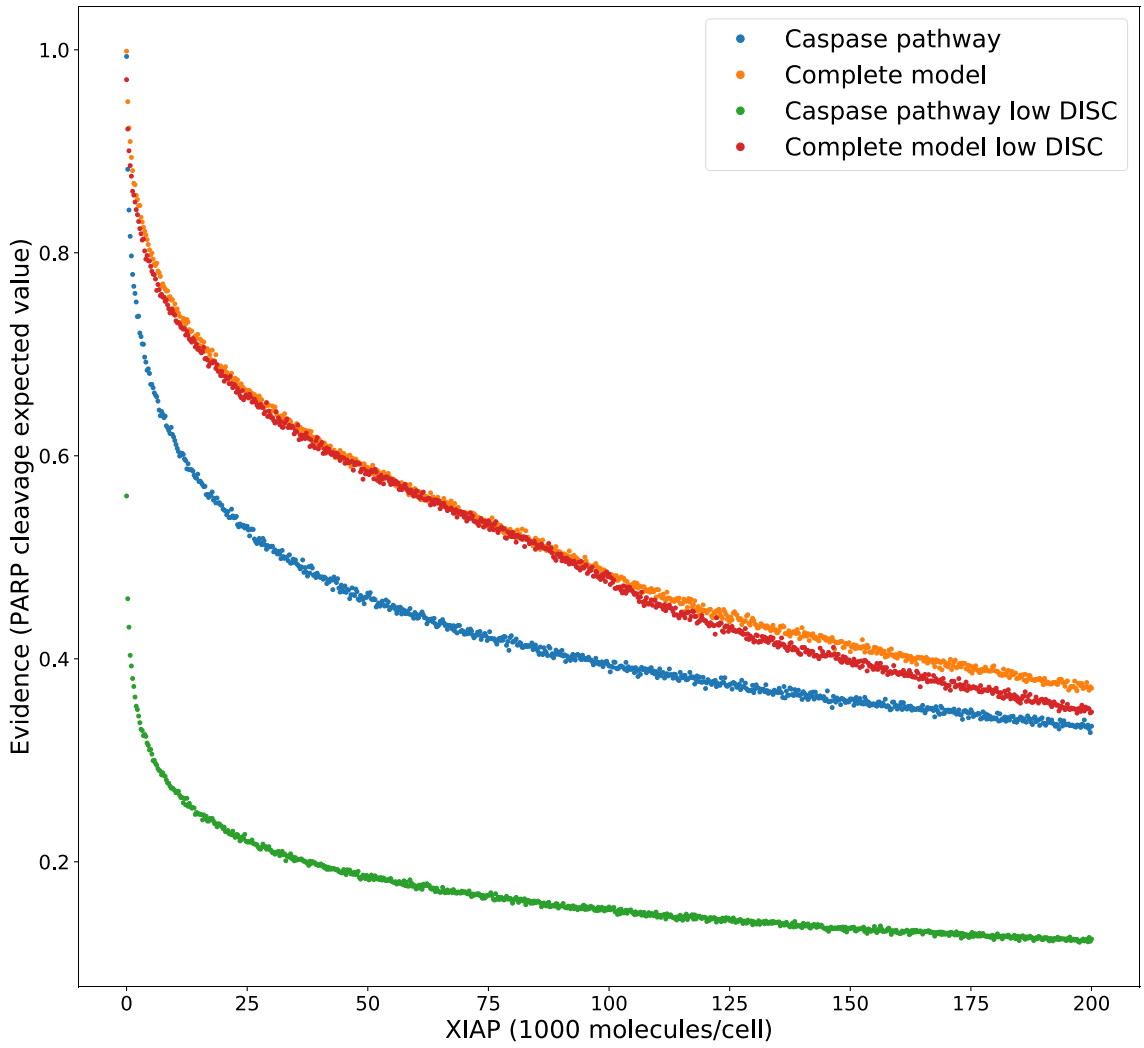
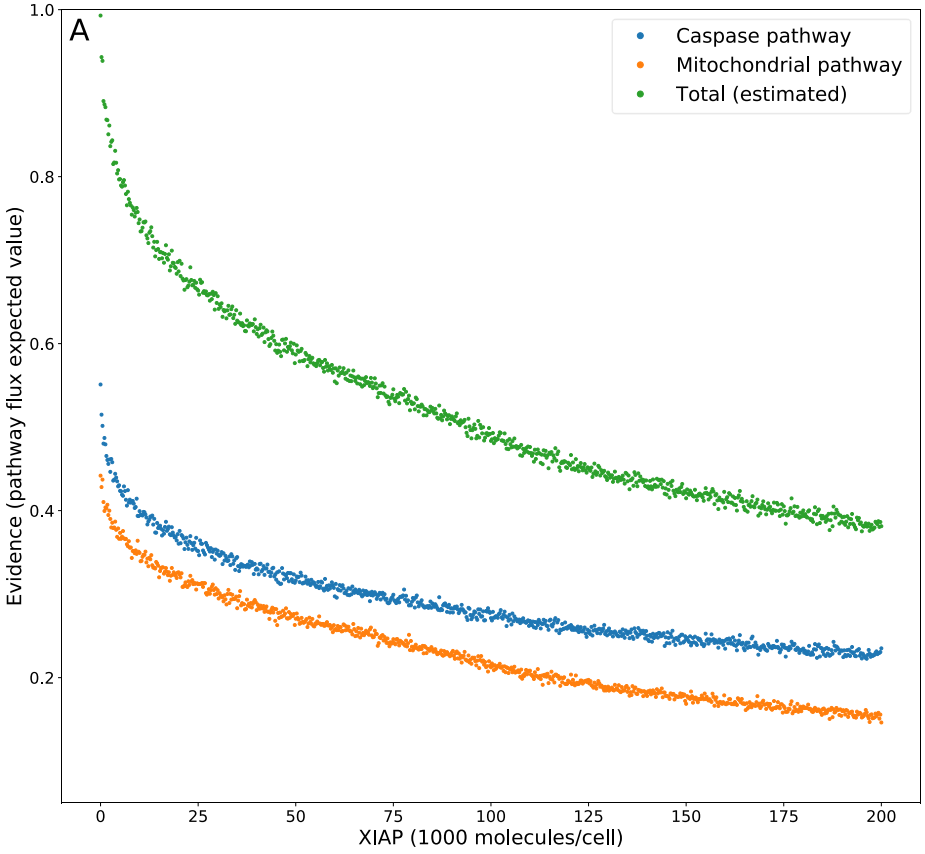


Figure 3.6. Expected PARP cleavage for the caspase pathway and complete network under both low and high DISC conditions (100 and 130,000 molecules per cell of FADD and Caspase-8 respectively).

Figures 3.7A and 3.7B display the expected flux through the caspase pathway and complete network for the high and low DISC values respectively. At higher DISC values, signal flux through the caspase pathway is consistently higher than flux through the mitochondrial pathway, despite the greater effect on the caspase only pathway. At lower DISC values the signal flux through the mitochondrial pathway far outstrips the flux through the caspase pathway. This is interesting in the context of the hypothesis that mitochondrial activation is downstream of caspase activation in Type I cells and upstream in Type II [92]. If a weaker signal does indeed push the signal through the mitochondrial pathway the initial activation of Caspase-8 would be weak and the amplifying activity of the mitochondria would ramp up before either Caspase-8 or Caspase-3 became fully active. On the other hand, strong initial activation that pushes the signal through the caspase pathway would have the opposite effect. Also notable is the nearly identical trajectories of the total signal flux through the low and high DISC models. The average difference over the range of XIAP was only 0.011. This is consistent with the hypothesis that both Type I and Type II cells respond equally well to receptor mediated apoptosis.

Overall these results set up three scenarios. On one end, high DISC formation and low XIAP results in the independence of apoptosis from the mitochondrial pathway. This behavior is consistent with Type I cells like SKW6.4 and H9 cell lines [92]. On the other end of the spectrum is the case with low DISC formation (and by construction low Caspase-8 activity) and near complete dependence on the mitochondrial pathway. Such behavior is consistent with Type II cells. In between these two extremes is the case with high DISC formation (and Caspase-8) activation but still has a dependence on mitochondrial activity. Such behavior is consistent with MCF-7 cell that are known to have traits of both phenotypes. Based on these results, a hypothetical mechanism for this behavior is mitochondrial induced inhibition of XIAP and subsequent signaling through the caspase pathway.


A
Caspase pathway
Mitochondrial pathway
Total (estimated)
Evidence (pathway flux expected value)
1.0
0.8
0.6
0.4
0.2
0
25
50
75
100
125
150
175
200
XIAP (1000 molecules/cell)

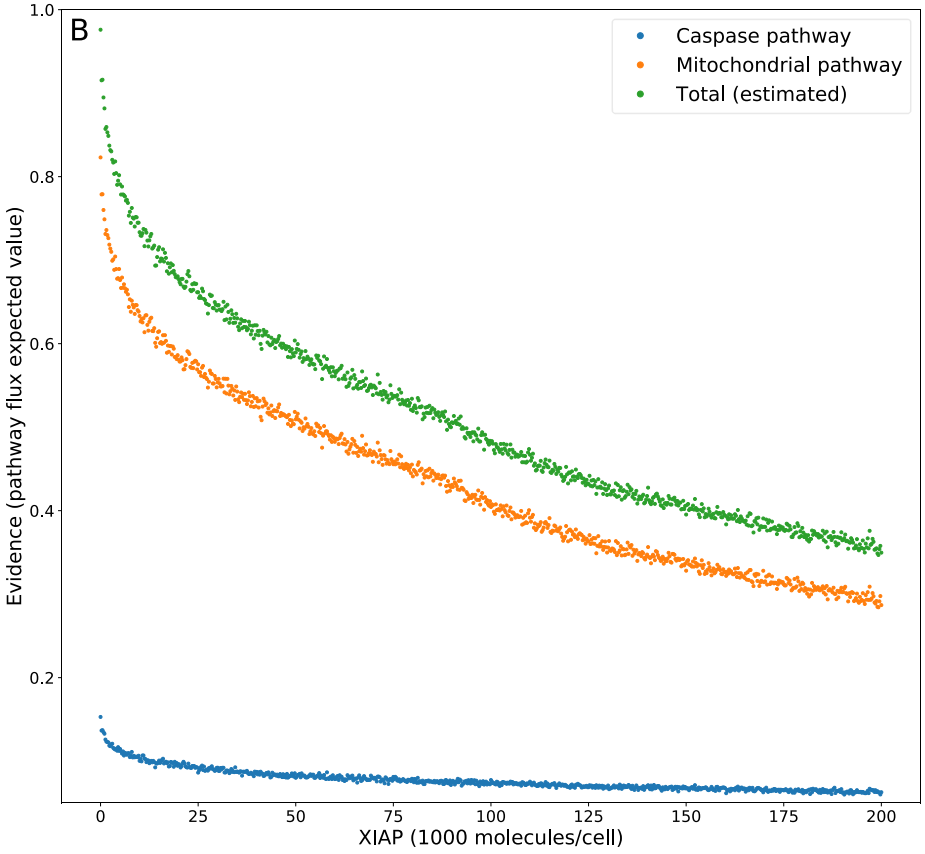
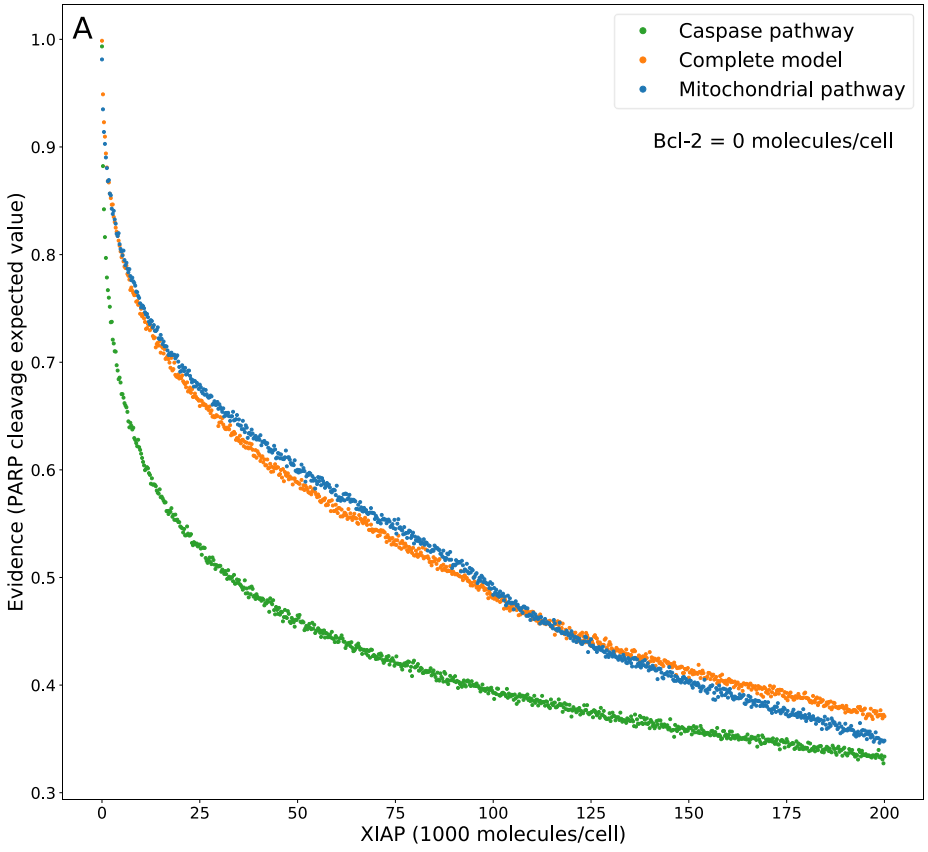

B
Caspase pathway
Mitochondrial pathway
Total (estimated)
Evidence (pathway flux expected value)
1.0
0.8
0.6
0.4
0.2
0
25
50
75
100
125
150
175
200
XIAP (1000 molecules/cell)

Figure 3.7. Signal flux through the caspase and mitochondrial pathways as well as the estimated total flux through the system. (A) Signal flux under high DISC conditions (FADD and Caspase-8 both at 130,000 molecules per cell). (B) Signal flux under low DISC conditions (FADD and Caspase-8 both at 100 molecules per cell).

3.4.4 Evidence Ratios and XIAP Influence on Type I/II Apoptosis Phenotype

Typical model selection methods calculate the evidence ratios, or Bayes factors, to choose a preferred model and estimate the confidence of that choice [82, 83]. When comparing the changes in the evidence of an outcome as regulatory conditions change, the changes in the evidence ratios can provide additional information about changing network dynamics under regulatory perturbations. To characterize the effect of XIAP on the choice of apoptotic phenotype, Type I or II, we calculated the evidence ratios (Figure 3.8B) for each value of XIAP between the caspase pathway and both the complete network and mitochondrial pathway (Figure 3.8A) with a fully active mitochondrial pathway (no Bcl-2 activity). In these calculations, the denominator represents the caspase pathway so that higher values favor a need for mitochondrial involvement. An interesting feature of both the complete and mitochondrial evidence ratio trends is the peak and reversal at a moderate level XIAP. This reflects the initially intense inhibition of the caspase pathway that decelerates relatively quickly as XIAP increases, and a steadier rate of increased inhibition on networks that incorporate the mitochondrial pathway. The ratios peak between 45,000 and 50,000 molecules of XIAP, more than double the value of its target molecule Caspase-3 at 21,000, and represents the optimal level of XIAP for the requirement of the mitochondrial pathway and attainment of a Type II phenotype. Given the near monotonic decline of the evidence trends of both pathways, representing increasing suppression of apoptosis, the peak and decline in the evidence ratios may represent a shift toward complete apoptotic resistance. This was shown experimentally in Aldridge et al. [87].


A
Caspase pathway
Complete model
Mitochondrial pathway
Bcl-2 = 0 molecules/cell
Evidence (PARP cleavage expected value)
XIAP (1000 molecules/cell)

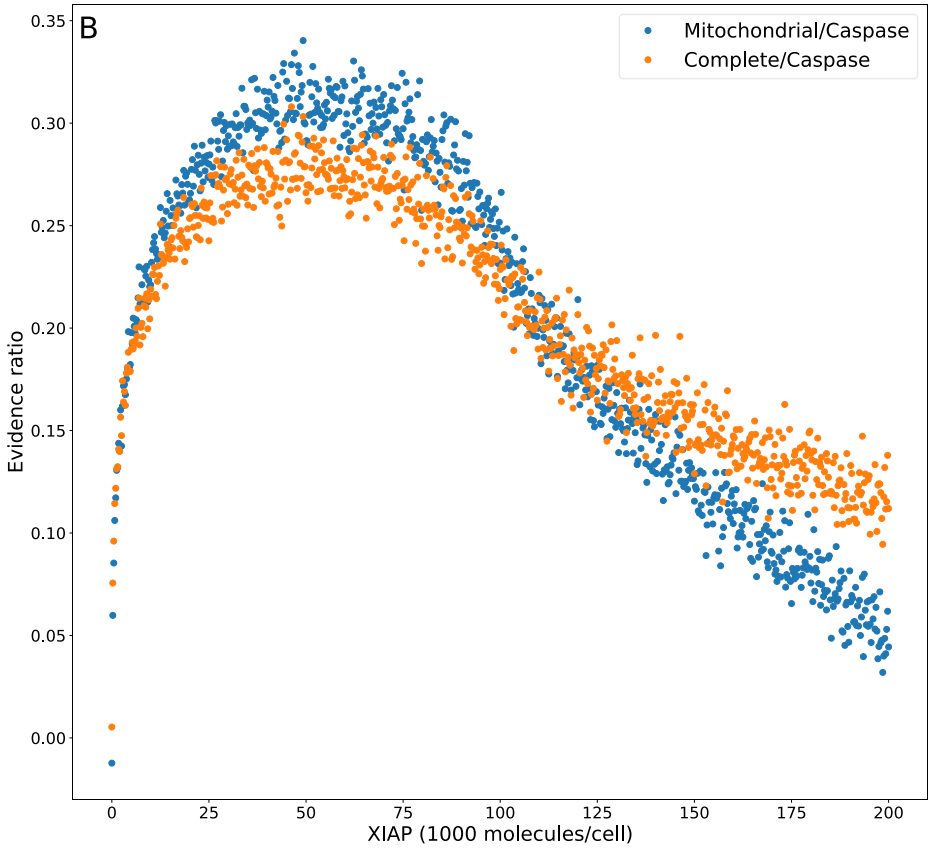
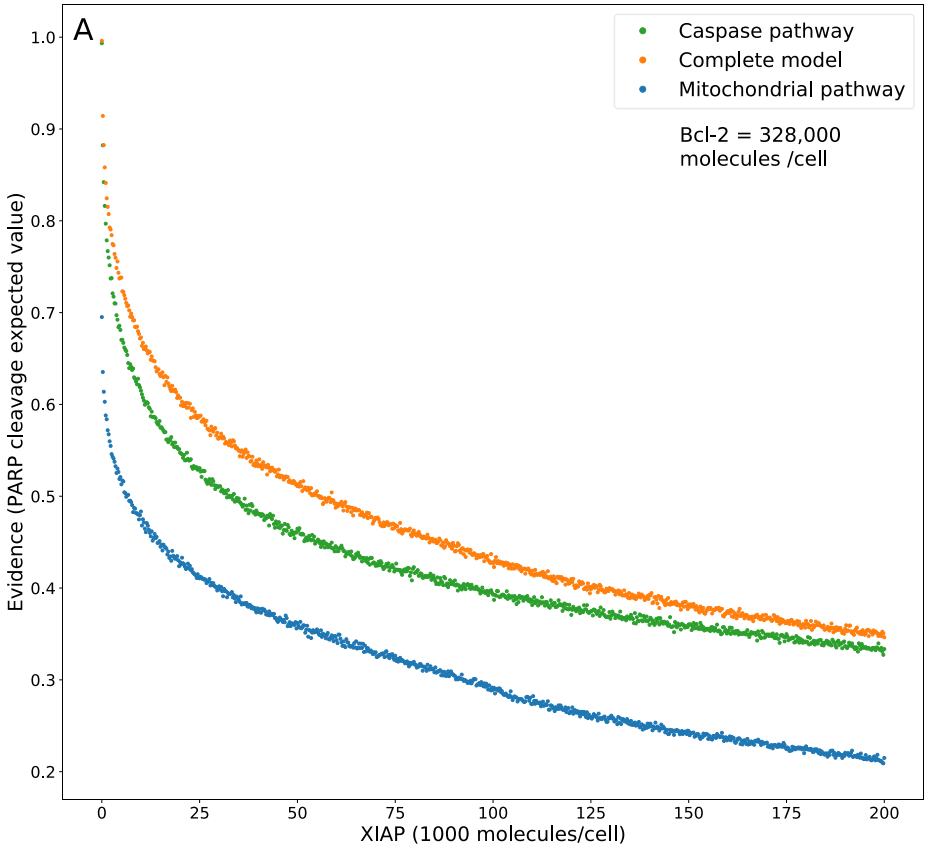

B
Mitochondrial/Caspase
Complete/Caspase
Evidence ratio
XIAP (1000 molecules/cell)

Figure 3.8. Evidence ratio calculations with no MOMP inhibition at increasing concentrations of XIAP. (A) Trace of expected PARP cleavage for the caspase pathway (green), mitochondrial pathway (blue), and complete network (orange). (B) Ratios of expected PARP cleavage from (A): mitochondrial/caspase (blue), complete/caspase (orange).

A common technique to study apoptosis is to knockdown Bid, overexpress Bcl-2, or otherwise shut down MOMP induced apoptosis through the mitochondrial pathway. This strategy was used in Jost et al. [71] to study the role of XIAP in apoptosis and in the work of Aldridge et al. [87]. Taking a similar approach, we set Bcl-2 levels to 328,000 molecules per cell, in line with experimental findings [78], to suppress MOMP activity and recreated the evidence and evidence ratio trends (Figures 3.9A and 3.9B). Under these conditions the evidence trend for the mitochondrial pathway drops well below that of the caspase pathway which is reflected in the Bayes factor trend as a shift into negative territory, an indication that the caspase pathway is favored. The evidence trend for the complete network under MOMP inhibition is shifted closer to that for the caspase pathway but continues to be more likely to execute apoptosis throughout the range of XIAP. The peak for the associated Bayes factor trend is flattened by roughly two-thirds suggesting increased XIAP levels are less likely to induce a transition to a Type II phenotype in a system with an already hampered mitochondrial pathway. Inhibition of MOMP to the point of annihilating any contribution from the mitochondrial pathway would result in uninformative mitochondrial pathway evidence values and a trend in mitochondrial/caspase ratio that is simply an inverted reflection of the caspase only evidence trend. The evidence trend for the complete network would be indistinguishable from that for the caspase pathway alone and the complete/caspase ratio trend would simply flatline. Isolation of active biologically relevant subnetworks and direct comparison under changing conditions using trends in Bayesian evidence enables the extraction of information regarding the pathway interactions and differential network dynamics.


A
Caspase pathway
Complete model
Mitochondrial pathway
Bcl-2 = 328,000
molecules /cell
Evidence (PARP cleavage expected value)
XIAP (1000 molecules/cell)

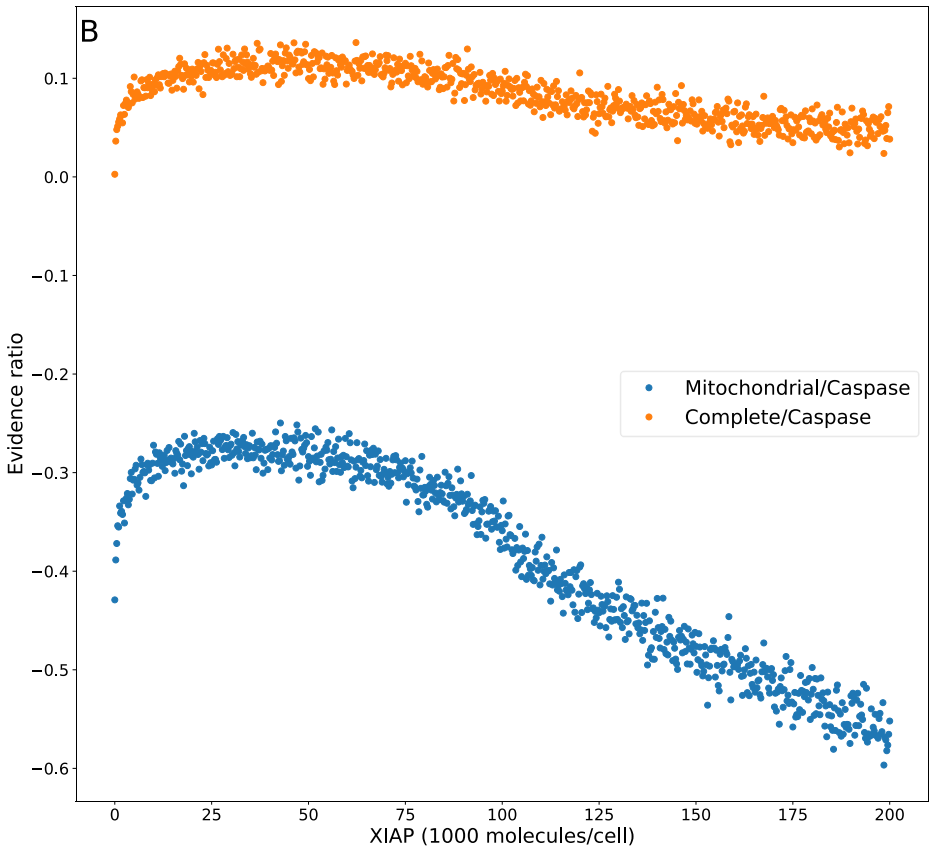

B
Mitochondrial/Caspase
Complete/Caspase
Evidence ratio
XIAP (1000 molecules/cell)

Figure 3.9. Evidence ratio calculations with MOMP inhibited by Bcl-2 at 328,000 molecules per cell and at increasing concentrations of XIAP. (A) Trace of expected PARP cleavage for the caspase pathway (green), mitochondrial pathway (blue), and complete network (orange). (B) Ratios of expected PARP cleavage from (A): mitochondrial/caspase (blue), complete/caspase (orange).

