

**DYNAMICAL SENSITIVITY ANALYSES OF KINETIC
MODELS IN BIOLOGY**

THANNEER MALAI PERUMAL
(B.E., NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, 2006)



A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMICAL AND BIOMOLECULAR ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2011

Acknowledgements

This dissertation is dream come true for me. The past five years spent towards realizing this dream has been a great learning experience. Many a times during this period, challenges have come my way, however, the joy of discovering something new at the end kept the fire going and made it every bit worth the effort. If not for all the people who stood by me during these challenging times, completing this thesis would have been an arduous task. Hence, this dissertation should rightfully start with this section where I take the opportunity to express my sincere gratitude towards all those instrumental in completion of this thesis.

Matha (Mother): First and foremost, I take this opportunity to thank my mother, **Mrs. Nirmala Perumal**, whose generous showering of love, care and affection have groomed me into a better human being. Her blessings and faith in me are the main driving force behind this research.

Pitha (Father): I like to thank my primary inspiration and childhood hero, my father, **Mr. Natarajan Perumal**, whose aspiration to become an engineer is reflected in me. His wise advice, emotional and moral support has made me the person I am today. The anecdotes of experiences from his life have principled me besides being a source of guiding light.

Guru (Teacher): I thank my thesis advisor **Prof. Dr. Rudiyan Gunawan**, without whom this thesis would have remained a dream. Rudi is not only a good teacher and mentor, but also a very good friend and philosopher. Only with an undergraduate degree in chemical engineering and little known biology, he accepted me into his group with his hands wide open. It is he who gave me knowledge, confidence, courage and motivation for this research. I express my feelings and infinite respect to this complete teacher using a divine Sanskrit saying “*Guru saakshaat parabrahma tasmai shree Guruvennamaha*”.

I am also thankful to **Prof. Dr. Saif A Khan**, who willingly accepted to advise me in the last days of my PhD. I am indebted to my PhD committee members, **Prof. Dr. Chiu, Min-Sen** and **Prof. Dr. Lee, Dong-Yup** for their constructive comments which further shaped this dissertation. I take this opportunity to thank all the external reviewers of my papers and my thesis, for their critical and constructive comments which improved the quality of this thesis.

It is also right time to thank all my teachers who inspired and mentored me to utilize their teachings in my research. Specifically, I express my sincere gratitude to **Prof. Dr. Sirish Lalji Shah**, **Prof. Dr. Lakshminarayanan Samavedham** and **Prof. Dr. Wei, Shou-Hu**. I take this opportunity to thank all my seniors, who taught me

both the subject and the nuances of research. To specially mention a few, **Dr. Rao Raghuraj**, **Dr. Suresh Selvarasu**, **Dr. Sundar Raj Thangavelu** and many others.

Deivam (God): I thank almighty for giving me good health, sound mind, loving family and wonderful friends.

Funding Agencies: I would like to acknowledge the **Singapore Millennium Foundation** and **NUS Research Fellowship** for providing financial support for my PhD. I would also like to show my appreciation towards all those funding agencies that have supported my research work. I sincerely thank **NUS, Singapore** and **ETH, Zurich** for providing me with state-of-art computational facilities and an outstanding research environment.

Support staff: I like to express my sincere gratitude to **Mr. Boey, Kok Hong**, the most sincere employee of ChBE, NUS, one who made lab life a cake-walk. I also like to thank **Ms. Fam, Hwee Koong, Samantha** for helping us with all the research related purchases. Special thanks to **Mr. Yeo Heng Hee** of high performance computing center, NUS for helping me use the MATLAB clusters. The two important ladies, one who needs special mention are **Ms. How Yoke Leng, Doris** and **Ms. Tan, Hui Ting**, who made things easy for us in departmental paper work.

Friends: I would like to thank my beloved lab members of Chemical And Biological Systems Engineering Lab (CABSEL), **Dr. Suresh Kumar Poovathingal**, **Mr. Sridharan Srinath**, **Mr. Ang Kok Siong**, **Mr. Lakshmi Narayanan Lakshmanan**, **Mr. Jia Genjie** and **Mr. Tam Zhi Yang**. Many arguments during lab meetings, lunch and tea times have resulted as constructive inputs for this thesis.

I would like to thank all my friends and house mates who made my stay in Singapore a rewarding and memorable one. My best buddies, **Mr. Thillaigovindan Jayaraj** and **Mr. Varun Rao** deserve special mention. Their wise words, thoughtful suggestions and never ending support have helped me sail through tough times.

Family: I would like to thank my family and extended family for their continual support in all my endeavors. Last but not the least; I would like to extend my sincere gratitude and thanks to the sole proof reader of this thesis, my better half, **Mrs. Deepa Subramaniyan**, who has been very much instrumental in this thesis completion. I know she is beside me in all the above acknowledgements, yet my heart longs for a special note. Her mere presence and support completes me.

I know that this list is not complete. Hence, I wish to sincerely acknowledge here, all the encouragement and support I received directly or indirectly from different persons at different times of my life.

Singapore, 2011

Table of Contents

Acknowledgements	I
Table of Contents	III
Summary.....	VI
Preface.....	X
List of Tables	XII
List of Figures.....	XIII
1 Introduction.....	1
1.1 Complexity, Robustness and Dynamics of Biological Systems	2
1.1.1 Cellular complexity.....	2
1.1.2 Robustness: An organizational principle	4
1.1.3 Dynamics of biological systems	7
1.2 Systems Biology: A Novel Paradigm in Chemical and Biological Engineering.....	7
1.2.1 Systems modeling in biology: An overview	8
1.2.2 Systems analysis tools in biology	13
1.3 Thesis Motivation and Objectives	18
2 Parametric Sensitivity Analysis	21
2.1 Theory of Classical Sensitivity Analysis.....	22
2.1.1 Local analysis	23
2.1.2 Global analysis.....	26
2.1.3 Hybrid analysis	29
2.2 Caveats of Classical Sensitivity Analysis.....	31
2.2.1 Case study: Simple network model.....	32
2.2.2 Parametric sensitivity analysis for dynamical systems: A caveat	33
2.3 Summary.....	36
3 Green's Function Matrix based Sensitivity Analysis.....	38
3.1 Introduction	39
3.2 Green's Function Matrix (GFM) based Sensitivity Analysis	40
3.3 Case Studies.....	44
3.3.1 Application to common network motifs	44
3.3.2 Application to FasL-induced cell death model of human Jurkat cells	50
3.4 Discussion.....	57
3.4.1 Experimental and biological relevance	57
3.4.2 Comparison with the classical sensitivity analysis	59
3.4.3 Applications of GFM analysis	61
3.4.4 Advantages and limitations of GFM analysis	64
3.5 Summary.....	65
4 Impulse Parametric Sensitivity Analysis	66
4.1 Introduction	67
4.2 Impulse Parametric Sensitivity Analysis (iPSA)	68
4.3 Case Studies.....	71
4.3.1 Simple network model	71
4.3.2 FasL-induced cell death of human Jurkat T-cell lines	72
4.4 Discussion.....	76

4.5 Summary	78
5 Pathway Parametric Sensitivity Analysis	80
5.1 Introduction	81
5.2 Pathway Parametric Sensitivity Analysis (pathPSA)	82
5.2.1 Mathematical models	82
5.2.2 Pathway identification.....	83
5.2.3 Model re-parameterization	84
5.2.4 Dynamical analysis	85
5.3 Case Study: Application to FasL-Induced Apoptotic Cell Death.....	86
5.4 Discussion.....	89
5.4.1 Experimental and biological relevance	89
5.4.2 Comparison with classical PSA, GFM and iPSA.....	91
5.5 Summary.....	92
6 Molecular Density Perturbation Analysis	94
6.1 Introduction	95
6.2 Molecular Density Perturbation (MDP) Analysis	97
6.3 Application to TRAIL-induced Cell Death Model of HeLa cell Population.....	100
6.3.1 GFM and iPSA analysis of TRAIL induced apoptosis of a single cell	103
6.3.2 eFAST analysis of TRAIL induced apoptosis.....	104
6.3.3 MDP analysis of TRAIL induced apoptosis.....	107
6.4 Discussion.....	110
6.4.1 Variability in the conclusion of MDP analysis.....	110
6.4.2 Relevance and Application.....	112
6.4.3 Comparison with GFM, iPSA and eFAST analyses	115
6.4.4 Applications of MDP analysis.....	116
6.5 Summary.....	116
7 Dimensionality Reduction of Large Kinetic Models.....	118
7.1 Introduction	119
7.2 Methods	121
7.2.1 Model reduction using parametric sensitivity analysis (PSA).....	121
7.2.2 Model reduction using dynamic sensitivity analysis (iPSA and GFM).....	122
7.2.3 Combined methods for model reduction	124
7.3 Examples	127
7.3.1 Alkane pyrolysis model.....	127
7.3.2 Natural gas combustion – GRI-Mech 3.0.....	130
7.3.3 <i>Ab initio</i> kinetic model of industrial ethane pyrolysis.....	132
7.4 Summary.....	134
8 Thesis Outlook.....	136
8.1 Dynamical Sensitivity Analyses of Kinetic Models in Biology	137
8.1.1 Motivation.....	137
8.1.2 Sensitivity analyses	138
8.1.3 Model reduction	139
8.1.4 Contributions.....	140
8.1.5 Advantages and limitations	141
8.2 Recommendations for Future Work	143
9 Bibliography	147
Appendix A	158
Appendix B	159
Appendix C	168

Appendix D	173
Appendix E	179
Appendix F	188
Curriculum Vita.....	194

Summary

Tremendous growth of biological data in the post genomic era has given rise to the recognition of biological complexity. Such complexity often limits human intuition in understanding how functional regulation is accomplished in a cell. The complexity of a typical cellular network has been argued to provide robustness to common perturbations, but at a cost of fragility to rare changes. Despite its obvious benefits, robustness property can turn into an Achilles heel if the cellular mechanisms that confer this property are hijacked, such as in cancer. Thus, the understanding of cellular complexity, robustness and its tradeoffs can greatly benefit the drug discovery efforts for human diseases.

Cellular complexity has motivated the use of mathematical representations to describe biological networks. By way of systems analysis, these mathematical models can elucidate the mechanisms that are responsible for giving the observed functional behavior. To this end, the current dissertation presents novel systems analyses, to elucidate the cellular mechanisms that give rise to the functional dynamics, robustness and fragility in a cell, thereby providing a better understanding of complex diseases and guiding principles for systems oriented drug design.

Most of the published systems analyses tools used for understanding the dynamics of deterministic and probabilistic models in systems biology relates the property of robustness of system output (in)sensitivity and these analyses share a common feature, which they rely on perturbations on the system parameters and quantification of the resulting change in system output. One of such commonly used tool for analyzing deterministic ordinary differential equation (ODE) models is the local parametric sensitivity analysis (PSA), which maps out the parametric dependence of system behavior. It is shown here that the dynamical aspect of cellular regulation is not immediately apparent from the associated parametric sensitivity coefficients and more importantly, that they can even lead to incorrect conclusions in understanding system dynamics. Briefly, the reason stems from the fact that the perturbations are realized on time-invariant system parameters, which means that the perturbations are persistent (analogous to a step change). Thus the dynamical information of when a parameter perturbation matters is not available.

In order to overcome the above said drawback, for analyzing the dynamics of ODE models, three novel local dynamic sensitivity methods have been developed in this dissertation. These analyses are called the Green's function matrix (GFM) analysis, the impulse parametric sensitivity analysis (iPSA) and the pathway parametric sensitivity analysis (pathPSA). Even though these methods are formulated

in lines of the classical PSA, i.e. based on cause-effect relationship, they differ in the manner with which the perturbations are introduced on the system. The GFM sensitivities introduce time-varying perturbations on the system states, while the iPSA introduce impulse perturbations on the system parameters, also at different times. Thus, the sensitivity coefficients from these analyses can reveal a dynamic, molecule-by-molecule or parameter-by-parameter insight on how system behaviour and complementarily, how (an impulse) signal propagates through the network. Coincidentally, the iPSA coefficients can be calculated directly from the GFM coefficients, by simple matrix multiplications. These methods provide the necessary information to understand how system dynamics is achieved, i.e. which species (molecules) and which parameters (reactions) are essential and when they become important, respectively.

Cells often rely not on a single molecule or reaction, rather on a group of molecules and reactions that gives rise to the observed behavior. Hence, it is equally important to analyse pathways, comprising a (minimum) set of molecules and reactions that can (independently) carry out cellular response to stimuli. To this end, another novel local dynamic sensitivity analysis, called pathway parametric sensitivity analysis (pathPSA) is developed. The pathPSA combines the structural and dynamic information of the network. Hence, the analysis comprises two steps: (i) identification of functional pathways (structure), and (ii) introducing perturbations on the parameters associated with each pathway (dynamics). Like the GSM and iPSA, the pathPSA coefficients give dynamical insights on how the cellular functional behavior is accomplished and how the robustness/fragility is controlled by, in this case, pathways. i.e., which pathways are important and when they are important. The caveats of classical PSA and the efficacies of all the above said dynamical analysis tools (the GFM, iPSA and pathPSA) are demonstrated through applications to an ODE model of synthetic network motifs and a FasL-induced programmed cell death model of Jurkat T cell lines.

Recent advances in single cell assays have revealed the cell-to-cell variability which can explain phenotypic heterogeneity, even in isoclonal cell population. Hence, the understanding cellular dynamics and analyzing robustness under such variability need a global analysis framework. Existing global analysis methods such as eFAST, DGSM and Glocal analysis, only employ perturbations in a single characteristic (e.g. mean and variance) of the molecular or parametric distribution, and these perturbations are also realized at a single time point. On the contrary, a novel global dynamic sensitivity analysis is developed in this dissertation with a similar aim as the new dynamical analyses above. In this analysis, perturbations are effected on the marginal density functions of molecules at different times, hence the name: molecular density perturbation (MDP) analysis. The efficacy of this method is compared with

the GFM and eFAST analysis through an application to a population model of TRAIL-induced programmed cell death in Hela cell line.

Analyzing biological complexity and robustness properties can help in a better understanding of the mechanisms by which network functions are accomplished. Based on kinetic models and perturbation analyses, such as the ones presented here, dynamical insights on the functional regulation and signal propagation in the cellular network of interest can be obtained. The methods developed here are compared against related existing analyses; using well established signaling models of both synthetically generated and those from literature (e.g., programmed cell death models). The analyses give a species-by-species or parameter-by-parameter account of how a network function or output is accomplished and also where an induced perturbation travels. The analysis also gives experimental suggestions to tune the model. In practice, the results can guide model identification and reduction of cellular systems and suggest experimentally testable hypotheses. In addition, the biological knowledge gained can assist drug discovery efforts in the identification of potential drug targets, the understanding of drug efficacy and specificity, and finally the optimization of drug dosing and timing. The development of these methods represents a concrete step towards robustness-based drug design through systems biology.

Finally, one of the major applications of PSA is in reduction of large scale reaction kinetic models. In contrast to the conventional PSA based model reduction, this dissertation presents five new techniques that are based on the dynamical information obtained from the iPSA and GFM sensitivities. The reduction of model dimensionality is carried out by either eliminating the reactions or species to which a selected model outputs show low sensitivities. Also, three combined model reduction algorithms based on both the iPSA and GFM sensitivities are presented. The efficacies of the methods are demonstrated through three industrially-relevant applications, including alkane pyrolysis, natural gas combustion, and industrial steam cracking of ethane.

Dynamical Sensitivity Analysis for Biological Systems

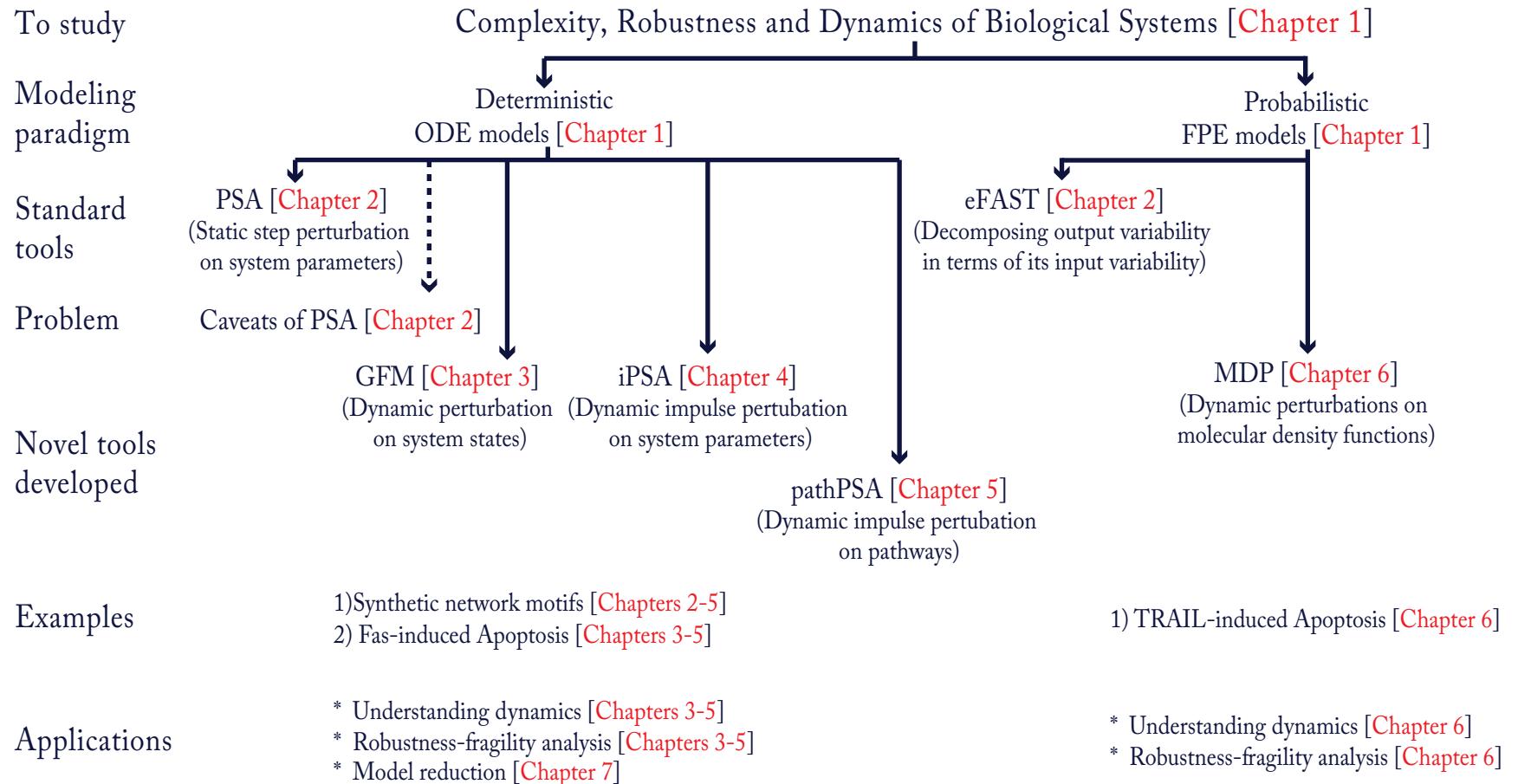


Figure S.1. Thesis Summary. The thesis details the development of dynamical sensitivity analyses for kinetic models in biology and the application of these analyses to a variety of examples ranging from biology to chemical engineering.

Preface

Systems biology applies systems theory concepts to study the interactions and functional behaviors in molecular biology, in which mathematical models are used to understand the dynamics of biological systems. The interdisciplinary nature of this research necessitates a fair amount of understanding of biology and mathematics. Hence the goal here is to give enough information in order to make this thesis, for the most part, self contained for readers from both fields. Admittedly, the information contained here is incomplete. However, the reader can gain enhanced understanding of biological robustness and its properties, and should be able to better appreciate the typical complexity of biological systems, and the usefulness of mathematical analysis in unraveling this complexity.

This dissertation presents novel dynamic sensitivity analysis tools, for both deterministic and population (stochastic) models in systems biology. It is organized into eight chapters. At the beginning of each chapter, a synopsis, which situates the chapter in the context of the thesis and summarizes its content, is presented. Chapters 1 and 2 introduce the premise and explain the existing problems in analyzing cellular dynamics. Chapters 3 to 6 introduce different dynamic sensitivity analyses, which majorly differ in the nature of their perturbations. The efficacies of all these methods are evaluated using cellular signaling pathway models (mainly involving apoptotic cell death). Chapter 7 gives one of the applications of these dynamic sensitivities for model reduction and Chapter 8 concludes the thesis. A more detailed organization of this dissertation is as follows:

- **CHAPTER 1** – In this chapter, I introduce the complexity and robustness characteristics of biological systems, and present the need to understand system dynamics. Later, a brief review on different modeling paradigms and analysis tools used to understand cellular dynamics is presented.
- **CHAPTER 2** – In this chapter, I introduce and demonstrate the caveats of the classical parametric sensitivities (both local and global) in analyzing the dynamics of ordinary differential equation (ODE) models in systems biology.
- **CHAPTER 3** – To overcome the caveats of local parametric sensitivities, a novel dynamic sensitivity analysis based on time-varying perturbations on molecular concentrations of the ODE models, called Green's function matrix (GFM) analysis, is developed.
- **CHAPTER 4** – To complement the GFM analysis, a novel dynamic sensitivity analysis based on impulse perturbations on parameters of the ODE models, called impulse parametric sensitivity analysis (iPSA), is developed in this chapter.
- **CHAPTER 5** – Here, a pathway based dynamic sensitivity analysis method, called pathway parametric sensitivity analysis (pathPSA), which introduces parametric perturbations on the elementary signaling modes (ESMs) of the ODE models, to quantify the system dynamics is developed.
- **CHAPTER 6** – A global dynamic sensitivity analysis, which introduces time-varying perturbations on molecular density functions, named as molecular density perturbation (MDP) analysis, is developed.
- **CHAPTER 7** – Based on the dynamic sensitivities presented in Chapters 3 and 4, GFM and iPSA, five new algorithms for reducing detailed chemical kinetic models are developed in this chapter.
- **CHAPTER 8** – In this chapter, I present the summary of findings, show the contribution of the current dissertation and provide further recommendations for future work.