3.4.5 Precision vs Computational Cost

Increasing the precision of the evidence estimates, and tightening the evidence trendlines, is accomplished by increasing the number of live points in the nested sampling algorithm. The trade-off is an increase in the number of evaluations required to reach the termination of the algorithm and an accompanying increase in total computation time. Figures 3.10A and 3.10B display the required number of evaluations for the caspase pathway and complete network at population sizes of 500, 1000, 2000, 4000, 8000, and 16,000, when run with the PARP cleavage objective function. For both models the number of evaluations roughly doubles for every doubling in population size. Of note here is the higher number of required evaluations for the lower parameter model. The caspase pathway has only 22 parameters and required an average of 64,612 evaluations at a population size of 16,000 while the complete network, with its 56 parameters required only 53,652 evaluations, on average (Table 3.1). Figures 3.10C and 3.10D are the average estimated errors calculated by the MultiNest algorithm over each population size for the caspase and complete networks respectively. As expected, error estimates fall roughly as n^(-1/2) [45], signifying clear diminishing returns as the number of live points is increased. The average CPU process times, as estimated by Python's time.clock() method, are given in Figures 3.10E and 3.10F for the caspase and complete networks respectively. Despite the greater number of required evaluations for the caspase network the average clock times for the complete network is significantly higher. At a population of 16,000 the caspase network had an average clock time of 11,964 seconds compared to 76,981 for the complete network.

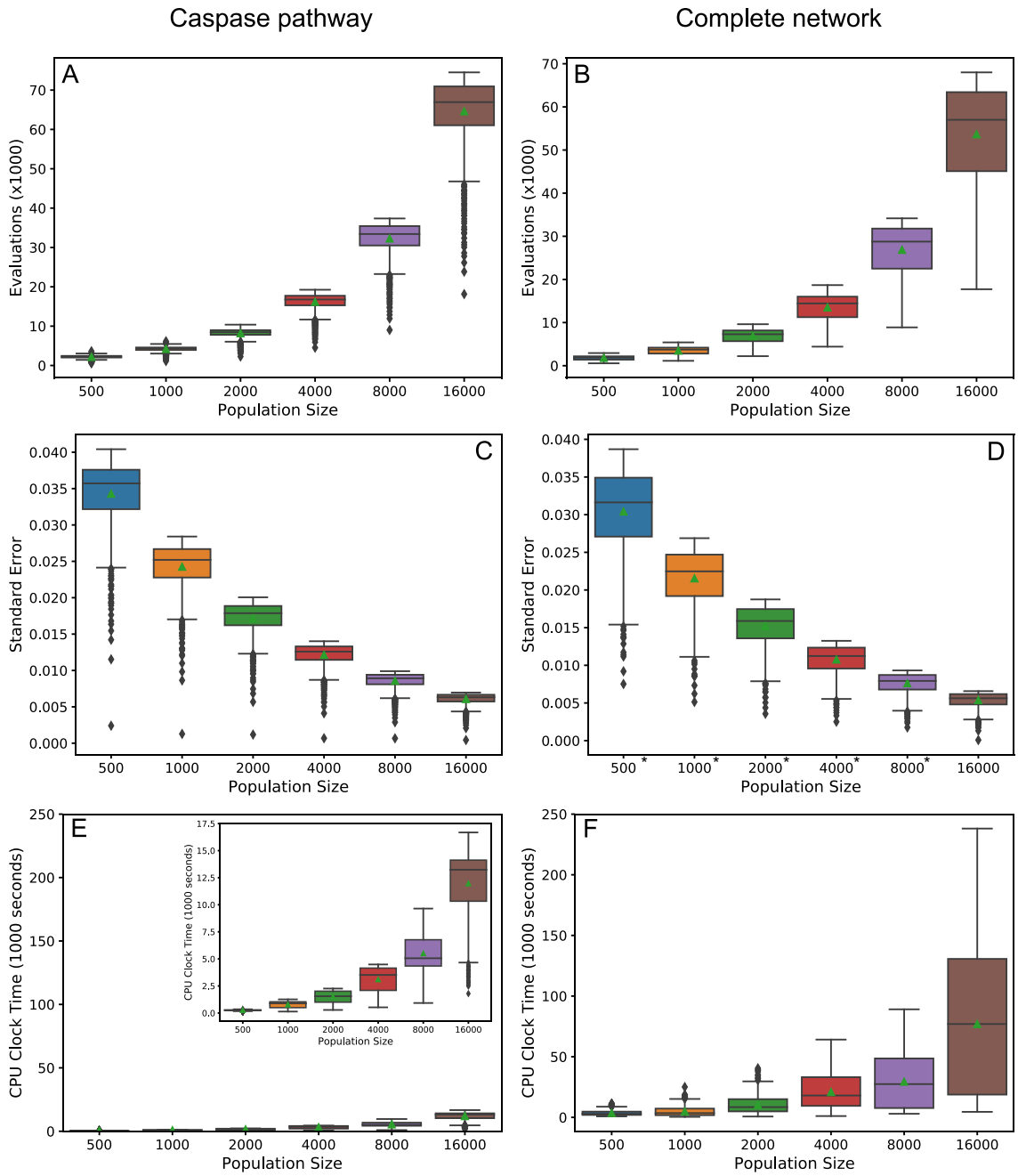
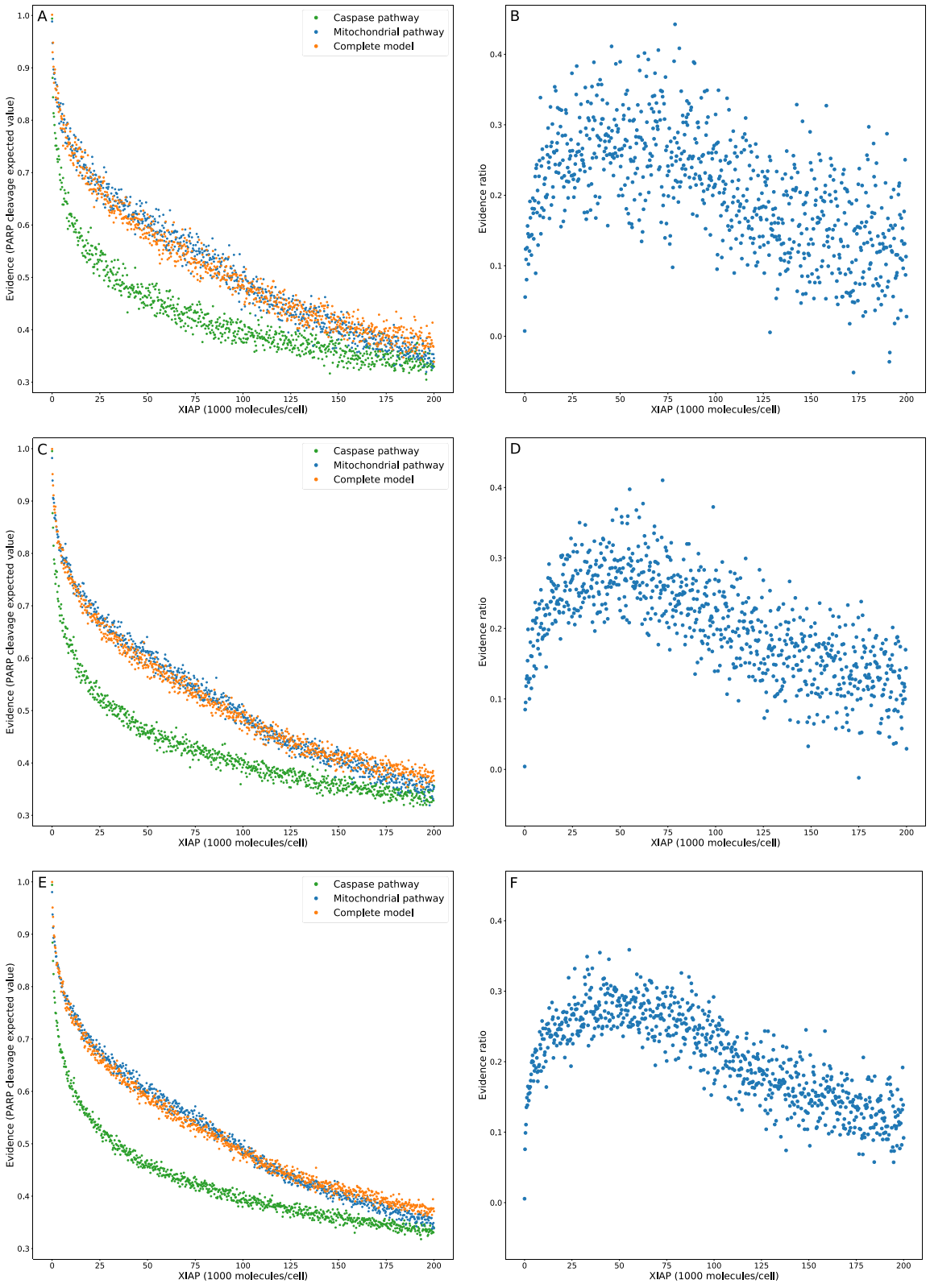


Figure 3.10. Precision vs. computational cost. (A) and (B) Average number of evaluations before termination of the MultiNest algorithm over a range of population sizes for the caspase pathway and complete network respectively. (C) and (D) Average of error estimates from MultiNest for each population size and the caspase and complete networks. (E) and (F) Average estimated CPU clock time over each population size for the caspase and complete networks respectively. *MultiNest was unable to estimate the error at XIAP = 0.

Table 3.1. Averages for the standard error, clock time, and number of evaluations before termination for different nested sampling population sizes.

Average standard error		Average clock time		Average evaluations	
live points	Caspase	live points	Caspase	live points	Caspase
500	0.0342985	500	255.2357428	500	2189.952559
1000	0.0242612	1000	769.5212609	1000	4200.756554
2000	0.0171671	2000	1484.313521	2000	8199.962547
4000	0.0121464	4000	3101.324894	4000	16171.84395
8000	0.008586	8000	5499.347441	8000	32270.97628
16000	0.006073	16000	11963.5571	16000	64612.04494
live points	Mitochondrial	live points	Mitochondrial	live points	Mitochondrial
500	0.0306757	500	2112.236401	500	1775.50437
1000	0.0217125	1000	4136.864189	1000	3424.536829
2000	0.015362	2000	8130.265223	2000	6710.838951
4000	0.0108627	4000	16362.12229	4000	13273.40075
8000	0.0076846	8000	33172.76319	8000	26401.47191
16000	0.0054326	16000	38408.39177	16000	52729.67915
live points	Complete	live points	Complete	live points	Complete
500	0.0304191	500	3583.213421	500	1805.554307
1000	0.0215387	1000	4863.895905	1000	3499.827715
2000	0.0152105	2000	10403.40514	2000	6836.940075
4000	0.0107683	4000	20917.64258	4000	13509.73658
8000	0.0076136	8000	29435.60684	8000	26868.33583
16000	0.0053777	16000	76980.86868	16000	53651.61798

Ultimately, the choice of population size for the methods we have laid out here will depend on the networks to be compared, the objective function, and how well the evidence trends must be resolved in order to make inferences about network dynamics. For example, at a population size of 500 the evidence trend for the caspase pathway is clearly discernable from the mitochondrial pathway and the complete network, but the latter two are largely overlapping (Figure 3.11A). At higher population levels, however, two distinct mitochondrial and complete trends become apparent (Figure 3.11K). If Bayes factor trends are desired then the choice of population size must take into consideration the amplification of the noise from both trends (see Figures 3.11(B, D, F, H, J, L) for complete/caspase Bayes factor trends).


A
Caspase pathway
Mitochondrial pathway
Complete model
Evidence (PARP cleavage expected value)
XIAP (1000 molecules/cell)
B
Evidence ratio
XIAP (1000 molecules/cell)
C
Caspase pathway
Mitochondrial pathway
Complete model
Evidence (PARP cleavage expected value)
XIAP (1000 molecules/cell)
D
Evidence ratio
XIAP (1000 molecules/cell)
E
Caspase pathway
Mitochondrial pathway
Complete model
Evidence (PARP cleavage expected value)
XIAP (1000 molecules/cell)
F
Evidence ratio
XIAP (1000 molecules/cell)

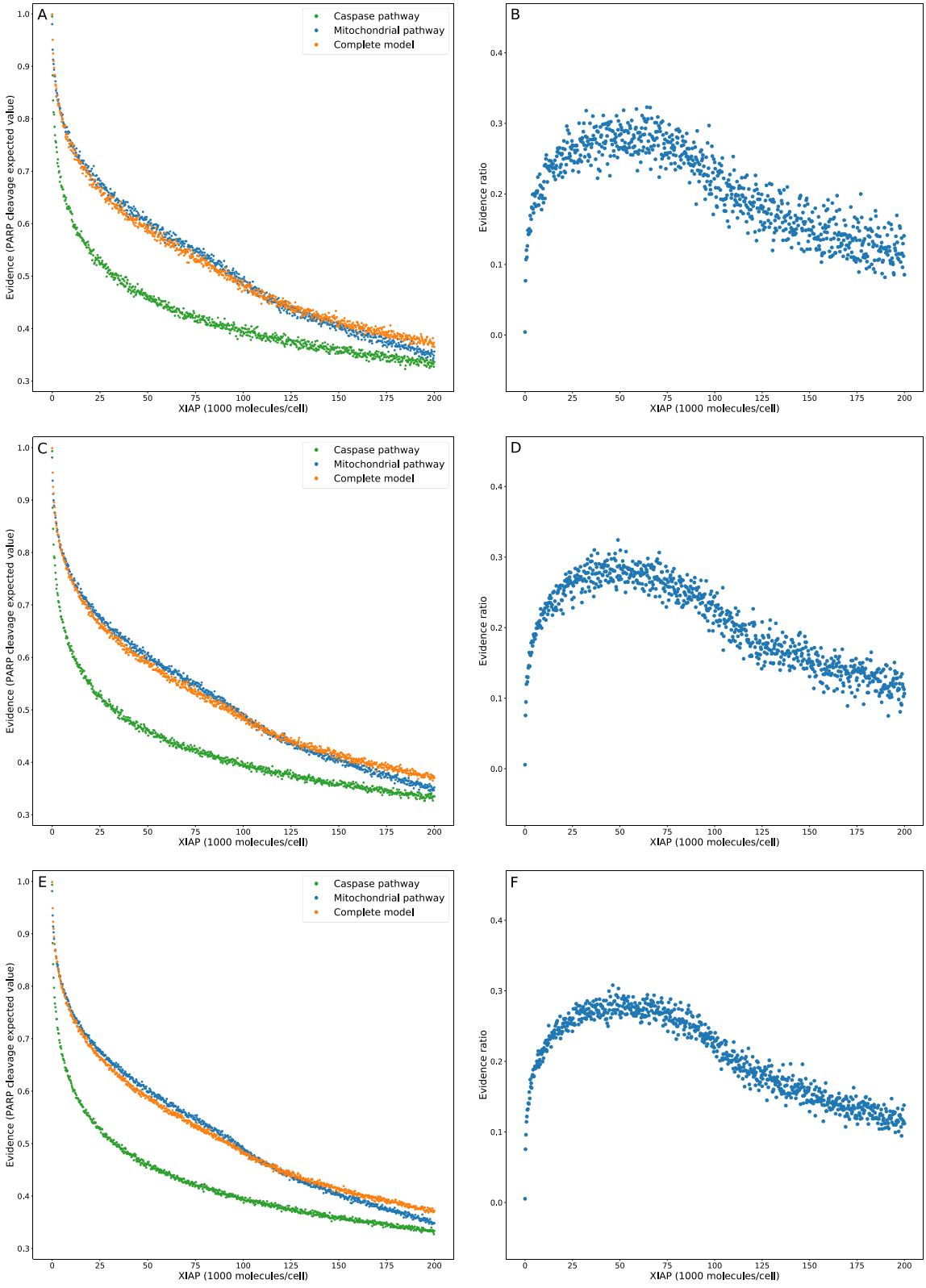

A
Caspase pathway
Mitochondrial pathway
Complete model
Evidence (PARP cleavage expected value)
XIAP (1000 molecules/cell)
B
Evidence ratio
XIAP (1000 molecules/cell)
C
Caspase pathway
Mitochondrial pathway
Complete model
Evidence (PARP cleavage expected value)
XIAP (1000 molecules/cell)
D
Evidence ratio
XIAP (1000 molecules/cell)
E
Caspase pathway
Mitochondrial pathway
Complete model
Evidence (PARP cleavage expected value)
XIAP (1000 molecules/cell)
F
Evidence ratio
XIAP (1000 molecules/cell)

Figure 3.11. Evidence and evidence ratio plots over a range of XIAP concentrations at increasing population levels for the nested sampling algorithm. (A), (C), (E), (G), (I), and (K) Evidence plots for the caspase pathway (green), mitochondrial pathway (blue), and complete network (orange) with nested sampling population levels of 500, 1000, 2000, 4000, 8000, and 16,000 respectively. (B), (D), (F), (H), (J), and (L) Complete/caspase Bayes factor plots derived from the respective evidence plots in (A), (C), (E), (G), (I), and (K).

Chapter 4

Bayesian Evidence-Based Analysis of the Extrinsic Apoptosis Signaling Network

4.1 Summary

This chapter serves as a follow-up application chapter to chapters 2 and 3. It extends the probabilistic analysis of the extrinsic apoptosis system. Apoptosis, a type of programmed cell death, is essential to normal development and homeostatic but also implicated in numerous diseases when misregulated. In many cancer types, for example, cells are, or become, resistant to apoptosis and overcoming this resistance has been the focus of much recent research. Thus, it is vitally important not only to identify key regulators of apoptosis but to predict the phenotypic changes they induce. Unfortunately, making such predictions in rigorous fashion is challenging with current computational methods. Here, we extend our previous analysis of the extrinsic apoptosis system using the methods described in Chapter 2 and applied in Chapter 3. In short, we calculate the Bayesian evidence over a proposed distribution of parameters for an objective function that represents some quantity of interest. This is equivalent to calculating the expected value of that quantity given what we know about the system. The expected value is calculated over a range of initial conditions for selected regulatory proteins and for various in silico experimental setups, allowing for exploration of network dynamics, differential effects on different subnetworks, and possible explanations of phenotypic control. We apply these methods to three well studied regulatory axes of extrinsic apoptosis: initiation of the apoptotic signal at the death inducing signaling complex (DISC), regulation of Caspase-3 by the apoptosis regulator XIAP, and regulation of mitochondrial outer membrane permeabilization by the family of Bcl-2 proteins. Where available we verify our results with existing experimental evidence.

4.2 Introduction

Apoptosis is a type of programmed cell death and a critical process for normal development and homeostasis [94]. Its importance is made manifest by its role in numerous illnesses including cancer and neurodegenerative and autoimmune diseases [95, 96]. The role of apoptosis in cancer has been a particularly prominent research topic. Evasion of apoptosis by cancer cells, often by upregulation of antiapoptotic regulators like Bcl-2, is thought to be responsible for the resistance to chemotherapy/radiotherapy, as these treatments act by initiation of apoptotic pathways [97, 98]. This has led to an abundance of research identifying the regulatory and effector elements of apoptosis and targeting them to initiate or prime an apoptotic response [66, 99, 100]. Although this has led to the discovery of much of the apoptotic regulatory machinery, the phenotypic response to changes in those regulators has not been fully defined.

Alterations in cellular phenotype due to changes in regulatory conditions are typically difficult to predict. Such behavior emerges from the interactions between many components of a system and variations in expression of one, or a few, regulatory elements can have complex effects that result in unexpected outcomes [1, 2]. Characterization of systems with such complexity via experimental methods is an arduous and expensive task. Hence, computational simulation methods are often used in conjunction with experimentation to better advance the understanding of these systems [3]. A particularly important topic with regard to apoptotic phenotype concerns the conditions under which apoptosis is independent of the mitochondria (Type I phenotype) or dependent on it (Type II) [71, 93, 101, 102, 103]. Computational studies on this question include the work of Aldridge et al. using Lyapunov exponents and bifurcation diagrams, and Raychaudhuri and Raychaudhuri using Monte Carlo-based methods [87, 88]. Unfortunately, confidence in computational predictions requires a precision in model parameter values that is often lacking.

In the previous chapters we devised a probabilistic approach to the computational analysis of such complex systems. We used Bayesian methods from the field of model selection and multimodel inference, in particular nested sampling, to compare the effect of changes

in the apoptosis regulator XIAP on various pathways and processes of the extrinsic apoptosis system. Nested sampling is designed to estimate the model evidence, or fit to data, given a likelihood function for a statistical model. (See Chapters 2 and 3) The evidence estimate is calculated by integrating the likelihood function over the prior distribution. For the mechanistic apoptosis model, we used an objective function representing apoptosis and plausible parameter ranges for biological reactions. In short, instead of calibrating the parameters of a model and predicting quantitative values for targets within the system, we integrate out the parameters and produce expected values for those targets.

In this follow-on work we have expanded the application of these methods on three major regulatory axes of apoptosis to further investigate targets of control over the modes of apoptotic execution (Figure 4.1). First we examine various conditions that favor a Type I or II phenotype, or regulatory changes that could transition a cell from one phenotype to the other. Two mechanistic processes are examined in this case: death-inducing signaling complex (DISC) driven activation of apoptotic signaling, and the interplay between the apoptotic regulator XIAP and its target Caspase-3. The DISC consists of a ligand bound receptor from the tumor necrosis factor (TNF) family, a scaffold protein such as FADD, and Caspase-8, which is recruited to the complex and cleaved into an active form that propagates the apoptotic signal [52-54]. XIAP binds and ubiquitinates Caspase-3, which conveys the apoptotic signal via cleavage of the protein PARP, marking it for degradation and permanently deactivating it [72, 73]. For a Type II phenotype, amplification via signaling through the mitochondria is paramount. Thus, mitochondrial outer membrane permeabilization (MOMP) is the third regulatory axes studied here. MOMP, governed by the Bcl-2 family of proteins, amplifies and propagates the apoptotic signal by allowing the release of proapoptotic factors from the mitochondria [62, 104]. Many pro- and antiapoptotic Bcl-2 proteins have been identified. Here we analyze the effects of the antiapoptotic proteins Bcl-2, Bcl-xl and Mcl-1, as well as the inhibitors of those proteins Bad and Noxa [63-68].

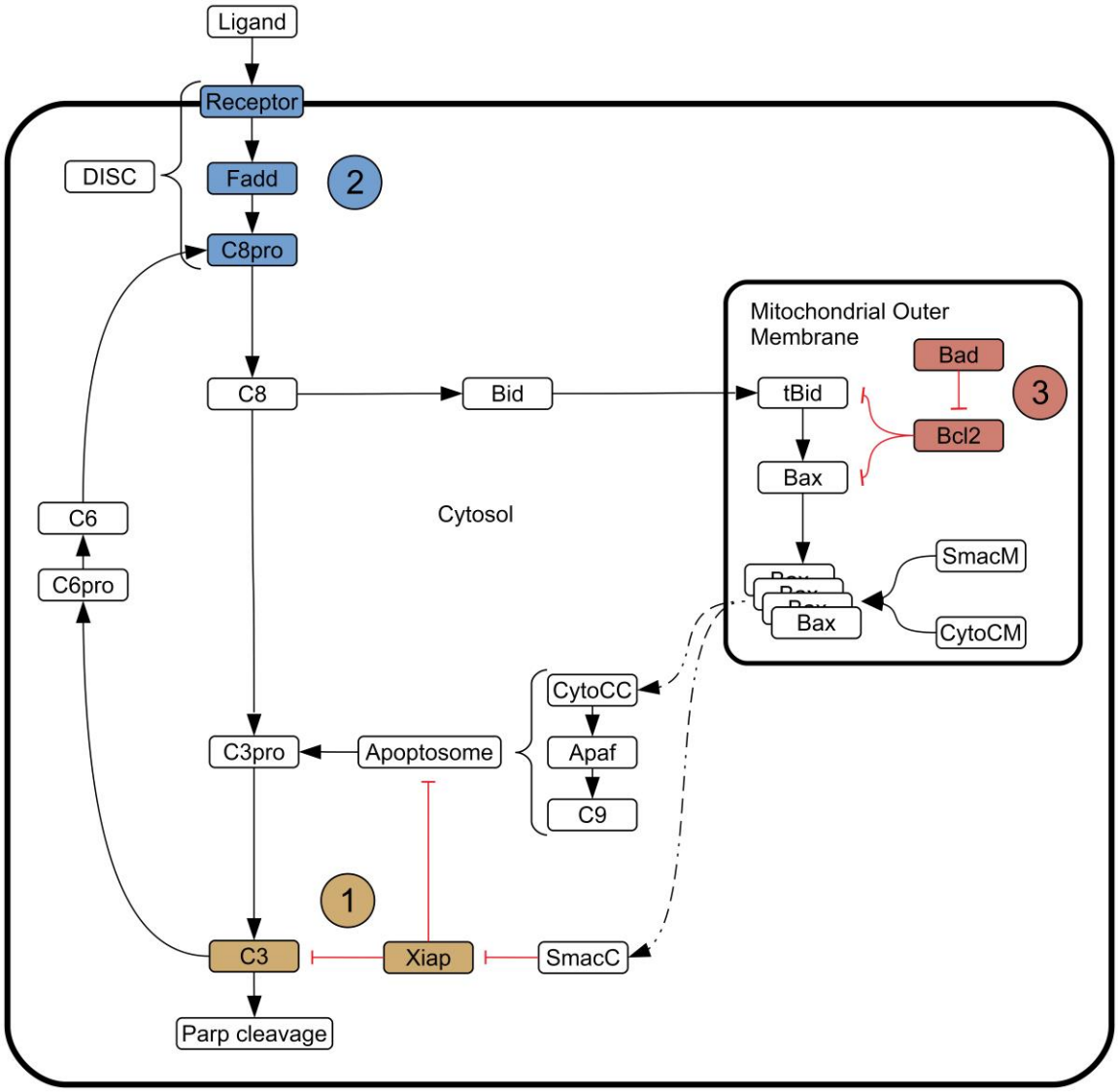

Ligand
Receptor
DISC
Fadd
2
C8pro
C8
Bid
Mitochondrial Outer Membrane
Bad
3
tBid
Bcl2
Bax
C6
Cytosol
C6pro
SmacM
Bax
CytoCM
CytoCC
C3pro
Apoptosome
Apaf
C9
1
C3
Xiap
SmacC
Parp cleavage

Figure 4.1. Schematic of the extrinsic apoptosis model and axes of regulation. A full description of the model can be found in Chapter 2. Here we focus on three major regulatory axes of this network, axes that have all been the subject of previous experimental and/or computational studies. ① We first consider the interplay between the apoptosis regulator XIAP and its target, the apoptosis effector Caspase-3. The ratio of these two components have previously been investigated as potentially determining the apoptotic phenotype of the cell. ② Next we consider components of the DISC including the TNF receptor, the scaffold protein FADD, and the initiator Caspase-8. Weak activation of Caspase-8 is thought to necessitate amplification of the signal via the mitochondrial pathway [93]. Receptor expression in particular has been implicated as the DISC component responsible for this phenotype. ③ Lastly, we consider regulation of MOMP and the Bcl-2 family components that control it. Here we primarily examine the effects of variations in expression levels of antiapoptotic proteins such as Bcl-2 that bind to and inhibit Bcl-2 family effectors of MOMP, and proapoptotic proteins such as Bad that bind and inhibit the antiapoptotic components.

4.3 Results

The general strategy in this chapter closely parallels that found in Chapter 3. In a multi-model inference approach the extrinsic apoptosis network is broken down into relevant subnetworks and the Bayesian evidence for each is estimated with respect to an objective function that represents successful apoptosis over a range of regulatory conditions. Here the proportion of PARP that is cleaved at the end of a simulation is our proxy for a successful apoptotic signal that will result in cell death (see Section 2.2). As regulatory conditions change each subnetwork will respond in a unique way, produce distinct trends in the Bayesian evidence for apoptosis, and provide insights into the regulatory control of apoptosis. In chapter 3 we focused primarily on the role of the apoptosis regulator XIAP in signal transduction and determination of apoptotic phenotype. In this work we broaden that scope and focus on three purported regulatory axes of apoptosis, all of which have been studied experimentally. First, we again consider the regulator XIAP but this time in conjunction with its target Caspase-3. The ratio of these two network components has been previously suggested as the key factor in determining the phenotype of the cell [87]. A cell with a Type I phenotype is able to achieve apoptosis after knockdown of the mitochondrial pathway while a Type II phenotype depends on it. It thus suffices to compare the Bayesian evidence (expected values for indicators of apoptosis) for the complete network to that of the caspase only pathway when testing the effects of regulatory perturbation on cellular phenotype. Another regulatory axis thought to govern the Type I/II apoptotic phenotype of a cell is the DISC. A weak initial apoptotic signal from the DISC is thought to require amplification of that signal from the mitochondrial pathway leading to a Type II phenotype [93]. In this case the Bayesian evidence for the caspase and complete networks are compared under changing concentrations of the DISC components with a particular focus on the concentration of the receptor [105]. The last regulatory axis we consider is at the mitochondria itself. Amplification of the apoptotic signal is effected by the export of proapoptotic factors across the mitochondrial membrane. This process is governed by members of the Bcl-2 family of proteins. Here we vary pro- and antiapoptotic members of that family

and evaluate the changes in Bayesian evidence for achieving apoptosis using the complete network. In addition to the multimodel inference approach a pathway targeted approach was used to estimate the signal flux through the caspase and mitochondrial under varying conditions at the DISC regulatory axis, providing a complementary view of signal transduction that incorporates cross-talk between pathways (see Section 2.1).

4.3.1 XIAP/Caspase-3 Concentration Ratios and Type I/II Phenotype Outcome

The role of XIAP in determining the Type I/II apoptotic phenotype of a cell was experimentally examined in the work of Jost et al. [71]. They studied the differences between Type II hepatocytes and Type I lymphocytes under various regulatory conditions and made several discoveries, such as opposing directions of XIAP levels in the two cell types after FasL treatment, that led to the conclusion that XIAP was the key regulator of apoptotic phenotype. In addition to the finding for XIAP alone, Jost et al. also found that XIAP was stabilized via association with active Caspase-3, but that the activation of Caspase-3 in wild type hepatocytes was substantially less than in wild type or Bid deficient thymocytes, and non-existent in Bid deficient hepatocytes. From these observations they suggested that the ratio of XIAP to effector caspases, both active and inactive, may govern the Type I/II choice of apoptotic phenotype. Aldridge et al. picked up on this suggestion with their Lyapunov exponent and phase diagram analysis on the XIAP:Caspase-3 ratio and how it affects type I/II phenotype determination [87]. They applied the method on a Bcl-2 overexpressed apoptosis model over a range of initial values for XIAP and Caspase-3 and defined a dividing line, called a separatrix, within the phase diagram that determined the boundary between the Type I and Type II phenotypes. They concluded that a high XIAP:Caspase-3 ratio implied a preference for the Type II phenotype and low ratios implied a preference for Type I. They further supported this finding by showing that several cell lines, of Type I or II phenotype, had XIAP:Caspase-3 ratios that placed them correctly on the phase diagram.

To test this hypothesis using the Bayesian evidence methods laid out in Chapter 2 we applied the MultiNest nested sampling algorithm and our custom objective functions to the caspase pathway and the complete network models for extrinsic apoptosis over a range of concentrations for both XIAP and Caspase-3. The initial value for XIAP was varied from 0 to 100,000 molecules per cell in increments of 4000 while the value for Caspase-3 was varied from 2500 to 400,000 in increments of 2500. The population of parameter sets was set to a size of 32,000.

Analyzing an extrinsic apoptosis model where Bcl-2 has been overexpressed to the point of entirely shutting down the mitochondrial pathway is equivalent to analyzing a model of the caspase cascade in isolation. The evidence landscape for PARP cleavage on this model is displayed in Figure 4.2A-B and the trends over increasing values of XIAP and Caspase-3 are clear and expected. Increasing Caspase-3, the effector caspase for PARP cleavage, increases the evidence for achieving apoptosis, and increasing XIAP, the inhibitor of Caspase-3, decreases it. Thus, a lower XIAP:Caspase-3 ratio implies a higher probability that the isolated caspase pathway, representative of a Type I phenotype, will achieve apoptosis and a higher ratio implies a lower probability. If we assume that a lower probability of Type I effected apoptosis implies a higher probability for Type II, then the caspase pathway evidence landscape is analogous to the phase diagram in Aldridge et al. [87].

Unfortunately, the complete network has a qualitatively identical evidence landscape where higher and lower XIAP:Caspase-3 ratios imply a lower and higher probability of achieving apoptosis respectively (Figure 4.3A-B). In other words, a network with a fully active mitochondrial pathway is affected in exactly the same manner as a network without it. This should not be surprising, as both pathways converge at Caspase-3 and will be inhibited by XIAP regardless of which apoptotic route is taken.

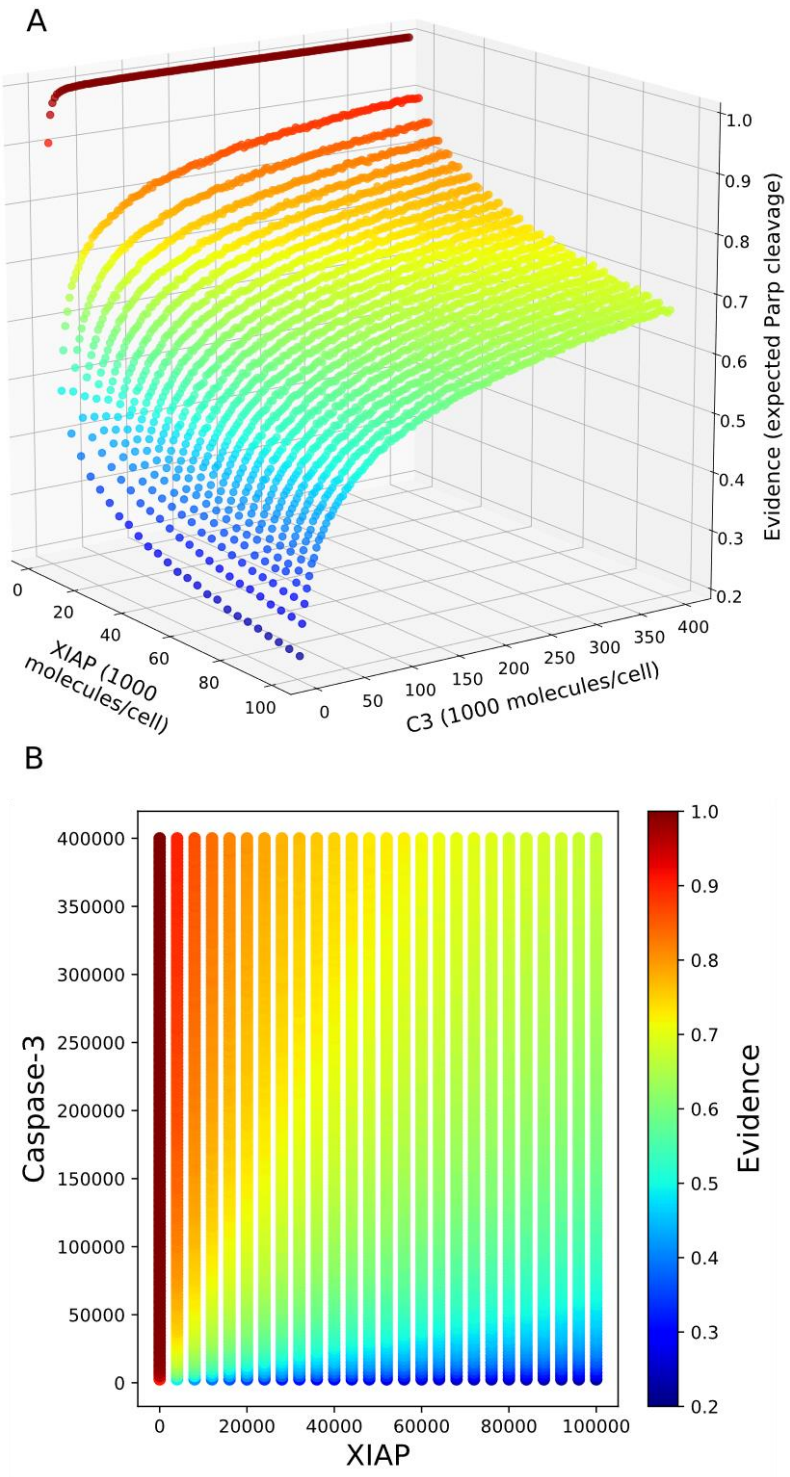


Figure 4.2. XIAP vs Caspase-3 (C3) evidence values for the caspase pathway over a range of XIAP and Caspase-3. XIAP was varied from 0 to 100,000 in increments of 4000 molecules per cell and Caspase-3 was varied from 2500 to 400,000 in increments of 2500. (A) 3D plot of the evidence values (B) 2D plot for the same values.

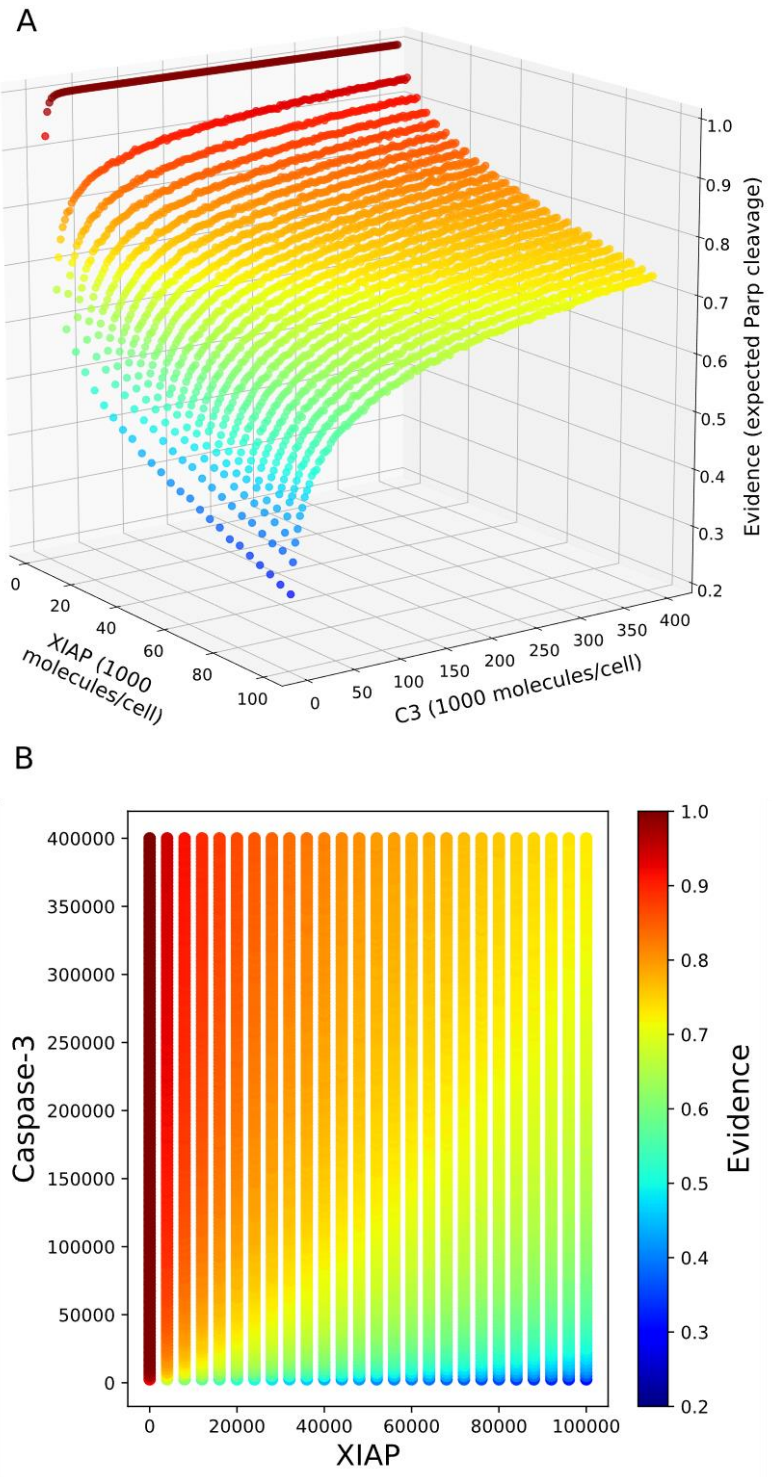


Figure 4.3. XIAP vs Caspase-3 (C3) evidence values for the complete network over a range of XIAP and Caspase-3. Again, XIAP was varied from 0 to 100,000 in increments of 4000 molecules per cell and Caspase-3 was varied from 2500 to 400,000 in increments of 2500. (A) 3D plot of the evidence values (B) 2D plot for the same values.

Nonetheless, there are additional regulatory elements in the complete network, such as Smac inhibition of XIAP, that lead to subtle differences in the shape of these evidence landscapes. At low levels of Caspase-3 we see the same pattern as in section 3.4.4 above; increased XIAP, from low levels, inhibits the caspase pathway more than the complete network, but this inhibition levels off more quickly. And like in that previous work this is reflected in the evidence ratio landscape as an initial spike, indicating an increasing chance of achieving apoptosis through the complete network, and an eventual reversal (Figure 4.4A-B). As the Caspase-3 levels increase, the advantage to the complete network provided by XIAP is dampened and the pattern flattens out considerably. In general, increasing Caspase-3 levels appears to push the evidence ratio landscape toward parity (a Bayes factor level of zero) between the two networks for all XIAP levels above zero. At an XIAP level of zero the nested sampling algorithm quickly converges to parameter sets that result in full PARP cleavage and results in evidence ratios near zero for most levels of Caspase-3, an uninformative edge case. If the ratio between XIAP and Caspase-3 served as a major factor in determination of apoptotic phenotype we would expect to see resistance to apoptotic inhibition from a high XIAP:Caspase-3 ratio in the complete network, with its fully active mitochondrial pathway, relative to the Caspase-3 only pathway. We do not see this under the conditions used here. There is no discernable shift in the evidence ratios that would indicate a role for the XIAP:Caspase-3 ratio in Type I/II determination (Figure 4.4B). Note that cell types with high and low XIAP/Caspase-3 ratios would still give preference to Type II and I phenotypes respectively, despite appearing not to be a deciding factor. In all, these results suggest that low levels of Caspase-3 may leave the Type II pathway as the only route to apoptosis and in such a case, XIAP acts to modulate the response with moderate levels of XIAP being ideal for a Type II phenotype.

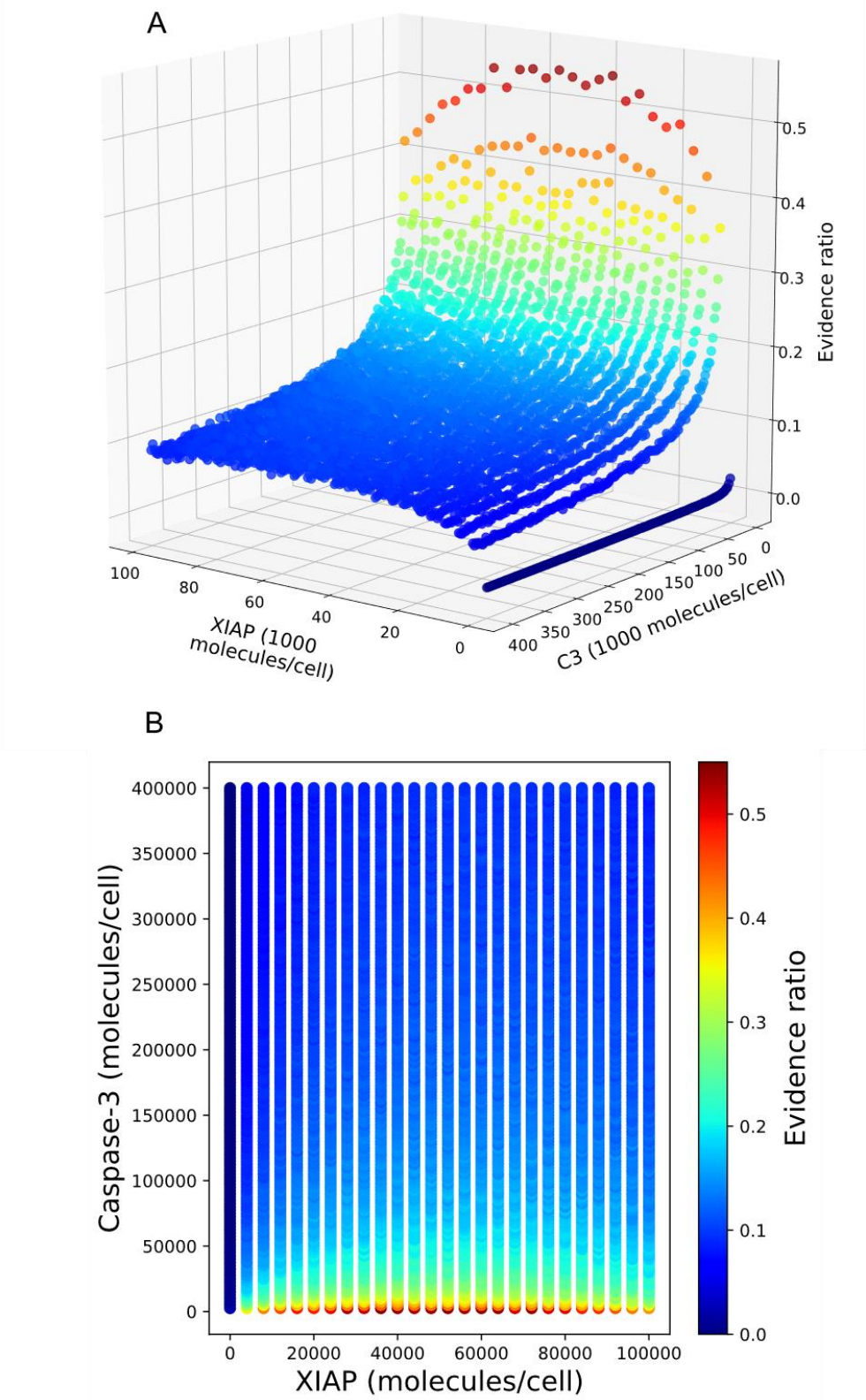


Figure 4.4. Evidence ratios (complete/caspase) from the values in Figures 4.2 and 4.3 for XIAP vs Caspase-3 (C3). (A) 3D plot of the ratios (B) 2D plot for the same values.

4.3.2 Effect of Receptor Count and DISC Formation on Apoptosis Subnetworks

The discovery of the two-phenotype system for extrinsic apoptosis originated with Scaffidi et al. when they made a number of observations regarding differences between cell types they deemed Type I and II (SKW6.4/H9 and CEM/Jurkat cells respectively) [93]. They found rapid activation of Caspase-8 and -3 in Type I cell lines compared to Type II, Bcl-2 mediated apoptotic protection in Type II cells but not in Type I, and relatively weak DISC formation in Type II cells. Their work ultimately led to the hypothesis that restricted activation of Caspase-8 at the DISC required amplification through the mitochondria for apoptosis to proceed. Meng et al. studied the components of the DISC in Type I and II cell lines and found comparable expression of Caspase-8, Fadd, and cFlip between the two phenotypes, but a lower expression of cell surface Fas in Type II cells [105]. They hypothesized that this was the distinguishing characteristic between the two cell types. They evaluated this hypothesis by up- or downregulating the Fas receptor in Type II and I cells respectively; attempting to convert them from one phenotype to the other. For Type II to I conversion they transfected Type II Jurkat cells with either nothing (an empty vector), Fas, Bcl-xl, or Fas and Bcl-xl; and then stimulated each variation with CH.11 (an anti-Fas antibody). Only the Bcl-xl only transfected cells were protected. All others, including the Fas/Bcl-xl overexpressed cells, showed similar levels of cell death, quantified via microfluorimetry, indicating a Type I route to apoptosis. Conversely, they transfected Type I A498 cells with either an empty vector, Fas shRNA, Bcl-xl, or Fas shRNA and Bcl-xl and proceeded to show that the Fas shRNA transfected cells had gained protection from CH.11 when also overexpressing Bcl-xl; indicating conversion to a Type II phenotype.

To evaluate the hypothesis that receptor count dictates the apoptotic phenotype of the cell, we ran the nested sampling algorithm at a relatively high resolution (number of initial values) with receptor counts from 100 to 100,000 molecules per cell, in increments of 100, on both the caspase only and complete networks. High resolution was used here to better capture the faster changes seen in low receptor count cases, as we will see below. This was repeated on four backgrounds representing various levels of DISC formation. To attain a

general representation of DISC formation efficacy, Fadd and Procaspase-8 (C8), the other two components of the DISC, were set, in tandem, to levels of 100, 1000, 10,000, and 100,000 molecules per cell. We used a population size of 16,000 parameter sets to compute the evidence estimates. Note that computational resources will generally necessitate a balance, or trade-off, between precision, driven by population size, and resolution. Bcl-2 was absent here to allow for a fully active mitochondrial pathway and XIAP was set to 0 to accurately gauge the effect of lower receptor/DISC activity on shifting to a Type II pathway in isolation from other regulators that are purported to elicit the same effect. All log-evidence values and associated standard deviations can be found in Figure 4.8 at the end of this section.

The progressive reduction in DISC formation has clearly divergent effects on the two networks. At high levels of Fadd/Procaspase-8, both networks attain high evidence scores, indicating a high aggregate level of PARP cleavage, across the range of receptor counts (Figure 4.5A). As Fadd/Procaspase-8 levels are decreased the evidence values for the caspase pathway fall at a much greater rate (Figures 4.5B-D).

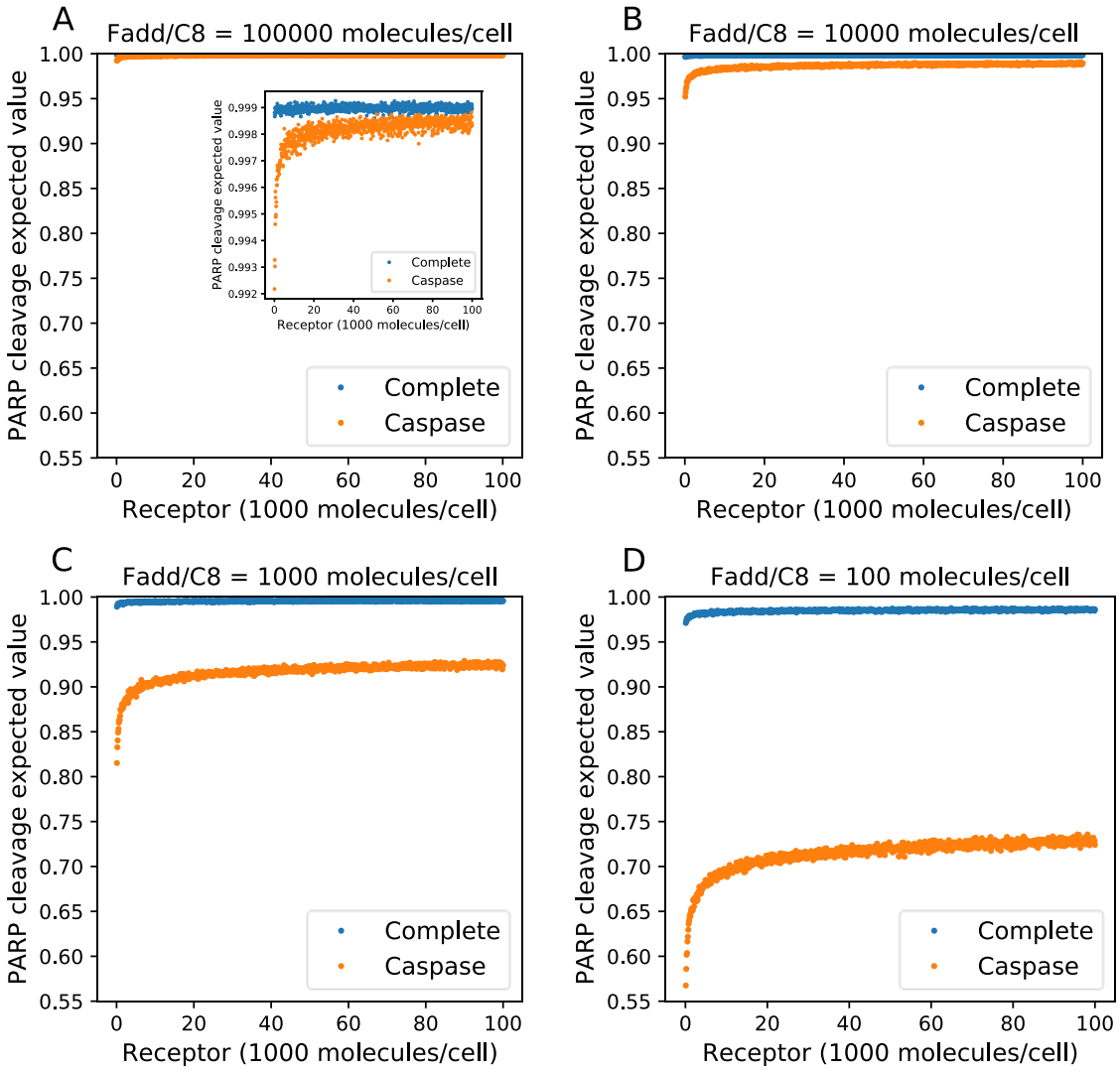


Figure 4.5. Receptor and Fadd/Procaspase-8 level effects on the complete network and caspase pathway evidence for achieving apoptosis (expected values of PARP cleavage). (A-D) Caspase (orange) and Complete (blue) expected PARP cleavage over receptor levels from 100 to 100,000 in increments of 100 and at Fadd/Procaspase-8 levels of (A) 100,000, (B) 10,000, (C) 1000, and (D) 100.