List of Tables

Table Number	Title	Page Number
3.1	Green's Function Matrix (GFM) analysis of simple network motifs	45
6.1	Variability in the top ten MDP coefficients, calculated with respect to five replicated experiments.	110
6.2	Ranking correlation between distance metrics, type and magnitude of perturbations	111
7.1	Reduction of alkane pyrolysis model using PSA, iPSA and GFM	128
7.2	Reduction of alkane pyrolysis model using combined iPSA and GFM	130
7.3	Reduction of natural gas combustion - GRI Mech 3.0 model	132
7.4	Reduction of ethane cracking model	133
7.5	Reduction of ethane cracking model (change in reactions)	134

List of Figures

Figure Number	Title	Page Number
S.1	Thesis Summary	X
1.1	Synthetic network motifs	3
1.2	Spectrum of modeling in biology	9
1.3	Construction of the uncertainty system	14
1.4	Two-parameter bifurcation diagram	16
2.1	Global and local analysis	27
2.2	A simple network model	32
2.3	Local parametric sensitivity analysis of x_6 activation under x_1 stimulus	33
2.4	Illustration of perturbation-response in parametric sensitivity analysis	35
3.1	Green's Function Matrix (GFM) analysis	41
3.2	A heat map of the GFM coefficient	42
3.3	GFM analysis of networks without feedback	43
3.4	<i>In silico</i> validation of the GFM analysis	47
3.5	Network model of FasL-induced programmed cell death in Jurkat T cell line	50
3.6	Parametric sensitivity analysis of caspase-3 activation by a FasL stimulus	52
3.7	Green's Function Matrix (GFM) analysis of caspase-3 activation	53
3.8	Signal progression of FasL impulse through the programmed cell death network	55
3.9	Model reduction of FasL-induced apoptosis pathway.	62
4.1	Impulse Parametric Sensitivity Analysis (iPSA)	68
4.2	Impulse parametric sensitivity analysis of x_6 activation under x_1 stimulus	71
4.3	Local parametric sensitivity analysis of the programmed cell death model	72
4.4	The iPSA of caspase-3 activation	73
4.5	The iPSA of caspase-3 at the final output time	75
4.6	Impulse signal progression of J1_f through Fas-induced apoptosis network	76
5.1	Dendrogram of pathPSA coefficients	87
5.2	PathPSA of caspase-3 activation	89
6.1	A heat map of the MDP coefficients	99
6.2	Network model of TRAIL-induced cell death in Hela cell lines	101
6.3	Green's Function Matrix (GFM) analysis of cPARP activation	102

	by a constant TRAIL stimulus in a single cell	
6.4	Extended Fourier Amplitude Sensitivity Test (eFAST) of cPARP activation in a cellular population	104
6.5	Extended Fourier Amplitude Sensitivity Test (eFAST) of cPARP activation in apoptotic and non-apoptotic cell population	105
6.6	Molecular Density based Perturbation (MDP) analysis of cPARP activation by a constant TRAIL stimulus in cellular population	107
6.7	Comparative MDP analysis of cPARP activation by a constant TRAIL stimulus in apoptotic and non-apoptotic cell population	109
6.8	<i>In silico</i> validation of and hypothesis generation using the MDP analysis of cPARP activation under constant TRAIL Stimulus	113
7.1	Sensitivity analysis	121
7.2	Top – Down (TD) and Bottom – Up (BU) algorithms	125
7.3	Low temperature alkane pyrolysis model	128
7.4	Reduction of low temperature alkane pyrolysis model	129
7.5	Isobaric-Isothermal natural gas combustion (GRI Mech 3.0) model	131
7.6	Reduction of isobaric-isothermal natural gas combustion (GRI Mech 3.0) model	131
7.7	Industrial steam cracking of ethane at steady state in a PFR	132
8.1	Thesis outlook and future work	146

CHAPTER 1

1 Introduction

Synopsis:

This chapter introduces the concepts of complexity, robustness and dynamics of biological systems and emphasizes the need to understand them. An overview of modeling paradigms in systems biology and the associated systems analysis tools used to gain understanding of the underlying cellular mechanisms are also presented. The advantages and limitations of the state-of-the-art analyses tools are discussed and how this thesis contributes to overcome some of these limitations is shown. Finally, the motivation behind and the objectives of the dissertation are presented.

1.1 Complexity, Robustness and Dynamics of Biological Systems

From the early 19th century, many problems in physics and chemistry were tackled by the reductionist approach [1], where the comprehension of a system is done by detailed characterization of its parts. This success led to the use of the same approach in the mid 19th century to study biological systems [2]. However, other than delineating the parts list for the contemporary molecular biology, reductionist approach often provided only incomplete explanation about biological behavior [3]. This failure can be attributed to a number of factors, of which the majority owes to the complexity found in the typical biological systems, for example neuronal interconnections in human brain and the diversity of our immune system.

1.1.1 Cellular complexity

Complex systems have been used to describe many problems in modern biology, chemistry, physics, engineering and others [4]. Complexity is a qualitative phenomenon, generally describing systems whose properties cannot be understood by decomposing and analyzing their components alone. A simple phosphorylation and dephosphorylation reaction of a protein dynamics as shown in Figure 1.1(a) is a good example for a complex system. The signal-response characteristics of this system cannot be inferred only from the components characteristics (i.e., from S: signal strength, R: de-phosphorylation level and RP: phosphorylation level) alone. Instead, the system also needs the kinetic rate information for both the phosphorylation and dephosphorylation reactions to ascertain the emerging signal-response behavior (i.e., based on the choice of rate kinetics the signal-response characteristics can be either

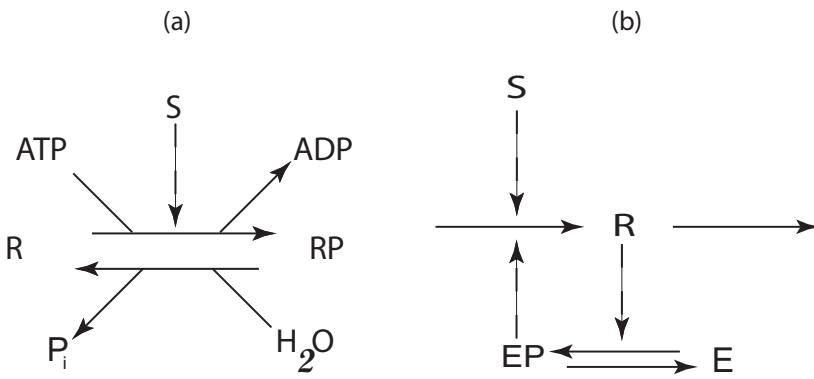


Figure 1.1. Synthetic network motifs. (a) Simple phosphorylation – dephosphorylation motif, (b) Mutual activation motif.

hyperbolic or sigmoidal) [5]. Hence in short, complex systems are systems where the “*whole is greater than the sum of its parts*” [6]. Examples of complex systems can be readily found among natural and engineered systems, including biological networks (i.e., from genetic to protein to metabolic), signaling pathways, combustion reactions, ecosystems, the world-wide-web, and even the propagation of pandemic infections [6]. Ideas of complex systems are making their way into many fields including social sciences, political sciences, finance, biology and medicine. In fact, the theoretical study of complex systems has become a separate field of science in itself [7].

Biological complexity arises because of the large number of components and the even larger number of interactions between them. These interactions take place across multiple organizational levels with varying scales of time and space [8]. For example, a simplified network consisting of four genes, as shown in Figure 1.1(b), interacting in a nonlinear fashion, can give rise to a very complex behavior. The network shows fascinating emergent properties across different time scales, generates distinct output behavior depending upon the input signals, and the presence of feedback loop even makes the system behave like a bistable switch [5]. Hence, even though the behavior of the components of biological systems (i.e., genes, proteins, etc...) is simple, it is the system design and the interactions among them that make it more complex.

Many human diseases, including cancer, diabetes, and neural disorders, are manifestation of the (mal)functioning of the system components in biology. Hence, an understanding of the system complexity is hoped to shed light on these diseased states and could possibly influence the design of new drugs that are intended to treat these complex diseases [9]. However, as noted earlier, even to maintain homeostasis, biological systems can and need to constantly change and adapt themselves within a range of feasible parameters [6]. As a consequence, these systems may have evolved to possess some of the characteristic features such as: (i) self-organization: the ability of the system to organize themselves (without any external parameters) that are far different from being random, (ii) adaptability: the ability of the system to change its functionality in response to the surrounding, (iii) robustness: the ability of the system to maintain its functionality against perturbations, (iv) emergent behavior: the ability of the system to exhibit behaviors that are markedly different from its basic constituents, and (v) chaotic behavior: the ability of the system to generate unpredictable temporal behavior [10].

1.1.2 Robustness: An organizational principle

Robustness is a ubiquitous property of biological systems and has been recognized as a fundamental organizational principle in the evolution of cellular functions [11-13]. Robust behaviour has been observed in many biological systems, such as in λ -phage switching [14], bacterial chemotaxis [15], and *Drosophila* circadian rhythms [16]. Robustness is often thought as a desirable property of network interconnections, which is selected through evolution. Cellular complexity often appears to arise mainly from robustness as one of the design goal [17]. Studies have suggested that many biological networks show a robust behavior to variations in biomolecular

concentrations and such network structures are conserved across organisms [18-20]. Hence, understanding the connection between complexity and robustness became an active research topic [21-27].

The definition of robustness is domain specific. In different organizational levels of biology, different types of robustness have been reported, ranging from genotypic to phenotypic to environmental to functional robustness [28]. This thesis concentrates particularly on functional robustness, a property that allows a system to maintain its functions despite internal and external perturbations [12]. In this case, in order to define robustness, one must identify three major components: system, function, and perturbations [29] and to perform robustness analysis, one needs to specify the system, its function (or behavior) that remains unchanged, and the type of uncertainties and disturbances for which this invariance property holds [17, 30].

General characteristic properties of a robust system are adaptation, parametric insensitivity and graceful degradation [18]. Some of the major network characteristics shared among robust systems are (i) motifs and modules, allowing for a distributed information processing and decoupling of sub-systems against perturbations [31], (ii) systems control mechanisms, such as feedback and feedforward loops [13], and (iii) alternative fail safe mechanisms, such as redundancy, diversity and purging [12]. A commercial jet airplane makes a good analogous engineering system to understand the above said biological robustness mechanisms [12]. Many commercial planes have an autopilot mode, operated through an automatic flight control system (AFCS), that maintains a flight path (direction, altitude, and velocity of flight) against variations in the conditions of external environment (such as winds, clouds, etc.). This control employs mechanisms, such as negative feedback control, in which deviations from the defined flight path are automatically corrected. AFCS is generally composed of three

modules with the same functions, thereby creating redundancy. However, each module is designed differently to avoid a common mode failure, analogous to, heterogeneity. The three computers are made to be modular, so that failure in one module does not affect the functions of other parts of the system. This type of mechanism is implemented using digital technologies that decouple low-level voltages from digital signal (ON/OFF of pulses), thereby preventing noise from influencing its functions. Although this is a simplified explanation of a man made robust system, the basic concepts are also relevant in biological systems. For instance, cancer can be viewed as a robust system with its genetically heterogeneous cell population, providing a high level of redundancy, and each cancer cell possesses the ability to counteract the action of drugs [9, 32]. Hence, an in depth understanding of the dynamics of the acquired robustness property of these diseases will also provide us with methods to control them.

Though a commercial plane is robust against its component failure and perturbations in external conditions, it is also extremely fragile against highly improbable events such as total power failure [29]. Similarly, despite its obvious benefits, robustness property can turn into an Achilles heel if the cellular mechanisms that confer this property are hijacked, such as in cancer and diabetes [12, 18, 19, 32]. The acquired robustness property gives these diseases the ability to adapt to drug actions and develop drug resistance. Hence, how robustness and fragility of biological systems are exploited during disease onset and progression is one of the major factors which will open up a new avenue for drug discovery research [9].

1.1.3 Dynamics of biological systems

On a separate note, dynamics is a prominent feature of many important biological processes (e.g., oscillations in cell cycle and circadian rhythm [33, 34], switching behaviour in programmed cell death [35], and adaptation in chemotaxis [36]). Many cellular properties and functions, including robustness, are accomplished through intricate regulatory networks that control cellular processes from mRNA transcription to post-translational protein activity. Here, the word “regulatory” implicitly describes an active dynamical response to internal and external stimuli, in which an orchestra of cellular processes, from signaling to gene expression, maybe set in motion in response to a perturbation or a stimulus. For example, caspase dependent programmed cell death (apoptosis) may activate a chain of processes in response to a death signal, including the expression of pro-apoptotic genes, the inhibition of anti-apoptotic proteins, mitochondrial remodeling and subsequent release of pro-death proteins, and the cleavage of various caspases [37]. Thus, understanding cellular dynamics, to gain insights on the mechanisms that give rise to and control the dynamic behaviour, has become a prime concern [38-42]. But as seen earlier, the classical reductionist approach of biology does not immediately answer such questions. Hence, this requires a shift to “what to look for” in biology, in which a top-down approach should complement a bottom-up-approach [18].

1.2 Systems Biology: A Novel Paradigm in Chemical and Biological Engineering

The challenge of understanding emergent behaviors from complex cellular networks has motivated the use of quantitative (mathematical) representations of cellular networks and the analysis of such models to discover the basis of cellular phenotypes,

to predict the effects of network changes on cellular behaviour and to ultimately design networks with new and desired properties and functions [43, 44]. Systems biology approach attempts to understand biological systems much like engineering systems, through the identification of their structure and dynamics, the control and the design of cellular circuits with desired properties [13, 45]. Mathematical models of various cellular networks, such as cell cycle [46], circadian rhythm [47], and metabolism [48], have been developed from experimental data through reverse engineering. The analysis of these models of native biological and pathological systems has greater impacts such as:

1. for better understanding of pathology and malfunction in order to provide potential therapeutic targets for treatment and diseases [9, 49, 50],
2. to design biological systems having desired properties not existing in nature [51],
3. to engineer organisms to produce valuable compounds [52].

1.2.1 Systems modeling in biology: An overview

Many dynamical modeling paradigms have been used to represent cellular networks ranging from deterministic models to stochastic models, as shown in Figure 1.2. A deterministic model is one in which every set of variable states is uniquely determined by its parameters and by sets of initial states of these variables in the model. Examples of such models include Boolean models and ordinary differential equations (ODE) [53]. On the other hand, probabilistic models are models in which the states are random variables described by probability distributions. In this case, examples include Fokker-Planck equations (FPE) and chemical master equations (CME) [54]. The source of randomness can either be external or internal or both [55]. As a consequence, two identical systems starting from the same initial condition may

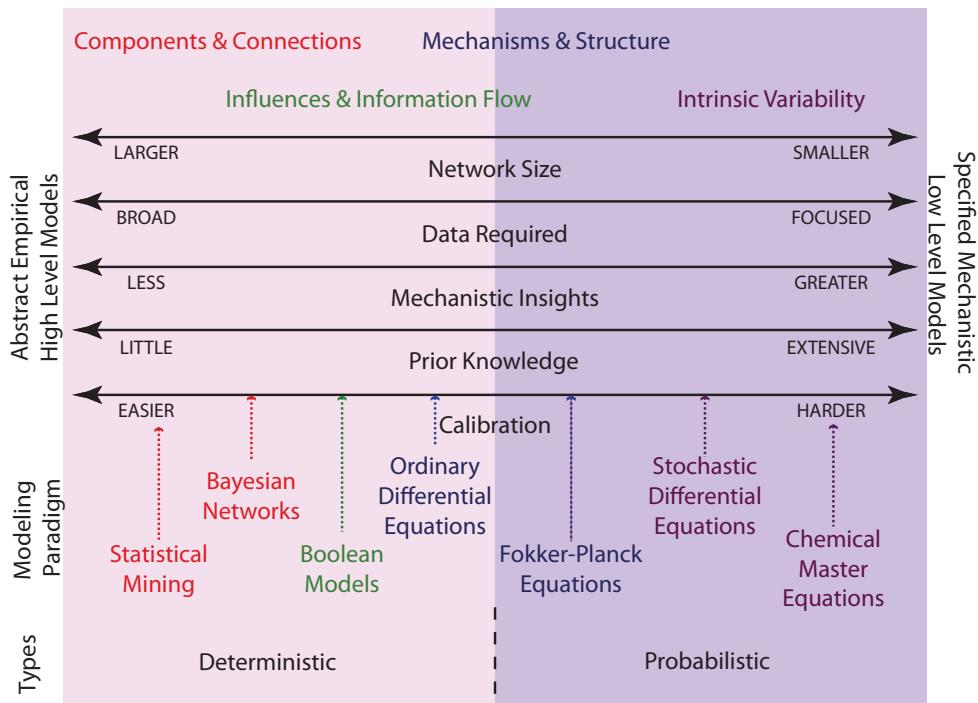


Figure 1.2. Spectrum of modeling in biology. Different paradigms of systems modeling, spanning from high level or abstract statistical models to detailed chemical master equations are depicted.

follow different trajectories due to its random (stochastic) nature. As seen in Figure 1.2, different modeling paradigms span different levels of abstraction, giving varying levels of insights about the system and some of them are detailed below.

Boolean Models (deterministic): Boolean models characterize the influence and information flow among system components using simple Boolean logics [53]. In a Boolean model, each variable of a system (e.g. gene activity) can assume two values, such as active (ON/1) or inactive (OFF/0), and the dynamics (i.e., changes with time) are based on a set of logical rules/Boolean functions which govern the relationship among the components of the network [25]. Mathematically, the dynamics of a Boolean network describing a regulatory system is given by

$$\mathbf{x}(t+1) = \mathbf{b}(\mathbf{x}(t)) \quad (1.1)$$

where $\mathbf{x} \in \mathbf{R}^n$ denotes the state variable representing the mRNA or protein levels (ON/OFF) and $\mathbf{b}(\mathbf{x}(t))$ is the vector valued Boolean rule, describing the relationship

between $\mathbf{x}(t)$ and $\mathbf{x}(t+1)$. Although such models are the simplest abstraction of dynamical biological systems, they can provide qualitative information on the structure of interaction among genes and proteins [25]. Boolean models have been used to describe biological networks such as mammalian cell cycle [56], gap-gene genetic network involved in the fruit fly *Drosophila melanogaster* segmentation [57], and morphogenesis in *Arabidopsis thaliana* [58].

ODE Models (deterministic): The most widespread formalism to model dynamical systems in science and engineering, including biology, makes use of ordinary differential equations (ODE). In systems biology, these are typically mechanistic models, which describe the concentrations of biomolecules, called states. Mathematically, an ODE model is written as [59, 60]

$$\frac{d\mathbf{x}(t, \hat{\mathbf{p}}, \mathbf{x}_0)}{dt} = \mathbf{g}(\mathbf{x}, \hat{\mathbf{p}}); \quad \mathbf{x}(t_0, \hat{\mathbf{p}}) = \mathbf{x}_0 \quad (1.2)$$

where the state $\mathbf{x} \in \mathbb{R}^n$ is typically the concentration vector of biomolecular species, such as mRNAs and proteins, while the function \mathbf{g} is the constitutive, often nonlinear, rate equation. The right hand side of the ODE captures the generation and consumption of biomolecules due to a variety of processes in the cell (e.g. transcription, translation, phosphorylation and dephosphorylation, *etc*). The rates of which depend on a set of kinetic parameters that are consolidated in the vector $\hat{\mathbf{p}} \in \mathbb{R}^m$, representing the relevant kinetics, energy, and/or transport coefficients that appear in the constitutive equations. Since the initial conditions \mathbf{x}_0 can be treated in the same way as model parameters, the aggregate vector $\mathbf{p} \in \mathbb{R}^{m+n}$ is used here to denote the combined parameters and initial conditions, i.e. $\mathbf{p} = [\hat{\mathbf{p}}^T \quad \mathbf{x}_0^T]^T$. This formulation is general enough to describe most systems of interest in biology, such as metabolic,

signaling, and genetic regulatory models [61, 62], with examples including models of apoptosis [63] , cell cycle pathway [64] and circadian rhythm [47].

Fokker-Planck Equations (FPE) Models (probabilistic): Model parameters $\hat{\mathbf{p}}$ and the initial conditions \mathbf{x}_0 in (1.2) are usually derived from (noisy) experimental data and therefore have inherent uncertainty. In other words, the parameters and initial conditions are random and can be described by density functions $f_{\hat{\mathbf{p}}}(\hat{\mathbf{p}})$ and $f_{\mathbf{x}_0}(\mathbf{x}_0)$, respectively. Consequently, under these conditions, $f_{\mathbf{x}}(t, \mathbf{x})$ satisfies the following partial differential equation , given by

$$\frac{\partial}{\partial t} f_{\mathbf{x}}(t, \mathbf{x} | t_0, \mathbf{x}_0) + \frac{\partial}{\partial \mathbf{x}} (\mathbf{g} \cdot f_{\mathbf{x}}(t, \mathbf{x} | t_0, \mathbf{x}_0)) = 0 \quad (1.3)$$

where $f_{\mathbf{x}}(0, \mathbf{x} | t_0, \mathbf{x}_0) = f_{\mathbf{x}}(t_0, \mathbf{x}_0)$ are the joint pdf of initial conditions. This formulation in (1.3) takes a specific form of Fokker-Planck equation with only drift velocity \mathbf{g} , and no diffusion matrix [65, 66]. Only for very simple systems that (1.3) have an analytical solution, specifically where the separation of variables is possible or if the velocity \mathbf{g} is constant [67]. In most cases, the solution is obtained using very simple but computationally intensive Monte-Carlo (MC) algorithms or by a complex but computationally efficient transformation of (1.3) to a Schrodinger equation [67]. Some of the examples of FPE models in systems biology include a model to characterize the cell migration process and a model to understand the apoptotic signaling in Hela cells [66].

Chemical Master Equation (CME) Models (probabilistic): CME model formulation describes biological processes in a cell as random events due to the low concentration of participants (nM level). As differential equations assume that concentrations of molecules vary continuously and deterministically, these assumptions no longer apply in many cases of cellular processes [68]. In a cell, some

molecules, such as the components of gene transcription, only exist at low copy number (~ 10 s of molecules), thus violating the continuity assumption [68]. As a consequence, the same system starting from the same initial conditions may follow different trajectories due to the intrinsic random nature. In CME, discrete \mathbf{X} number of molecules are taken as state variables, and a joint probability distribution $P(t, \mathbf{x} | t_0, \mathbf{x}_0)$ is introduced to express the probability that the cell contains X_i molecules of the i^{th} species at time t , given the initial condition \mathbf{X}_0 at t_0 [69, 70]. The time evolution of the function $P(t, \mathbf{x})$ can be expressed as follows:

$$P(t + \Delta t, \mathbf{x} | t_0, \mathbf{x}_0) = P(t, \mathbf{x} | t_0, \mathbf{x}_0) + \left(\sum_{j=1}^m \left(\alpha_j(\mathbf{x} - \mathbf{v}_j) P(t, \mathbf{x} - \mathbf{v}_j | t_0, \mathbf{x}_0) - \alpha_j(\mathbf{x}) P(t, \mathbf{x} | t_0, \mathbf{x}_0) \right) \right) \Delta t \quad (1.4)$$

where m is the number of reactions that can occur in the system, \mathbf{v}_j is the state-change vector representing the change in molecular states caused by the reaction j , $\alpha_j(\mathbf{x})$ is the propensity function describing the probability that reaction j will occur in the interval $[t, t + \Delta t]$ given that the system is in the state \mathbf{X} at time t [71]. In the thermodynamic limit as the number of molecules become large and when the initial variability is avoided, the stochastic solution of (1.4) will approach a deterministic solution [72]. Though the master equation provides an intuitively clear picture of the stochastic process by accounting for the intrinsic fluctuations, its solution is computationally challenging, for example using the stochastic simulation algorithm (SSA) [70]. Nevertheless, this approach has gained increasing popularity with examples including models of phage λ -infected *E.coli* cells [73] and in circadian rhythms [74].

1.2.2 Systems analysis tools in biology

As discussed in the previous section, a mathematical model, which represents the biological pathway, provides an avenue to systematically investigate a cellular network, through the application of systems analyses [12, 75, 76]. Different modeling frame works (Boolean, ODE, Fokker-Plank and CME) necessitate the use of different types of analysis tools. Deterministic ODE models and probabilistic FPE models are of particular interest in this work. In fact, the analysis of ODE models is an active area of research in the field of systems biology and traditionally in mathematics and engineering as well, where existing methods range from structured singular value (SSV) to sensitivity analysis and bifurcation analysis (for more examples of analyses, see [21, 22, 25, 26, 77-81]). Below, some of these methods are outlined. In essence, these methods investigate the effects of perturbations in the model parameters or inputs on the system output behaviour, in order to ascertain the role of various cellular processes. The agreement and/or disagreement between model analysis and experimental findings or biological knowledge will give feedback about the validity of the model and to the design of future experiments, closing the loop in the iterative approach of systems biology [44].

Structured Singular Value (SSV) analysis: As described earlier, the underlying robust principles are believed to be universal in both biological and sophisticated engineering systems [20], motivating the use of tools from control systems theory for analyzing complex biological systems. SSV analysis has been developed for analyzing robustness in engineering systems with rigorous mathematical foundation and is readily available in standard software packages [82]. In this analysis, for a given bounded set of internally or externally occurring uncertainties (perturbations), robustness is checked by the loss of stability and performance in a system (output

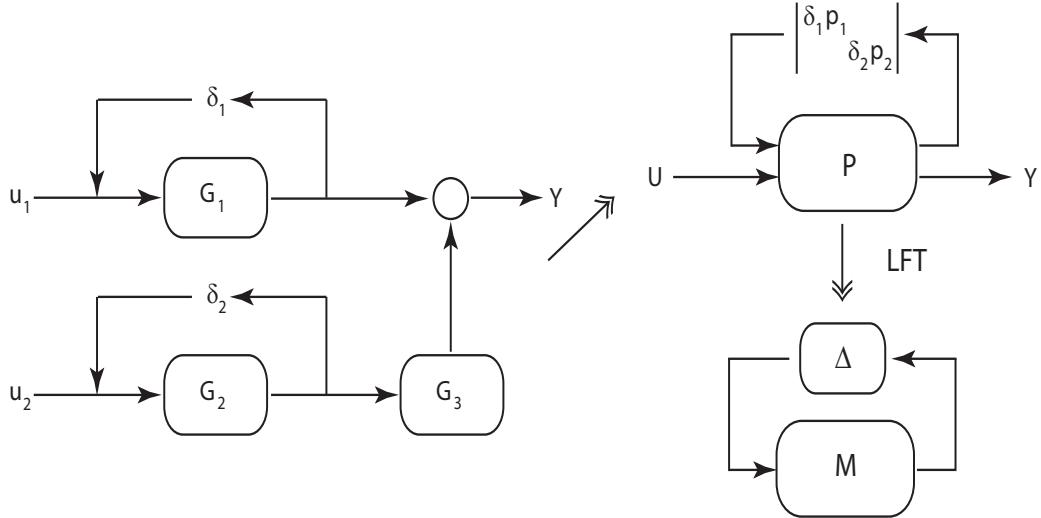


Figure 1.3. Construction of the uncertainty system. An example of an uncertain system that is shaped into the $\mathbf{P}\Delta$ block system, where Δ is the block-diagonal matrix, $[\delta_1 \mathbf{p}_1 \ \delta_2 \mathbf{p}_2]$. This system can be lifted via linear fractional transformation to the $\mathbf{M}\Delta$ block structure and tested for robust stability. Closing the input/ output channels of the $\mathbf{P}\Delta$ block structure, one can test for robust performance.

behavior). In other words, a system is robust if certain stability and performance criteria are satisfied for all feasible perturbations Δ [26].

SSV analysis can be thought as a generalization of the Nyquist stability criterion for single variable system [82]. In this case, robust stability (RS) of the system M with respect to the uncertainty (or perturbations) Δ , is checked by computing μ , defined as follows

$$\mu(M)^{-1} \triangleq \min_{\Delta} \{\bar{\sigma}(\Delta) \mid \det(I - M\Delta) = 0 \text{ for structured } \Delta\} \quad (1.5)$$

where $\bar{\sigma}(\Delta)$ defines the maximum singular value of the perturbation Δ on the system M . Thus, $\mu \geq 1$ in (1.5) implies $\bar{\sigma}(\Delta) < 1$, which means that there exist a permissible perturbation in which the system becomes stable (this is related to the singularity of $I - M\Delta$). When the uncertainties relates to model parameters, the analysis searches for a combination of parameter perturbations which cause the system to become unstable. Precise calculation of μ is a NP-hard problem [82], but the bounds for μ can be computed efficiently using off-the-shelf software. The robust

performance of $M\Delta$ can be calculated in a similar fashion, requiring only a minor modification on M and Δ .

The SSV analysis is a frequency based analysis which considers the system response to a general class of input perturbations, i.e. static and dynamic [83]. However, the underlying theory behind this analysis assumes that the system is linear. Nonlinear systems commonly found in biology, have to be linearized first before the application of the SSV analysis. To apply SSV analysis, the network and its uncertainty will also need to be rearranged into $M\Delta$ block diagram. Figure 1.3 shows an example in which a (nominal) system with two inputs, an output and uncertainty about two of the model parameters can be shaped into the standard $M\Delta$ block for SSV analysis, in which the uncertainties are collected into the Δ block. Such analysis has been applied to study the robustness of biological systems, including Fas-induced apoptosis network [83], mitotic control in *Xenopus* frog eggs [76], and circadian pacemaker in *Drosophila* [76].

Bifurcation analysis: Bifurcation analysis, which has been extensively used in many fields, from engineering to chemistry to mathematics, can provide a diagram showing the dependence of system qualitative behavior on model parameters. This analysis describes the change in behavior of a system, for example from oscillatory behavior to stable steady state, when one or more model parameters are varied. The boundary in the parameter space across which these changes occur are called bifurcation points or loci [84]. The minimum distance between a nominal operating point and (the nearest) bifurcation loci in the parameter space, i.e., the minimum distance to bifurcation, has been used to measure robustness in biological networks [75]. This distance characterizes the minimum parameteric perturbation magnitude needed to change the

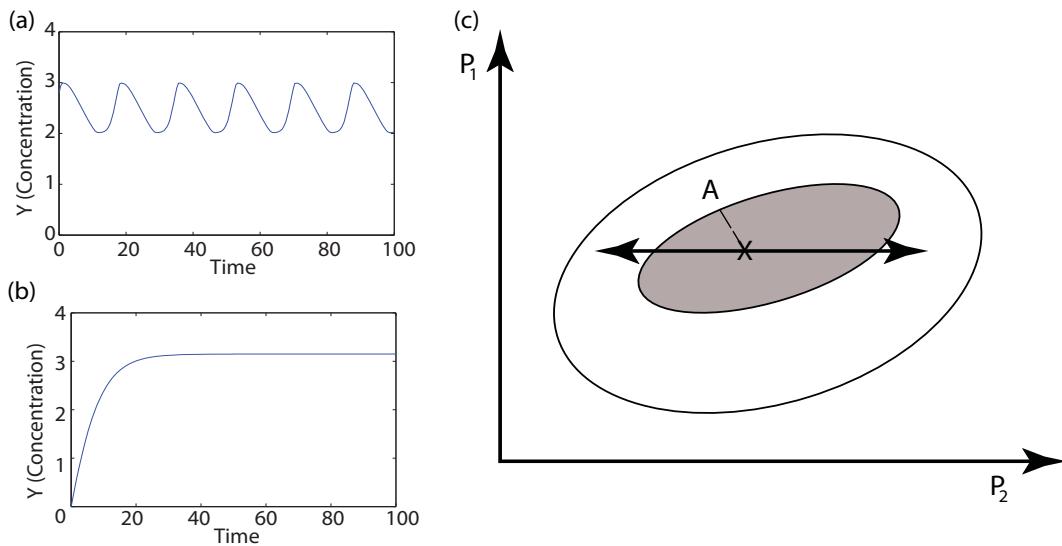


Figure 1.4. Two-parameter bifurcation diagram. Schematic representation of output profile and behavior loci of a sample system of one output and two parameters. (a) Output profile of the system for the parameter values represented by X in parameter space. (b) Output profile of the system for the parameter values represented by A in parameter space. (c) Behavior loci for a two parameter system. Distance AX is called as distance to bifurcation.

functional behavior, and thus directly correlates with robustness, i.e. larger the distance, more robust is the functional behavior.

Consider a hypothetical system with an output variable y , e.g., a concentration level, and two parameters p_1 and p_2 . For the nominal parameter values, the system is oscillating as shown in Figure 1.4(a). When parameter values are changed, system may lose the oscillating output, as illustrated in Figure 1.4(b). The transitions of the qualitative behavior of the system (oscillating to non-oscillating) for various values of parameters p_1 and p_2 are delineated in the form of a bifurcation diagram. Figure 1.4(c) shows an example of bifurcation loci for such a system, where the shaded and non-shaded regime represents oscillating and non-oscillating behaviors, respectively. The arrows show a perturbation (changes) on one of the system parameter, p_2 . As seen in Figure 1.4(c), by increasing or decreasing the parameter value (p_2) from the operating point, the system behavior changes from oscillating to non-oscillating.

The minimum distance $X - A$ in Figure 1.4(c) between the operating point and the nearest bifurcation loci gives a measure of the robustness of the system oscillatory

behavior. This distance characterizes the robustness of the system behaviour against given perturbation [85]. A system is then said to be globally robust if the feasible set of perturbations are of smaller magnitudes than the minimum distance to bifurcation. Though useful and powerful, bifurcation analysis is limited to the analysis of steady state behavior of a system. Software tools are available to map the bifurcation points of an ODE model of biochemical networks [86]. Some biological applications of this analysis include the analysis of Xenopus cell cycle oscillator [75], mitotic control in frog eggs [87], and cell cycle progression of budding yeast [88].

Parametric Sensitivity analysis: Sensitivity analysis of mathematical models is a well developed and used method in the science and engineering literature [89-95]. This analysis has also become a powerful tool to analyze mathematical models of biochemical pathways. It is the most common *in silico* analysis in systems biology due to its ease of computation and intuitive interpretation of the results. The sensitivity coefficients directly give information regarding the contribution of individual parts in a system to the desired system output behaviour. Low sensitivities indicate the property of robustness, and conversely high sensitivities implicate fragilities.

In ODE models, the system behaviour is often described by the state or output time profiles (such as concentration). As mentioned above, the model parameters include the physicochemical parameters of the system (such as those related to reaction kinetics, transport properties, etc.) as well as initial conditions, operating conditions, and geometric parameters of the systems. These physicochemical parameters are measured experimentally or estimated theoretically and therefore, have uncertainties. On the other hand, the operating conditions may change in time for a variety of reasons, such as varying receptor concentration and environmental

conditions like temperature. The parametric sensitivity analysis investigates the state changes with respect to perturbation in the parameter values [83]. One version of the sensitivity analysis uses the parametric sensitivity coefficients, defined as

$$S_{i,j} = \frac{\text{Change in concentration at time } t}{\text{Change in kinetic parameter}} = \frac{\partial x_i(t)}{\partial p_j} \quad (1.6)$$

where x_i is the i^{th} state in an ODE model with n states and p_j is the j^{th} kinetic parameter of an ODE model with m parameters, as given by (1.2). These indices are obtained directly by differentiating, and solved simultaneously by integrating the ODE models in (1.2). A more detailed explanation of this method is given in Chapter 2. Parametric sensitivity analysis has found wide applications in the analysis of biochemical networks, including programmed cell death [21, 83, 96-99], budding yeast cell cycle control [34], IL-6 signalling pathway [38], circadian rhythm model in Neurospora [33], coupled MAPK and PI3K signal transduction pathway [100], and many others.

1.3 Thesis Motivation and Objectives

As seen earlier, the complexity of cellular networks often limits human intuition in understanding how a functional regulation is accomplished in a cell. It is clear that the control of network functions cannot reside in a single gene or protein, and might well be dispersed over many components [42, 101]. A system is more than an assembly of genes and proteins and so its properties cannot be fully understood by merely drawing diagrams of their interconnections [102]. Although such a diagram represents an important first step, it is analogous to a static roadmap [101]. What one really seeks to know is the traffic patterns and the conditions under which they emerge. One would like to know how the existing molecular information can be used to understand the control of cellular behavior [18]. The classical reductionist approach of biology does

not immediately answer such questions. Since the underlying principles of complexity and robustness are universal in both biological and engineering systems, control systems theory from engineering can aid in understanding biological complexity, robustness and its tradeoffs in dynamical cellular systems. The marriage between the two fields has given birth to the field of systems biology, which focuses on the emergence of cellular functional behaviour (such as robustness) from the interactions of many biological components [18].

Unraveling biological complexity and understanding the mechanism of robustness and system dynamics has become an active field of research in systems biology [5, 16, 77, 103, 104]. The accomplishment of such goals will have tremendous impact on drug discovery research, synthetic biology and in biotechnology [19]. To this end, mathematical models coupled with quantitative analysis tools, as presented here, have been used. However, most of the methods presented in Section 1.2.2 investigate the effects of perturbations on model parameters or inputs on the system output behaviour to ascertain the importance and role of various cellular processes. As these methods focus solely on static perturbations on model parameters (including initial conditions), the dynamical aspect of cellular regulation is not immediately apparent from these analyses. In contrast, this dissertation presents novel sensitivity analyses methods that can dynamically illustrate molecule-by-molecule or reaction-by-reaction or even pathway-by-pathway insights, in producing the observed functional behaviour of the system. This is achieved by introducing dynamical perturbations on the system. Since the perturbations are dynamic in nature, the analyses presented in this work not only indicate which perturbations are important, but also point to when these perturbations matter.

The results of the dynamical sensitivity analyses presented in this dissertation can be directly validated in experiments and may even suggest experimentally testable hypotheses. In addition to understanding the underlying mechanism of cellular systems and robust-yet-fragile behavior, the knowledge gained will have applications ranging from model identification and validation to model reduction. Also, these analyses can assist drug discovery efforts in the identification of potential drug targets, the understanding of drug efficacy and specificity, and finally the optimization of drug dosing and timing. As more *in silico* models of cellular networks emerge from the field of system biology, the proposed analyses will provide the enabling technology for the use of these models to benefit human health and medicine through robustness-based systems-oriented drug design. Finally, as an additional application of the sensitivity analyses, this dissertation also details novel model reduction algorithms based on dynamic sensitivities for reducing large kinetic models of chemical and biological systems

Objectives: Therefore, based on the above discussion, the specific objectives of the current dissertation are:

1. To show the caveats of classical parametric sensitivity analysis in understanding the system dynamics,
2. To develop systems analysis methodologies based on dynamic sensitivities that can directly elucidate the cellular mechanisms behind robust-yet-fragile dynamical behavior,
3. To apply the novel dynamic sensitivities to reduce detailed kinetic models of chemical and biological systems.

CHAPTER 2

2 Parametric Sensitivity Analysis[†]

Sensitivity analysis for modelers?

Would you go to an orthopedist who didn't use X-ray?

-----Jean-Marie Furbringer

Synopsis:

Classical parametric sensitivity analysis (PSA) has become one of the most commonly used tools in computational systems biology, in which sensitivity coefficients are used to study the parametric dependence of biological models. As seen in the previous chapter, many of these models describe dynamical behaviour of biological systems. Subsequently, PSA has been used to elucidate important cellular processes that regulate this dynamics. Hence, this chapter introduces three kinds of parametric sensitivity analyses: local, global and hybrid. Also, a careful interpretation of parametric perturbations used in the PSA is presented here to explain the issue of using this analysis in inferring dynamics. Based on a synthetic switch activation example, this chapter shows that the PSA coefficients are not suitable in inferring the mechanisms by which dynamical behaviour arises and in fact, may even lead to incorrect conclusions.

[†] Excerpts of this chapter are part of the following publication:

Perumal TM and Gunawan R. (2011) Understanding dynamics using sensitivity analysis: caveat and solution. *BMC Syst. Biol.*, 5(1): 41. PMID: 214060955

2.1 Theory of Classical Sensitivity Analysis

Parametric sensitivity analysis (PSA) has found widespread applications in analyzing models of both scientific and engineering systems. It has become a must have tool in the computational systems biologists' arsenal. In systems modeling of biology, PSA has found wide applications, such as for model calibration and identifiability, model validation and reduction, identification of bottlenecking processes, elucidation of mechanisms of complex cellular behaviour, and investigation of cellular robustness [13, 105]. In most common applications of this analysis, one computes sensitivity coefficients or metrics, which generally reflects the relationship between change in model output and the perturbation on system parameters that causes this change. A few notable examples of PSA applications in dynamic biological models include programmed cell death [21, 83, 96-98], budding yeast cell cycle control , IL-6 signalling pathway [38], circadian rhythm models [16, 33, 103], and coupled MAPK and PI3K signal transduction pathway [100].