This is exemplified in the average difference in evidence values, across receptor counts, for sequential levels of DISC formation (Figure 4.6A). Reduction in receptor count has a similarly divergent effect with evidence values generally falling much faster for the caspase pathway. At high Fadd/Procaspase-8 levels receptor count reduction has little effect on the caspase pathway and no detectable effect on complete network. As Fadd/Procaspase-8 levels decrease, receptor count reduction produces a pronounced decrease in evidence values for the caspase pathway but a much more muted decrease for the complete network. In general, the decrease in evidence values accelerates as receptor counts fall, particularly at very low levels of receptor, an effect that is amplified by lower levels of Fadd/Procaspase-8. This results in a range of evidence values for the caspase pathway that broadens much faster than the complete network as Fadd/Procaspase-8 levels fall (Figure 4.6B).

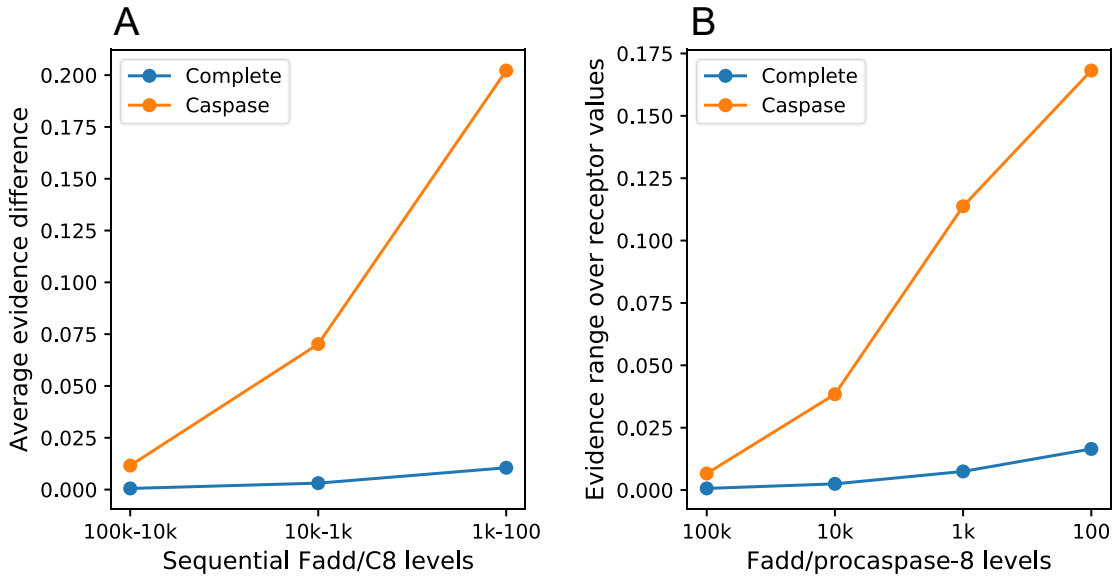


Figure 4.6. Statistics for the average evidence difference and the range of evidence values over increasing receptor concentrations for sequential/increasing Fadd/Caspase-8 levels. (A) The difference between evidence averages, over the receptor range, between successive levels of Fadd/Procaspase-8 for each pathway. (B) Evidence range over receptor levels for each pathway and at each level of Fadd/Procaspase-8.

The combined effect from the decrease in Fadd/Procaspase-8 and receptor is apparent in the Evidence ratio values (Figure 4.7). There is a clear shift higher, favoring the complete network, as Fadd/Procaspase-8 are decreased and a trend higher with decreasing receptor that is also amplified as Fadd/Procaspase-8 levels drop. The complete network, with its active mitochondrial pathway, consistently exhibits far greater resistance to downgraded apoptotic signaling as components of the DISC are reduced in concentration and the reduced signal for the caspase network model clearly accelerates as those components fall. These observations support the hypotheses that a general reduction in DISC formation efficiency, and a specific reduction in receptor count, favor the Type II pathway. The hypothesis put forward that the receptor is the deciding factor in the Type I/II phenotype decision is certainly plausible given these results. A further comparison of all three components individually might also be instructive. In addition, there appears to be a synergistic effect between the DISC components as implied by the amplified affect that receptor reduction has at lower Fadd/Procaspase-8 levels.

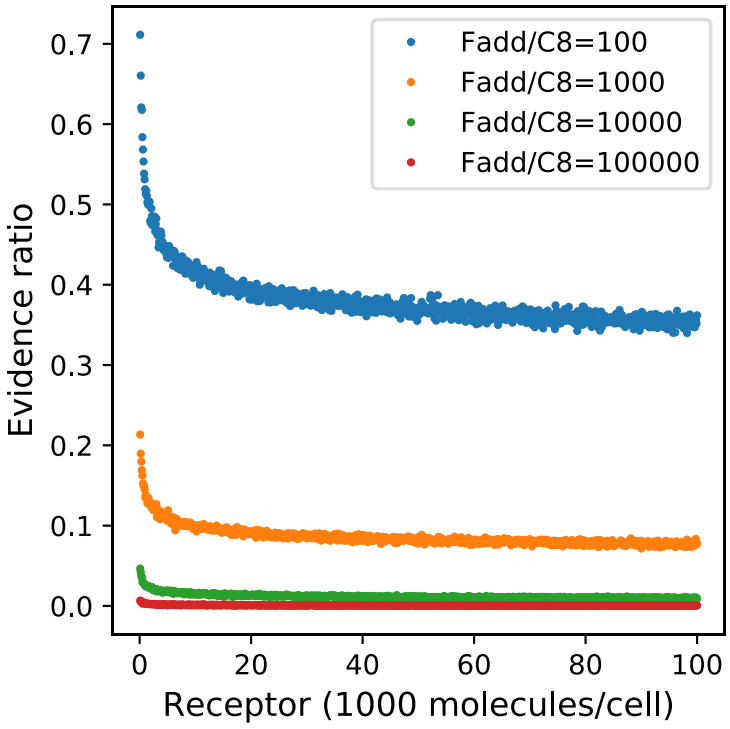


Figure 4.7. Evidence ratio values over the range of receptor concentrations at each level of Fadd/Procaspase-8: 100,000 (red), 10,000 (green), 1000 (orange), and 100 (blue). A higher ratio implies a preference for the Type II pathway.

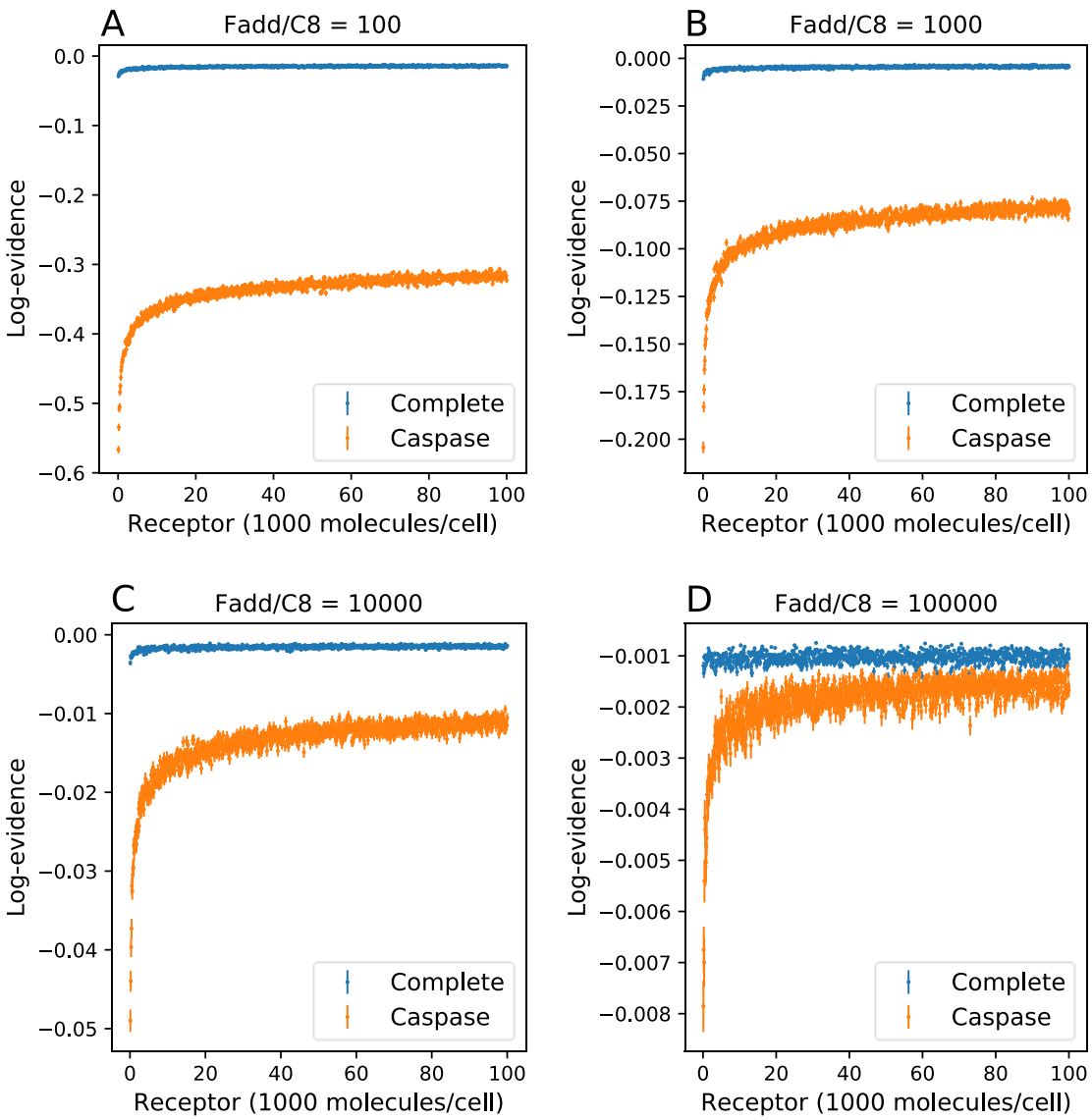


Figure 4.8. Associated log-evidence values with standard error bars for the evidence values in Figure 4.5.

4.3.3 Effect of Receptor Count and DISC Formation on Apoptosis Signal Flux

An analysis of the signal flux over increasing DISC formation and receptor counts clearly indicates a shift toward mitochondrial pathway signal route dominance as DISC components decrease (Figure 4.9), similar to the results in section 3.4.3. At a FADD/Caspase-8 level of 100 the mitochondrial pathway carries nearly 80% of the signal for most levels of receptor. The caspase pathway becomes dominant at a FADD/Caspase-8 level of 10,000 and receptor levels also around 10,000. Noteworthy is the symmetry between the two pathways and the nearly full signal throughput for all combinations of values tested. This further supports the hypothesis that the signal is equally effective for Type I and Type II cells.

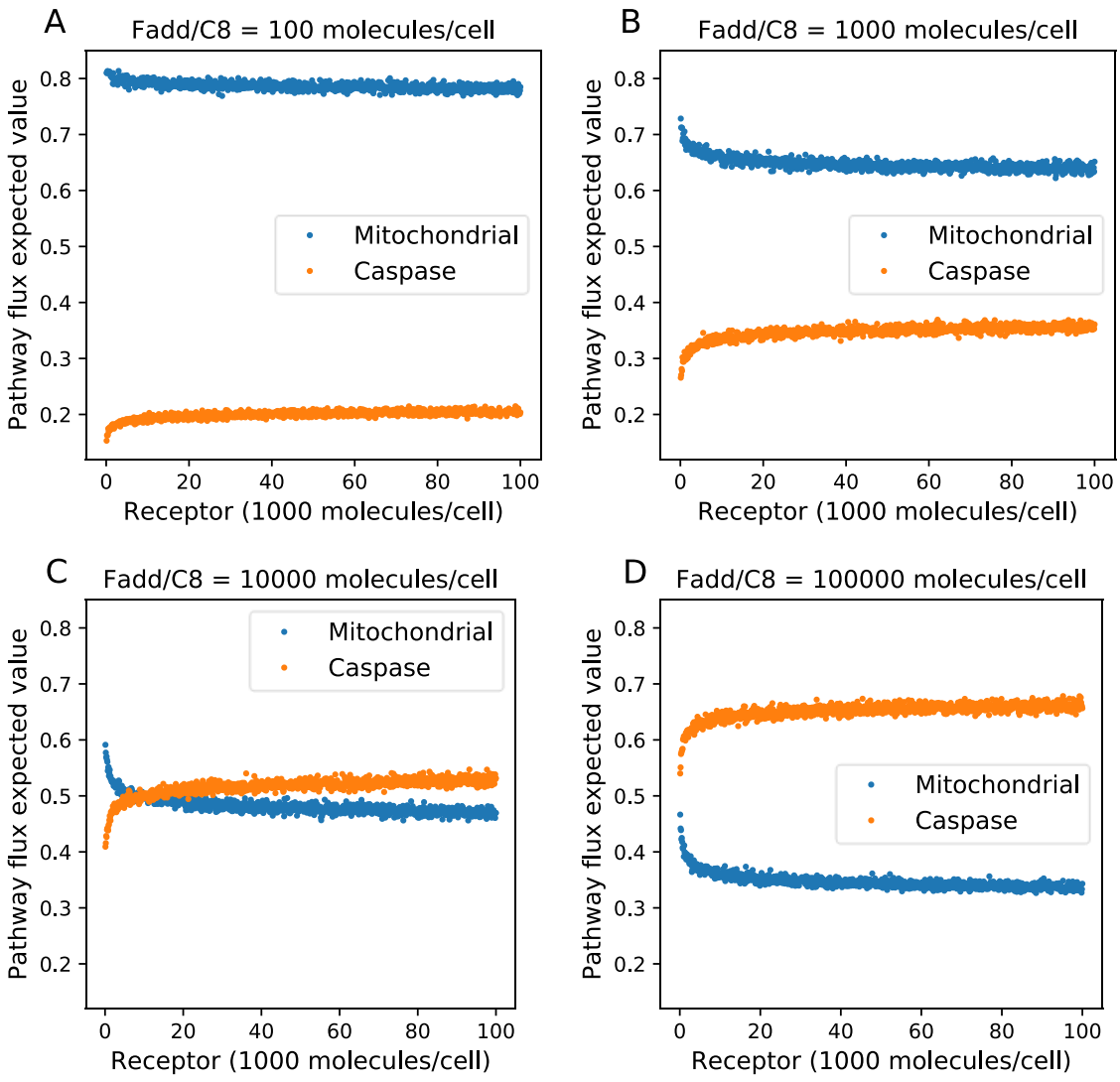


Figure 4.9. Signal flux through the caspase and mitochondrial pathways over increasing receptor concentration and at increasing levels of Fadd/Procaspase-8: (A) 100,000 molecules per cell of Fadd and Procaspase-8, (B) 10,000, (C) 1000, and (D) 100.

4.3.4 Incomplete Apoptosis Recovery from MOMP Sensitizers using Common Priors

The Type I and II pathways are respectively defined as independent and dependent of MOMP induced apoptosis [93]. As such, regulation of MOMP is paramount for proper apoptotic signaling in cells with a Type II phenotype. MOMP is regulated by a family of related proteins. The first of these, Bcl-2, was characterized by Vaux et al. to be involved in the proliferation/survival of certain cancer cell lines [106]. Further research on Bcl-2 would unravel the role of Bcl-2 in apoptosis and lead to the discovery of the Bcl-2 family of proteins that all share some level of domain homology and which form the regulatory network for MOMP [68]. This family of proteins can be broken down into three categories based on the presence or absence of the four Bcl-2 homology (BH) domains [63]. The antiapoptotic members, such as the Bcl-2 protein, as well as Bcl-xl, Mcl-1, and others, contain all four BH3 domains. The BH1-BH3 domains are present in the effector proteins Bax and Bak which are responsible for export of proapoptotic factors from the mitochondria and are activated by BH3-only proteins like Bid, Bim, Bad, Noxa, and Puma [107]. Work by Letai et al. revealed that BH3-only proteins can be further divided into activator and sensitizer proteins, completing the generally accepted view of Bcl-2 interactions [63]. Briefly, BH3-only activator proteins like Bid and Bim directly induce oligomerization of the effector proteins Bax and Bak, which subsequently form pores in the mitochondrial outer membrane [62, 63, 65, 66, 108]. The antiapoptotic proteins bind and inhibit the activator and effector proteins, disrupting the interaction between the two [63, 64, 65, 66, 109]. Finally, the sensitizers like Bad, Noxa, and Puma bind and sequester the antiapoptotic proteins allowing the activators to work [66, 110].

The role of Bcl-2, and related proteins, in the evasion of apoptosis for many cancers, and the illumination of the MOMP regulatory network, have made the mitochondrial pathway an attractive therapeutic target. Development of antagonists of the antiapoptotic proteins Bcl-2, Bcl-xl, etc., is the focus of much of the research and the first small-molecule inhibitor of Bcl-2, Venclexta (AbbVie), was approved in June 2018 for certain cases of chronic lymphocytic leukemia (CLL) [66, 101, 111]. Such antagonists mimic the action of the

sensitizer proteins, inhibiting the inhibitors of apoptosis. Unfortunately, development is complicated by the heterogenous binding affinities between the various sensitizers and antiapoptotic proteins [66]. Bad, for example, has a strong affinity for Bcl-2 and Bcl-xl but not for Mcl-1. Noxa on the other hand binds strongly only to Mcl-1. To identify potentially apoptosis inducing agents, Certo et al. developed an assay to predict, for a given cell line, which antiapoptotic proteins mitochondrial permeabilization depends on [99]. In this method, mitochondria are isolated and exposed to various BH3 peptides, the binding fragments of the BH3-only proteins. The amount of cytochrome c released from the mitochondria is then quantified via ELISA. With the effective BH3-proteins identified, and a table of known binding partners for those proteins, the acting antiapoptotic proteins can be predicted. They demonstrated the method using several cell lines including IL-3 dependent murine FL5.12 cells that are protected by Bcl-2 upon IL-3 removal, and murine hybridoma 2B4 cells that were overexpressed with either Bcl-2 or Mcl-1. They also found that isolated mitochondria from murine, Bcl-2 dependent, leukemia cells released much higher amounts of cytochrome c when exposed to BH3 peptides with high Bcl-2 affinity, suggesting that these cells are "primed for death", i.e. the antiapoptotic proteins are largely bound to BH3-only activators that can be displaced with the introduction of BH3-only sensitizers. The method has since had various improvements and has been tested against many more cell lines including acute lymphoblastic leukemia, diffuse large B cell lymphoma, CD4/CD8 positive thymocytes, and more [100, 112, 113, 114, 115].

To examine the regulation of the mitochondrial pathway and ascertain conditions under which it might make a worthwhile therapeutic target we estimated the evidence for this pathway over a range of regulatory conditions. For simplicity of analysis and computational austerity we initially limited the MOMP regulatory network to four proteins: the BH3-only activator Bid, the antiapoptotic Bcl-2, the BH3-only sensitizer Bad, and the effector Bax. In all runs Bcl-2 was varied from 0 to 200,000 molecules per cell, well within experimentally reported figures, at increments of 10,000, and Bad, the Bcl-2 antagonist, was varied from 0 to 400,000, much higher than reported levels, also in increments of

10,000. We sought to reflect conditions under which the Type II pathway is preferred over Type I, conditions for which MOMP regulation is most relevant. Thus, elements of the DISC, the receptor, Fadd, and Procaspase-8, were set to a low value of 100, and XIAP was set to a moderate value of 42,000. Note that the evidence score for the caspase only pathway is constant when changing only those regulators specific to the mitochondrial pathway. At these levels for the DISC components and XIAP, the caspase pathway evidence value was estimated to be 0.188. See Table 2.1 for a list of all other initial values. The evidence values resulting from this setup are displayed in figure 4.10.

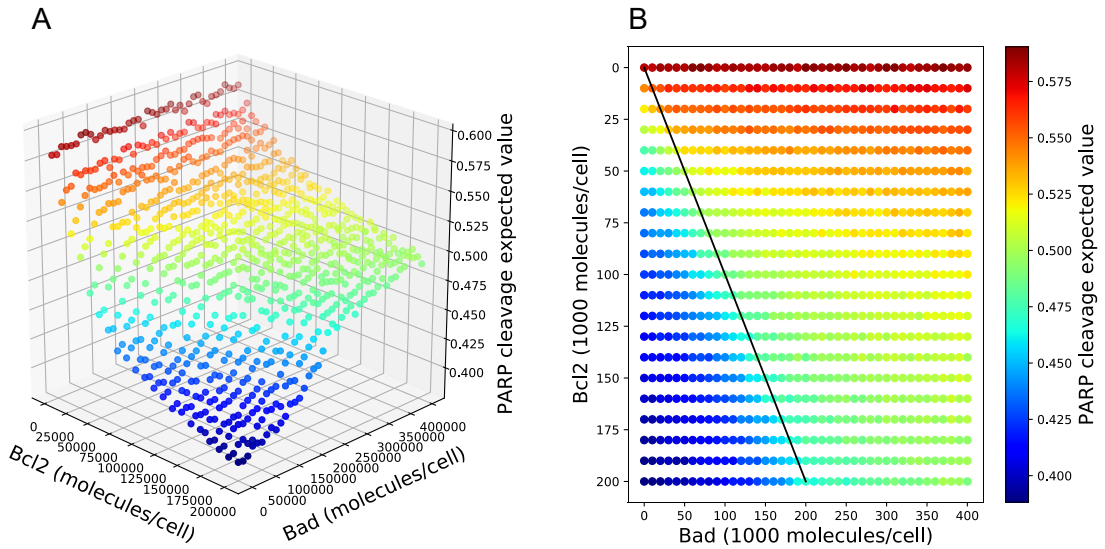


Figure 4.10. Bayesian evidence characteristics of Bcl-2 and Bad regulation of MOMP. Concentrations ranged from 0-200,000 molecules per cell for Bcl-2 and 0-400,000 for Bad in increments of 10,000 for both. Receptor, Fadd, and Procaspase-8 levels were all set to 100. (A) The evidence (expected PARP cleavage) values over the given ranges for Bcl-2 and Bad. (B) Flat representation of (A). The black line represents parity between Bcl-2 and Bad.

The complete network evidence values exceed those for the caspase pathway for all values of Bcl-2 and Bad, putting all evidence ratios above 0 and indicating a preference for the Type II phenotype throughout. In the absence of Bad, increasing levels of Bcl-2 lowers the expected values for the proportion of PARP that's cleaved from a high of 0.590 to a low of 0.388. This decrease in cleavage as Bcl-2 increases decelerates throughout. With Bcl-2 at a level of 200,000, increasing Bad, as expected, increases the PARP cleavage expected values. This increase initially accelerates until it reaches about 200,000, equal to Bcl-2, where it levels out. The evidence estimate reaches a high of around 0.500 at high levels of Bad for a recovery of only 55%. This reflects a general pattern in which apoptotic signal recovery induced by Bad is most efficacious up to parity between Bad and Bcl-2, after which additional increases in Bad effects diminishing returns and never fully recovers the signal (Figure 3.10B). Assuming a base evidence value of 0.58, the value estimated when Bcl-2 and Bad are 0, statistics for percent recovery at parity and at maximum levels of Bad, as well as the percentage of the total gain in evidence obtained at parity, are given in Figure 4.11.

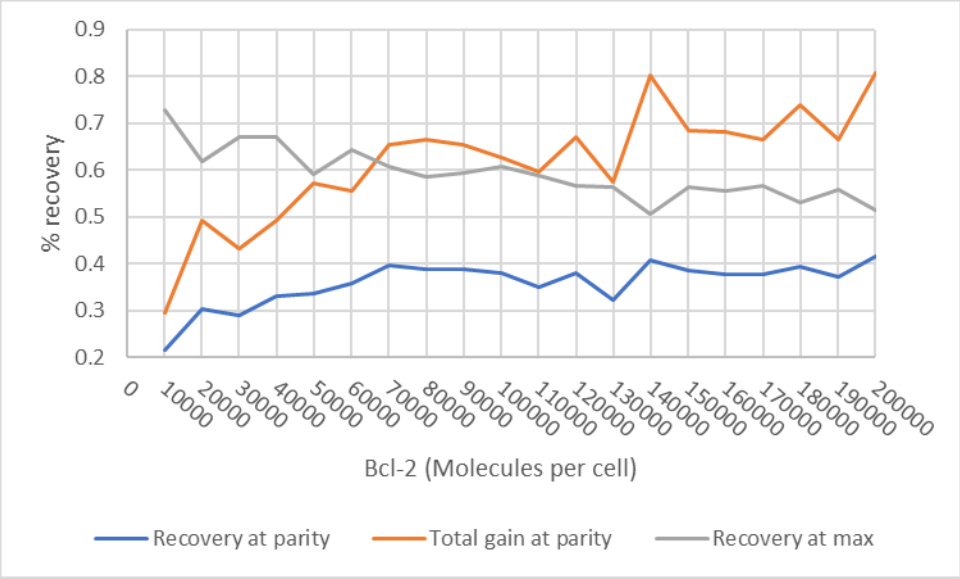


Figure 4.11. Statistics for the recovery of apoptosis when inhibiting Bcl-2 with Bad. Grey: % recovery obtained at the maximum Bad value of 400,000 molecules per cell. Blue: % recovery when Bad is at parity with Bcl-2. Orange: % of recovery found at the maximum Bad level that can be obtained with Bad at parity with Bcl-2.

Recovery at the maximum Bad level of 400,000 is 73.0% when Bcl-2 levels are at 10,000 and generally decreases as Bcl-2 increases. Thus, for the given priors, even with low Bcl-2 and high Bad levels, Bcl-2 is substantially retarding the Type II pathway's ability to achieve apoptosis. Signal recovery at parity ranges from 21.5% to 41.5% and is consistently just under the 40% mark for Bcl-2 levels of 70,000 and higher. The percent of total gain (recovery at the maximum level of Bad) found at parity ranges from 29.6% to 80.1% and generally increases with increasing Bcl-2. Thus, most of the observed apoptotic recovery induced by Bad is obtained when Bad levels are equal to or less than Bcl-2. Increasing Bad beyond this point has a muted effect.