Sensitivity analysis is the most common model analysis tool for which many off-the-shelf software packages exist that provide an integrated and user-friendly computational platform for model simulations and analyses (e.g., MATLAB [106] and XPPAUT [107]). The PSA of ODE models can be readily done using software packages such as SimBiology toolbox of MATLAB [108], PottersWheel [109], Gepasi [110], Copasi [111], JDesigner/Jarnac [112], JSim [113], BioSens [114],SBML-SAT [115], and SensSB [116]. These and other softwares for sensitivity analysis have been summarized in the review articles by Alves. et al. [117] and Klipp. et al. [118]. Regardless of the tools used, the interpretations of the sensitivity metrics obtained are intuitive; parameters with large sensitivity magnitude are deemed to be important and hence, considered to be the controlling factors in the system functional

regulation. Consequently, one of the common uses of PSA in systems biology is to infer the importance (or lack of importance) of cellular processes or pathways and to provide mechanistic explanations for biological behaviour [38-42].

Depending on the magnitude of perturbations, sensitivity analyses can be classified into three types, local (infinitesimal perturbation), global (finite perturbation) and hybrid (combination of local and global methods) analyses. These are discussed in detail below.

2.1.1 Local analysis

The solution to the non-linear ODE model of (1.2), $\mathbf{x}(t, \mathbf{p})$, is the nominal solution. A finite parameter perturbation of magnitude $\Delta\mathbf{p}$ on the system will change the nominal solution $\mathbf{x}(t, \mathbf{p})$ to the perturbed solution $\mathbf{x}(t, \mathbf{p} + \Delta\mathbf{p})$. This perturbed solution of the states can be written using the Taylor series expansion as follows

$$\mathbf{x}(t, \mathbf{p} + \Delta\mathbf{p}) = \mathbf{x}(t, \mathbf{p}) + \sum_{j=1}^m \frac{\partial \mathbf{x}(t, \mathbf{p})}{\partial p_j} \Delta p_j + \dots \quad (2.1)$$

where the partial derivatives $\partial \mathbf{x}(t, \mathbf{p}) / \partial p_j$ are the first order sensitivity coefficients, respectively. The first order coefficients describe the linear change in state vector \mathbf{x} at any time t with respect to an infinitesimal perturbation on parameter p_j . In general, the parametric perturbation can be introduced at any time $\tau (t_0 \leq \tau \leq t)$ [93]. Mathematically, the first order sensitivity of state x_i with respect to parameter p_j is given by

$$S_{i,j}(t, \tau) = \frac{\partial x_i(t)}{\partial p_j(\tau)} = \frac{\text{change in the } i\text{-th state at time } t}{\text{perturbation on the } j\text{-th parameter at time } \tau} \quad (2.2)$$

But in the classical PSA, the perturbation time τ is commonly taken to be the initial time t_0 and hence, the argument τ is typically dropped out of (2.2) and the sensitivity coefficients only carry a single time dependence on the observation time t [38-40, 42, 105, 119-121]. Since the magnitude of perturbations considered are infinitesimally small, the sensitivity coefficients will depend on the nominal or baseline parameter values. Therefore, the sensitivities above are considered to be local.

A variety of approaches have been developed to solve for the above said sensitivity coefficients for ODE models, including finite difference method, direct differential method (DDM) and Green's function method [90-92, 94].

Finite Difference Method (FDM) [94]: The simplest way to calculate sensitivity coefficients in (2.2) is to apply finite difference approximation. Using only the first two terms of the Taylor series expansion in (2.1), one can compute the (first order) sensitivities by

$$\begin{aligned}\frac{\partial \mathbf{x}(t, \mathbf{p})}{\partial p_j} &= \frac{\mathbf{x}(t, p_j + \Delta p_j) - \mathbf{x}(t, p_j)}{\Delta p_j} + O(\Delta p_j) \\ \frac{\partial \mathbf{x}(t, \mathbf{p})}{\partial p_j} &\approx \frac{\mathbf{x}(t, p_j + \Delta p_j) - \mathbf{x}(t, p_j)}{\Delta p_j}\end{aligned}\tag{2.3}$$

Therefore, using FDM to calculate the sensitivity coefficients, one will need to solve the model twice, once using the nominal and another using the perturbed parameter values. FDM is particularly useful as it can handle any type of models (even when the derivative is not defined) as long as model outputs can be produced when the parameters are given. However, the computed sensitivities are only an approximation, whose accuracy depends upon the choice of the perturbation magnitude Δp_j , i.e. smaller Δp_j should be used to reduce truncation error in the Taylor series, but not too small such that the difference in \mathbf{x} due to the perturbation is still larger than the simulation error.

Direct Differential Method (DDM) [94]: DDM is an exact method for computing the sensitivities of ODE models. In order to obtain, the sensitivity coefficients as defined in (2.2), the ODE model in (1.2) is directly differentiated with respect to the model parameters to obtain the differential equation describing these parametric sensitivity coefficients [94], as follows:

$$\frac{d}{dt} \frac{\partial \mathbf{x}(t)}{\partial \mathbf{p}} = \mathbf{J}(t) \frac{\partial \mathbf{x}(t)}{\partial \mathbf{p}} + \frac{\partial \mathbf{g}}{\partial \mathbf{p}}, \quad \frac{\partial \mathbf{x}(0)}{\partial \mathbf{p}} = [\mathbf{0}_{n \times m} \quad \mathbf{I}_{n \times n}] \quad (2.4)$$

where the matrix \mathbf{J} is called the Jacobian matrix, $\mathbf{0}_{n \times m}$ and $\mathbf{I}_{n \times n}$ are $n \times m$ zero and $n \times n$ identity matrices, respectively. Here, equation (2.4) is solved simultaneously with (1.2) to obtain the sensitivity coefficients in (2.2). DDM method is the most commonly used method for computing local sensitivity values, since it provides complete information about each sensitivity index as a function of the independent variables. However, when the number of parameters m far exceeds the number of states, DDM requires solving $(n \times 1) + (n \times m)$ coupled ODEs. Hence may become stiff and expensive to solve. A more computationally efficient method is presented below. In this dissertation, DDM is used to compute the local sensitivity coefficients.

Green's Function Method (GFM) [122]: In biological systems number of system states is usually smaller than parameters, i.e., $n < m$, and hence a more efficient way to solve (2.4) is by first solving the homogeneous ODE portion of (2.4) (i.e. without the second term on the right hand side). This problem corresponds to the following Green's function problem

$$\frac{d\mathbf{G}(t, \tau)}{dt} = \mathbf{J}(t)\mathbf{G}(t, \tau); \quad \mathbf{G}(\tau, \tau) = 1 \quad (2.5)$$

where \mathbf{J} is the Jacobian matrix and $\mathbf{G}(t, \tau) = \partial \mathbf{x}(t)/\partial \mathbf{x}(\tau)$ is the Green's function matrix. Later, the sensitivity coefficients can be obtained using the following integral transformation:

$$\frac{\partial \mathbf{x}(t)}{\partial p_j} = \mathbf{G}(t, 0) \cdot \boldsymbol{\delta} + \int_0^t \mathbf{G}(t, \tau) \frac{\partial \mathbf{g}(\tau)}{\partial p_j} d\tau \quad (2.6)$$

where each element δ_κ in the vector $\boldsymbol{\delta}$ is a Kronecker delta function.

Adjoint Sensitivity Analysis [123]1: Another efficient way to calculate local parametric sensitivity coefficients is to use an adjoint method. Here, an adjoint problem of equation (2.5) is formulated as follows:

$$\frac{d\lambda(t)}{dt} = -\mathbf{J}^T(t)\lambda(t); \quad \lambda(t_f) = \mathbf{J}(t_f) \quad (2.6)$$

The evaluation of $\lambda(t)$ from equation (2.6) requires only one forward integration from t_0 to t_f of nominal model in (1.2) to compute $\mathbf{x}(t)$ and one backward integration from t_f to t_0 of the n -dimensional adjoint model. Having $\lambda(t)$, parametric sensitivity coefficients are calculated using the following integration scheme:

$$\frac{\partial \mathbf{x}(t)}{\partial p_j} = \int_0^{t_f} \lambda(t) \frac{\partial \mathbf{g}(t)}{\partial p_j} dt \quad (2.6)$$

2.1.2 Global analysis

Global analysis considers (large) finite perturbations to parameters and hence accounting for the (hyper surface) functional mapping from the input (perturbations) to the output (behavior) space, as shown in Figure 2.1. In this context, the local analysis above is only a first order approximation to such mapping, in which the sensitivity coefficients are the tangent vector at the nominal parameter values. As seen earlier in Section 1.2.1, large variations among cells originating from genomic and stochastic variations [73, 124] and the nonlinearity of typical cellular systems, may

limit the usefulness of local (linearized) analysis results. Hence, the first order local sensitivities may give an incomplete portrayal of the underlying mapping between the perturbations and system response. In a typical global analysis, the dependence of model output to parameters is evaluated by sampling parameter values from the feasible set and computing model output behavior for each parameter set. The global sensitivity coefficients are defined to reflect the variation of model outputs within the range of permissible parameter values. A few examples of global sensitivity methods are given below, while more detailed descriptions of similar analyses can be found elsewhere [125].

Stochastic Sensitivity Analysis [126]: This method is based on the assumption that all the input parameters and initial conditions are random variables whose probability density functions (pdf) are known. By doing so, the ODE model in (1.2) can be written as a Fokker-Planck model, which provides the pdf of $f_x(t, \mathbf{x})$. Here, the global sensitivity of the system output can be computed from the expected value of

$$E[\mathbf{X}] = \int \mathbf{x} f_x(t, \mathbf{x}) dx \quad (2.7)$$

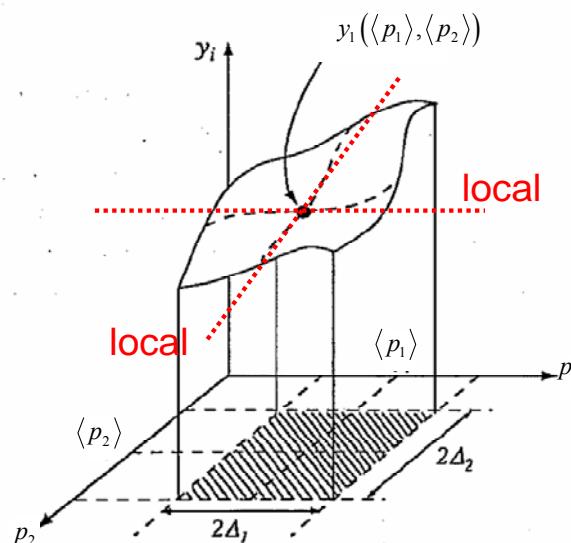


Figure 2.1. Global and local analysis. Response surface over the domain of change of two input parameters.

Stochastic sensitivity analysis has been widely applied in analyzing complex chemical kinetics [126-129], but seldom used to analyse systems biological models.

Extended Fourier Amplitude Sensitivity Test (eFAST) [130]: This variance based method was originally developed by Cukier and coworkers [130] and later improved by Saltelli and coworkers [131]. This method studies how the variation in system output can be apportioned to different inputs. In a k -dimensional unit hypercube Ω^k , high dimensional model representation (HDMR) of the marginal pdf of output $f_{X_i}(t, x_i)$ is given by

$$f_{X_i}(t, x_i) = f_{X_i}(t_0, x_i) + \sum_{k \neq i} f_{X_k}(t_0, x_k) + \sum_{k \neq i} \sum_{l > k, l \neq i} f_{X_k, X_l}(t_0, x_k, x_l) + \dots \quad (2.8)$$

in which each individual term is also square integrable over the domain of existence [132]. It has been proven that if each term in (2.8) has a zero mean, the HDMR decomposition of $f_{X_i}(t, x_i)$ is unique and every term in the decomposition is orthogonal to each other, giving

$$\begin{aligned} f_{X_i}(t_0, x_i) &= E[X_i] \\ f_{X_k}(t_0, x_k) &= E[X_i | X_k] - E[X_i] \\ f_{X_k, X_l}(t_0, x_k, x_l) &= E[X_i | X_k, X_l] - E[X_i | X_k] - E[X_i | X_l] - E[X_i] \end{aligned} \quad (2.9)$$

Hence, taking variance of $f_{X_i}(t, x_i)$ in (2.8) and substituting (2.9), gives the ANOVA-HDMR decomposition of $V[f_{X_i}(t, x_i)]$ as

$$\begin{aligned} V[f_{X_i}(t, x_i)] &= \sum_{k \neq i} V[E[X_i | X_k]] + \\ &\quad \sum_{k \neq i} \sum_{l > k, l \neq i} V[E[X_i | X_k, X_l] - E[X_i | X_k] - E[X_i | X_l] - E[X_i]] + \dots \end{aligned} \quad (2.10)$$

Finally, dividing both sides of (2.10) by $V[f_{X_i}(t, x_i)]$ will give the first order and total sensitivity indices as

$$S_{ik} = \frac{V[E[X_i | X_k]]}{V[f_{X_i}(t, x_i)]} \quad (2.11)$$

$$tS_{ik} = 1 - \frac{V[E[X_i | X_{\sim k}]]}{V[f_{X_i}(t, x_i)]}$$

Here, the first-order sensitivity index represents the direct (individual effects) contribution of each input variability to that of the output, while the total sensitivity index represents the total contribution of any input variance to the output variance, including the first-order and all higher order contributions due to interactions [132]. Thus, high sensitivity magnitudes point the importance of an input to the output. There exist a few algorithms that calculate the global sensitivity indices in (2.11) [132], but in the present dissertation, the eFAST algorithm of Saltelli and coworkers [132, 133] was used.

2.1.3 Hybrid analysis

It is clearly seen from the previous sections that global analyses are generally more computationally intensive than local analyses, whereas local analyses may not reflect the system behavior under all possible parameter sets. Hence, hybrid methods, like Derivative based Global Sensitivity Measures (DGSM) [134] and ‘Glocal’ analysis [135], have been proposed to combine local and global sensitivity analysis methods, to efficiently analyse the model behaviour under all possible parameters sets. These analyses use the global methods to search for all possible parameter sets and local methods to evaluate the system behavior under this parametric uncertainty.

Derivative based Global Sensitivity Measures (DGSM) [134]: In this method, random samples of the pdf of initial conditions, $f_x(t_0, \mathbf{x})$ and parameters, $f_p(t_0, \hat{\mathbf{p}})$ in (1.3) are generated and used as nominal models for calculating local sensitivity

coefficients, as in (2.2). From the mean and variance of the resulting distribution of local sensitivity coefficients, global sensitivities are calculated as follows

$$\begin{aligned} E\left[\frac{\partial \mathbf{X}}{\partial p_j}\right] &= \int \frac{\partial \mathbf{X}}{\partial p_j} f_{\frac{\partial \mathbf{X}}{\partial p_j}}\left(t, \frac{\partial \mathbf{X}}{\partial p_j}\right) dx; \quad V\left[\frac{\partial \mathbf{X}}{\partial p_j}\right] = \int \left(\frac{\partial \mathbf{X}}{\partial p_j} - E\left[\frac{\partial \mathbf{X}}{\partial p_j}\right]\right)^2 dx \\ G_j &= \frac{E^2\left[\frac{\partial \mathbf{X}}{\partial p_j}\right] + V^2\left[\frac{\partial \mathbf{X}}{\partial p_j}\right]}{\sum_j E^2\left[\frac{\partial \mathbf{X}}{\partial p_j}\right] + V^2\left[\frac{\partial \mathbf{X}}{\partial p_j}\right]} \end{aligned} \quad (2.12)$$

where G_j is the sensitivity coefficient. Assuming that local sensitivity coefficient calculations are as expensive as the model solutions, this analysis is computationally more efficient than the variance based global methods. Even though this analysis is very simple in implementation, it may fail to predict the changes in system behavior for global parametric variations of systems with multiple distinctive output peaks (multimodality).

Glocal analysis [135]: In this analysis the sensitivity is related to the volume of (high-dimensional) parameter space, which can produce model output behavior that is consistent with the biology of the system. The calculation of such volume is achieved using a Monte Carlo, approach guided by principal component analysis in order to allow efficient sampling. A ‘local’ analysis is then performed to determine the sensitivity for each of the samples drawn from the previously identified volume of parameter space to five different kinds of perturbations, including concentration perturbations and molecular noise. This has been applied to study the robustness of cyanobacterial circadian oscillator [135].

2.2 Caveats of Classical Sensitivity Analysis

In many applications, classical PSA is used to generate parameter ranking based on the magnitude of sensitivity coefficients, either taken at a specific time or using consolidated sensitivity metrics, such as time-integral or average or norm of sensitivity coefficients [99, 120, 136], for example using:

$$\begin{aligned} [S_{\text{inf}}]_{ij} &= \max_k (|S_{ij}(t_k)|); \quad [S_{\text{int}}]_{ij} = \int_{t_0}^t S_{ij}(\hat{t}) d\hat{t}; \quad [S_{t_k}]_{ij} = S_{ij}(t = t_k) \\ [S_{\text{FIM}}]_{ij} &= (\text{FIM})_{jj} \quad \text{where FIM} = \sum_{t_k=t_0}^t \mathbf{S}_{i[1..n]}^T(t_k) \mathbf{S}_{i[1..n]}(t_k) \end{aligned} \quad (2.13)$$

where the indices i and j again denote the i^{th} state and j^{th} parameter, and S_{inf} , S_{FIM} , S_{int} and S_{t_k} are the sensitivity metrics based on infinite norm [120], Fisher information matrix [136], time integral [99] and sensitivity magnitude at a particular time, respectively. . The parameter ranking is subsequently used to conclude about the important mechanisms or the property (such as robustness) of the biological system behaviour [38-42].

This section shows that the dynamical aspects of cellular functional regulation cannot be inferred from the sensitivity coefficients of PSA, neither directly nor as any of the consolidated sensitivity metrics shown above. More importantly, the corresponding parameter rankings from PSA can give erroneous inference about the controlling mechanisms. Briefly, the reason stems from the fact that in classical PSA, perturbations are introduced on system parameters, which are time-invariant or static. In other words, these parametric perturbations are persistent and their effects on the system behaviour are integrated over time. Therefore, while PSA can indicate which parameter perturbations are important, it does not point to when these perturbations matter. This problem is illustrated using a local PSA of a synthetic network model.

Although the illustration here was done using local sensitivity analysis, the same issue generally applies to global PSA.

2.2.1 Case study: Simple network model

To illustrate the shortcoming of local PSA in analyzing system dynamics, consider a simple six state model involving three reactions with Michealis-Menten (MM) kinetics, as shown in Figure 2.2(a) (model parameters, rate kinetics and initial concentrations are given in Table A.1). In this network, the activation of x_6 followed a switch-like dynamics in response to the stimulus x_1 , as illustrated in Figure 2.2(b) (nominal). The model describes two pathways that contribute to x_6 activation: a direct x_2 pathway and an indirect x_2, x_3 , and x_5 pathway.

In this example, *in silico* knock-out experiments were performed by removing each pathway individually in order to assess the dominance of one pathway over the other in x_6 activation. Both full network and knock-out (KO) simulations were performed under a stimulus of $x_1(t_0 = 0) = 1$. As illustrated in Figure 2.2(b), while the

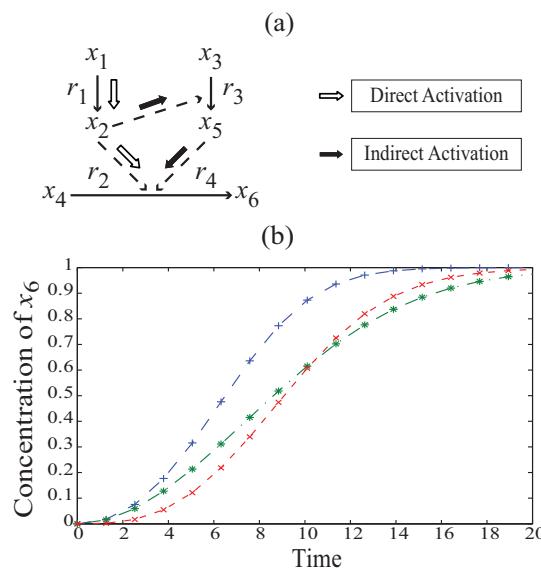


Figure 2.2. A simple network model. (a) A simple network with 6 states and 4 reactions. Straight arrows connect substrate to product and the dotted arrows indicate enzymatic activity. Details of model equations and parameter values are given in Table A.1. (b) Activation of x_6 under a constant stimulus of $x_1 = 1$: complete network (+), indirect pathway knock-out (*) and direct pathway knockout (x).

initial x_6 activation in the indirect pathway KO remained the same as that of the original model, the switch-like activation was much less pronounced. On the other hand, the original switching behaviour was preserved in the direct pathway KO, but the switching time was delayed due to a slower initial activation. Taken together, these KO simulations suggested that the x_6 activation is mainly accomplished through the indirect pathway, while the direct pathway contributes mainly to the initial x_6 activation.