Although some cell lines have been shown to be dependent on only a single antiapoptotic protein, the general case is much more complicated with a network of regulators governing MOMP execution [31, 65]. Binding affinities between these regulators often dictate inhibitory targets. Bcl-2, for example, preferentially binds to the effector Bax while Mcl-1 prefers Bak and Bcl-xl will inhibit both [65, 68, 115]. Sensitizers are likewise specific to their binding partners. Bad for example binds strongly to both Bcl-2 and Bcl-xl, while Noxa is specific to Mcl-1 [65, 66, 68, 115]. The limited and diminishing effect that sensitizers have on antiapoptotic proteins appears to hold for these more complex regulatory networks. To examine the effect of additional regulators on the likelihood of achieving apoptosis through the mitochondrial pathway, a MOMP specific model was used that included Bak, Bcl-xl, Mcl-1 and Noxa (Figure 4.12A). Because the run time is correlated to the size of the model, we focused on the MOMP portion of the network and eliminated Caspase-6 as well as any components downstream of Cytochrome c and Smac including APAF, Caspase-9, XIAP,

Caspase-3, and PARP. The objective function was then switched to export of Smac from the mitochondria. Values for exported Smac were evaluated for varying levels of the sensitizers Bad and Noxa (Figures 4.12B). Note that Bid is also responsible for embedding Bax into the mitochondrial membrane in this model and that all three antiapoptotic proteins inhibit Bid. All non-sensitizer Bcl-2 family proteins were set to values in line with levels estimated from Type II Jurkat cells in Dai et al. (2017). They are as follows: Bid: 171,000, Bax: 5000, Bak: 35,000, Bcl-2: 157,000, Bcl-xl: 113,000, and Mcl-1: 57,000. The range for Bad was 0-600,000 at increments of 20,000 and for Noxa was 0-120,000 at increments of 6000. Low values for the DISC components and a moderate XIAP value were again used here.

In the absence of antiapoptotic and sensitizer proteins, the Type II model evidence under these conditions was estimated to be 0.702 (Figure 3.12B, gray plane). With the addition of the antiapoptotic proteins the evidence for Smac export drops to about 0.224. Addition of Noxa, alone, at 120,000 molecules per cell raises the evidence to roughly 0.276, a recovery of only 10.9%. With Noxa at 60,000 roughly in line with Mcl-1, the evidence is about 0.271, representing about 90.4% of the total gain found at 120,000. Addition of Bad alone has a greater effect with the evidence rising to about 0.346 at 600,000 molecules per cell, for a 25.5% recovery. At 280,000 molecules per cell, 10,000 more than the combined Bcl-2 and Bcl-xl values, the evidence was estimated to be 0.306, 67.2% of the total gain at 600,000. The difference in effectiveness between Noxa and Bad may reflect the relative quantities of their respective targets but also the regulatory topology of the network. Addition of both sensitizers yields additional gains in apoptotic recovery. With Noxa at 120,000 and Bad at 600,000 the evidence rises to 0.438, a recovery of 44.8%, and at 60,000 for Noxa and 280,000 for Bad the evidence is roughly 0.380, 72.9% of the maximum total gain. These results mirror those above, Noxa and Bad exert the majority of their effects by the time they reach parity with their targets and have only a partial apoptotic recovery for the given generic prior ranges.

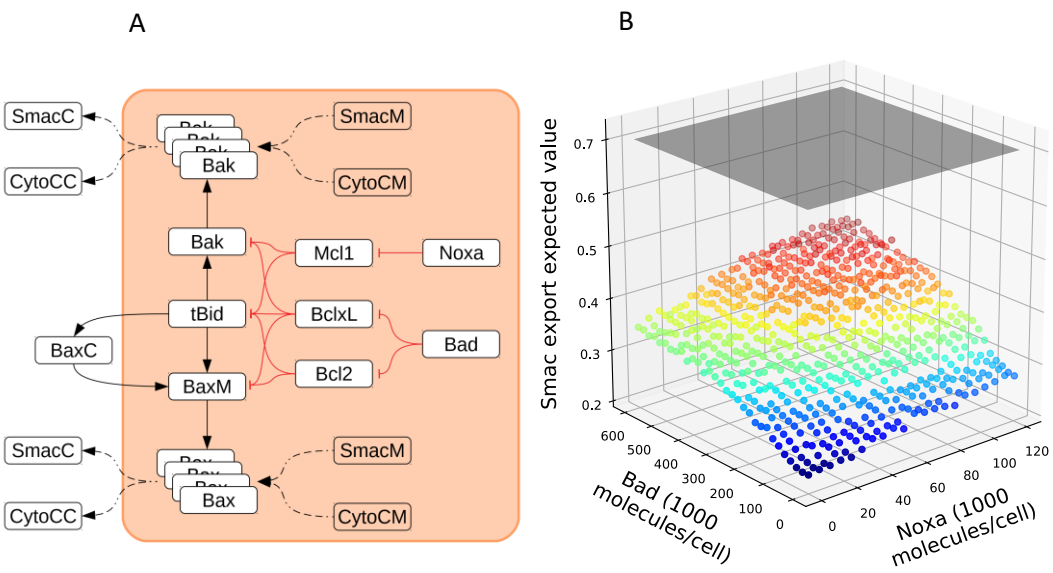


Figure 4.12. Bayesian evidence characteristics of a complex MOMP regulatory network. (A) An expanded asymmetrical MOMP regulatory network. (B) Expected Smac export values for the Type II pathway over a range of 0-120,000 molecules per cell at increments of 6000 for Noxa and a range of 0-600,000 at increments of 20,000 for Bad. Bid, Bax, Bak, Bcl-2, Bcl-xl, and Mcl-1 were set to 171,000, 5000, 35,000, 157,000, 113,000, and 57,000 respectively. The grey plane represents the evidence estimation when Bcl-2, Bcl-xl, Mcl-1, Bad, and Noxa are set to 0.

4.3.5 Near Full Apoptosis Recovery from MOMP Sensitizers using Adjusted Priors

Although the results of the previous analysis for MOMP regulation are directionally consistent with experimentation, the lack of full apoptotic recovery, even with sensitizer levels in excess of normal ranges, is not [66, 111]. This can be explained by the generic prior parameter ranges used in those experiments: log ranges of [-8.0, -4.0] for the forward binding rate and [-4.0, 0.0] for the reverse. To get a more accurate estimation of the extent of apoptotic inhibition by Bcl-2 and recovery via Bad we adjusted the reverse binding rates to more accurately reflect known dissociation constants (K_D) for MOMP regulatory binding reactions Bcl-2:Bid, Bcl-2:Bax, and Bad:Bcl-2 [116]. We assumed a rough average K_D estimate of 10 for all three interactions and thus, set the reverse binding rate to [-7.0, -3.0].

The resulting estimates of PARP cleavage over the ranges of Bcl-2 and Bad are displayed in Figure 4.13B, with the original results displayed in 4.13A on the same scale as a reference.

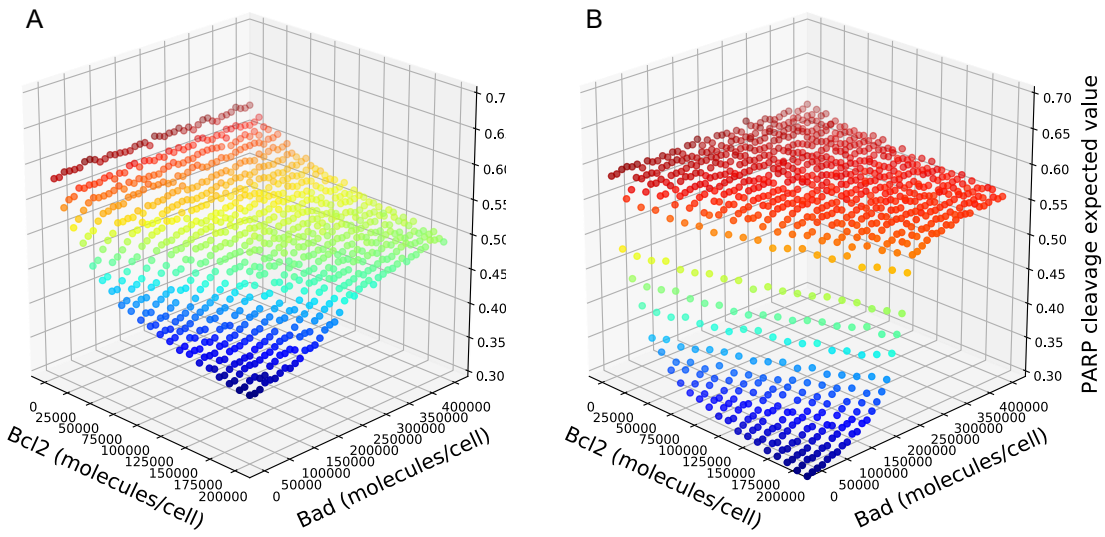


Figure 4.13. Bayesian evidence characteristics of MOMP regulation with binding rate parameters adjusted to more accurately reflect experimentally derived binding affinities for the interactions Bcl-2:Bid, Bcl-2:Bax, and Bad:Bcl-2. Bcl-2 was ranged from 0 to 200,000 molecules per cell and Bad was ranged from 0 to 400,000. Both were increased in increments of 10,000. (A) The original evidence (expected PARP cleavage) values over the given ranges for Bcl-2 and Bad. (B) The evidence values given the adjusted rate parameters.

Statistics for apoptosis recovery from the sensitizer Bad are below in Figure 4.14. The higher binding affinity of Bcl-2 with its targets clearly drives a stronger antiapoptotic effect with the expected PARP cleavage falling to 0.274 in the absence of Bad. Recovery upon addition of Bad also sees a stronger reaction. Recovery at maximum Bad is greater than 90% for all values of Bcl-2 and the percent recovery at parity with Bcl-2 is consistently in the mid 70's. As with the previous case, the percentage of the total PARP cleavage recovered at parity with Bcl-2 hovers around 80%.

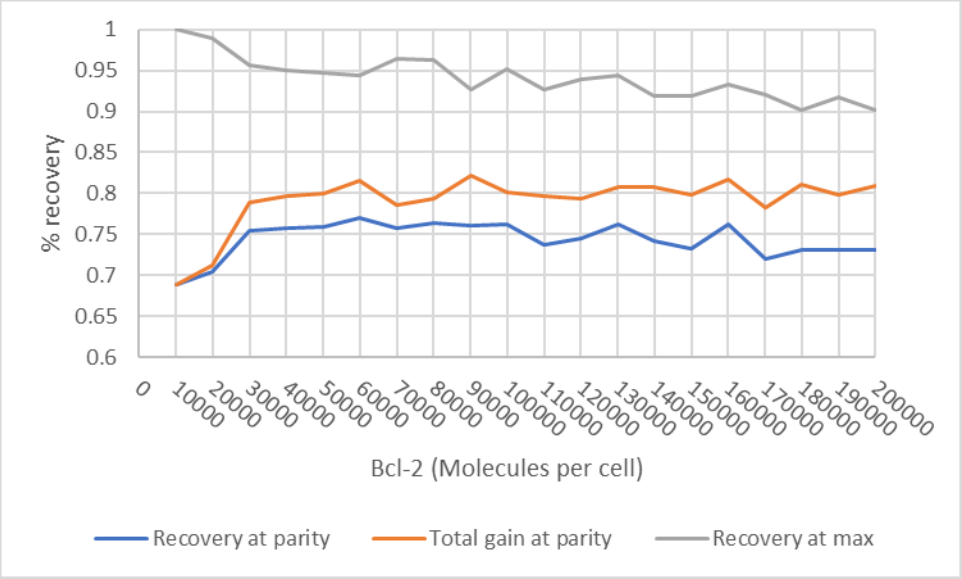


Figure 4.14. Statistics for the recovery of apoptosis when binding affinities are closer to experimentally defined K_D values. Grey: % recovery obtained at the maximum Bad value of 400,000 molecules per cell. Blue: % recovery when Bad is at parity with Bcl-2. Orange: % of recovery found at the maximum Bad level that can be obtained with Bad at parity with Bcl-2.

Chapter 5

HypBuilder: Automatic Generation of Ensembles of Physicochemical Models

5.1 Summary

HypBuilder is designed to quickly construct ensembles of mechanistic models for use in model selection or in silico experiment sweeps and to alleviate the arduous task of manual model construction. The use and features of the software with many small examples are detailed. Two larger examples are also described. The first is the recapitulation of an existing manually constructed model of extrinsic apoptosis. The second is the mass construction of 1450 variants of the mitochondrial outer membrane permeabilization. The HypBuilder software can be found at https://github.com/LoLab-VU/HypBuilder.

5.2. Introduction

Quantitative computational modeling of biological systems has become a valuable tool for explaining biological phenomena when used in conjunction with experimental methods. The modeling process starts with the construction of a reaction topology that is thought to reasonably represent the system in question and calibration of model parameters to fit simulation outcomes to experimental data. Hypotheses are then made based on changes in simulated outcomes upon perturbation of the model. A number of platforms for the construction of quantitative executable models have been developed. These range from graphical network designers such as CellDesigner and Vcell to rule-based modeling systems like Kappa and BioNetGen [27-30, 117, 118]. However, the growth in development and utilization of modeling tools has led to problems regarding the reuse, sharing, and modification of these models.

Addressing these issues has been the focus of several recent projects. The systems biology markup language (SBML), for example, is an XML-based modeling format designed

for interoperability between disparate modeling ecosystems, enabling the easy exchange and reuse of models [119]. Model exchange and reuse are further promoted with the existence of an accessible storehouse of previously constructed models. That is the purpose of the BioModels database, a publicly available repository for storing curated and annotated biological models [120]. It's often the case that models are not reused as is but are modified to address particular scientific inquiries. Modifying models, particularly those defined at the top level by systems of differential equations, can be arduous and error prone. Rule-based modeling systems like BioNetGen and Kappa mitigate much of the issue by defining models based on the interactions between various components of the system which are then automatically translated into the equations necessary for simulation [26]. PySB takes this a step further and allows more complicated patterns of interactions to be defined in a hierarchical manner which further allows for various components of models to be broken out and recombined in new ways [31]. Yet, despite these advances, construction of large models can still be time-consuming and error prone, and the construction of large numbers of models necessary for the purposes of model selection or in silico experimental sweeps are impractical. In such cases a method for the large-scale production of automatically generated executable mechanistic models is necessary.

In this work we introduce HypBuilder, software for the automatic generation of mechanistic rule-based models. HypBuilder is a Python based tool that generates models in the PySB format from a user provided list of components and reactions, and an expandable library of rule sets defined on those reactions. This both simplifies the input needed to create a model and permits large numbers of models to be generated from a single input file via combinations of select reactions. This approach has the potential to expedite the large-scale production of mechanistic models.

5.3. Software and Usage

5.3.1 General Workflow

Figures 5.1A and 5.1B display the steps taken by HypBuilder to generate models and the components necessary to carry them out. Input is a simple CSV file containing all necessary information for the construction of one or many kinetic models in the PySB format. HypBuilder then takes the reactions designated in that file and references a molecular interaction library for corresponding sets of reaction rules. These rules are combined, in a way that is also designated in the input file, and each combination serves as a basis for a model. HypBuilder then calls the relevant PySB classes for model construction and export. The result is a collection of automatically generated executable PySB model files.

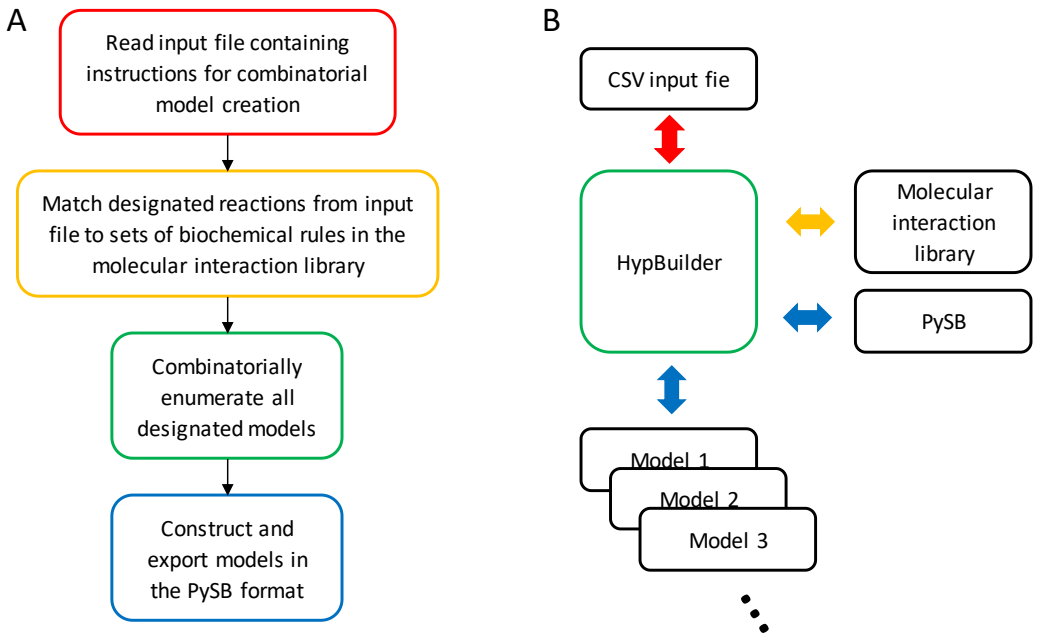


Figure 5.1. HypBuilder workflow and processing.

5.3.2 Input File

The input file takes the form of simple CSV-based list of model components, reactions, and special instructions. At its simplest, it is a list of monomers with initial values, and a list of reactions (Figure 5.2A). The list of monomers is headed by the term “model components” while the list of required reactions is simply headed by “required reactions”. A reaction consists of a reaction name and a list of all participants along with their roles in the reaction in brackets. For example, the catalysis reaction in Figure 5.2A has three components, A(), B() and C(), with the roles of catalyzer, substrate, and product respectively. HypBuilder takes the information from the input file, references the molecular reaction library, and constructs the corresponding PySB model (Figure 5.2B).

```
model components
A, 1
B, 1
C, 1

required reactions

catalysis, A()[catalyzer], B()[substrate], C()[product]
```

5.3.3 Run File

The run file simply calls the ModelAssembler class from HypBuilder and passes in a Molecular Interaction Library (Section 5.3.5) and a model input file. An option for listing monomer-based or species-based observables can also be input here and is detailed below. Other global model formatting options could be added here in the future.

```
from HypBuilder import ModelAssembler

ModelAssembler('library.txt', 'input.csv')
```

5.3.4 Output

HypBuilder output consists of one or more executable PySB models with the follow basic components. Please reference [31] and http://pysb.org/ for information on the PySB ecosystem and documentation. Below is the model generated from the input file from Section 5.3.2.

```
Model()
Monomer('A', ['B'])
Monomer('C')
Monomer('B', ['A'])

Parameter('catalysis_0_A_catalyzer_B_substrate_C_product_2kf',
1.0)
Parameter('catalysis_0_A_catalyzer_B_substrate_C_product_1kr',
1.0)
Parameter('catalysis_1_A_catalyzer_B_substrate_C_product_1kc',
1.0)
Parameter('A_0', 1.0)
Parameter('C_0', 1.0)
Parameter('B_0', 1.0)

Observable('A_obs', A())
Observable('C_obs', C())
Observable('B_obs', B())

Rule('catalysis_0_A_catalyzer_B_substrate_C_product',
      A(B=None) + B(A=None) | A(B=1) % B(A=1),
      catalysis_0_A_catalyzer_B_substrate_C_product_2kf,
      catalysis_0_A_catalyzer_B_substrate_C_product_1kr)
Rule('catalysis_1_A_catalyzer_B_substrate_C_product',
      A(B=1) % B(A=1) >> A(B=None) + C(),
      catalysis_1_A_catalyzer_B_substrate_C_product_1kc)

Initial(A(B=None), A_0)
Initial(C(), C_0)
Initial(B(A=None), B_0)
```

5.3.4.1 Monomers

HypBuilder cross references the list of monomers with the list of reactions, and the associated rule templates from the reaction library, to correctly write the PySB monomer objects. For example, because the catalyzer A() in the input file (Figure 5.2A) must first bind the substrate B(), both the A and B monomers have binding sites corresponding to the

other. In the absence of explicitly naming the binding sites (Section 5.3.6) the sites are automatically named for the binding partner they are intended for. An important departure from standard rule-based modeling in PySB and BioNetGen is the lack of support for state sites in HypBuilder. Although the use of state sites, such as designating a protein as phosphorylated or unphosphorylated, can make models more compact and manual model building more efficient for the experienced builder, they greatly complicate the automatic generation of models. In HypBuilder a separate state is simply represented by a distinct monomer. Whether state sites are used, or each state is represented by a monomer, the resulting system of equations is mathematically equivalent. The only downside is the need to list a monomer for every state of every component in the model, a simple task in HypBuilder.

5.3.4.2 Parameters

The parameters representing initial values are generated from the values listed after each model component. Ranges of initial values can also be defined (Section 5.3.6.1). Reaction rate parameters can be defined by listing them after the reaction participants. In example 5.2A there are no such parameters listed, in which case the parameter value will be given a placeholder value of 1.0. This is done when the parameters are expected to be sampled as in the case of Bayesian evidence-based model selection. Note that if parameters are listed they must be in the same order designated in the Molecular Interaction Library. Parameter names for reactions are descriptive of the reaction and reaction components as can be seen in Figure 5.2B. Parameters are numbered by the corresponding rules in the reaction. In Figure 5.2, for example, the single rule for catalytic binding is bidirectional, so the parameters for the forward and reverse binding rates are labeled 0. The rule representing catalysis is then labeled 1. The postfix of the parameter is representative of the type, direction, degree, and variability of the reaction. ‘2kf’ for example represents a second-degree forward reaction, while ‘1kr’ represents a first-degree reverse reaction and first-degree catalytic reactions are represented as ‘1kc’ Initial values, and any other parameter that is explicitly defined, are simply represented with an appended ‘_0’.

5.3.4.3 Observables

There are two settings for observables in HypBuilder. In the default setting HypBuilder writes down an observable for every monomer. If we want observables for every combination of monomers, such as individual and various bound species, we add ‘species’ to the argument list.

```
from HypBuilder import ModelAssembler

ModelAssembler('library.txt', 'input.csv', 'species')
```

The example input file above creates one bound species, monomer A bound to B, in addition to the unbound monomers. Thus, there are four observables defined when the ‘species’ argument is specified. The names for these observables mirror the species they designate and include each monomer involved, the binding sites, and binding status’ of each monomer in the species.

```
Observable('A_B_None_obs', A(B=None))
Observable('C__obs', C())
Observable('B_A_None_obs', B(A=None))
Observable('A_B_1__B_A_1_obs', A(B=1) % B(A=1))
```

5.3.4.4 Rules

After the monomers are cross referenced with the list of reactions to correctly fill out the monomer binding sites, HypBuilder uses those monomers in the construction of the model rules. HypBuilder reads in the reactions from the input file and finds the corresponding template rules from the molecular interaction library. Monomers are then substituted into those templates based on their designated roles in the reaction. Multiple rules can be contained in a single reaction template, as is the case for the catalyze reaction in section 5.3.5. In that case two rules are created, a rule governing the reversible binding of the catalyzer (monomer A) to the substrate (monomer B), and a rule for the catalysis step, converting substrate to product (monomer C). The names of the rules are descriptive, mirroring the

names of the corresponding parameters. They contain the name of the reaction and each participant along with its role in the reaction They are numbered based on the order of rules in the reaction. By default, any binding sites not involved in a reaction are assumed to be unbound. Strategies to build more complex rules can be found in Section 5.3.6 below.

5.3.4.5 Initials

By default, model initials are constructed for each unbound monomer with the parameter values defined in the input file. However, initials for bound species can also be designated (Section 5.3.6.8).

5.3.5 Molecular Interaction Library

PySB uses several levels of increasing complexity to build up a model [31]. At the base level is a system of ordinary differential equations (ODEs) representing the mass action kinetics of the system. These ODEs are generated by rules that define interactions between various molecules in the model. Combinations of rules can then be combined into macros that represent simple biological processes. For example, PySB's catalyze macro is made up of rules for the reversible binding of enzyme to substrate and for the catalysis step in the reaction. There exists a library of macros within PySB to aid in manual model construction. Macros can then be further built up into larger groups of molecular components and pathway structures that can be assembled into models. A rule/macro library is ideal for the task of automatic model generation but the macro library in PySB is designed to aid in manual model construction and is not suited for model generation in a combinatorial manner. Thus, we have designed a stand-alone easily extensible rule/macro library format for use with HypBuilder.

The molecular interaction library is a list of molecule types (substrate, product, etc.), the reactions they participate in (catalysis, phosphorylation, etc.), and the associated rule template(s) for those reactions. The '+++' lines signify the end of a molecule while the '$$$' lines signify the end of a reaction within a molecule. Note that while every molecule type

designated in a reaction from the input file must be represented in the library, only one must list the reaction and contain the reaction template(s). For the catalysis reaction the product entry holds this information. The catalyzer and product are tied to the reaction via the template(s) contained in the product's library entry. Also note the flexibility afforded in putting together groups of rules. An alternative to the product entry holding both rule templates is to split it up and associate rules with those specific monomers that participate in them. In such a scenario the substrate molecule holds the substrate binding reaction and associated binding template between it and the catalyzer, and the product holds the catalytic step and its associated template. The templates are generic rules with molecule types in place of specific monomers that HypBuilder substitutes with the appropriate monomer. The placement of each monomer in the rule is governed by the designated role given to them in the input file reactions. Molecule types can be associated with any number of reactions and monomers can be labeled as any number of molecules types within various reactions.

```
molecule: catalyzer
+++

molecule: substrate
reaction: substrate_binding
template: catalyzer(substrate=None) + substrate(catalyzer=None)
      <> catalyzer(substrate=1) % substrate(catalyzer=1)
$$$
+++

molecule: product
reaction: catalytic_step
template: catalyzer(substrate=1) % substrate(catalyzer=1)
      >> catalyzer(substrate=None) + product()
$$$
reaction: catalysis
template: catalyzer(substrate=None) + substrate(catalyzer=None)
      <> catalyzer(substrate=1) % substrate(catalyzer=1)
template: catalyzer(substrate=1) % substrate(catalyzer=1)
      >> catalyzer(substrate=None) + product()
$$$
+++
```

A more complicated example from the molecular interaction library is displayed below. This reaction describes the rules for pore formation, a sequence of binding reactions between like molecules up to the status of tetramer. There are three rules here one for each increase in the molecules complexity. Note the binding sites in these rules. Although one can specify the site names for a given reaction in the input file, it is also possible to do so globally from the library. This method is less flexible in that the site names are simply numbered to reflect multiple binding sites for the same molecule type. A single call to the pore() reaction will automatically include four identical molecule types in the rule set. Also note that this reaction is self-contained in that it requires no named interacting partners.

```
molecule: pore
reaction: pore_formation
direction: self
template: pore(pore_1=None, pore_2=None)
      + pore(pore_1=None, pore_2=None)
      <> pore(pore_1=None, pore_2=1)
      % pore(pore_1=1, pore_2=None)
template: pore(pore_1=None, pore_2=None)
      + pore(pore_1=None, pore_2=1)
      % pore(pore_1=1, pore_2=None)
      <> pore(pore_1=3, pore_2=1)
      % pore(pore_1=1, pore_2=2)
      % pore(pore_1=2, pore_2=3)
template: pore(pore_1=None, pore_2=None)
      + pore(pore_1=3, pore_2=1)
      % pore(pore_1=1, pore_2=2)
      % pore(pore_1=2, pore_2=3)
      <> pore(pore_1=4, pore_2=1)
      % pore(pore_1=1, pore_2=2)
      % pore(pore_1=2, pore_2=3)
      % pore(pore_1=3, pore_2=4)
$$$
+++
```

5.3.6 Features

A number of features have been built into HypBuilder both for the large-scale construction of executable models and the construction of more complex rule sets that may be needed. These features are described here.

5.3.6.1 Initial Value Ranges

One of the more likely reasons for the construction of many variations of a model is to simply vary the initial conditions and observe how that affects the outcome of a simulated experiment. HypBuilder has three ways to create variations of a model with different initial values as displayed in the following 'model components' section of a HypBuilder input file.

```
model components

A, 1,2                  # simple list of initial values
B, 3:3-4                # number:start-stop
C, 5-6:0.5              # start-stop:increment
D, 1,2,2:3-4,5-6:1      # combinations
```

The first way is to simply list the desired initial values, as with component A above, which can be either 1 or 2. The other two ways involve providing ranges of initial values to HypBuilder. For component B the number of desired initial values is followed by an (inclusive) range for which that number of values will be evenly spaced. 3:3-4 designates 3, 3.5 and 4 as valid values. The third option, demonstrated for component C, is to list an inclusive range followed by an increment length. For example, 5-6:0.5 will designate 5, 5.5 and 6 as values. All of these methods can be combined as was done for component D above where 1,2,2:3-4,5-6:1 designates 1, 2, 3, 4, 5, and 6 as valid initial values. HypBuilder will construct a complete executable model for every combination of initial values. For the component list above 108 models will be constructed. In the future we will include the option to define a function that enumerates initial values over a predefined range.

5.3.6.2 Optional Reactions

The primary reason for the development of HypBuilder is the large-scale production of executable kinetic models with various reaction topologies for the purposes of model selection and exploration of network dynamics under various regulatory conditions and in silico experiments. To accomplish this task HypBuilder constructs combinations of

reaction sets base on a list of reactions designated ‘optional reactions’ in the input file. In the following reaction list is one required reaction that must exist in every model and three optional reactions that do not.

```
required reactions

inhibition, D()[inhibitor], E()[inh_target]

optional reactions

inhibition, C()[inhibitor], D()[inh_target]
inhibition, A()[inhibitor], C()[inh_target]
inhibition, B()[inhibitor], C()[inh_target]
```

There are eight possible combinations of optional reactions and HypBuilder will thus create eight models to account for them. Caution must be taken when designating reactions as optional. The 2^n models for the n optional reactions can quickly blow up. And although creation of models with HypBuilder is relatively fast, analysis of 2^n models may be infeasible. It should also be noted that monomers, observables, and initials for components that are not used in a particular model will not be include in the model. The resulting model for the required reaction-only case is shown below.

```
Model()

Monomer('E', ['D'])
Monomer('D', ['E'])

Parameter('inhibition_0_D_inhibitor_E_inh_target_2kf', 1.0)
Parameter('inhibition_0_D_inhibitor_E_inh_target_1kr', 1.0)
Parameter('E_0', 1.0)
Parameter('D_0', 1.0)

Observable('E_obs', E())
Observable('D_obs', D())

Rule('inhibition_0_D_inhibitor_E_inh_target',
D(E=None) + E(D=None) | D(E=1) % E(D=1),
inhibition_0_D_inhibitor_E_inh_target_2kf,
inhibition_0_D_inhibitor_E_inh_target_1kr)

Initial(E(D=None), E_0)
Initial(D(E=None), D_0)
```

5.3.6.3 Grouped Reactions

One way to limit the number of models is through reaction grouping. There are currently two grouping methods in HypBuilder. Any reactions grouped together in the input file must be included together in any model they appear in. Below are the 'reactions' sections of an input file including both methods.

```
required reactions

inhibition, A()[inhibitor], B()[inh_target]

optional reactions

inhibition, B()[inhibitor], C()[inh_target], {disjoint:1}
inhibition, D()[inhibitor], E()[inh_target], {disjoint:1}
inhibition, A()[inhibitor], C()[inh_target], {group:2}
inhibition, B()[inhibitor], D()[inh_target], {group:2}
inhibition, A()[inhibitor], D()[inh_target]
```

Grouped reactions are labeled as such and numbered within curly brackets after the reaction definition. The first two optional reactions are grouped together via the 'disjoint:1' label. In HypBuilder, a disjoint group is one in which no other optional reactions can be included. Thus, there is only one model that includes those two reactions (The required reaction is, of course, also included). The second two reactions are grouped together via the 'group:2' label. These two reactions must be grouped together in any model but can also be grouped with other, non-disjoint reactions. They can thus be included with, or without, the fifth optional reaction providing two additional models. That leaves the required reaction alone and the required reaction plus the fifth optional reaction as the last two possible models for a total of five, in contrast to the 32 models that would have been generated without grouping.

5.3.6.4 Data Nodes

One of the targeted applications for the large-scale production of mechanistic models is model selection which, in some fashion, requires fitting the models to data. The combinatorial methodology of model construction in HypBuilder can potentially create disconnected models with subnetworks that do not reach the component for which data is available. Such models are useless and should be eliminated In HypBuilder this is done by designating components as data targets. Consider the following input file.

```
model components

A, 1
B, 1
C, 1
D, 1
E, 1, {d}

required reactions

inhibition, D()[inhibitor], E()[inh_target]

optional reactions

inhibition, C()[inhibitor], D()[inh_target]
inhibition, A()[inhibitor], C()[inh_target]
inhibition, B()[inhibitor], C()[inh_target]
```

Component E has been labeled a data target with the lowercase ‘d’ in curly brackets. HypBuilder will take this information and eliminate any model with components that cannot reach this node. The eight possible networks for this input file are displayed in Figure 5.2. Note that the bottom four are all missing the reaction from component C to component D and are thus redundant with respect to explaining the data for component E. Three of those networks, those in the red box, have subnetworks that fail to reach the data node and will thus be eliminated from consideration by HypBuilder.

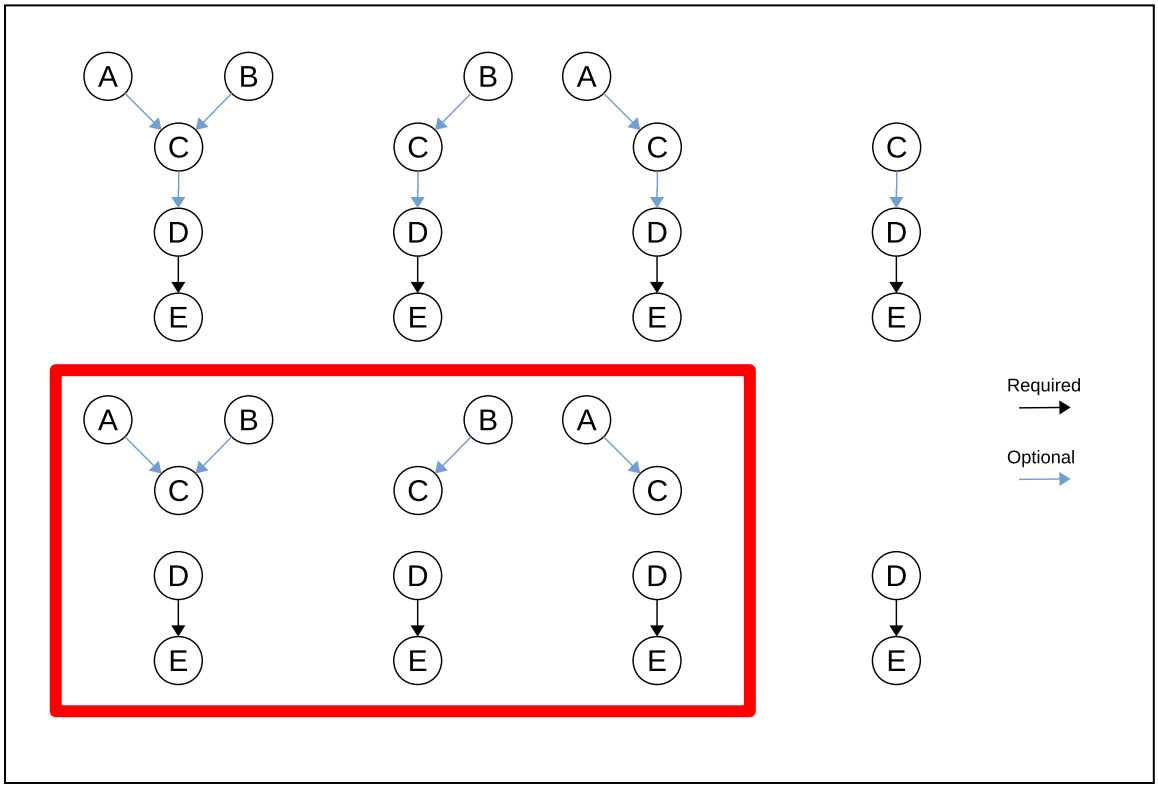


Figure 5.2. All possible networks created for the Data node example. Those in the red box have a data node (E) that cannot be reached by every other node in the network and will thus be eliminated from consideration.

5.3.6.5 Don't Write Don't Care

When manually writing rules for a PySB model there exists the concept of 'don't write don't care'. This comes about when existing sites on monomers that participate in the rule are left absent, rendering the state of that site irrelevant within that rule. In HypBuilder, any (binding) sites not defined in a rule template are, by default, assumed to be unbound (set to 'None'). Although this simplifies the information needed to write the input file and/or entries in the library, it eliminates the 'don't write don't care' approach and the flexibility it provides. Thus, we have implemented a means to include the concept in Hyp-Builder constructed models. Consider the following input file.

```
model components

A, 1
B, 1
C, 1
D, 1

required reactions

inhibition, A()[inhibitor], B()[inh_target]
inhibition, B()[inhibitor], C()[inh_target], {dwdc:B(A)|dwdc:C(D)}
inhibition, C()[inhibitor], D()[inh_target]
```

Note the two ‘dwdc’ labels in curly brackets after the second reaction. These are declarations that a binding site on a model component should be left blank within this reaction. Thus, the binding site for A on component B, and the binding site for D on component C will be absent in the rules generated for this reaction, as is displayed below.

```
Rule('inhibition_0_A_inhibitor_B_inh_target',
      A(B=None) + B(A=None, C=None)
      | A(B=1) % B(A=1, C=None),
      inhibition_0_A_inhibitor_B_inh_target_2kf,
      inhibition_0_A_inhibitor_B_inh_target_1kr)
Rule('inhibition_0_B_inhibitor_C_inh_target',
      B(C=None) + C(B=None)
      | B(C=1) % C(B=1),
      inhibition_0_B_inhibitor_C_inh_target_2kf,
      inhibition_0_B_inhibitor_C_inh_target_1kr)
Rule('inhibition_0_C_inhibitor_D_inh_target',
      C(B=None, D=None) + D(C=None)
      | C(B=None, D=1) % D(C=1),
      inhibition_0_C_inhibitor_D_inh_target_2kf,
      inhibition_0_C_inhibitor_D_inh_target_1kr)
```

5.3.6.6 Binding Sites

The default behavior of HypBuilder with regard to binding site designations is to simply name the site after the binding partner that targets it. To override this behavior, the new site names are listed in parentheses after the appropriate model component and in the appropriate reaction as displayed in the example here.

```
required reactions

bind, A()[binder], C(comp_site)[binder_target]
bind, B()[binder], C(comp_site)[binder_target]
bind, C()[binder], D(C_1)[binder_target]
bind, C()[binder], D(C_2)[binder_target]
pore_formation, C(s1:s2)[pore]
```

The resulting monomers and rules for these reactions are given below. Overriding the automatic naming allows for much more flexibility in model construction. For example, monomers A and B now competitively bind to the same site on monomer C ('comp_site'), and monomer C now has two distinct binding sites ('C_1' and 'C_2') on monomer D. The two binding sites used in the pore formation reaction for monomer C have also been over-written. When multiple sites are given new names, they must be listed in order with a colon in between them. The nth name in the list will always overwrite the nth binding site for a reaction.

```
Monomer('A', ['C'])
Monomer('C', ['comp_site', 'D', 's1', 's2'])
Monomer('B', ['C'])
Monomer('D', ['C_1', 'C_2'])
```

```
Rule('bind_0_A_binder_C_binder_target_comp_site',
    A(C=None) + C(comp_site=None, D=None, s1=None, s2=None)
    | A(C=1) % C(comp_site=1, D=None, s1=None, s2=None),
    bind_0_A_binder_C_binder_target_comp_site_2kf,
    bind_0_A_binder_C_binder_target_comp_site_1kr)
Rule('bind_0_B_binder_C_binder_target_comp_site',
    B(C=None) + C(comp_site=None, D=None, s1=None, s2=None)
    | B(C=1) % C(comp_site=1, D=None, s1=None, s2=None),
    bind_0_B_binder_C_binder_target_comp_site_2kf,
    bind_0_B_binder_C_binder_target_comp_site_1kr)
Rule('bind_0_C_binder_D_binder_target_C_1',
    C(comp_site=None, D=None, s1=None, s2=None)
    + D(C_1=None, C_2=None)
    | C(comp_site=None, D=1, s1=None, s2=None)
    % D(C_1=1, C_2=None),
    bind_0_C_binder_D_binder_target_C_1_2kf,
    bind_0_C_binder_D_binder_target_C_1_1kr)
Rule('bind_0_C_binder_D_binder_target_C_2',
    C(comp_site=None, D=None, s1=None, s2=None)
    + D(C_1=None, C_2=None)
    | C(comp_site=None, D=1, s1=None, s2=None)
    % D(C_1=None, C_2=1),
    bind_0_C_binder_D_binder_target_C_2_2kf,
    bind_0_C_binder_D_binder_target_C_2_1kr)
Rule('pore_formation_0_C_pore_s1_s2',
    C(comp_site=None, D=None, s1=None, s2=None)
    + C(comp_site=None, D=None, s1=None, s2=None)
    | C(comp_site=None, D=None, s1=None, s2=1)
    % C(comp_site=None, D=None, s1=1, s2=None),
    pore_formation_0_C_pore_s1_s2_2kf,
    pore_formation_0_C_pore_s1_s2_1kr)
Rule('pore_formation_1_C_pore_s1_s2',
    C(comp_site=None, D=None, s1=None, s2=None)
    + C(comp_site=None, D=None, s1=None, s2=1)
    % C(comp_site=None, D=None, s1=1, s2=None)
    | C(comp_site=None, D=None, s1=3, s2=1)
    % C(comp_site=None, D=None, s1=1, s2=2)
    % C(comp_site=None, D=None, s1=2, s2=3),
    pore_formation_1_C_pore_s1_s2_2kf,
    pore_formation_1_C_pore_s1_s2_1kr)
Rule('pore_formation_2_C_pore_s1_s2',
    C(comp_site=None, D=None, s1=None, s2=None)
    + C(comp_site=None, D=None, s1=3, s2=1)
    % C(comp_site=None, D=None, s1=1, s2=2)
    % C(comp_site=None, D=None, s1=2, s2=3)
    | C(comp_site=None, D=None, s1=4, s2=1)
    % C(comp_site=None, D=None, s1=1, s2=2)
    % C(comp_site=None, D=None, s1=2, s2=3)
    % C(comp_site=None, D=None, s1=3, s2=4),
    pore_formation_2_C_pore_s1_s2_2kf,
    pore_formation_2_C_pore_s1_s2_1kr)
```

5.3.6.7 Reaction Sequences

HypBuilder currently supports a method to quickly string together sequences of reactions. By designating an order to a sequence of reactions the result of reaction n will be forced onto reaction n+1. For example, consider the following set of input reactions.

```
required reactions

bind, A()[binder], B()[binder_target], {s1:1}
bind, B()[binder], C()[binder_target], {s1:2}
bind, C()[binder], D()[binder_target], {s1:3}
```

Designating these reactions s1:1, s1:2, s1:3 forces these binding reactions to be ordered as is displayed in the associated rule set.

```
Rule('bind_0_A_binder_B_binder_target',
      A(B=None) + B(A=None, C=None)
      | A(B=1) % B(A=1, C=None),
      bind_0_A_binder_B_binder_target_2kf,
      bind_0_A_binder_B_binder_target_1kr)
Rule('bind_0_B_binder_C_binder_target',
      B(A=ANY, C=None) + C(B=None, D=None)
      | B(A=ANY, C=1) % C(B=1, D=None),
      bind_0_B_binder_C_binder_target_2kf,
      bind_0_B_binder_C_binder_target_1kr)
Rule('bind_0_C_binder_D_binder_target',
      C(B=ANY, D=None) + D(C=None)
      | C(B=ANY, D=1) % D(C=1),
      bind_0_C_binder_D_binder_target_2kf,
      bind_0_C_binder_D_binder_target_1kr)
```

Note that in the second rule the binding site for A on component B is labeled ‘ANY’. ‘ANY’, in the language of PySB, means that that site must be bound for that rule to be executed. Because that binding site is specific to component A we have enforced a specific order on these reactions. In the third rule we have that B must be bound to C before C can bind to D. But because B must also be bound to A before it can bind to C we have a sequence of binding reactions that result in a complex of all four components. Multiple overlapping sequences of reactions can also be defined. Because state sites are not supported in

HypBuilder, only binding sites are affected. Designating such reaction sequences will not work in competitive binding cases and thus should only be used on a limited basis. A more proper way to construct sequences of reactions is to write it into the molecular reaction library. Reaction sequencing will likely be removed in the future.

5.3.6.8 Initial Binding

A situation that may come up is the need to begin a simulation with monomers that are already bound, so initial values for these bound species must therefore be defined in the model. HypBuilder supports the construction of such models. Consider the following input file.

```
model components

A, 10
B, 10
C, 50
D, 100
E, 100

required reactions

bind, A()[binder], B()[binder_target], {f:5}
bind, C()[binder], E()[binder_target], {f:all}
bind, D()[binder], E()[binder_target], {f:all}
```

Here, f:5 in the first reaction indicates that an initial value for the bound species consisting of the monomers A and B should be five, leaving 5 each of A and B. The second and third binding reactions are both followed by the f:all tag in curly brackets, indicating the maximum possible binding. Both reactions have a binder (C and D) with the same target (E) and there is more combined C and D than E, thus we expect E to be completely consumed while C and D compete for it. HypBuilder handles this iteratively. At each iteration it will first enumerate the total number of possible binding pairs for each species, 5000 for C % E and 10000 for D % E. It then randomly chooses one of these species based on the percentage of the total number of pairs. Thus, at the first iteration C % E has a 5000/(5000

+ 10000) = 33% chance of getting chosen. Once one is chosen the number of possible binding pairs is updated and the next iteration is executed. This continues until there are no possible binding pairs left. The initial values and their parameters for the above input file are displayed below.

```
Parameter('C_E_0', 39.0)
Parameter('D_E_0', 61.0)
Parameter('A_B_0', 5.0)
Parameter('A_0', 5.0)
Parameter('C_0', 11.0)
Parameter('B_0', 5.0)
Parameter('E_0', 0.0)
Parameter('D_0', 39.0)

Initial(C(E=1) % E(C=1, D=None), C_E_0)
Initial(D(E=1) % E(C=None, D=1), D_E_0)
Initial(A(B=1) % B(A=1), A_B_0)
Initial(A(B=None), A_0)
Initial(C(E=None), C_0)
Initial(B(A=None), B_0)
Initial(E(C=None, D=None), E_0)
Initial(D(E=None), D_0)
```

5.3.6.9 Text

Any additional information required to be in every model can be include by simply adding to the input file under the header of ‘text’, like so.

```
Text

“add text here”
“add more text”
“and so on.”
```

5.4 Examples

5.4.1 Extrinsic Apoptosis Reaction Model

A family of Extrinsic Apoptosis Reaction Models (EARM), is described in [31] and can be found at https://earm.readthedocs.io/en/latest/index.html. The models are a test bed for hypothesis exploration of the extrinsic apoptosis system. Here we demonstrate the recapitulation of one of those models, EARM 2.0 Embedded, with HypBuilder. We consider the comparative ease of automatic model construction versus manual construction, and the extent of the library needed to build the model. The entire HypBuilder input file for EARM 2.0 Embedded is in appendix X.

The number of monomers in the HypBuilder model is 39 versus 23 for the manually constructed model. This reflects the fact that each monomer state in the manually constructed model is a separate monomer for the automatically constructed one, as we will shortly demonstrate. The number of rules is identical between the manual and automatically generated models and the structure of each, as well as the underlying mass-action ODEs, is effectively identical. Examples of the catalytic reaction for Caspase-8 induced cleavage of Bid from the HypBuilder input file, the associated automatically constructed model, and the manually constructed model are given below. Note that the input file for this reaction, and it's associated parameters is far shorter, and less complicated to write, than the associated rules and parameters it generates, or those for the manually constructed case. Overall the character count for the EARM input file is ~4900, while the automatic and manually generated models have character counts of ~39,400 and 23,500 respectively. Thus, writing this single model in HypBuilder results in more than a 5x reduction in required coding if the necessary library entries already exist. The absence of state sites in the HypBuilder version is apparent in the cleaved Bid protein. In the second automatically generated rule the monomer representing the un-cleaved protein (BidU) is destroyed and the monomer for the cleaved protein (BidT) is created. For the analogous manually created rule the single Bid monomer simply changes state from "state=U" to "state=T". Mathematically these two rules are equivalent. When both full models are constructed with the same

parameters they are undistinguishable (Figure 5.3). The molecular reaction library needed to automatically construct this model via HypBuilder requires only 18 molecule types and 10 reaction types (Appendix I).

HypBuilder input file

```
catalysis, C8A()[catalyzer], BidU()[substrate], BidT()[product],
7.8693418477379699e-06, 1.3478527255190856e-05, 0.20070834648071373
```

Automatically constructed rules

```
Rule('catalysis_0_C8A_catalyzer_BidU_substrate_BidT_product',
      C8A(BidU=None, BAR=None, C3pro=None) + BidU(C8A=None)
      | C8A(BidU=1, BAR=None, C3pro=None) % BidU(C8A=1),
      catalysis_0_C8A_catalyzer_BidU_substrate_BidT_product_2kf_0,
      catalysis_0_C8A_catalyzer_BidU_substrate_BidT_product_1kr_0)

Rule('catalysis_1_C8A_catalyzer_BidU_substrate_BidT_product',
      C8A(BidU=1, BAR=None, C3pro=None) % BidU(C8A=1)
      >> C8A(BidU=None, BAR=None, C3pro=None) + BidT(),
      catalysis_1_C8A_catalyzer_BidU_substrate_BidT_product_1kc_0)

Parameter('catalysis_0_C8A_catalyzer_BidU_substrate_BidT_prod-
uct_2kf_0', 7.86934184773797e-06)
Parameter('catalysis_0_C8A_catalyzer_BidU_substrate_BidT_prod-
uct_1kr_0', 1.3478527255190856e-05)
Parameter('catalysis_1_C8A_catalyzer_BidU_substrate_BidT_prod-
uct_1kc_0', 0.20070834648071373)
```

Manually constructed rules

```
Rule('bind_C8A_BidU_to_C8ABidU',
      C8(bf=None, state='A') + Bid(bf=None, state='U')
      <> C8(bf=1, state='A') % Bid(bf=1, state='U'),
      bind_C8A_BidU_to_C8ABidU_kf,
      bind_C8A_BidU_to_C8ABidU_kr)

Rule('catalyze_C8ABidU_to_C8A_BidT',
      C8(bf=1, state='A') % Bid(bf=1, state='U')
      | C8(bf=None, state='A') + Bid(bf=None, state='T'),
      catalyze_C8ABidU_to_C8A_BidT_kc)

Parameter('bind_C8A_BidU_to_C8ABidU_kf', 7.8693418477379699e-06)
Parameter('bind_C8A_BidU_to_C8ABidU_kr', 1.3478527255190856e-05)
Parameter('catalyze_C8ABidU_to_C8A_BidT_kc', 0.20070834648071373)
```

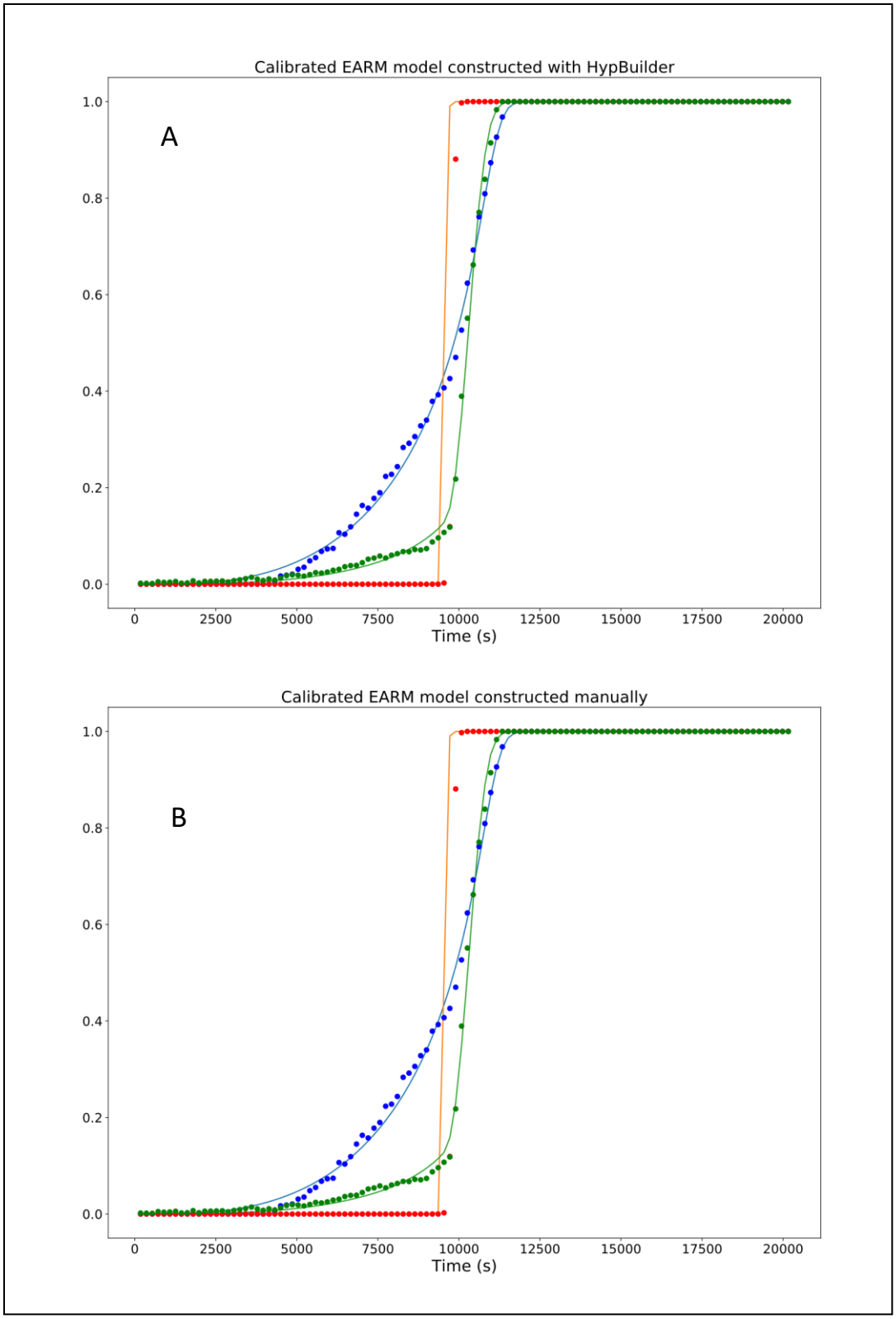


Figure 5.3. Comparison of manually and automatically constructed EARM models. (A) the automatically constructed model. (B) the manually constructed model.