2.2.2 Parametric sensitivity analysis for dynamical systems: A caveat

DDM method presented in Section 2.1.1 was also used to study the pathway dominance in this simple network. These sensitivity coefficients describe the change in system output (state trajectory) at time t with respect to (an infinitesimal) perturbation on the system parameter values at time τ . Here, the PSA was performed for the same stimulus $x_1(0)=1$ with $\tau=0$ and the sensitivity coefficients were

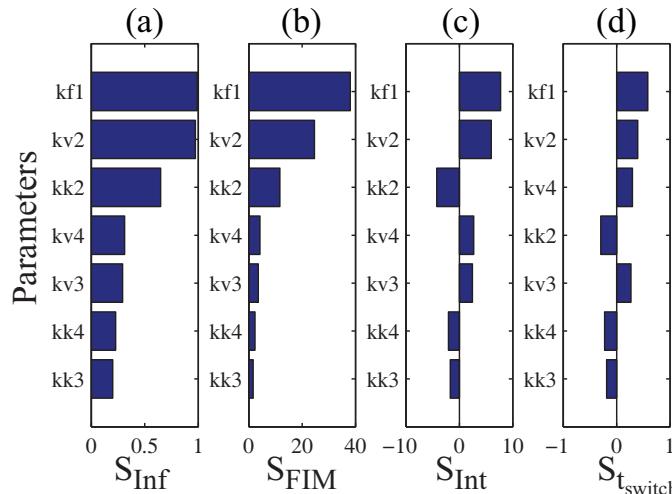


Figure 2.3. Local parametric sensitivity analysis of x_6 activation under x_1 stimulus. The bar graphs show the consolidated sensitivity metrics of x_6 with respect to model parameters based on (a) infinite norm, (b) Fisher Information Matrix (FIM) and (c) time integrated sensitivity coefficients; and (d) the sensitivity magnitudes at switching time ($t = 7.12$ time units). The parameter numbers refer to the reactions as shown in Figure 2.2, where the subscripts f denotes the forward rate constant and k and v denotes the rate constants of Michealis-Menten kinetics.

computed for the time range of 0–15 time units. The sensitivities of x_6 with respect to all model parameters are ranked in Figure 2.3 using consolidated sensitivity metrics, i.e. infinite norm [120] (Figure 2.3(a)), FIM [136] (Figure 2.3(b)), and time-integral [99] (Figure 2.3(c)), and using sensitivity magnitudes at switching time ($t = 7.12$ time units; Figure 2.3(d)). The conclusion from these rankings was the same: (1) the largest sensitivity was associated with the kinetics of x_1 conversion to x_2 and (2) the direct pathway (r_2) parameters have larger sensitivities than those from the indirect pathway (r_4), suggesting larger influence of the direct pathway on the x_6 activation. Hence, the conclusion from the PSA is in direct contradiction with the findings from *in silico* KO experiments. The above discrepancy between the KO and classical PSA results can be explained by taking a closer look at the way parametric sensitivity coefficients in (2.2) are calculated:

$$S_{i,j}(t, \tau) = \int_{\tau}^t \dot{S}_{i,j}(\hat{t}) d\hat{t} = \int_{t_0}^t \dot{S}_{i,j}(\hat{t}) H(\hat{t} - \tau) d\hat{t} \quad (2.14)$$

where $\tau = t_0$ is the usual perturbation time, $\dot{S}_{i,j}(\hat{t})$ is the time derivative of sensitivity coefficient as shown in (2.4) and $H(t)$ is the Heaviside step function. In this case, since model parameters are static or time-invariant, the parametric perturbations in the PSA consist of step changes in the parameter values, as depicted in Figure 2.4(a). Hence, the sensitivity coefficients at time t represent an integrated or accumulated change in the states from τ to t due to a persistent parameter change started at time τ (see Figure 2.4(b)). Indeed, substituting the full equation of $\dot{S}(t)$ from (2.4) in (2.14) gives

$$S_{i,j}(t, \tau) = \int_{t_0}^t \left[\mathbf{J}_{i,[1..n]} S_{[1..n],j}(\hat{t}) + \frac{\partial g_i}{\partial p_j}(\hat{t}) \right] H(\hat{t} - \tau) d\hat{t} \quad (2.15)$$

Here, one sees two terms in the integrand that contribute to the sensitivity coefficients at time t : (1) the first is related to the (integrated) sensitivities that are carried over from the initial perturbation time τ and (2) the second accounts for the instantaneous rate changes due to the parametric perturbations that still persist at time \hat{t} . Thus, in the PSA, a large sensitivity magnitude of $S_{i,j}(t, \tau)$ indicates the importance of the j -th parameter in time window of τ and t , during which the perturbation is applied to the system. Hence, the use of these coefficients to infer the dynamical importance of parameters is inappropriate and can even be misleading.

The above said reason is responsible for the PSA of the simple network model rendering an incorrect conclusion regarding direct versus indirect pathway activating x_6 . As seen in the *in silico* KO experiments, the direct pathway regulates the initial activation of x_6 , while the actual switching is carried out by the indirect pathway. In the PSA of this model ($\tau = 0$), the early importance of the direct pathway and also the reaction r_1 persisted beyond the initial times in the sensitivity coefficients due to the aforementioned integrated effect. In this case, the importance of the indirect pathway was not apparent from the parameter sensitivity rankings in the background of large (integrated) sensitivities with respect to r_1 and the direct pathway. Correspondingly,

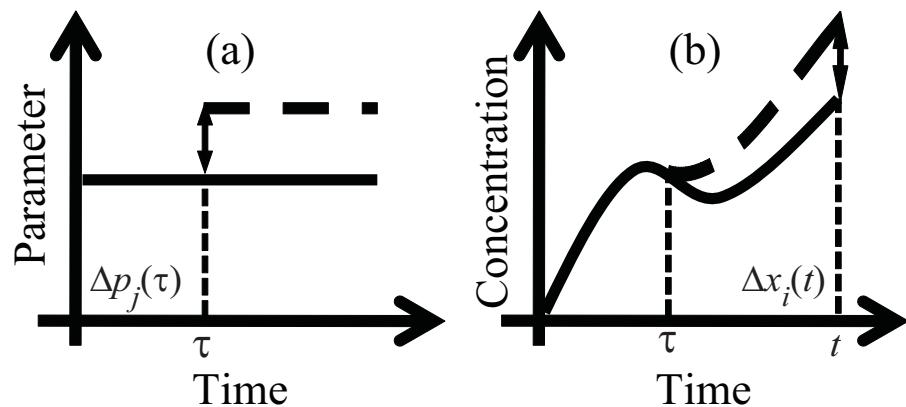


Figure 2.4. Illustration of perturbation-response in parametric sensitivity analysis. Solid lines represent the nominal and the dashed lines show the perturbed trajectory, respectively. Figures are not drawn to scale.

the use of any time-consolidated sensitivity metrics will only worsen this problem.

2.3 Summary

Sensitivity analysis has become one of the must have tools in the computational systems biologist arsenal. It addresses the manner in which the functional output behavior depends on the system parameters. As seen in Section 2.1, based on the type of parametric variations, sensitivity analysis falls in to three categories: local, global and hybrid. Local analysis provides an asymptotic approach by analyzing the local linearized change of system output with respect to an infinitesimal (asymptotically small) parametric variation at a given point in the parameter space (a nominal operating condition). On the other hand, global analysis provides an exploratory approach by addressing the change in model behavior over a wide range of parameter values. Complementing both these analyses, hybrid methods combine both the local and global methods to analyse the parametric dependence of system output behavior under uncertainty. Even though this chapter introduces and provides interpretations of some of these methods, for a more detailed explanation and applications of these methods, readers are referred to more detailed reviews [94, 105, 125].

While classical parametric sensitivity analysis provides a powerful tool to understand the parametric dependence of biological behaviour, its suitability in inferring mechanisms of dynamic behaviour has not been properly addressed. Using both, a rigorous theoretical formalism and a synthetic network model, the problem of using the classical PSA in identifying the controlling mechanisms of a dynamical system is illustrated. The issue mainly arose from the information needed to do this inference, where one needs to know not only which parameters are critical, but also when they matter. However, the persistent parametric perturbations in standard PSA are incapable of providing this information as the sensitivity coefficients represent an

integrated effect. Of course, this does not mean that the PSA of dynamical models is incorrect, but rather indicates the fact that the interpretation of the sensitivity coefficients should be carefully managed. In particular, a large sensitivity magnitude with respect to a parameter suggests the importance of this parameter in the time period between the perturbation time τ and the state observation time t , but not its dynamic importance.

CHAPTER 3

3 Green's Function Matrix based Sensitivity Analysis[‡]

Synopsis:

The use of mathematical models to understand the functional regulations in a cell and the caveat of classical (parametric) sensitivity analysis in understanding the underlying system dynamics have been shown in Chapters 1 and 2, respectively. In order to overcome the caveat of PSA, this chapter introduces a dynamical sensitivity analysis based on the use of the Green's Function Matrix (GFM) as sensitivity coefficients with respect to concentration perturbations. In contrast to the classical (parametric) sensitivity analysis, the GFM analysis gives a dynamical, molecule-by-molecule insight on how system behaviour is accomplished and complementarily how (impulse) signal propagates through the network. The efficacy of the method is demonstrated through applications to common network motifs and a FasL-induced programmed cell death model in Jurkat T cell line.

[‡] Excerpts of this chapter are part of the following publications:

Perumal TM, Wu Y, and Gunawan R. (2009) Dynamical analysis of cellular networks based on the Green's function matrix. *J Theor Biol*, 261(2) 248-59. PMID:19660478

Perumal TM, Yan W and Gunawan R. (2008) Robustness analysis of cellular systems for *in silico* drug discovery, *In Proceedings of the 17th World Congress - The International Federation of Automatic Control (IFAC)*, Seoul, Korea, July 6-11, PP: 12607-12612

3.1 Introduction

As seen earlier in Chapter 1, the complexity of cellular networks often limits human intuition in understanding functional regulations in a cell from static network diagrams. To this end, mathematical models of ordinary differential equations (ODE) have commonly been used to simulate dynamical behaviour of cellular networks, to which a quantitative model analysis can be applied in order to gain biological insights. Unfortunately, as shown in Chapter 2, the dynamical aspect of cellular regulation may not be immediately apparent from the application of existing ODE model analyses tools, like the classical parametric sensitivity analysis (PSA). The reason stems from the fact that the perturbations considered are on system parameters, which are static variables and hence are persistent in nature, thereby showing an integrated effect on the system output. Also they are realized at single time point, which is usually the initial time.

In contrast, the analysis presented here, is able to dynamically illustrate the sequence of molecules that participate in producing the observed functional behavior of the system. This analysis is based on the Green's Function Matrix (GFM), which is a well-known method to solve non-homogeneous differential equations [137], including those for classical PSA [94]. Specifically, the analysis uses the GFM as sensitivity coefficients with respect to impulse perturbations on the state variables at different initial times, and thereby reveals dynamic information about the actively participating states. Though this aspect of the GFM has been previously introduced to analyze ODE models of combustion kinetics [138], it was not been well characterized and understood to analyse the dynamics of deterministic ODE models in systems biology. Here, the application of the GFM analysis on a model of FasL-induced programmed cell death in human cancer T cell line Jurkat [98] revealed the key

regulators of caspase-3 activation and the FasL death signal propagation in the network. The information gained from the GFM analysis can be useful in the drug discovery research to identify potential drug targets, to understand drug efficacy and specificity, and to optimize drug dosing and timing.

3.2 Green's Function Matrix (GFM) based Sensitivity Analysis

The GFM analysis is built on the same lines of PSA, i.e., cause-effect relationship. It was shown earlier in Section 2.1.1, the Green's function method can be used to calculate the classical parametric sensitivity coefficients, in which the homogeneous part of (2.4) (the Green's function problem) is computed first and subsequently the sensitivity coefficients are obtained by an integral transformation of the non-homogeneous part [122]. Here, the GFM coefficients themselves better portray the dynamical behavior of the system [138]. As explained earlier, the GFM coefficients in essence reflect the model sensitivities to perturbations on concentrations at time τ and hence the $(i,j)^{\text{th}}$ element of the GFM $\mathbf{S}^x(t, \tau)$ corresponds to

$$\bar{S}_{i,j}^x(t, \tau) = \frac{\partial x_i(t)}{\partial x_j(\tau)} = \frac{\text{change in the } i\text{-th state at time } t}{\text{perturbation on the } j\text{-th state at time } \tau} \quad (3.1)$$

where x_i and x_j are the i^{th} and j^{th} state in an ODE model and the observation time t is larger than or equal to the perturbation time τ . In other words, the GFM coefficients quantify the relative change in the i^{th} system state at time t caused by an impulse perturbation in the j^{th} state at time τ . Figure 3.1(a) illustrates the cause-effect relationship between an impulse perturbation on x_j at time τ and its corresponding effect on x_i at time t . These coefficients are formulated directly from and solved simultaneously with (1.2). Like in the classical PSA, the GFM analysis is also

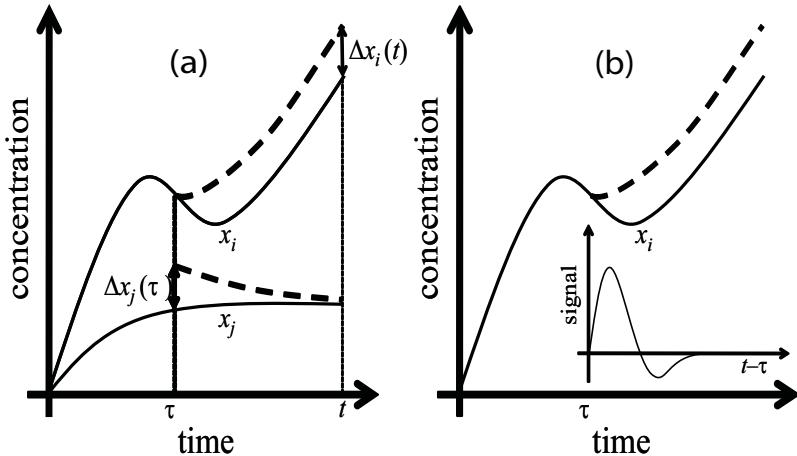


Figure 3.1. Green's Function Matrix analysis (figures are not drawn to scale) (a) The $(i,j)^{\text{th}}$ GFM coefficient describes the (normalized) ratio between the relative change Δx_i in the i^{th} molecular concentration at time t and the causative impulse perturbation Δx_j applied to the j^{th} molecular concentration at time τ . The dashed lines show the concentrations after the perturbation Δx_j is introduced. (b) A signal progressing from x_j through the molecule x_i is defined as the impulse change in Δx_i at time t relative to the impulse perturbation Δx_j at time τ . In this example, the early signal is positive since the perturbation initially leads to an increase in Δx_i , but becomes negative later and then disappears over time as Δx_i reached its steady state value.

reported as the normalized sensitivities and henceforth the notation $S^x(t, \tau)$ shall

refer to:

$$S_{i,j}^x(t, \tau) = \frac{\partial x_i(t)}{\partial x_j(\tau)} \frac{x_j(\tau)}{x_i(t)} = \frac{\partial \log x_i(t)}{\partial \log x_j(\tau)} \quad (3.2)$$

Each $(i,j)^{\text{th}}$ element of the GFM can be visualized as a heat map, as shown in Figure 3.2. The positive (negative) sensitivities in the heat map illustrate the dynamical aspect of the perturbation-output relationship: if the states denote molecular concentrations, then an impulse increase in the j^{th} molecular concentration x_j at time τ will cause an eventual increase (decrease) in the i^{th} molecular level x_i at time t . The lower right half of the plot is null as the system is causal, i.e. the perturbation will not cause any changes in the states at times $t < \tau$.

Other dynamical information that can be extracted from the GFM relates to how the effect of a perturbation propagates in the network. By defining this perturbation as a signal, the signal progression through molecules or nodes is measured by the time incremental (impulse) change in the molecular concentration x_i

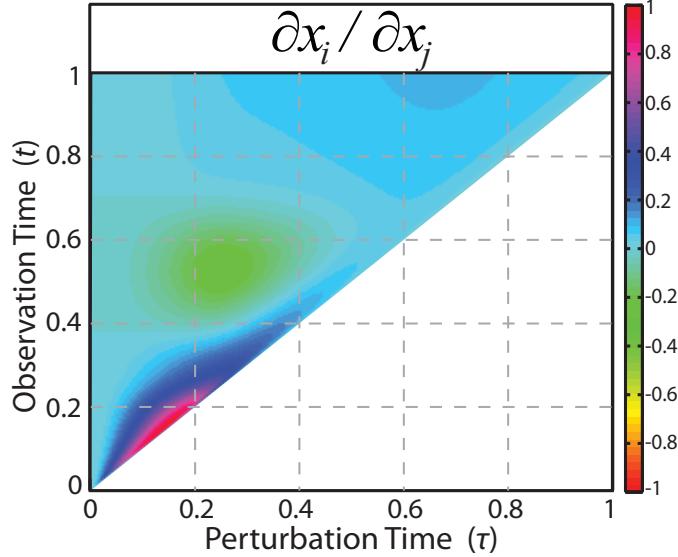


Figure 3.2. A heat map of the GFM coefficient. The heat map visualizes the $(i,j)^{\text{th}}$ GFM coefficient, showing the change in the x_i level with respect to a perturbation on the x_j . The x -axis of these plots represents the perturbation time τ at which an impulse perturbation is introduced on x_j , while the y -axis represents the observation time t at which the level of x_i changes.

at time t relative to the original perturbation, as illustrated in Figure 3.1(b). Hence, when time tends to infinity, the impulse signal asymptotically reaches zero.

Mathematically, the (propagated) signal starting from x_j at time τ is defined as:

$$\lim_{\Delta t \rightarrow 0} \frac{\Delta x_i(t + \Delta t) - \Delta x_i(t)}{\Delta t} \frac{1}{\Delta x_j(\tau)} \quad (3.3)$$

where $\Delta x_i(t) = x_i^\Delta(t) - x_i(t)$ and $x_i^\Delta(t)$ refers to the perturbed state trajectory. Taking the limit of perturbation magnitude to zero (infinitesimal perturbation) and reformulating the numerator terms, one can obtain a signal propagation measure according to:

$$\lim_{\substack{\Delta t \rightarrow 0 \\ \Delta x_j(\tau) \rightarrow 0}} \frac{\Delta x_i(t + \Delta t) - \Delta x_i(t)}{\Delta t} \frac{\Delta x_j(\tau)}{\Delta x_j(\tau)} = \lim_{\Delta t \rightarrow 0} \frac{S_{i,j}^x(t + \Delta t, \tau) - S_{i,j}^x(t, \tau)}{\Delta t} = \frac{dS_{i,j}^x(t, \tau)}{dt} \quad (3.4)$$

Thus, the evaluation of the signal propagation measure involves a simple matrix multiplication between Jacobian and the GFM, as given in (2.5).

The calculation of the GFM coefficients from the combined ODEs of (1.2) and (2.5) for different τ 's is computationally expensive if done directly. Since (1.2) is a linear time-varying (LTV) system and the GFM is the fundamental matrix of this system, the computational cost can be reduced significantly by taking advantage of the semigroup property, in which

$$\mathbf{S}^x(t_2, t_0) = \mathbf{S}^x(t_2, t_1)\mathbf{S}^x(t_1, t_0) \quad (3.5)$$

for any $t_0 \leq t_1 \leq t_2$ [139]. Thus, the complete information of $\mathbf{S}^x(t, \tau)$ can be constructed sequentially from the solutions of $\mathbf{S}^x(\tau_k + \Delta\tau, \tau_k)$ for all time of perturbations τ_k 's, i.e. one only needs to solve (2.6) for one time step $\Delta\tau$ from each τ_k . In a modular system without feedback, like those in Figure 3.3, the GFM analysis of the overall network can be reconstructed from the analysis of individual subsystems. For example, consider two modules A and B interacting by means of the molecule y_1 , as shown in Figure 3.3(a). The GFM coefficients of the overall system

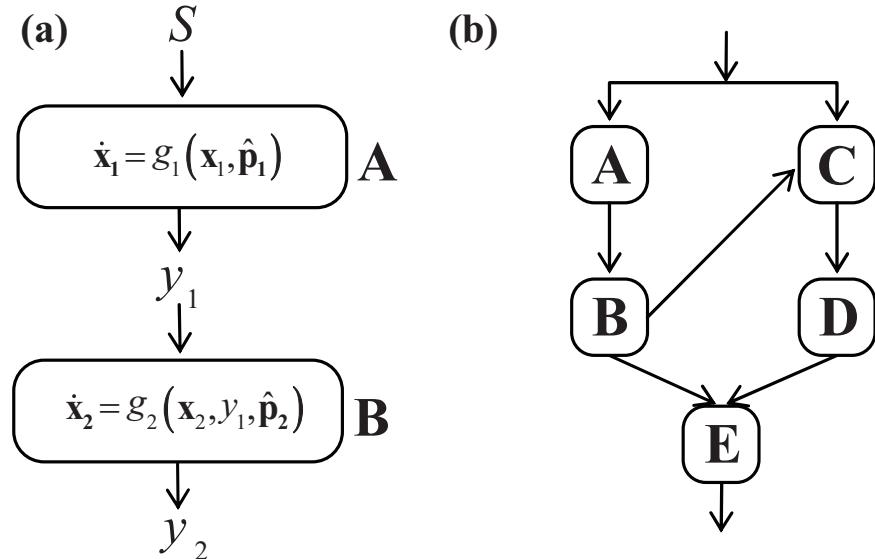


Figure 3.3. GFM analysis of networks without feedback. The GFM of the overall system can be reconstructed from the analysis of each modular sub-network. (a) Let \mathbf{x}_1 and \mathbf{x}_2 be the state vectors of systems A and B in series, respectively, in which the communication between the systems is accomplished by y_1 . The GFM of the output y_2 in response to the stimulus S can be calculated from the GFM of A and B according to (3.5). (b) The same analysis can also be done for networks with cross talks.

can be computed from the GFM of A and B using the chain rule of differentiation as follows:

$$\frac{\partial \mathbf{x}_2(t)}{\partial \mathbf{x}_1(\tau)} = \int_{\tau}^t \frac{\partial \mathbf{x}_2(t)}{\partial y_1(t')} \frac{\partial y_1(t')}{\partial \mathbf{x}_1(\tau)} dt' \quad (3.6)$$

where \mathbf{x}_1 and \mathbf{x}_2 are the state vector of the ODEs describing module A and B, respectively. As existing modeling efforts in systems biology typically focus on subsystems of the whole cellular networks, the GFM based analysis of subsystems, possibly done by separate research laboratories, can be integrated easily. On the other hand, the traditional (parametric) sensitivity analysis of the combined networks will require solving the sensitivity coefficients of the full system, which may involve a significant computational cost. The same integration is also possible in general nonlinear systems with cross-talks, as shown in Figure 3.3(b). Unfortunately, if there exists any feedback in the network, such reconstruction is no longer possible.