5.4.2 Direct, Indirect, and Embedded MOMP Models

Mitochondrial Outer Membrane permeabilization is the key step in mitochondrial activation of apoptosis and three distinct models have been proposed for its execution (Figure 5.4). In the 'direct' model BH3-only activators, like Bid, act on and activate the effector pore-forming proteins like Bax and Bak. The anti-apoptosis regulators such as Bcl-2 then inhibit the activators. In the displacement (indirect) model the antiapoptotic regulators bind and inhibit the effectors rendering them inactive until an activator displaces the antiapoptotic protein by binding the antiapoptotic proteins. Finally, the embedded model combines properties of both the direct and indirect models. In this model the activators directly activate the effectors and the antiapoptotic proteins inhibit both the activators and the effector proteins.

For model selection and multimodel inference purposes one may want to compare different regulatory networks and for that, a separate model is required for each. Here we have used HypBuilder to automatically construct 805 Direct/Embedded models and 645 Indirect models using two HypBuilder input files (Appendix II). In each case all combinations of the inhibitory interactions (orange connections in Figure 5.4) were produced. The Direct/Embedded models contain direct interactions for activators and effector proteins (blue connections) while the Indirect models do not. Both cases take advantage of the 'data node' feature in HypBuilder to eliminate models with subnetworks that are disconnected from a node for which we have data, in this case SmacC. Both also make use of the 'text' feature in order to add increasing amounts of the activator Bid, roughly in line with data on that protein. In addition, the Indirect models must start out in a state in which the antiapoptotic proteins are bound to the effectors. Here we use the 'initial binding' feature in HypBuilder to randomly fill out the binding partners. In all, 1450 executable PySB models were generated with a minimal input effort. The molecular interaction library used here is the same as that for the EARM case above. All 1450 models can be found at https://github.com/LoLab-VU/BIND.

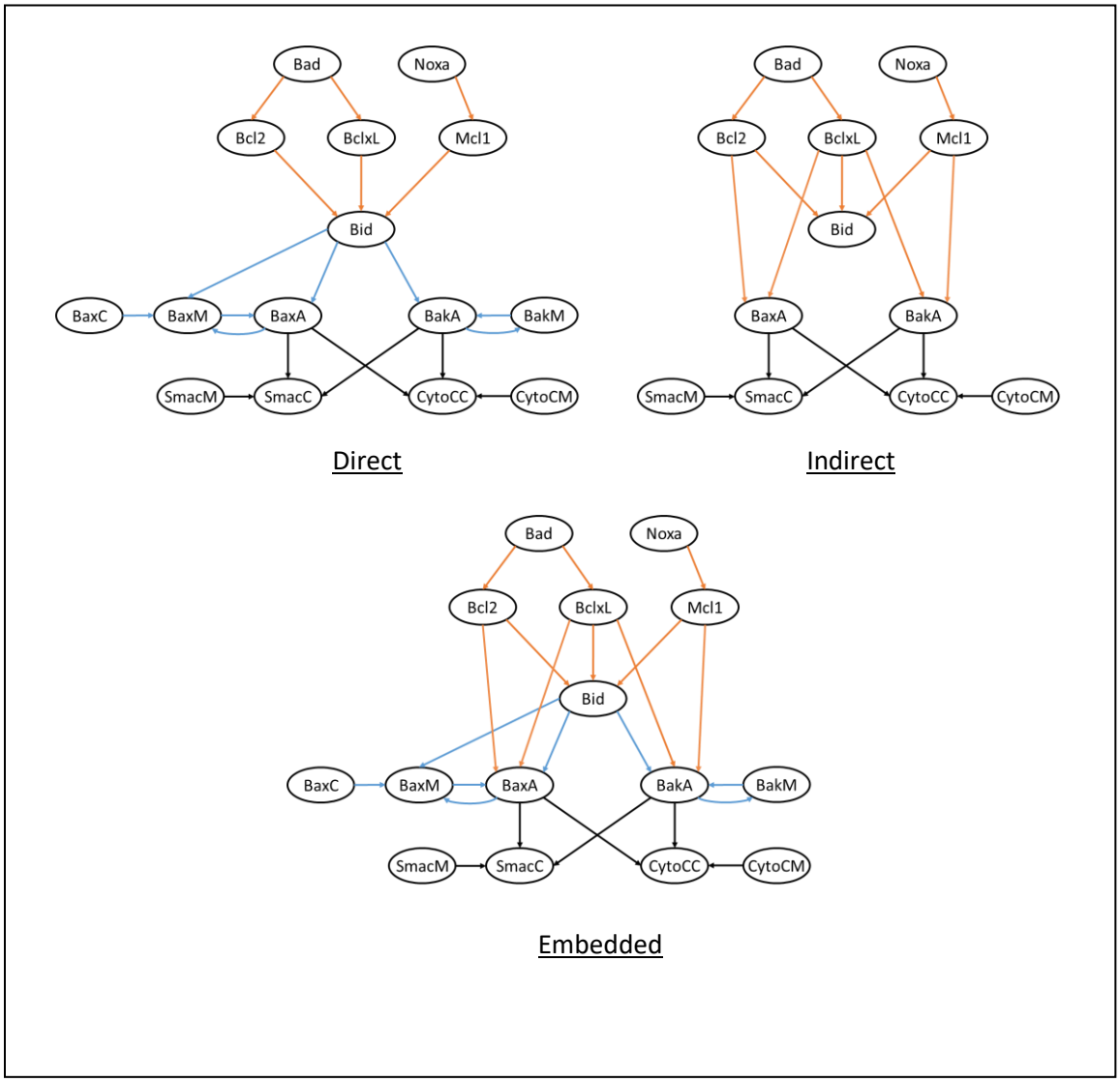


Figure 5.4. MOMP regulatory networks for the Direct, Indirect, and Embedded models.

Chapter 6

Discussion and Future Directions

6.1 Discussion

6.1.1 Bayesian Evidence-Based Exploration of Network Dynamics

Characterization of information flow through biological networks, the interactions between various pathways or network components, and shifts in phenotype upon regulatory perturbations is an exceedingly difficult task. Although comparative analysis of signal flow within a network is possible with current computational methods, the dependence of physicochemical models on unknown parameters makes the computational examination of each network component highly dependent on costly experimentation.

To take advantage of the enormous amount of existing knowledge encoded in these physicochemical reaction networks without the dependence on explicit parameter values we take a probabilistic approach to the inference of changes in network dynamics. By integrating an objective function that represents a simulated outcome over parameter distributions that represent the current knowledge of the parameters, we obtain the likelihood of attaining that outcome given all available information. Qualitative inference of network behavior for various in silico experimental setups and regulatory conditions are then attainable without explicit knowledge of every parameter value. The utility of the method was evident when applied to the regulation of extrinsic apoptosis. Networks that incorporate an active mitochondrial pathway displayed a higher resistance to apoptotic inhibition from increasing levels of XIAP, consistent with experimental evidence that XIAP induces a Type II phenotype [71]. Also in line with experimental evidence [93] are the results that suggest low/high signal initiation is consistent with Type II/I phenotype respectively and that both phenotypes achieve apoptosis equally well.

A limitation of the method is the computational cost. A number of factors affect the run time of the algorithm including size of the model, the objective function, and the desired precision. For example, the average estimated clock time of 76,981 seconds for the complete network under the PARP cleavage objective function and at a population size of 16,000 equates to nearly 714 CPU days for that entire run. Although the run time for the evidence calculation can vary greatly, it is highly correlated to the size of the model as was seen with the equivalent caspase pathway run at an average of 11,964 seconds. Fortunately, reducing the resolution (the number of sets of initial values for which an evidence value is estimated) and the precision (the population size) can drastically reduce the cost and in many cases the method will still be viable. One aspect of the method that is severely restrictive is the number of model components that can be varied in the same run since the computational cost increases exponentially with each additional variable. Reasonable parameter ranges must also be chosen. Information regarding the parameters can be incorporated into the evidence calculations by adjusting the range and shape of the priors. Here we used generic but biologically plausible ranges with uniform distributions in log space and produced results that were qualitatively consistent with previous experimental results.

In this work we have demonstrated a probabilistic approach to the qualitative analysis of the network dynamics of physicochemical models. It is designed to incorporate all available knowledge of the reaction topology, and the parameters on that topology, and calculate the likelihood of achieving an outcome of interest. Inferences on network dynamics are then made by repeating this calculation under changing regulatory conditions and various in silico experiments. We tested the method against a model of the extrinsic apoptosis system and produced results that were consistent with several lines of experimental research. To our knowledge this is the first attempt at such a probabilistic analysis of network dynamics for physicochemical models. We believe this method will prove valuable for the large-scale exploration of those dynamics, particularly when parameter knowledge and data are scarce.

6.1.2 Bayesian Evidence-Based Analysis of Extrinsic Apoptosis Regulatory Axes

Extrinsic apoptosis is a critical cellular process, important therapeutic target, and has been the focus of much study. Nonetheless, its regulation is not completely understood. In Chapter 4 we extended our characterization of the regulation of extrinsic apoptosis at three major regulatory axes: DISC driven activation of Caspase-8, XIAP induced inhibition of Caspase-3 activation, and the network of Bcl-2 family proteins that regulate MOMP.

We first considered the ratio of XIAP to its target of inhibition Caspase-3. We found that the effect of XIAP on the choice of phenotype (low XIAP being consistent with Type I while high levels are consistent with Type II) is considerably more pronounced at low levels of Caspase-3. The variation of Caspase-3 has its own effect on the dynamics of the system. As levels of Caspase-3 fall, the likelihood of achieving apoptosis through either the caspase pathway or the complete model drops, but this effect, particularly with moderate levels of XIAP, is less pronounced for the complete model. Thus, low levels of Caspase-3 favor a Type II pathway. Although differences in the topology of the models, and the resulting expected values for PARP cleavage, could result in a XIAP:Caspase-3 ratio dependent bias in pathway dominance, as predicted in Aldridge et al. [22], we observed no evidence of that under the conditions tested here. Overall, the general effect of both XIAP and Caspase-3 is the same for both models, but subtle differences create a more complex composite response to changes in the quantities of these two proteins. The combined effect dictates that the Type II pathway is favored under conditions with low Caspase-3 and moderate XIAP levels. Although both XIAP and Caspase-3 affect both the caspase pathway and the complete model in qualitatively the same way, it appears that the inclusion of mitochondrial involvement can help the system overcome impediments to apoptosis, which supports the hypothesis that MOMP is an apoptotic signal amplification mechanism [93].

Analysis of the effect of receptor count on the choice of pathway appears to support the hypothesis from Meng et al. [105] that a low receptor count favors the Type II pathway and could potentially be the determining factor. As with XIAP and Caspase-3, receptor count affects both networks in the same qualitative manner but to different degrees. A low

receptor count generally makes achieving apoptosis more difficult in both cases but affects the caspase pathway to a greater extent. The effect of receptor count is also synergistic with lower levels of the other DISC components Fadd and Procaspase-8, supporting the original hypothesis from Scaffidi et al. [93] that weak Caspase-8 activation at the DISC is characteristic of a Type II phenotype. The analysis of signal flux for the DISC components appears to indicate that lower DISC formation shifts the signal flux from the caspase to the mitochondrial pathway. This is in line with the analysis in Chapter 3 and supports a number of conclusions in Scaffidi et al. such as the order of activation of the various components and the equal effectiveness of both phenotypes for effecting apoptosis.

Analysis of MOMP regulation indicated that inhibition of antiapoptotic proteins with sensitizers, sensitizer peptides, or their mimetics would be incomplete using generic binding rate constants. The initial sampling run assumed the same range of parameter values for all binding reactions. Assignment of stronger binding affinities for the Bcl-2 protein regulators of MOMP, affinities more in line with experimentally derived K_D values, elicited both stronger antiapoptotic effects and a stronger recovery of apoptosis. Thus, existing knowledge can be used to shape and shift the prior distributions in order to obtain results that are more in line with experimental outcomes. The composition of the MOMP regulatory network and the quantities of its components will also greatly affect how much an inhibitor will increase the likelihood of apoptosis. Side effects, like the thrombocytopenia that results from Bcl-xl inhibition, must also be considered [121]. Such side effects could potentially be encoded into a computational model. Given quantitative information on all relevant MOMP regulatory proteins and the correct regulatory topology, the therapeutic targets that are most likely to result in the desired outcome could be hypothesized using the methods developed here. Follow up modeling with traditional methods along with experimentation could then be used for confirmation. A similar strategy could be used to target other extrinsic apoptotic signal regulators like XIAP and cFlip, or regulators of any other computationally modeled system.

Overall, the results produced by calculating expected values for cleaved PARP and pathway flux over changing conditions at the regulatory axes explored in this chapter proved to be consistent with experimental evidence, further solidifying our belief that this method will be a useful tool for the analysis of biological networks when rate parameters and the data to calibrate them are sparse.

6.1.3 Large Scale Construction of Physicochemical Models with HypBuilder

In Chapter 5 we presented HypBuilder, software for the large-scale production of executable mechanistic models in the PySB format. The capability to quickly and easily construct ensembles of models and model variants will enable model selection and in silico experiment sweeps at scales that were previously infeasible.

6.2 Future Directions

6.2.1 Expansion of Use Cases for the Probabilistic Analysis of Network Dynamics

In this work we have applied Bayesian-based estimation of expected values for quantities of interest to the extrinsic apoptosis network in the absence of time-series data. The method allowed us to explore network behavior, in a qualitative fashion, using generic parameter ranges instead of an explicit 'best fit' parameter set. However, we feel this approach *may* prove to be a more statistically rigorous analytical method than those currently in use. The general workflow would be to first use a Markov chain Monte Carlo method to fit posterior parameter distributions to available data, which can be done with readily available software like PyDREAM [37]. Those posteriors would then become priors for the subsequent evidence calculation. Follow up studies using this approach could explore a number of questions. For example, how much of the prior parameter space must we consider in order to make realistic predictions. Is a single best fit parameter set enough? Or an average of an ensemble of parameters sets consistent with experimental data [122, 123]? Or are there cases in which it would be prudent to consider the entirety of the prior distribution? How much data is necessary to foster accurate quantitative predictions is another

question. Less data means more parameter uncertainty in any calibration. In this work we've used our method without experimental data, relying only on generic ranges for our parameters. Although we view this as a 'first-pass' analysis, we saw very good qualitative consistency with experimental evidence. Additional data would narrow and shape the priors and further strengthen the predictive power. With more data our qualitative results would become more quantitative in nature. In general, the placement of data in the model affects which parameters become more constrained which, in turn, affects the questions one can confidently answer. A comparison, incorporating all these factors, with single parameter or ensemble methods would be instructive. Computation time would also need to be factored in. Rigor and accuracy must be weighed against the available resources and time constraints.

6.2.2 Exploration of Generated Biological Hypotheses

In addition to the consistency with existing experimental evidence, the analysis in Chapters 3 and 4 provided several novel hypotheses that could be explored further with additional experimentation. For example, the mechanism by which XIAP exerts phenotypic control over the apoptotic phenotype appears to simply be differential inhibition between the caspase pathway and the complete network. Another potential avenue of study is the potential mechanism by which the mitochondria effects amplification of the apoptosis signal. Under some conditions it appears to primarily work through inhibition of XIAP, subsequently allowing the signal through the direct caspase pathway. This mechanism appears to describe well the MCF-7 cell line and potentially represents a third extrinsic apoptosis phenotype where Type I is characterized by the complete independence of the mitochondria and Type II by full routing of the signal through the mitochondrial pathway. Control of the signal route by the DISC, and specifically the receptor, is another potential subject of study, as is the potential relationship between signal strength, XIAP, and the specific mechanism of apoptosis signal transduction. Lastly, our hypothesis that BH3 memetic compounds exert the majority of their effects up to doses in line with the concentration of

their targets could be experimentally tested, with potential significance with regard to the dosage of such drugs.

6.2.3 HypBuilder Development

HypBuilder development will be ongoing. The goal is to provide an interface for PySB that simplifies model construction and allows for large-scale model building while retaining all functionality that PySB provides. Full PySB functionality currently requires additions to the software such as support for Expressions, a means to add non-mass-action-based rate laws to a model.

BIBLIOGRAPHY

1. Bhalla, U. S. & Iyengar, R. Emergent Properties of Networks of Biological Signaling Pathways. *Science* **283**, 381–387 (1999).

2. Loscalzo, J. & Barabasi, A.-L. Systems biology and the future of medicine. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **3**, 619–627 (2011).

3. Kitano, H. Computational systems biology. *Nature* **420**, 206–210 (2002).

4. Osborn, O. & Olefsky, J. M. The cellular and signaling networks linking the immune system and metabolism in disease. *Nat Med* **18**, 363–374 (2012).

5. Harper, J. A., Yuan, J. S., Tan, J. B., Visan, I. & Guidos, C. J. Notch signaling in development and disease. *Clinical Genetics* **64**, 461–472 (2003).

6. Hornberg, J. J., Bruggeman, F. J., Westerhoff, H. V. & Lankelma, J. Cancer: A Systems Biology disease. *Biosystems* **83**, 81–90 (2006).

7. Kreeger, P. K. & Lauffenburger, D. A. Cancer systems biology: a network modeling perspective. *Carcinogenesis* **31**, 2–8 (2010).

8. Kolch, W., Halasz, M., Granovskaya, M. & Kholodenko, B. N. The dynamic control of signal transduction networks in cancer cells. *Nature Reviews Cancer* **15**, 515–527 (2015).

9. Balkwill, F. R., Capasso, M. & Hagemann, T. The tumor microenvironment at a glance. *J Cell Sci* **125**, 5591–5596 (2012).

10. Zou, J., Zheng, M.-W., Li, G. & Su, Z.-G. Advanced Systems Biology Methods in Drug Discovery and Translational Biomedicine. *BioMed Research International*

11. Xie, L. *et al.* Towards Structural Systems Pharmacology to Study Complex Diseases and Personalized Medicine. *PLOS Computational Biology* **10**, e1003554 (2014).

12. Wang, Z. & Deisboeck, T. S. Mathematical modeling in cancer drug discovery. *Drug Discovery Today* **19**, 145–150 (2014). https://www.hindawi.com/journals/bmri/2013/742835/ (2013) doi:10.1155/2013/742835.

13. Jia, J. *et al.* Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov* **8**, 111–128 (2009).

14. Rajasethupathy, P., Vayttaden, S. J. & Bhalla, U. S. Systems modeling: a pathway to drug discovery. *Current Opinion in Chemical Biology* **9**, 400–406 (2005).

15. Can the pharmaceutical industry reduce attrition rates? | Nature Reviews Drug Discovery. https://www.nature.com/articles/nrd1470.

16. Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* **9**, 203–214 (2010).

17. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology* **15**, 81–94 (2018).

18. Marusyk, A. & Polyak, K. Tumor heterogeneity: Causes and consequences. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1805**, 105–117 (2010).

19. Korcsmaros, T., Schneider, M. V. & Superti-Furga, G. Next generation of network medicine: interdisciplinary signaling approaches. *Int Bio (Cam)* **9**, 97–108 (2017).

20. Riel, V. & A.w, N. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Brief Bioinform* **7**, 364–374 (2006).

21. Kurz, F. T. *et al.* Network dynamics: quantitative analysis of complex behavior in metabolism, organelles, and cells, from experiments to models and back. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **9**, e1352 (2017).

22. Aldridge, B. B., Burke, J. M., Lauffenburger, D. A. & Sorger, P. K. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol* **8**, 1195–1203 (2006).

23. Le Novère, N. Quantitative and logic modelling of molecular and gene networks. *Nature Reviews Genetics* **16**, 146–158 (2015).

24. Machado, D. *et al.* Modeling formalisms in Systems Biology. *AMB Express* **1**, 45 (2011).

25. Saadatpour, A. & Albert, R. A comparative study of qualitative and quantitative dynamic models of biological regulatory networks. *EPJ Nonlinear Biomed Phys* **4**, 5 (2016).

26. Chylek, L. A. *et al.* Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **6**, 13–36 (2014).

27. Faeder, J. R., Blinov, M. L. & Hlavacek, W. S. Rule-Based Modeling of Biochemical Systems with BioNetGen. in *Systems Biology* (ed. Maly, I. V.) 113–167 (Humana Press, 2009). doi:10.1007/978-1-59745-525-1_5.

28. Harris, L. A. *et al.* BioNetGen 2.2: advances in rule-based modeling. *Bioinformatics* **32**, 3366–3368 (2016).

29. Boutillier, P. *et al.* The Kappa platform for rule-based modeling. *Bioinformatics* **34**, i583–i592 (2018).

30. Danos, V. & Laneve, C. Formal molecular biology. *Theoretical Computer Science* **325**, 69–110 (2004).

31. Lopez, C. F., Muhlich, J. L., Bachman, J. A. & Sorger, P. K. Programming biological models in Python using PySB. *Molecular Systems Biology* **9**, 646 (2013).

32. Moles, C. G., Mendes, P. & Banga, J. R. Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods. *Genome Res.* **13**, 2467–2474 (2003).

33. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by Simulated Annealing. *Science* **220**, 671–680 (1983).

34. Eberhart, R. & Kennedy, J. A new optimizer using particle swarm theory. in *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science* 39–43 (1995). doi:10.1109/MHS.1995.494215.

35. Whitley, D. A genetic algorithm tutorial. *Stat Comput* **4**, 65–85 (1994).

36. Vrugt, J. A. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling and Software* **75**, 273–316 (2016).

37. Shockley, E. M., Vrugt, J. A. & Lopez, C. F. PyDREAM: high-dimensional parameter inference for biological models in python. *Bioinformatics* **34**, 695–697 (2018).

38. Skilling, J. Nested sampling for general Bayesian computation. *Bayesian Anal.* **1**, 833–859 (2006).

39. Sivia, D. & Skilling, J. *Data Analysis: A Bayesian Tutorial*. (OUP Oxford, 2006).

40. Schwarz, G. Estimating the Dimension of a Model. *Ann. Statist.* **6**, 461–464 (1978).

41. Lartillot, N. & Philippe, H. Computing Bayes Factors Using Thermodynamic Integration. *Syst Biol* **55**, 195–207 (2006).

42. Feroz, F. & Hobson, M. P. Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis. *Monthly Notices of the Royal Astronomical Society* **384**, 449–463 (2008).

43. Feroz, F., Hobson, M. P. & Bridges, M. MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Mon Not R Astron Soc* **398**, 1601–1614 (2009).

44. Handley, W. J., Hobson, M. P. & Lasenby, A. N. PolyChord: nested sampling for cosmology. *Monthly Notices of the Royal Astronomical Society: Letters* **450**, L61–L65 (2015).

45. Handley, W. J., Hobson, M. P. & Lasenby, A. N. PolyChord: next-generation nested sampling. *Mon. Not. R. Astron. Soc.* **453**, 4385–4399 (2015).

46. Brewer, B. J. & Foreman-Mackey, D. DNest4: Diffusive Nested Sampling in C++ and Python. *Journal of Statistical Software* **86**, 1–33 (2018).

47. Brewer, B. J., Pártay, L. B. & Csányi, G. Diffusive Nested Sampling. *arXiv:0912.2380 [astro-ph, physics:physics, stat]* (2009).

48. Higson, E., Handley, W., Hobson, M. & Lasenby, A. Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation. *Stat Comput* **29**, 891–913 (2019).

49. Speagle, J. S. dynesty: A Dynamic Nested Sampling Package for Estimating Bayesian Posteriors and Evidences. *arXiv:1904.02180 [astro-ph, stat]* (2019).

50. Feroz, F., Hobson, M. P., Cameron, E. & Pettitt, A. N. Importance Nested Sampling and the MultiNest Algorithm. *arXiv:1306.2144 [astro-ph, physics:physics, stat]* (2013).

51. Ashkenazi, A. & Dixit, V. M. Death Receptors: Signaling and Modulation. *Science* **281**, 1305–1308 (1998).

52. Boldin, M. P. *et al.* A Novel Protein That Interacts with the Death Domain of Fas/APO1 Contains a Sequence Motif Related to the Death Domain. *J. Biol. Chem.* **270**, 7795–7798 (1995).

53. Kischkel, F. C. *et al.* Apo2L/TRAIL-Dependent Recruitment of Endogenous FADD and Caspase-8 to Death Receptors 4 and 5. *Immunity* **12**, 611–620 (2000).

54. Sprick, M. R. *et al.* FADD/MORT1 and Caspase-8 Are Recruited to TRAIL Receptors 1 and 2 and Are Essential for Apoptosis Mediated by TRAIL Receptor 2. *Immunity* **12**, 599–609 (2000).

55. Krueger, A., Schmitz, I., Baumann, S., Krammer, P. H. & Kirchhoff, S. Cellular FLICE-inhibitory Protein Splice Variants Inhibit Different Steps of Caspase-8 Activation at the CD95 Death-inducing Signaling Complex. *J. Biol. Chem.* **276**, 20633–20640 (2001).

56. Salvesen, G. S. & Dixit, V. M. Caspase activation: The induced-proximity model. *PNAS* **96**, 10964–10967 (1999).

57. Martin, D. A., Siegel, R. M., Zheng, L. & Lenardo, M. J. Membrane Oligomerization and Cleavage Activates the Caspase-8 (FLICE/MACHα1) Death Signal. *J. Biol. Chem.* **273**, 4345–4349 (1998).

58. Boatright, K. M. & Salvesen, G. S. Mechanisms of caspase activation. *Current Opinion in Cell Biology* **15**, 725–731 (2003).

59. Stennicke, H. R. *et al.* Pro-caspase-3 Is a Major Physiologic Target of Caspase-8. *J. Biol. Chem.* **273**, 27084–27090 (1998).

60. Li, H., Zhu, H., Xu, C. & Yuan, J. Cleavage of BID by Caspase 8 Mediates the Mitochondrial Damage in the Fas Pathway of Apoptosis. *Cell* **94**, 491–501 (1998).