3.3 Case Studies

3.3.1 Application to common network motifs

A complex molecular network is often made up of simpler network modules, i.e. a network of networks, connected by standard linkages [140]. In order to illustrate the property of the GFM coefficients, this analysis was first applied to examples of common network motifs in biology [5]. In these small networks, the system output R responds to a constant stimulus S (see Section B.1 for model equations, parameter values, and the ODE solver used). Table 3.1 describes the results of the GFM analysis in comparison to the classical parametric sensitivity analysis, when applied to three network motifs.

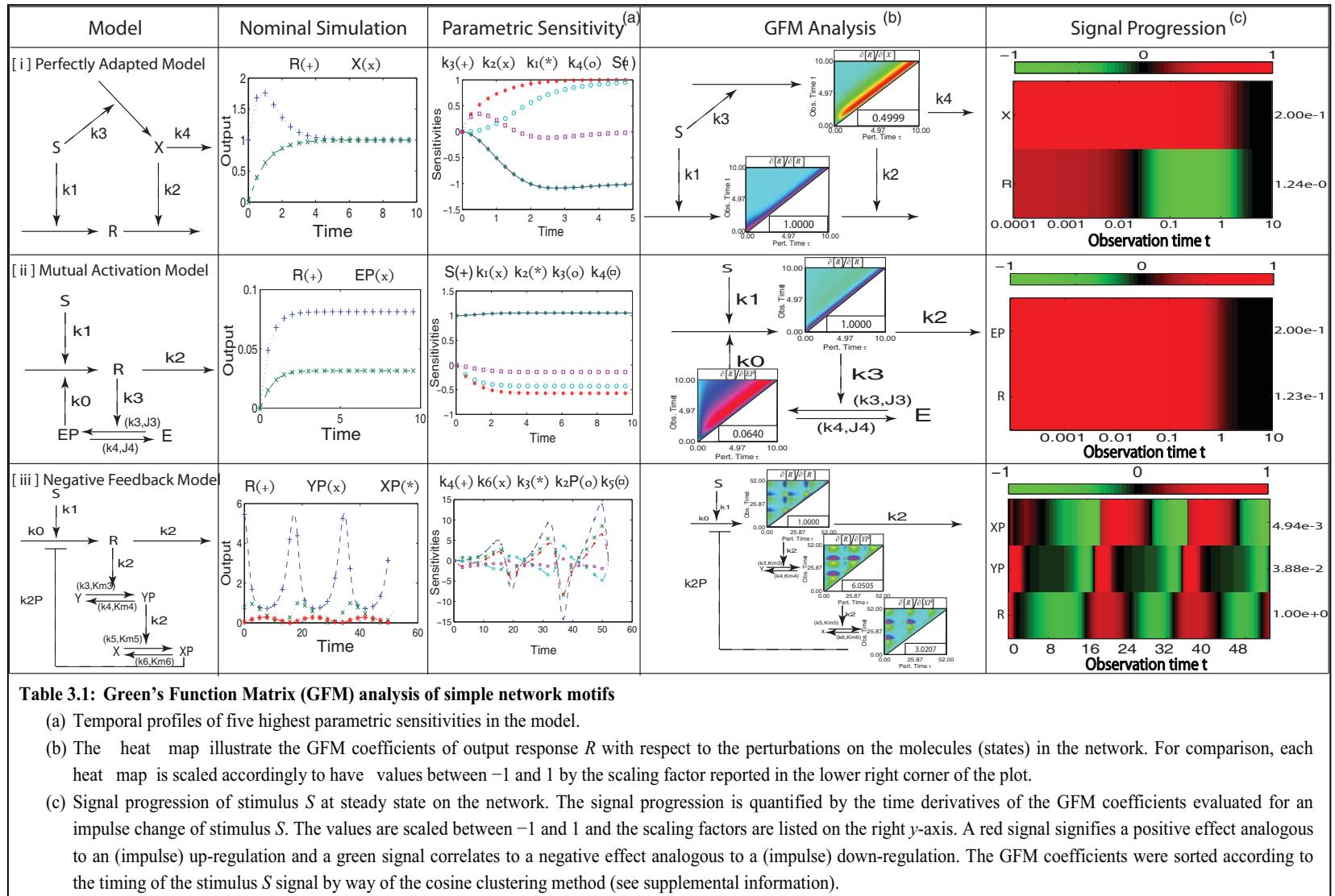


Table 3.1: Green's Function Matrix (GFM) analysis of simple network motifs

- (a) Temporal profiles of five highest parametric sensitivities in the model.
- (b) The heat map illustrate the GFM coefficients of output response R with respect to the perturbations on the molecules (states) in the network. For comparison, each heat map is scaled accordingly to have values between -1 and 1 by the scaling factor reported in the lower right corner of the plot.
- (c) Signal progression of stimulus S at steady state on the network. The signal progression is quantified by the time derivatives of the GFM coefficients evaluated for an impulse change of stimulus S . The values are scaled between -1 and 1 and the scaling factors are listed on the right y -axis. A red signal signifies a positive effect analogous to an (impulse) up-regulation and a green signal correlates to a negative effect analogous to a (impulse) down-regulation. The GFM coefficients were sorted according to the timing of the stimulus S signal by way of the cosine clustering method (see supplemental information).

3.3.1.1 Perfectly adapted signal-response model

The first and the simplest example is made up of synthesis and degradation processes (Table 3.1[i]). In most cases, this network motif is embedded in more complex pathways to generate adaptive signal-response curves [140]. For a constant stimulus S , the response R exhibits a perfect adaptation as shown in Table 3.1. Perfect adaptation means that although the transient response depends on the stimulus strength, the steady state magnitude of R is independent of S . The classical parametric sensitivity analysis suggested that the response R depends most strongly to the synthesis parameter (k_3) and the degradation parameter (k_2).

The GFM analysis of the response R with respect to perturbations on X and R itself revealed the underlying system dynamics. Perturbations of X at early times had lesser effect on R than those at latter times, indicating that the initial level of X was of lesser importance than the production of X due to the stimulus S . In addition, these perturbations had a negative effect on the levels of R due to the degradatory action by X . As expected, the analysis illustrated the perfect adaptation response in which perturbations on R and X only cause transient changes and the steady state level of R is insensitive to these perturbations.

3.3.1.2 Mutual activation model

The second network motif consists of synthesis, degradation, phosphorylation and dephosphorylation processes (Table 3.1[ii]). The network diagram and the simulated trajectory intuitively explained that in response to a constant stimulus S , the output molecule R is produced or activated and this R will in turn lead to higher production of itself through a positive feedback involving phosphorylation of the enzyme E to EP . Since R and EP mutually activate each other, this network motif is called a

mutual activation model. The parametric sensitivity analysis of this model showed that the response R is more sensitive to the stimulus S than to the positive feedback action of EP .

The GFM analysis of the response R with respect to perturbations on EP and R gave a counter-intuitive dynamic picture. According to the GFM analysis, an impulse perturbation on R initially gave an expected increase in R (positive sensitivities), but later led to a drop in the level of R (negative sensitivities). In order to validate this observation and to explain the cause, a perturbation on R was introduced in the simulation at time $\tau = 5$. As shown in Figure 3.4(a), this perturbation

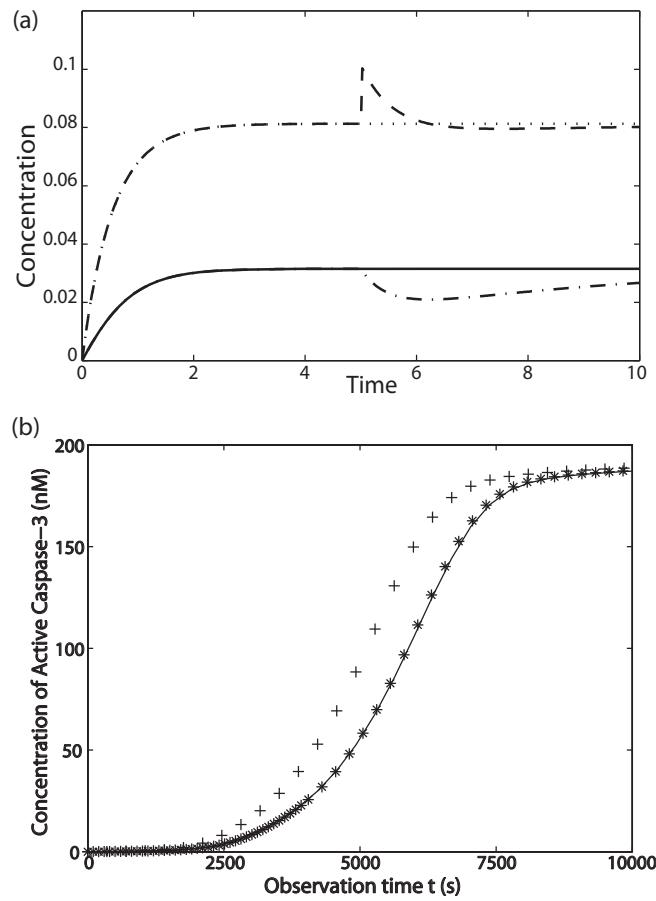


Figure 3.4. *In silico* validation of the GFM analysis. (a) Mutual activation model under signal stimulus $S = 15$. An impulse perturbation of 25% relative magnitude was introduced to R at $\tau = 5$ ($\Delta P = 0.02$ units). The responses R before (--) and after (···) perturbation are shown here with the EP profiles before (—) and after (—●) perturbation. (b) Caspase-3 Activation under 2nM FasL Stimulus. Impulse perturbations of 50% relative magnitude were introduced on pro-caspase-8 at $\tau = 0$ ($\Delta Procaspace-8 = 16.665$ nM) and $\tau = 4000$ seconds ($\Delta Procaspace-8 = 11.490$ nM). The active caspase-3 concentrations are plotted for nominal (—), $\Delta Procaspace-8$ at $\square \tau = 0$ (+), and $\Delta Procaspace-8$ at $\square \tau = 4000$ seconds (*).

resulted in a prolonged decrease in the phosphorylation of E , thereby reducing the amount of R produced by the positive feedback loop. The GFM coefficients of EP with respect to R further confirmed this conclusion (see Section B.1). The transition from fast and direct stimulus response to slow and indirect positive feedback was not apparent from the integrated parametric sensitivities.

3.3.1.3 Negative feedback oscillator model

Oscillations in biological systems are often produced by a negative feedback loop or by a combination of positive and negative feedback loops. The next network motif consists of simple synthesis-degradation and phosphorylation-dephosphorylation processes in a negative feedback fashion (Table 3.1[iii]). The synthesis of R induced by S phosphorylates Y , which in turn phosphorylates X . The phosphorylated XP subsequently acts to inhibit the synthesis of R , forming the negative feedback loop. For a given stimulus S , the system executes sustained oscillations of R , YP and XP . The parametric sensitivity analysis of this oscillatory system diverged with time due to period changes, as expected, thus requiring further treatment to decouple the period and amplitude effects [79].

The GFM analysis of R with respect to impulse perturbations on R , YP and XP showed positive and negative sensitivity oscillations, which was expected to be in a limit cycle oscillatory system. Here, impulse perturbations on the states like those considered in the GFM caused a phase shift (lead/lag), in which the perturbed trajectory R alternated from being higher and lower than the nominal response. Similar analysis had been used to compute the phase sensitivities in an oscillatory system, such as circadian rhythms [78]

3.3.1.4 Tracking the progression of stimulus signal

Network diagrams can be used as mental models to qualitatively illustrate how an impulse signal propagates in the network. However, such pictures lack quantitative and dynamic information as these diagrams are static in nature. In addition, complex network structures such as feedback/feedforward loops in biology will further complicate the deduction of signal progression from static diagrams. When kinetic ODE models and parameters of these networks are available, the time derivatives of the GFM coefficients can be used to visualize the signal propagation by measuring the transient impulse changes in the molecular concentrations (i.e. nodes in the network) caused by the initial signal, as described above and illustrated in Figure 3.1(b).

Table 3.1 (last column) also gives the dynamical progression of an impulse signal of stimulus S in the simple network motifs above. The signal progression plots elucidate the timing and magnitude of signal as it travels through the nodes (states, molecules) in the network. In the case of perfectly adapted system, the signal was immediately seen in R , while only progressed to X after some time delay. This was consistent with the GFM analysis above indicating that early perturbations of X did not matter as much as later. In the mutual activation model, the signal appeared in R and EP almost simultaneously. However, the signal disappeared from R after $t = 1$, while impinged on EP until $t = 10$. In other words, the direct activation of R by S dominated over the positive feedback early on, which was reversed at the later times. Finally, in negative feedback oscillator, the initial carrier of signal was R , followed by YP and then XP , as expected. Once the oscillations developed, the signal progression showed repeating patterns of the $R-YP-XP$ sequence.

In the above examples, the GFM analysis was able to illustrate dynamically the progression of system response to a stimulus and of signal from a stimulus. The results were much in agreement with the knowledge of how these network motifs function and are designed, demonstrating the validity and efficacy of this analysis in known, small networks. In the next section, the GFM analysis was applied to a more complex network involving the programmed cell death to deduce the dominant pathway in the signaling network.

3.3.2 Application to FasL-induced cell death model of human Jurkat cells

In this example, the states represent biomolecules involved in the programmed cell death (apoptosis) signalling of human Jurkat T lymphocytes induced by the death

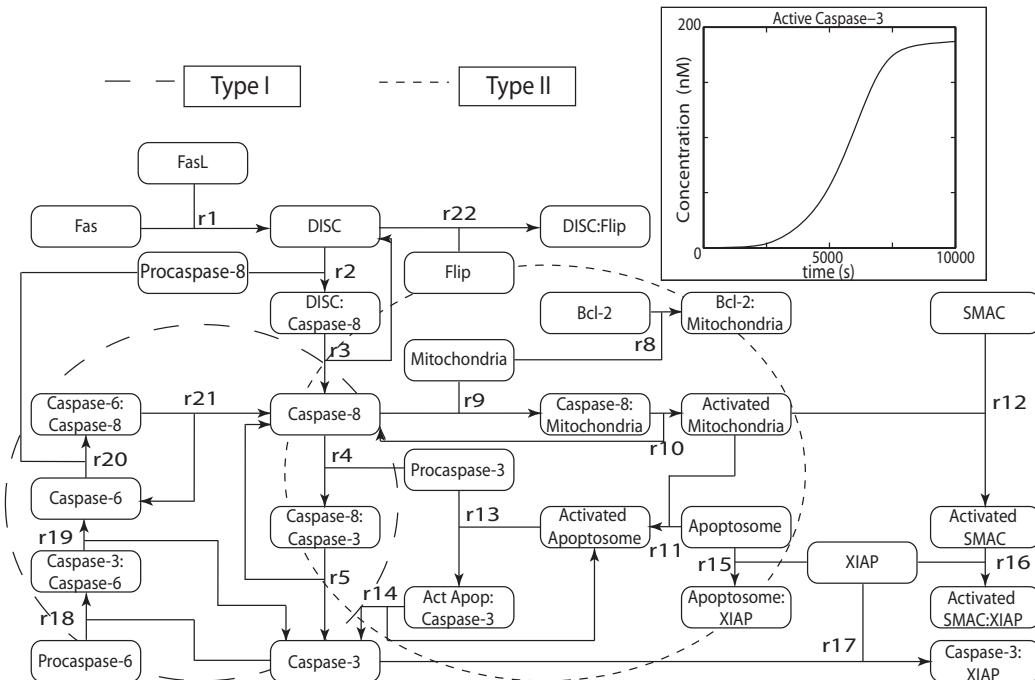


Figure 3.5. Network model of FasL-induced programmed cell death in Jurkat T cell lines. The type-I pathway involves a direct cleavage of pro-caspase-3 by caspase-8 to form an active caspase-3, while the type-II pathway describes a mitochondria-dependent activation of caspase-3. The rate equations and parameters are available in supplementary material (see Section B.2). The caspase-3 activation follows a switch-like behaviour in response to a constant FasL stimulus of 2 nM (see inset).

ligand FasL (see Figure 3.5). The model equations and parameters were obtained from a fit to experimental data [98](see Section B.2 for detailed reaction rates, kinetic parameters, type of solvers used and their settings). The output of interest was caspase-3, also known as the executioner caspase, which is a protease that cleaves many protein substrates [37]. In this network model, the activation of caspase-3 followed a switch-like response to a FasL stimulus as shown in Figure 3.5 (see inset). Although the model still lacks detailed descriptions of several key molecules in cell death signalling (e.g., Bax/Bad, Bid), it includes two major pathways for the caspase-3 activation: a direct caspase-8 (type-I) and a mitochondria-dependent pathway (type-II). The GFM analysis was used below to assess the dominance of one pathway over the other in this cell death model, as well as to study the signal propagation of a FasL stimulus through the network. The parametric sensitivity and GFM analysis was performed under a constant FasL stimulation ($\text{FasL} = 2\text{nM}$) over the time range of 0 to 10,000 seconds in order to allow the system to reach its new steady state. In this case, the model simulated an apoptotic cell, and the purpose of the analysis was to elucidate the key reactions and thereby predicting the important pathway in the regulation of the caspase-3 activation during programmed cell death (see Figure 3.5 inset).

3.3.2.1 Parametric sensitivity analysis

By selecting caspase-3 as the output of interest, the sensitivities of active caspase-3 level to perturbations on the model parameters in the network□ are given in Figure 3.6 in the order of decreasing importance. The sensitivity indices represent a consolidation of the parametric sensitivities over time using the Fisher Information Matrix [77, 141]. The analysis revealed that: (a) the formation of DISC complex and the subsequent activation of pro-caspase-8 had a very high sensitivity indicating the

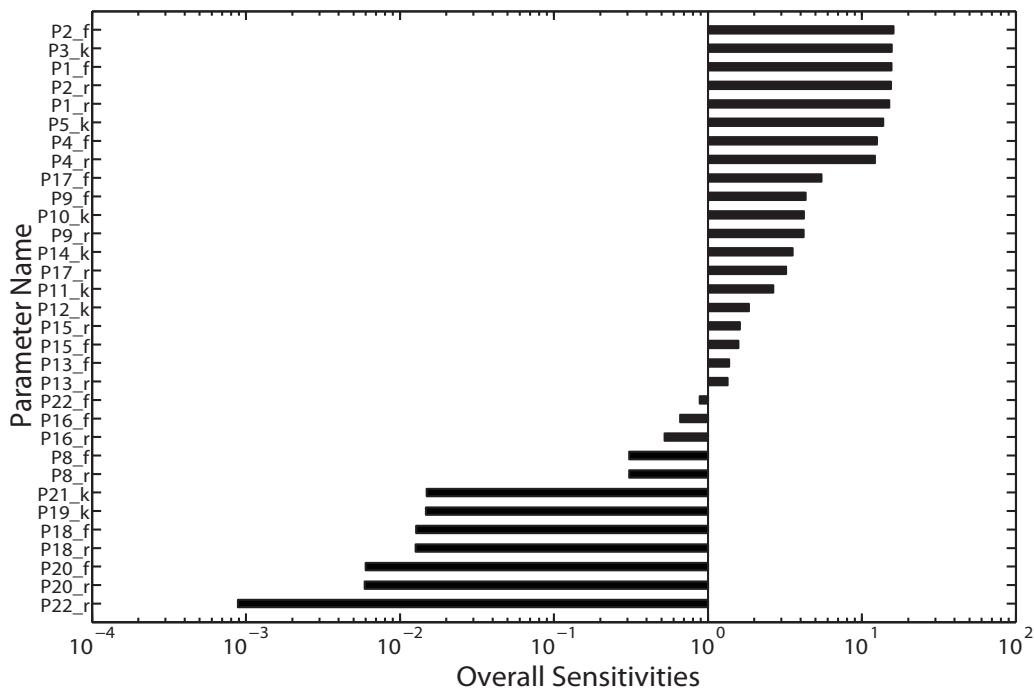


Figure 3.6. Parametric sensitivity analysis of caspase-3 activation by a FasL stimulus. The bar graph represents a consolidation of the parametric sensitivity of active Caspase-3 concentration over a time range of 0 – 10000 seconds based on the Fisher Information matrix. The parameter numbers refer to the reactions as shown in Figure 3.5, where *f* and *r* represents forward and backward rate constants for reversible reactions and *k* represents the rate constants for irreversible reactions. Detailed parameter values are given in Section B.2.

importance of caspase-8 production (at least initially), (b) the formation of caspase-8:procaspase-3 complex had higher sensitivity than that of apoptosome:procaspase-3 complex. Other sensitivity indices such as using the maximum absolute values or peak sensitivities at a given time (taken as switching time), gave the same conclusions (see Figure B.2 and B.3). Hence, the parametric sensitivity analysis suggested a Type-I dominant pathway in the cell death regulation of Jurkat cells.