61. Luo, X., Budihardjo, I., Zou, H., Slaughter, C. & Wang, X. Bid, a Bcl2 Interacting Protein, Mediates Cytochrome c Release from Mitochondria in Response to Activation of Cell Surface Death Receptors. *Cell* **94**, 481–490 (1998).

62. Desagher, S. *et al.* Bid-induced Conformational Change of Bax Is Responsible for Mitochondrial Cytochrome c Release during Apoptosis. *The Journal of Cell Biology* **144**, 891–901 (1999).

63. Kelekar, A. & Thompson, C. B. Bcl-2-family proteins: the role of the BH3 domain in apoptosis. *Trends in Cell Biology* **8**, 324–330 (1998).

64. Oltval, Z. N., Milliman, C. L. & Korsmeyer, S. J. Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programed cell death. *Cell* **74**, 609–619 (1993).

65. Leber, B., Lin, J. & Andrews, D. W. Embedded together: The life and death consequences of interaction of the Bcl-2 family with membranes. *Apoptosis* **12**, 897–911 (2007).

66. Letai, A. *et al.* Distinct BH3 domains either sensitize or activate mitochondrial apoptosis, serving as prototype cancer therapeutics. *Cancer Cell* **2**, 183–192 (2002).

67. Yang, E. *et al.* Bad, a heterodimeric partner for Bcl-xL and Bcl-2, displaces bax and promotes cell death. *Cell* **80**, 285–291 (1995).

68. Kale, J., Osterlund, E. J. & Andrews, D. W. BCL-2 family proteins: changing partners in the dance towards death. *Cell Death Differ* **25**, 65–80 (2018).

69. Tewari, M. *et al.* Yama/CPP32β, a mammalian homolog of CED-3, is a CrmA-inhibitable protease that cleaves the death substrate poly(ADP-ribose) polymerase. *Cell* **81**, 801–809 (1995).

70. Nicholson, D. W. *et al.* Identification and inhibition of the ICE/CED-3 protease necessary for mammalian apoptosis. *Nature* **376**, 37–43 (1995).

71. Jost, P. J. *et al.* XIAP discriminates between type I and type II FAS-induced apoptosis. *Nature* **460**, 1035–1039 (2009).

72. Huang, Y. *et al.* Structural Basis of Caspase Inhibition by XIAP: Differential Roles of the Linker versus the BIR Domain. *Cell* **104**, 781–790 (2001).

73. Suzuki, Y., Nakabayashi, Y. & Takahashi, R. Ubiquitin-protein ligase activity of X-linked inhibitor of apoptosis protein promotes proteasomal degradation of caspase-3 and enhances its anti-apoptotic effect in Fas-induced cell death. *PNAS* **98**, 8662–8667 (2001).

74. Shiozaki, E. N. *et al.* Mechanism of XIAP-Mediated Inhibition of Caspase-9. *Molecular Cell* **11**, 519–527 (2003).

75. Zou, H., Li, Y., Liu, X. & Wang, X. An APAF-1·Cytochrome c Multimeric Complex Is a Functional Apoptosome That Activates Procaspase-9. *J. Biol. Chem.* **274**, 11549–11556 (1999).

76. Adrain, C., Creagh, E. M. & Martin, S. J. Apoptosis-associated release of Smac/DIABLO from mitochondria requires active caspases and is blocked by Bcl-2. *The EMBO Journal* **20**, 6627–6636 (2001).

77. Cowling, V. & 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).

78. Dai, H. *et al.* Measurement of BH3-only protein tolerance. *Cell Death Differ* **25**, 282–293 (2018).

79. Eissing, T. *et al.* Bistability Analyses of a Caspase Activation Model for Receptor-induced Apoptosis. *J. Biol. Chem.* **279**, 36892–36897 (2004).

80. Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M. & Sorger, P. K. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459**, 428–432 (2009).

81. Buchner, J. *et al.* X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *A&A* **564**, A125 (2014).

82. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. (Springer-Verlag, 2002). doi:10.1007/b97636.

83. Symonds, M. R. E. & Moussalli, A. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behav Ecol Sociobiol* **65**, 13–21 (2011).

84. Albeck, J. G., Burke, J. M., Spencer, S. L., Lauffenburger, D. A. & Sorger, P. K. Modeling a Snap-Action, Variable-Delay Switch Controlling Extrinsic Cell Death. *PLOS Biology* **6**, e299 (2008).

85. Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M. & Sorger, P. K. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459**, 428–432 (2009).

86. Eydgahi, H. *et al.* Properties of cell death models calibrated and compared using Bayesian approaches. *Molecular Systems Biology* **9**, 644 (2013).

87. Aldridge, B. B., Gaudet, S., Lauffenburger, D. A. & Sorger, P. K. Lyapunov exponents and phase diagrams reveal multi-factorial control over TRAIL-induced apoptosis. *Molecular Systems Biology* **7**, 553 (2011).

88. Raychaudhuri, S. & Raychaudhuri, S. C. Monte Carlo Study Elucidates the Type 1/Type 2 Choice in Apoptotic Death Signaling in Healthy and Cancer Cells. *Cells* **2**, 361–392 (2013).

89. Aitken, S. & Akman, O. E. Nested sampling for parameter inference in systems biology: application to an exemplar circadian model. *BMC Systems Biology* **7**, 72 (2013).

90. Pullen, N. & Morris, R. J. Bayesian Model Comparison and Parameter Inference in Systems Biology Using Nested Sampling. *PLOS ONE* **9**, e88419 (2014).

91. Xu, T.-R. *et al.* Inferring Signaling Pathway Topologies from Multiple Perturbation Measurements of Specific Biochemical Species. *Sci. Signal.* **3**, ra20–ra20 (2010).

92. MacKay, D. J. C. & Kay, D. J. C. M. *Information Theory, Inference and Learning Algorithms*. (Cambridge University Press, 2003).

93. Scaffidi, C. *et al.* Two CD95 (APO-1/Fas) signaling pathways. *The EMBO Journal* **17**, 1675–1687 (1998).

94. Vaux, D. L. & Strasser, A. The molecular biology of apoptosis. *PNAS* **93**, 2239–2244 (1996).

95. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).

96. Thompson, C. B. Apoptosis in the pathogenesis and treatment of disease. *Science* **267**, 1456–1462 (1995).

97. Brown, J. M. & Attardi, L. D. The role of apoptosis in cancer development and treatment response. *Nat Rev Cancer* **5**, 231–237 (2005).

98. Fesik, S. W. Promoting apoptosis as a strategy for cancer drug discovery. *Nat Rev Cancer* **5**, 876–885 (2005).

99. Certo, M. *et al.* Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. *Cancer Cell* **9**, 351–365 (2006).

100. Deng, J. *et al.* BH3 Profiling Identifies Three Distinct Classes of Apoptotic Blocks to Predict Response to ABT-737 and Conventional Chemotherapeutic Agents. *Cancer Cell* **12**, 171–185 (2007).

101. Scaffidi, C. *et al.* Differential Modulation of Apoptosis Sensitivity in CD95 Type I and Type II Cells. *J. Biol. Chem.* **274**, 22532–22538 (1999).

102. Rudner, J. *et al.* Type I and type II reactions in TRAIL-induced apoptosis – results from dose–response studies. *Oncogene* **24**, 130–140 (2005).

103. Özören, N. & El-Deiry, W. S. Defining Characteristics of Types I and II Apoptotic Cells in Response to TRAIL. *Neoplasia* **4**, 551–557 (2002).

104. Green, D. R. & Reed, J. C. Mitochondria and Apoptosis. *Science* **281**, 1309–1312 (1998).

105. Meng, X. W. *et al.* High Cell Surface Death Receptor Expression Determines Type I Versus Type II Signaling. *J. Biol. Chem.* **286**, 35823–35833 (2011).

106. Vaux, D. L., Cory, S. & Adams, J. M. Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells. *Nature* **335**, 440–442 (1988).

107. Lomonosova, E. & Chinnadurai, G. BH3-only proteins in apoptosis and beyond: an overview. *Oncogene* **27**, S2–S19 (2008).

108. Westphal, D., Dewson, G., Czabotar, P. E. & Kluck, R. M. Molecular biology of Bax and Bak activation and action. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1813**, 521–531 (2011).

109. Wei, M. C. *et al.* Proapoptotic BAX and BAK: A Requisite Gateway to Mitochondrial Dysfunction and Death. *Science* **292**, 727–730 (2001).

110. Shamas-Din, A., Brahmbhatt, H., Leber, B. & Andrews, D. W. BH3-only proteins: Orchestrators of apoptosis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1813**, 508–520 (2011).

111. Souers, A. J. *et al.* ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nature Medicine* **19**, 202–208 (2013).

112. Del Gaizo Moore, V., Schlis, K. D., Sallan, S. E., Armstrong, S. A. & Letai, A. BCL-2 dependence and ABT-737 sensitivity in acute lymphoblastic leukemia. *Blood* **111**, 2300–2309 (2008).

113. Ryan, J. & Letai, A. BH3 profiling in whole cells by fluorimeter or FACS. *Methods* **61**, 156–164 (2013).

114. Touzeau, C. *et al.* BH3 profiling identifies heterogeneous dependency on Bcl-2 family members in multiple myeloma and predicts sensitivity to BH3 mimetics. *Leukemia* **30**, 761–764 (2016).

115. Letai, A. Apoptosis and Cancer. *Annual Review of Cancer Biology* **1**, 275–294 (2017).

116. Ku, B., Liang, C., Jung, J. U. & Oh, B.-H. Evidence that inhibition of BAX activation by BCL-2 involves its tight and preferential interaction with the BH3 domain of BAX. *Cell Res* **21**, 627–641 (2011).

117. Funahashi, A., Morohashi, M., Kitano, H. & Tanimura, N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. in (2003). doi:10.1016/s1478-5382(03)02370-9.

118. Moraru, I. I. *et al.* Virtual Cell modelling and simulation software environment. *IET Systems Biology* **2**, 352–362 (2008).

119. Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).

120. Le Novère, N. *et al.* BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res* **34**, D689–D691 (2006).

121. Qiao, J. et al. Imbalanced expression of Bcl-xL and Bax in platelets treated with plasma from immune thrombocytopenia. Immunol Res 64, 604–609 (2016).

122. Battogtokh, D., Asch, D. K., Case, M. E., Arnold, J. & Schüttler, H.-B. An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of Neurospora crassa. PNAS 99, 16904–16909 (2002).

123. Brown, K. S. & Sethna, J. P. Statistical mechanical approaches to models with many poorly known parameters. Phys. Rev. E 68, 021904 (2003).

APPENDICES

Appendix I: HypBuilder version of the Extrinsic Apoptosis Reaction Model 2.0 Embedded and the molecular reaction library needed to construct it.

EARM 2.0 Embedded for construction with HypBuilder

```
model components

L, 3000
R, 200
DISC, 0
C8A, 0
C8pro, 20000
C6A, 0
C6pro, 10000
BidU, 40000
BidT, 0
BidM, 0
flip, 100
BAR, 1000
SmacM, 100000
SmacC, 0
SmacA, 0
CytoCM, 500000
CytoCC, 0
CytoCA, 0
ApafI, 100000
ApafA, 0
C9, 100000
Apop, 0
C3A, 0
C3pro, 10000
C3ub, 0
XIAP, 100000
PARPU, 1000000
PARPC, 0
BaxC, 80000
BaxM, 0
BaxA, 0
BclxLC, 20000
BclxLM, 0
BakM, 20000
BakA, 0
Bcl2, 20000
Mcl1, 20000
Bad, 1000
Noxa, 0
```

required reactions

dimerization, L()[subunit_a], R()[subunit_b], DISC()[dimer], 4.8610717169040878e-09, 6.9984651157346155e-05, 0.0001004038401598357
catalysis, DISC()[catalyzer], C8pro()[substrate], C8A()[product], 1.652610985853028e-06, 0.00025515949741149136, 8.1035438307450976
catalysis, C8A()[catalyzer], BidU()[substrate], BidT()[product], 7.8693418477379699e-06, 1.3478527255190856e-05, 0.20070834648071373
inhibition, flip()[inhibitor], DISC()[inh_target], 2.0857207094671021e-06, 0.072993081874092425
inhibition, BAR()[inhibitor], C8A()[inh_target], 2.450754439292685e-05, 3.176107099863868e-05
equilibration, SmacC()[equil_a], SmacA()[equil_b], 0.68667658617411753, 0.00077853445949019686
equilibration, CytoCC()[equil_a], CytoCA()[equil_b], 0.020412343795425145, 0.0025071607534728909
catalysis, CytoCA()[catalyzer], ApafI()[substrate], ApafA()[product], 4.5927585852137639e-07, 2.8999962220310582e-05, 0.040599717918061738
conversion, C9()[subunit_d], ApafA()[subunit_c], Apop()[complex], 3.213196857547145e-07, 0.011684477645560146
catalysis, Apop()[catalyzer], C3pro()[substrate], C3A()[product], 7.8037691005466737e-10, 0.00032802481823310378, 1.1684643810480766
inhibition, XIAP()[inhibitor], Apop()[inh_target], 2.0704535531682789e-08, 0.0013859740873431536
inhibition, SmacA()[inhibitor], XIAP()[inh_target], 0.00040944017917514478, 0.0080051134954001509
catalysis, C8A()[catalyzer], C3pro()[substrate], C3A()[product], 1.9167619114268919e-06, 0.00010186532359752733, 2.7002149431488567
catalysis, XIAP()[catalyzer], C3A()[substrate], C3ub()[product], 1.8247861512403385e-07, 1.2921875123372521e-05, 0.099504980631579837
catalysis, C3A()[catalyzer], PARPU()[substrate], PARPC()[product], 1.889526802914492e-06, 0.00011387916767652634, 13.768204206283402
catalysis, C3A()[catalyzer], C6pro()[substrate], C6A()[product], 2.1334646851306905e-08, 0.0015363069332776634, 13.054054125378238
catalysis, C6A()[catalyzer], C8pro()[substrate], C8A()[product], 5.0560841288642224e-08, 9.1977396575087457e-05, 0.18488607021092679
equilibration, BidT()[equil_a], BidM()[equil_b], 0.026956531627050594, 1.1783387334989619e-05
equilibration, BaxC()[equil_a], BaxM()[equil_b], 0.0003405873853519951, 0.049750332184525257
equilibration, BclxLC()[equil_a], BclxLM()[equil_b], 0.068400564450976198, 0.14118753123479355
catalysis, BidM()[catalyzer], BaxM()[substrate], BaxA()[product], 3.9736775946970058e-09, 0.021482642730885865, 1.8203763301683924
catalysis, BidM()[catalyzer], BakM()[substrate], BakA()[product], 1.9408461301064605e-09, 0.00018094808868168103, 0.011177441271776072
self_catalyze, BaxA()[self_catalyzer], BaxM()[self_substrate], 2.0149060250585873e-06, 0.040311275999739105, 0.012192887363178946
self_catalyze, BakA()[self_catalyzer], BakM()[self_substrate], 5.3939824626322183e-06, 0.002166628073982069, 0.26475473643878961
inhibition, Bcl2()[inhibitor], BidM()[inh_target], 1.0277423337470143e-06, 2.7905757973936867
inhibition, BclxLM()[inhibitor], BidM()[inh_target], 1.6367225985110056e-07, 0.026264964906836046
inhibition, Mcl1()[inhibitor], BidM()[inh_target], 1.9428423606297654e-06, 0.0021648330129469106
inhibition, Bcl2()[inhibitor], BaxA()[inh_target], 1.4531081303335626e-08, 0.071917529283017381
inhibition, BclxLM()[inhibitor], BaxA()[inh_target], 1.2086000195405126e-06, 0.019684909992630065
inhibition, BclxLM()[inhibitor], BakA()[inh_target], 6.2674545719585416e-06, 0.067298588389140615
inhibition, Mcl1()[inhibitor], BakA()[inh_target], 1.2661363465041514e-05, 0.0002157216265888175
inhibition, Bad()[inhibitor], Bcl2()[inh_target], 5.3931067181433528e-05, 0.44013378070735282
inhibition, Bad()[inhibitor], BclxLM()[inh_target], 4.1877529969310183e-05, 0.16263349657520218
inhibition, Noxa()[inhibitor], Mcl1()[inh_target], 6.1197384627538123e-07, 0.01966269798058537
pore_formation, BaxA()[pore], 0.0067429287574002307, 1.6822795105298417e-05, 0.00029378841202344446, 0.000144616, 0.000085651, 0.002585238
pore_formation, BakA()[pore], 1.6054164088484109e-05, 0.00023979936345611174, 5.7415804556553368e-06, 0.021389853, 1.509873703e-05, 1.538796535e-05
transport, BaxA()[pore], CytoCM()[cargo_M], CytoCC()[cargo_C], 0.000106028, 0.0013725161741816323, 727.3165367902302
transport, BaxA()[pore], SmacM()[cargo_M], SmacC()[cargo_C], 2.375382573e-06, 0.00036936772401878048, 83.182122265178961
transport, BakA()[pore], CytoCM()[cargo_M], CytoCC()[cargo_C], 0.354927953e-06, 0.051294040280582424, 1.0588991714689346
transport, BakA()[pore], SmacM()[cargo_M], SmacC()[cargo_C], 2.40631696e-05, 0.00061884308859702392, 16.739408086871173

Molecular interaction library required for automatic EARM model construction

```
molecule: pore
reaction: pore_formation
template: pore(pore_1=None, pore_2=None)
      + pore(pore_1=None, pore_2=None)
      <> pore(pore_1=None, pore_2=1)
      % pore(pore_1=1, pore_2=None)
template: pore(pore_1=None, pore_2=None)
      + pore(pore_1=None, pore_2=1)
      % pore(pore_1=1, pore_2=None)
      <> pore(pore_1=3, pore_2=1)
      % pore(pore_1=1, pore_2=2)
      % pore(pore_1=2, pore_2=3)
template: pore(pore_1=None, pore_2=None)
      + pore(pore_1=3, pore_2=1)
      % pore(pore_1=1, pore_2=2)
      % pore(pore_1=2, pore_2=3)
      <> pore(pore_1=4, pore_2=1)
      % pore(pore_1=1, pore_2=2)
      % pore(pore_1=2, pore_2=3)
      % pore(pore_1=3, pore_2=4)
$$$
+++

molecule: cargo_M
+++

molecule: cargo_C
reaction: transport
template: pore(pore_1=4, pore_2=1, cargo_M=None)
      % pore(pore_1=1, pore_2=2, cargo_M=None)
      % pore(pore_1=2, pore_2=3, cargo_M=None)
      % pore(pore_1=3, pore_2=4, cargo_M=None)
      + cargo_M(pore=None)
      <> pore(pore_1=4, pore_2=1, cargo_M=None)
      % pore(pore_1=1, pore_2=2, cargo_M=None)
      % pore(pore_1=2, pore_2=3, cargo_M=None)
      % pore(pore_1=3, pore_2=4, cargo_M=5)
      % cargo_M(pore=5)
template: pore(pore_1=4, pore_2=1, cargo_M=None)
      % pore(pore_1=1, pore_2=2, cargo_M=None)
      % pore(pore_1=2, pore_2=3, cargo_M=None)
      % pore(pore_1=3, pore_2=4, cargo_M=5)
      % cargo_M(pore=5)
      >> pore(pore_1=4, pore_2=1, cargo_M=None)
      % pore(pore_1=1, pore_2=2, cargo_M=None)
      % pore(pore_1=2, pore_2=3, cargo_M=None)
      % pore(pore_1=3, pore_2=4, cargo_M=None)
      + cargo_C()
$$$
+++
```

```
molecule: inhibitor
+++

molecule: inh_target
reaction: inhibition
template: inhibitor(inh_target=None)
      + inh_target(inhibitor=None)
      <> inhibitor(inh_target=1)
      % inh_target(inhibitor=1)
$$$
+++

molecule: subunit_a
+++

molecule: subunit_b
+++

molecule: dimer
reaction: dimerization
template: subunit_a(subunit_b=None)
      + subunit_b(subunit_a=None)
      <> subunit_a(subunit_b=1)
      % subunit_b(subunit_a=1)
template: subunit_a(subunit_b=1)
      % subunit_b(subunit_a=1)
      >> dimer()
$$$
+++

molecule: subunit_c
+++

molecule: subunit_d
+++

molecule: complex
reaction: conversion
template: subunit_c()
      + subunit_d()
      <> complex()
$$$
+++
```

```
molecule: self_substrate
reaction: self_catalyze
template: self_catalyzer(self_substrate=None)
      + self_substrate(self_catalyzer=None)
      <> self_catalyzer(self_substrate=1)
      % self_substrate(self_catalyzer=1)
template: self_catalyzer(self_substrate=1)
      % self_substrate(self_catalyzer=1)
      >> self_catalyzer(self_substrate=None)
      + self_catalyzer(self_substrate=None)
$$$
+++

molecule: self_catalyzer
+++

molecule: catalyzer
+++

molecule: substrate
+++

molecule: product
reaction: catalysis
template: catalyzer(substrate=None)
      + substrate(catalyzer=None)
      <> catalyzer(substrate=1)
      % substrate(catalyzer=1)
template: catalyzer(substrate=1)
      % substrate(catalyzer=1)
      >> catalyzer(substrate=None)
      + product()
$$$
+++

molecule: equil_a
+++

molecule: equil_b
reaction: equilibration
template: equil_a()
      <> equil_b()
$$$
+++
```

Appendix II: HypBuilder version of Indirect and Direct/Embedded models.

Indirect Models

```
model components

Bid, 0.000001
SmacM, 100000
SmacC, 0, {d}
CytoCM, 500000
CytoCC, 0
BaxA, 5000
BclxL, 113000
BakA, 35000
Bcl2, 157000
Mcl1, 57000
Bad, 39000
Noxa, 6000


required reactions

pore_formation, BaxA()[pore]
pore_formation, BakA()[pore]
transport, BaxA()[pore], SmacM()[cargo_M], SmacC()[cargo_C]
transport, BakA()[pore], SmacM()[cargo_M], SmacC()[cargo_C]
transport, BaxA()[pore], CytoCM()[cargo_M], CytoCC()[cargo_C]
transport, BakA()[pore], CytoCM()[cargo_M], CytoCC()[cargo_C]

optional reactions

inhibition, Bcl2()[inhibitor], Bid()[inh_target]
inhibition, BclxL()[inhibitor], Bid()[inh_target]
inhibition, Mcl1()[inhibitor], Bid()[inh_target]
inhibition, Bcl2()[inhibitor], BaxA()[inh_target], {f:all}
inhibition, BclxL()[inhibitor], BaxA()[inh_target], {f:all}
inhibition, BclxL()[inhibitor], BakA()[inh_target], {f:all}
inhibition, Mcl1()[inhibitor], BakA()[inh_target], {f:all}
inhibition, Bad()[inhibitor], Bcl2()[inh_target]
inhibition, Bad()[inhibitor], BclxL()[inh_target]
inhibition, Noxa()[inhibitor], Mcl1()[inh_target]


text
"from sympy import Piecewise"
"Monomer('Timer')"
"Parameter('Timer_rate_0', 1.0)"
"Parameter('Timer_0', 0.0)"
"Observable('Timer_obs', Timer())"
"Rule('Timer_increment', None >> Timer(), Timer_rate_0)"
"Parameter('a_0', 47080.3299)"
"Parameter('b_0', 6.44553438)"
"Parameter('c_0', 20235.8565)"
"Parameter('UB_0', 171000)"
"Expression('bid_rate', Piece-
wise(((UB_0*b_0*c_0*((Timer_obs/a_0)**b_0)*(((Timer_obs/a_0)**b_0) + 1)**(-
(c_0+1)))/Timer_obs, Timer_obs > 0.0), (0.0, True)))"
"Rule('Bid_increment', None >> Bid(BaxM=None, BaxC=None, Mcl1=None, Bcl2=None,
BclxL=None, BakM=None), bid_rate)"
"Initial(Timer(), Timer_0)"
```

MOMP Direct/Embedded models

```
model components

Bid, 0.000001
SmacM, 100000
SmacC, 0, {d}
CytoCM, 500000
CytoCC, 0
BaxC, 5000
BaxM, 0
BaxA, 0
BclxL, 113000
BakM, 35000
BakA, 0
Bcl2, 157000
Mcl1, 57000
Bad, 39000
Noxa, 6000

required reactions

catalysis, Bid()[catalyzer], BaxC()[substrate], BaxM()[product]
catalysis, Bid()[catalyzer], BaxM()[substrate], BaxA()[product]
catalysis, Bid()[catalyzer], BakM()[substrate], BakA()[product]
self_catalyze, BaxA()[self_catalyzer], BaxM()[self_substrate]
self_catalyze, BakA()[self_catalyzer], BakM()[self_substrate]
pore_formation, BaxA()[pore]
pore_formation, BakA()[pore]
transport, BaxA()[pore], SmacM()[cargo_M], SmacC()[cargo_C]
transport, BakA()[pore], SmacM()[cargo_M], SmacC()[cargo_C]
transport, BaxA()[pore], CytoCM()[cargo_M], CytoCC()[cargo_C]
transport, BakA()[pore], CytoCM()[cargo_M], CytoCC()[cargo_C]

optional reactions

inhibition, Bcl2()[inhibitor], Bid()[inh_target]
inhibition, BclxL()[inhibitor], Bid()[inh_target]
inhibition, Mcl1()[inhibitor], Bid()[inh_target]
inhibition, Bcl2()[inhibitor], BaxA()[inh_target]
inhibition, BclxL()[inhibitor], BaxA()[inh_target]
inhibition, BclxL()[inhibitor], BakA()[inh_target]
inhibition, Mcl1()[inhibitor], BakA()[inh_target]
inhibition, Bad()[inhibitor], Bcl2()[inh_target]
inhibition, Bad()[inhibitor], BclxL()[inh_target]
inhibition, Noxa()[inhibitor], Mcl1()[inh_target]


text
"from sympy import Piecewise"
"Monomer('Timer')"
"Parameter('Timer_rate_0', 1.0)"
"Parameter('Timer_0', 0.0)"
"Observable('Timer_obs', Timer())"
"Rule('Timer_increment', None >> Timer(), Timer_rate_0)"
"Parameter('a_0', 47080.3299)"
"Parameter('b_0', 6.44553438)"
"Parameter('c_0', 20235.8565)"
"Parameter('UB_0', 171000)"
"Expression('bid_rate', Piece-
wise(((UB_0*b_0*c_0*((Timer_obs/a_0)**b_0)*(((Timer_obs/a_0)**b_0) + 1)**
(-(c_0+1)))/Timer_obs, Timer_obs > 0.0), (0.0, True)))"
"Rule('Bid_increment', None >> Bid(BaxM=None, BaxC=None, Mcl1=None, Bcl2=None,
BclxL=None, BakM=None), bid_rate)"
"Initial(Timer(), Timer_0)"
```