3.3.2.2 Green's function matrix (GFM) analysis

Again, by selecting caspase-3 as the output of interest, the sensitivities of active caspase-3 level to perturbations on the different molecules in the network at different τ s are illustrated in Figure 3.7. Each subplot represents an element in the row of GFM corresponding to the active caspase-3. The time range on the *t*-axis (ordinate)

between 5000 to 7000 seconds is of particular interest as these times correlates with the switching of caspase-3 activation by FasL stimulus.

In contrast to the conclusion reached by the parametric sensitivity analysis, a careful observation of the GFM coefficients in Figure 3.7 revealed a dominant mitochondria-dependent (type-II) pathway in the activation of caspase-3. Specifically, the analysis painted the following dynamical picture in regard to this activation (again by focusing on the t-axis between 5000 to 7000 seconds): (a) the formation of DISC and the subsequent activation of caspase-8 were the initial cell death response; (b) active caspase-8 next helped to “activate” the mitochondria by membrane

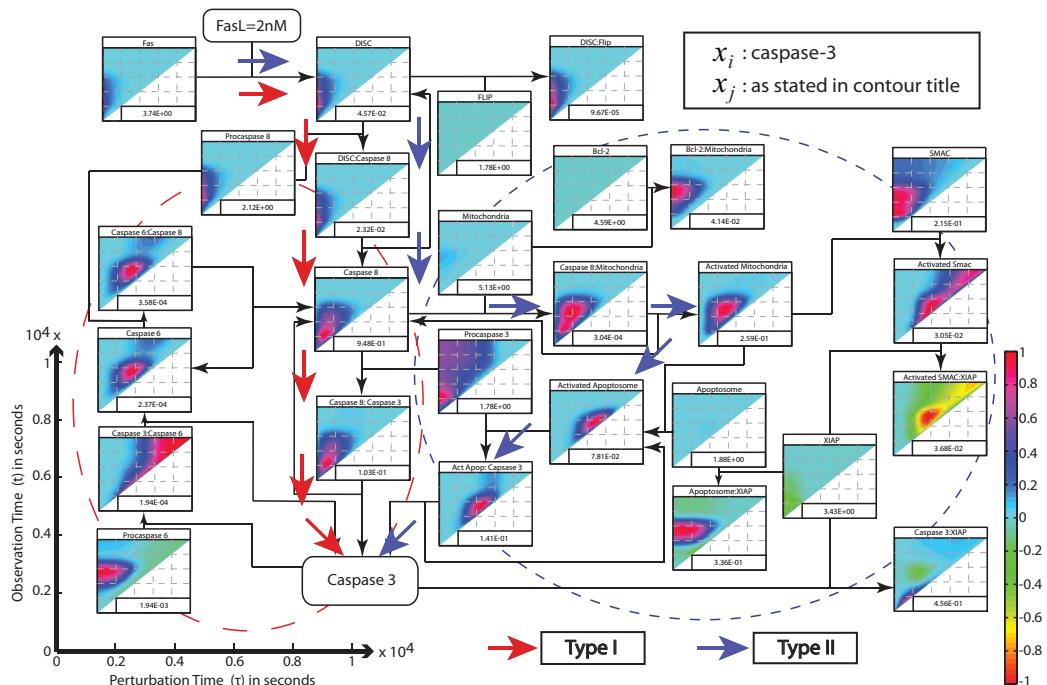


Figure 3.7. Green’s Function Matrix (GFM) analysis of caspase-3 activation. The heat maps represent the GFM coefficients of caspase-3 active level with respect to the perturbations on different molecules in the network $\partial[\text{caspase-3}]/\partial x_j$. For comparison, each heat map is scaled to have values bounded by ± 1 by the scaling factor (absolute maximums) reported in the lower right corner of the plot. The red arrows illustrate the type-I regulation of caspase-3 activation, while the blue arrows show the type-II mitochondrial-pathway. As expected, the analysis showed that the upstream molecules (Fas, DISC, pro- and caspase-8) constituted the early responders to FasL stimulus, as seen in the localization of high sensitivities around low τ values. During the caspase-3 switch (5000 – 7000 seconds), type-II molecules (activated mitochondria, apoptosome, apoptosome:pro-caspase-3 complex) were directly implicated to be the key regulators by the (diagonal) location and magnitude of the peak sensitivities. On the other hand, the type-I pathway was mostly responsible for the early activation of caspase-3 (~2000 seconds), which was again illustrated by the location of the peak sensitivities. There were only two inhibitors of apoptosis: XIAP and Bcl-2 (not taking into account molecular complexes).

permeabilization; and finally (c) this led to the formation of apoptosome and the eventual cleavage of pro-caspase-3 to active caspase-3. In addition, the analysis suggested that type-I pathway had a more dominant role only in the early caspase-3 activation ($0 \leq t \leq 3000$ seconds) based on the t -axis location of the peak sensitivities with respect to caspase-8 and caspase-8:pro-caspase-3 complex. This conclusion from the GFM analysis is further supported by experimental evidence showing that knock-outs of the mitochondrial pathway abolished FasL-induced cell death in Jurkat T cell line [142].

The most prominent inhibitor of apoptosis in the model was Bcl-2 and XIAP, as indicated by the large negative sensitivities. Bcl-2 action however was limited within the first second of the FasL stimulus (see Figure B.4). Indeed, both Bcl-2 and XIAP are well known inhibitors of apoptosis [37]. In addition, the analysis also suggested a few molecules that have dual roles: pro- and anti-apoptotic. For example, procaspase-6 appeared to have a pro-apoptotic nature due to its role in a positive feedback loop with caspase-8, but at the same time exhibited anti-apoptotic behaviour due to its binding with cleaved caspase-3. The contribution from the positive feedback loop of caspase-6, -8, and -3 however was relatively insignificant in comparison to the mitochondrial pathway, as the (normalized) sensitivities were two orders of magnitude lower than those of type-II. Dual roles of other molecules in the model such as the caspase-3:XIAP and apoptosome:XIAP complexes were due to their subsequent dissociation into caspase-3 or other pro-apoptotic molecules, and hence were of little interest. In general, the sensitivities with respect to molecular complexes carry little physical significance as these molecules are either reaction intermediates or sometimes mathematical abstractions.

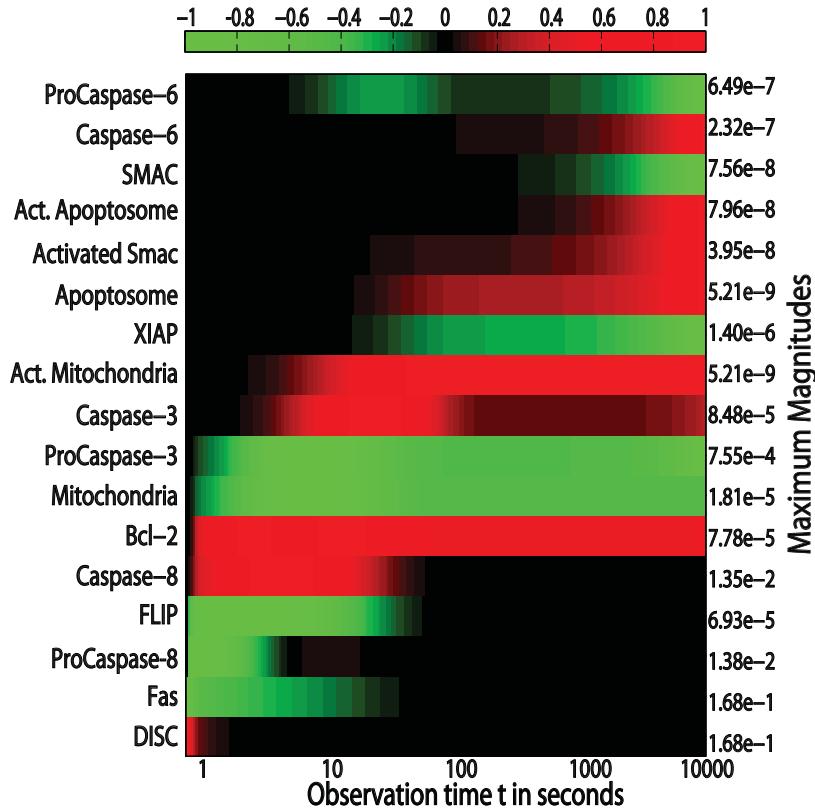


Figure 3.8. Signal progression of FasL impulse through the programmed cell death network. The signal progression is quantified by the time derivatives of the GFM coefficients under zero FasL background. The values are scaled to be between ± 1 and the scaling factors (respective maximum magnitudes) are shown on the right side of the figure. A red signal signifies a positive effect analogous to an (impulse) up-regulation and a green signal correlates to a negative effect analogous to a (impulse) down-regulation. The GFM coefficients were sorted according to the timing of the FasL signal by way of the cosine clustering method (see Section B.3).

As a validation of the analysis, *in silico* perturbation experiments were performed by increasing pro-caspase-8 concentration at two different times: $\tau = 0$ and $\tau = 4000$ seconds. In comparison to the nominal trajectory, Figure 3.4(b) shows a significant increase in the caspase-3 activation when the initial pro-caspase-8 level was increased by 50%. The same percentage increase at $\tau = 4000$ seconds however rendered little observable change. This finding is in agreement with the high and low sensitivities of caspase-3 to pro-caspase-8 perturbations at $\tau = 0$ and $\tau = 4000$ seconds, respectively (see Figure 3.7). Hence, even though GFM is a local analysis (i.e., the perturbations considered are infinitesimal), the results can predict the system response to finite perturbations.

3.3.2.3 Progression of FasL death signal

Figure 3.8 illustrates the dynamical progression of an impulse signal of FasL through different molecules in the cell death network under zero FasL background (i.e., FasL=0). This figure elucidates the timing, location, and magnitude of FasL signal as it travelled through the nodes (molecules) in the network. The initial carriers of FasL signal were Fas receptors and DISC (death inducing signalling complex) as expected, with DISC having a shorter duration (< 1s) due to the activity of FLIP. The next few nodes in the signal progression included pro-caspase-8, FLIP, and caspase-8 until approximately 50 seconds. Both type-I (direct caspase-8) and type-II pathway (mitochondria-dependent) were subsequently involved, but in different manners. Type-I pathway carried the first signal to caspase-3 (through caspase-8:caspase-3 complex), as indicated by the larger magnitude of signal progression than its type-II counterpart (e.g., caspase-8:mitochondria) (see Figure B.5). The early response of type-I pathway is consistent with the above analysis under constant FasL stimulation. However, this signal lasted only as long as caspase-8 signal persisted.

On the other hand, mitochondrial pathway underwent much prolonged signal duration, bringing about a second, delayed signal to caspase-3 beyond 4000 seconds. As seen in Figure 3.8, the signal traversed through the type-II route by way of mitochondrial activation (membrane permeabilization and release of pro-apoptotic species), apoptosome formation, and finally cleavage of pro-caspase-3 to caspase-3. Although the magnitude of signal progression through these nodes were lower than those of type-I (see Figure 3.8 right ordinate), the signal lasted for much longer to give an overall larger change in the actual concentrations. In addition, the signal progression through pro-caspase-6 had a similar pattern to the active caspase-3 as the two molecules form a complex. However, the positive feedback of caspase-6 did not

play a major role in carrying FasL signal as suggested by the low magnitudes of the signal progression measure. Again, this is consistent with the above analysis for a constant FasL stimulation.

3.4 Discussion

Evolution has given rise to large and complex cellular networks, making their understanding non-intuitive. As demonstrated in the application to small network motifs and the Jurkat cell death model, the GFM analysis can reveal how a given cellular output behaviour is regulated and how a network responds to an impulse perturbation on its states or molecules. The power of this analysis comes from the dynamical sensitivities with respect to state perturbations introduced at multiple time point's τ 's. Based on these dynamical sensitivities, one can obtain a step-by-step illustration of the events happening during a functional regulation or signal propagation. In contrast, the classical parametric sensitivities lack this dynamical information due to an integrated response to a step perturbation at initial time. In general, as more realistic and hence more complex network models emerge from systems biology efforts, the GFM analysis will complement other model analyses in explaining how network functionality and dynamics are accomplished.

3.4.1 Experimental and biological relevance

Sensitivity coefficients have become a common measure of robustness in systems biology, where low sensitivities indicate a robust behaviour, and high sensitivities point to a fragile output with respect to the specific perturbations under consideration [12, 13, 16, 22, 25, 26, 75, 96, 100, 143, 144]. As seen in Chapter 1, robustness has been argued to be an advantageous feature in life (cells, organisms, and population)

[145]. However, extreme robustness, by nature or engineering design, to common perturbations is known to pair with high fragility (the opposite of robustness) to rare events [146]. This robust-yet-fragile characteristic manifests in the large discrepancy between an overall insensitivity of cellular functions to most perturbations and the extreme sensitivity to few parameters [16].

Based on the relationship between robustness and sensitivity coefficients, the GFM analysis reveals the network fragility points in the form of key biological molecules, where such fragility points (as seen in Section 3.3.2.2) dynamically move from one molecule to the next during cellular regulation. Specifically, the GFM gives information about the molecules that are critical in regulating the system behaviour and portrays two dynamical aspects of the cause-effect relationship: the times at which the perturbations become important and at which the system output is significantly affected. One can obtain two biological insights from the GFM of a given network model. First, by choosing a particular output behaviour x_i and observing its sensitivities to perturbations on other molecules x_j 's in the network (a row of $S^x(t, \tau)$), large magnitudes of GFM coefficients will reveal the important molecules in the network that actively participate in bringing about the observed output behaviour (e.g. see Figure 3.2). In addition, the peak sensitivities to these key players will appear in sequence along the time axis τ (e.g., see Figure 3.7), which further illustrates the timing of how a particular output is regulated by these molecules. Such information can be useful in various applications from network validation and reduction to drug discovery.

Another biological insight comes from perturbing a chosen molecule x_j and observing its effect on the concentrations of other molecules x_i 's in the network (a column of $S^x(t, \tau)$), by defining an impulse perturbation as a signal, the selected

GFM coefficients will show how a signal initiated from a molecular species propagates through the network (e.g. see Figure 3.8). Such information should be useful in the understanding of drug efficacy and specificity.

3.4.2 Comparison with the classical sensitivity analysis

As noted above, the GFM analysis differs from the classical parametric sensitivity in that the perturbations are realized on the system states instead of on the model parameters. In fact, one can view the GFM analysis as a generalization of the classical parametric sensitivities with respect to initial concentrations in which the perturbations are introduced at different initial times. To aid the comparison between the two analyses, the ODE model in (1.2) is rewritten to include the model parameters in a generalized state vector \mathbf{z} , as follows:

$$\begin{aligned}\mathbf{z} &= \begin{bmatrix} \mathbf{x} \\ \mathbf{p} \end{bmatrix} \\ \frac{d\mathbf{z}}{dt} &= \begin{bmatrix} \frac{d\mathbf{x}}{dt} \\ \frac{d\mathbf{p}}{dt} \end{bmatrix} = \begin{bmatrix} \mathbf{g}(\mathbf{z}) \\ 0 \end{bmatrix}\end{aligned}\tag{3.7}$$

In this context, the parametric sensitivities is a subset of the GFM of \mathbf{z} , such that

$$\mathbf{S}(t) = \mathbf{S}_{[1\dots n; n+1\dots n+p]}^{\mathbf{z}}(t, t_0)\tag{3.8}$$

where $\mathbf{S}_{[1\dots n; n+1\dots n+p]}^{\mathbf{z}}$ is a submatrix of the GFM by including elements from rows 1, 2, ..., n and columns $n+1, n+2, \dots, n+p$. Since the parameters have zero dynamics, the impulse perturbation effectively results in a step change in the parameter value. Therefore, the parametric sensitivities reveal the integrated change in the system output due to a persistent perturbation in the parameters. Due to this integrated nature, the parametric sensitivities may not be suitable when used directly to infer the

dynamical importance of pathways, as shown in the application to the cell death model.

At a first glance, there are also similarities in the mathematical appearance of the GFM and the control coefficients in Metabolic Control Analysis (MCA), an analysis tool in metabolic engineering that reveals the distribution of control of metabolite concentration and fluxes over enzymes. Mathematically, the control coefficients in MCA are equivalent to parametric sensitivity analysis applied to kinetic metabolic models, in which the parameters represent enzyme activity or level. As in (3.7), a generalized state vector can be defined to include metabolite levels and enzyme activities, in which the enzyme activities have zero dynamics. Again, the concentration control coefficients in MCA can be obtained as submatrix of the GFM for which $\tau = t_0$. The GFM impulse perturbations also integrate to give a step change in the enzyme activities. Thus, the same fundamental differences exist between the GFM and MCA as between the GFM and the classical PSA.

Despite these similarities, the summation theorem does not generalize to the full GFM, i.e. the row sum $\sum_j S_{i,j}^x(t, \tau)$ does not necessarily equal to 0, even at steady state. The reason stems from the invalidity of the assumption taken in the summation theorem, where the change in the steady state concentrations due to a perturbation in concentration at the initial or intermediate time equals to zero (i.e. $(\Delta x_i)_{ss} = 0$) [147]. For example, consider a closed system with two molecules A and B , where A produces B through a reversible enzymatic transformation ($A \rightleftharpoons B$). In this system, appropriately chosen perturbations of the enzyme activities for the forward and reverse reactions can be introduced that do not cause any changes in the concentrations of A and B , giving the summation theorem in MCA. However, a

perturbation on the level of either A or B will give different equilibrium concentrations due to mass conservation in a closed system. Hence, considering the generalized state with enzyme activities as parameters, the summation and connectivity theorem is valid only on a portion of the GFM (e.g. $S_{[1\dots n; n+1\dots n+p]}^x$), but not necessarily for the complete GFM.

3.4.3 Applications of GFM analysis

The GFM analysis can be useful in many applications. For example, the agreement and disagreement between model analysis and experimental observations and/or biological knowledge can be used for model validation and the design of experiments. Here, the agreement between the analysis of the programmed cell death model and experimental evidence of the type-II regulation [142], gives support to this model validity, which was not given in the original publication [98] but confirmed by a subsequent analysis using a different method [148].

When a disagreement arises, the analysis gives experimentally testable predictions about the effect of a perturbation on the key molecular player(s). In the cell death example, the GFM suggested that the signaling follows a mitochondrial-dependent pathway. Indeed, many cancer cells up-regulate their anti-apoptotic proteins within the type-II pathway to evade apoptosis [149]. Hence, an interesting hypothesis is to reroute the cell death signaling in type-II cells through the type-I pathway. The GFM analysis of Jurkat model highlighted two molecules: procaspase-8 and Bcl-2 that have an early positive effect on type-I and an early negative effect on type II, respectively. This information suggests that by upregulating the available procaspase-8 and Bcl-2, one can increase the throughput of type-I and at the same time attenuate the contribution of type-II pathway. The GFM analysis of this mutant

network *in silico* confirmed that the cell-death occurs through type-I pathway (see Figure B.6 and B.7 for detail). The corresponding experiments using Jurkat cells may involve a viral transfection to overexpress procaspase-8 [150, 151] and knocking out Bcl-2 family proteins, Bax and Bad [142].

The GFM results can also be used to reduce model complexity. For example, based on the dominance of type-II pathway as revealed by the GFM analysis, the cell death model can be reduced by knocking out the type-I pathway and the caspase-6 feedback loop, as the caspase-3 switch was less sensitive to perturbations on these molecules. Figure 3.9 indicate that the reduced system still commits to cell death even in the absence of these pathways, while preserving the switch-like response. Since the contribution of the type-I pathway is non-zero, especially in the earlier times, the caspase-3 activation is slightly delayed. On the other hand, the removal of the type-II pathway significantly slows down the apoptosis and abolishes the switch characteristic, as shown in Figure 3.9.

As mentioned above, the results of the GFM analysis can also guide the drug discovery research to identify potential molecules for drug targets or drug cocktail.

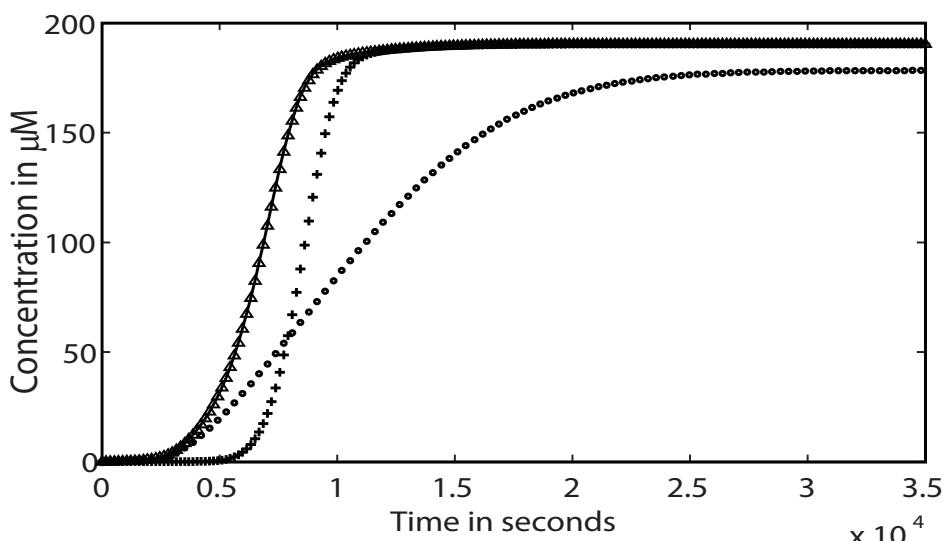


Figure 3.9. Model reduction of FasL-induced apoptosis pathway. The active caspase-3 concentrations are shown for the nominal (original) (—) and reduced models without type-I (+), type-II (o) or caspase-6 feedback pathway (Δ).

Multicomponent therapy using combination of drugs has increasingly gained more attention with various reported successes, such as salmeterol-fluticasone in asthma (Advair - GlaxoSmithKline) [152] and AZT-3TC in HIV infection (Combivir - GlaxoSmithKline) [153]. The selection of drug combinations in a clinical setting however had been done through deliberate mixing by rational design or happenstance. Therefore, a systematic method in multicomponent therapeutics is highly desirable, either by a large scale screening [154] or by *in silico* network analysis as recently outlined by Kitano [9]. The GFM analysis can help realize such systems-oriented drug discovery.

In the analysis of the cell death model, the stimulus is initiated from the Fas receptor, which belongs to the TNF (tumor necrosis factor) receptor family. In practice, several cancer therapeutic agents target TNF receptors to trigger programmed cell death, such as Apo2L/TRAIL from Genentech and Amgen [155]. In general, a large sensitivity magnitude in Figure 3.7 implies a synergistic action between a Fas receptor drug and the perturbation in the corresponding molecule. Specifically, this analysis suggests that the effectiveness of such drugs in Jurkat cells should depend on the type-II pathway. Therefore, the use of other drugs that increase the activity of type-II pathway should also boost the sensitivity of these cancer cells to the apoptotic insult by a Fas-receptor targeting agent. In addition, the GFM analysis also gives suggestion to the most effective timing (dynamics) of type-II targeting drugs, which should be delivered ~4000 seconds after the TNF drug. A more detailed model of the cell death network than the one used here, will be needed to provide more concrete suggestions of the target molecules.

In addition, as many new drugs are designed to target specific molecules in the network, a drug action can therefore be represented as a molecular perturbation. To

this end, the GFM analysis allows the visualization of the signal progression from a dose of such drug, as demonstrated in Figure 3.8. Hence, the drug efficacy and specificity can be studied by looking at the propagation of the drug “signal” through a disease network model. In particular, the overall drug effects depend not only on the magnitude but also on the duration of the signal. In the analysis of Jurkat cell death model, a signal initiated by Fas ligand or other related drugs will indeed reach the intended caspase-3 activation with the type-I pathway transmitting the early signal and the type-II being the dominant albeit slower signal carrier. Such information is not necessarily obvious from the static diagram or from the simulation of the concentrations.

3.4.4 Advantages and limitations of GFM analysis

As described in Section 3.2, the calculation of the GFM coefficients for different τ 's is computationally efficient by taking advantage of the semigroup property. Therefore, one only needs to solve for the GFM coefficients for a single time step $\Delta\tau$ starting from each τ_k , while the dynamic GFM coefficients for different (t, τ) pairs are calculated using inexpensive matrix multiplications. For example, the application of dynamic sensitivities to the Fas-induced apoptosis model with 28 species and 32 reversible reactions took less than 10 minutes, using a computer workstation with a dual-core CPU (Intel 6300 @ 1.86GHz) and 3GB RAM.

Like the standard parametric sensitivity analysis, the aforementioned results from the GFM also depend on the choice of the model parameters as well as the initial conditions. That is, the GFM analysis gives only local information about the cellular network based on infinitesimal perturbations. To account for cell to cell variability and other sources of uncertainty, a similar global analysis should consider a region in

the parameter or initial state space and improves upon the local analysis that focuses only at one choice of parameter set or initial states. The extension of the GFM analysis to global (finite) perturbations will be shown in Chapter 6.

3.5 Summary

In summary, biological complexity often necessitates the use of systems-oriented approach to establish the cause-effect relationship between a cellular network model and its experimentally observed property or function. Based on a kinetic model and impulse perturbations on the molecular concentrations, the GFM analysis offers dynamical insights on the functional regulation and signal propagation in the cellular network, for which the classical parametric sensitivity analysis may fail. In the application to the programmed cell death model, the GFM analysis gave a molecule-by-molecule account of how cell death regulation is carried out through the type-II pathway, in contrast to the type-I pathway dominance concluded from the parametric sensitivity analysis. Model reduction by deleting either type-I or type-II pathway confirmed the importance of type-II over type-I. Based on the results, an experimentally testable hypothesis was offered that reroutes the cell death regulation in type-II cells through the type-I pathway by the over expression of Bcl-2 and procaspase-8. Finally, the analysis can guide the drug discovery efforts in the identification of potential drug targets, the understanding of drug efficacy and specificity, and the optimization of drug dosing and timing. The development of this method represents a concrete step towards robustness-based drug design through systems biology.

CHAPTER 4

4 Impulse Parametric Sensitivity Analysis[§]

Synopsis:

To avoid the caveat of classical parametric sensitivity analysis (PSA) in analyzing system dynamics, a novel Green's function matrix (GFM) analysis that maps out the dynamic concentration dependence of system behaviour was developed in Chapter 3. In many systems biology applications, the dynamic parametric dependence of system behavior is often desired. To this end, this chapter details a novel sensitivity analysis, based on impulse perturbations to model parameters, called impulse parametric sensitivity analyses (iPSA). The iPSA gives a step-by-step mechanistic insight on how a particular output behavior is accomplished and complementarily how (an impulse) perturbation on the parameters propagates through the network. More specifically iPSA coefficients can reveal which parameters are critical and when they become important. The efficacy of this analysis is again demonstrated on the ODE model of FasL-induced cell death of human Jurkat T-cells.

[§] Excerpts of this chapter are part of the following publications:

Perumal TM and Gunawan R. (2011) Understanding dynamics using sensitivity analysis: caveat and solution. *BMC Syst. Biol.*, 5(1): 41. PMID: 214060955

Perumal TM and Gunawan R. (2011) Impulse Parametric Sensitivity Analysis, *In Proceedings of the 18th World Congress - The International Federation of Automatic Control (IFAC)*, Milan, Italy, PP: 9686-9690

4.1 Introduction

As seen from the previous chapters that complex systems are typically characterized by a large number of interacting chemical or biochemical species, often giving rise to non-intuitive behaviour. Also, mathematical models of these systems have been built, on which quantitative analyses can be applied to gain mechanistic understanding and subsequently to guide process optimization, design and control. Importantly, in systems modeling in biology, the classical PSA has become a powerful tool in investigating how system behaviour depends on model parameters [105]. But, as shown in Chapter 2, the dynamical aspect of cellular regulation is not immediately apparent from the classical PSA metrics due to the persistent nature of perturbations. More importantly they can even mislead the modelers in understanding system dynamics and controlling mechanisms. In order to gain dynamical insights, a method based on perturbations on concentrations at different times, called as the Green's function matrix (GFM) sensitivity analyses was presented in Chapter 3.

Nevertheless, it is also often desired to obtain the parametric dependence on system output behavior. This is because (kinetic) parameters are directly related to the processes of the system. Hence, complementary to the GFM analysis, this chapter contains detailed development of an impulse parametric sensitivity analysis (iPSA). The fundamental difference between the GFM and iPSA is in the type of perturbations introduced. The iPSA makes time-varying impulse perturbations to model parameters, instead of concentrations (model states). From the time-varying impulse perturbations, the iPSA can also give answers to the following questions about state dynamics: which are the important parameters and when do they become important? $\Delta p_j, \Delta \tau \rightarrow 0$. While both iPSA and GFM analyses can give a step-by-step

mechanistic insight on how particular system behaviour is accomplished and how (an impulse) perturbation propagates through the network, the choice of using one over the other depends on the purpose of the analysis. The efficacy of the iPSA analysis is also demonstrated through the same example as presented in Section 3.3.2: a FasL-induced programmed cell death model of Jurkat T-cells [98].

4.2 Impulse Parametric Sensitivity Analysis (iPSA)

As mentioned earlier, the iPSA introduces impulse perturbations to model parameters and quantifies the resulting change in system states, as illustrated in Figure 4.1(a-b).

In this case, the corresponding impulse sensitivity coefficient given by

$$iS_{i,j}(t, \tau) = \frac{\text{change in the } i\text{-th state at time } t}{\text{impulse perturbation on the } j\text{-th parameter at time } \tau} \quad (4.1)$$

reflects the change in the i^{th} state at time t due to an impulse perturbation on the j^{th} parameter at time τ . Since impulse perturbations on system parameters can cause an

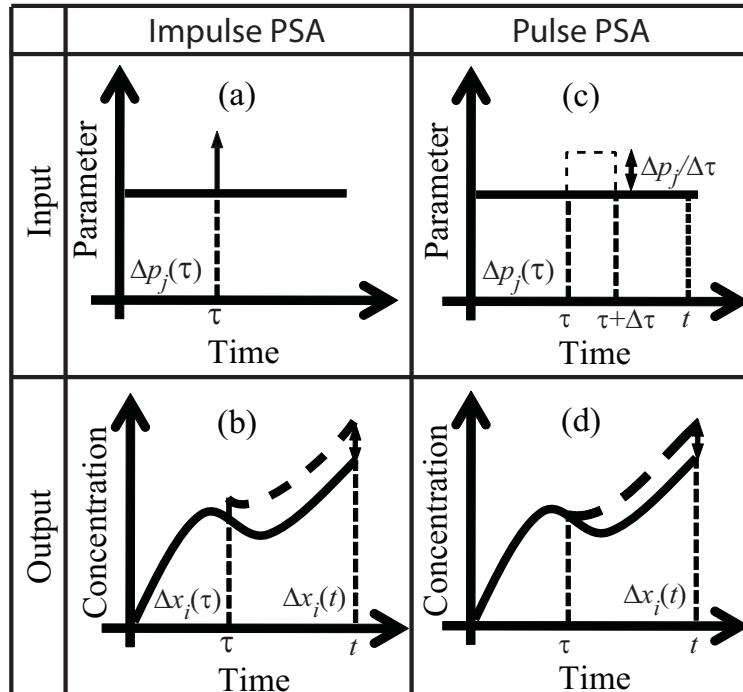


Figure 4.1. Impulse Parametric Sensitivity Analysis (iPSA). (a-b) Impulse perturbations on parameters cause an immediate and localized effect on the states. (c-d) Derivation of iPSA uses a pulse perturbation to approximate the impulse, in the limit $\Delta p_j, \Delta \tau \rightarrow 0$.

immediate and localized effect on system states at the time of perturbation, the dynamical inference of parametric importance can be obtained from these time varying impulse perturbations. Hence, the iPSA can give not only the important parameters, but also the time of importance on the state dynamics.

The derivation of the iPSA coefficients follows the illustration in Figure 4.1(c-d). Analogous to the PSA, the sensitivity coefficients of the iPSA are derived by quantifying the ratio between the change in the state x_i at time t and the causative pulse perturbation of size $\Delta p_j / \Delta \tau$ at time τ . The perturbation is exerted on the parameter p_j for duration of $\Delta \tau$. The derivation starts by quantifying the change in the states \mathbf{x} at time $\tau + \Delta \tau$ using a Taylor series expansion:

$$\Delta \mathbf{x}(\tau + \Delta \tau) = S_{[1..n],j}(\tau + \Delta \tau, \tau) \frac{\Delta p_j}{\Delta \tau} + O(\Delta p_j^2) \quad (4.2)$$

where $S_{[1..n],j}(\tau + \Delta \tau, \tau)$ is the vector of parametric sensitivity coefficients as given in (2.2). Subsequently, this change $\Delta \mathbf{x}(\tau + \Delta \tau)$ is propagated to the change in the state x_i at any time t using the GFM $\mathbf{S}^x(t, \tau + \Delta \tau)$ as in (3.1). Thus, the change $\Delta x_i(t)$ due to the pulse perturbation is given by

$$\Delta x_i(t) = S_{i,[1..n]}^x(t, \tau + \Delta \tau) \Delta \mathbf{x}(\tau + \Delta \tau). \quad (4.3)$$

Substitution of (4.2) in (4.3) gives

$$\Delta x_i(t) = S_{i,[1..n]}^x(t, \tau + \Delta \tau) S_{[1..n],j}(\tau + \Delta \tau, \tau) \frac{\Delta p_j}{\Delta \tau} + O(\Delta p_j^2). \quad (4.4)$$

Subsequently, by taking Taylor series expansion of the parametric sensitivities around the time of perturbation τ and dividing both sides by Δp_j , one will arrive with:

$$\begin{aligned}
\Delta x_i(t) &= S_{i,[1..n]}^x(t, \tau + \Delta\tau) \left[S_{[1..n],j}(\tau, \tau) + \dot{S}_{[1..n],j}(\tau, \tau) \Delta\tau + O(\Delta\tau^2) \right] \frac{\Delta p_j}{\Delta\tau} + O(\Delta p_j^2) \\
\frac{\Delta x_i(t)}{\Delta p_j} &= S_{i,[1..n]}^x(t, \tau + \Delta\tau) \left[\dot{S}_{[1..n],j}(\tau, \tau) + O(\Delta\tau) \right] + O(\Delta p_j) \\
\frac{\Delta x_i(t)}{\Delta p_j} &= S_{i,[1..n]}^x(t, \tau + \Delta\tau) \left[\frac{\partial g}{\partial p_j}(\tau) + O(\Delta\tau) \right] + O(\Delta p_j)
\end{aligned} \tag{4.5}$$

Finally, taking the limit as $\Delta p_j, \Delta\tau \rightarrow 0$ such that the pulse perturbation becomes an impulse, the iPSA coefficient is obtained as:

$$iS_{i,j}(t, \tau) = S_{i,[1..n]}^x(t, \tau) \frac{\partial g}{\partial p_j}(\tau) \tag{4.6}$$

By rewriting (4.6) as:

$$iS_{i,j}(t, \tau) = S_{i,[1..n]}^x(t, \tau) iS_{[1..n],j}(\tau, \tau) \tag{4.7}$$

one can further see that the impact of this impulse perturbation takes effect only at the perturbation time τ and that the consequence on the state trajectory is equivalent to perturbing the states themselves, similar to the GFM analysis presented in Chapter 3. Therefore, the calculation of the iPSA coefficients in (4.1) simply requires a matrix multiplication of the GFM coefficients in (3.1) with the non-homogeneous term $\partial g / \partial p_j$ at the initial time τ . Like in the PSA and GFM, the iPSA coefficients should also be normalized for comparison and parameter ranking purposes, according to:

$$i\bar{S}_{i,j}(t, \tau) = iS_{i,j}(t, \tau) \frac{p_j}{x_i(t)} \tag{4.8}$$

Since the iPSA involves two times, i.e., the time of perturbation τ and the time of observation t , like GFM coefficients in (3.2), each of $(i,j)^{\text{th}}$ element of the iPSA coefficients in (4.8) can also be visualized as a heat map. Similarly another dynamical information that can be extracted from the iPSA relates to how the effect of an impulse perturbation on one of the system parameter propagates in the network.

4.3 Case Studies

4.3.1 Simple network model

In the first case study, the efficacy of iPSA was studied using the simple network model presented in Section 2.2.1, under the same stimulus and for the same time range. Figure 4.2 shows the iPSA sensitivity coefficients of x_6 at the end of simulation time ($t = 15$ time units), with respect to the four most important parameters at different perturbation times (for the complete iPSA sensitivities, see Figure C.1). In agreement with the KO simulations in Section 2.2.1 and in contrast to the findings from the classical PSA in Section 2.2.2, the impulse sensitivities gave support to the dominance of the indirect pathway. Specifically, the results showed that x_6 activation:

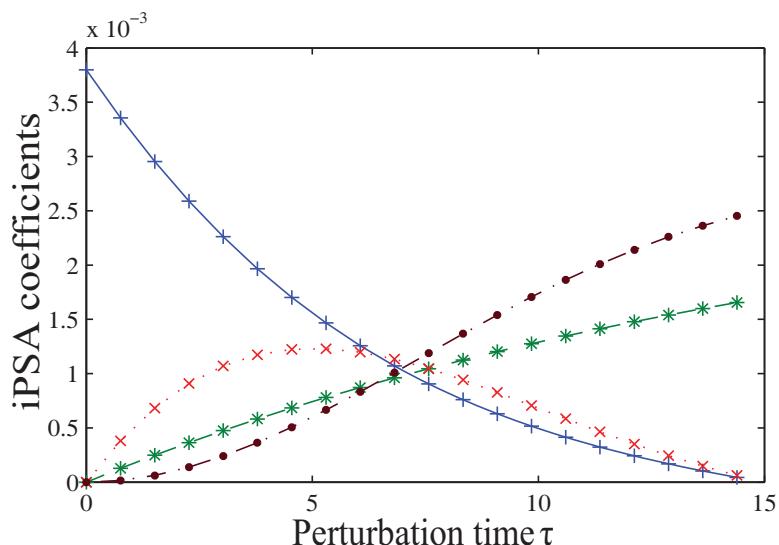


Figure 4.2 Impulse parametric sensitivity analysis of x_6 activation under x_1 stimulus. iPSA coefficients of x_6 with respect to impulse parametric perturbations in k_{f1} (+), k_{v2} (*), k_{v3} (x) and k_{v4}

(1) is initiated by r_1 (high initial sensitivity to parameter k_{f1}); (2) is accomplished mainly by the direct pathway prior to the switching time (sensitivity to r_2 is higher than to r_4 during these times); and (3) is subsequently carried by the indirect pathway (highest sensitivity to r_4 during switching times). A higher resolution analysis using heat maps of the complete iPSA coefficients with t and τ between 0 and 15 time units gave the same conclusion (see Figure C.2).

4.3.2 FasL-induced cell death of human Jurkat T-cell lines

In the second case study, a more complex biological network of FasL-induced programmed cell death in human Jurkat T-cell lines [98], as presented in Section 3.3.2, is considered. The iPSA was performed to assess the dominance of one pathway over the other, as well as to compare the dynamics of caspase-3 activation by both the pathways.

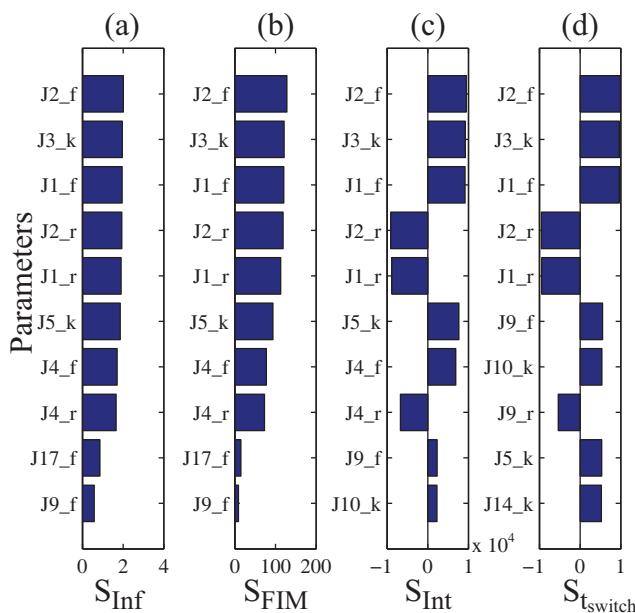


Figure 4.3. Local parametric sensitivity analysis of the programmed cell death model. (a-c) The bar graphs represent the ten largest sensitivity metrics of caspase-3 based on infinite norm, Fisher Information Matrix (FIM), time integrated sensitivity coefficients, and sensitivity magnitudes at the switching time ($t = 6060$ s). The complete parameter ranking is given in Figure C.3. The parameter numbers refer to the reactions shown in Figure 3.5, where the subscripts f and r denote forward and backward rate constants for reversible reactions and k denotes the rate constants for irreversible reactions.

The classical PSA and the iPSA analyses were calculated under a constant FasL stimulation (FasL = 2nM) over the time range of 10,000 seconds. The rankings

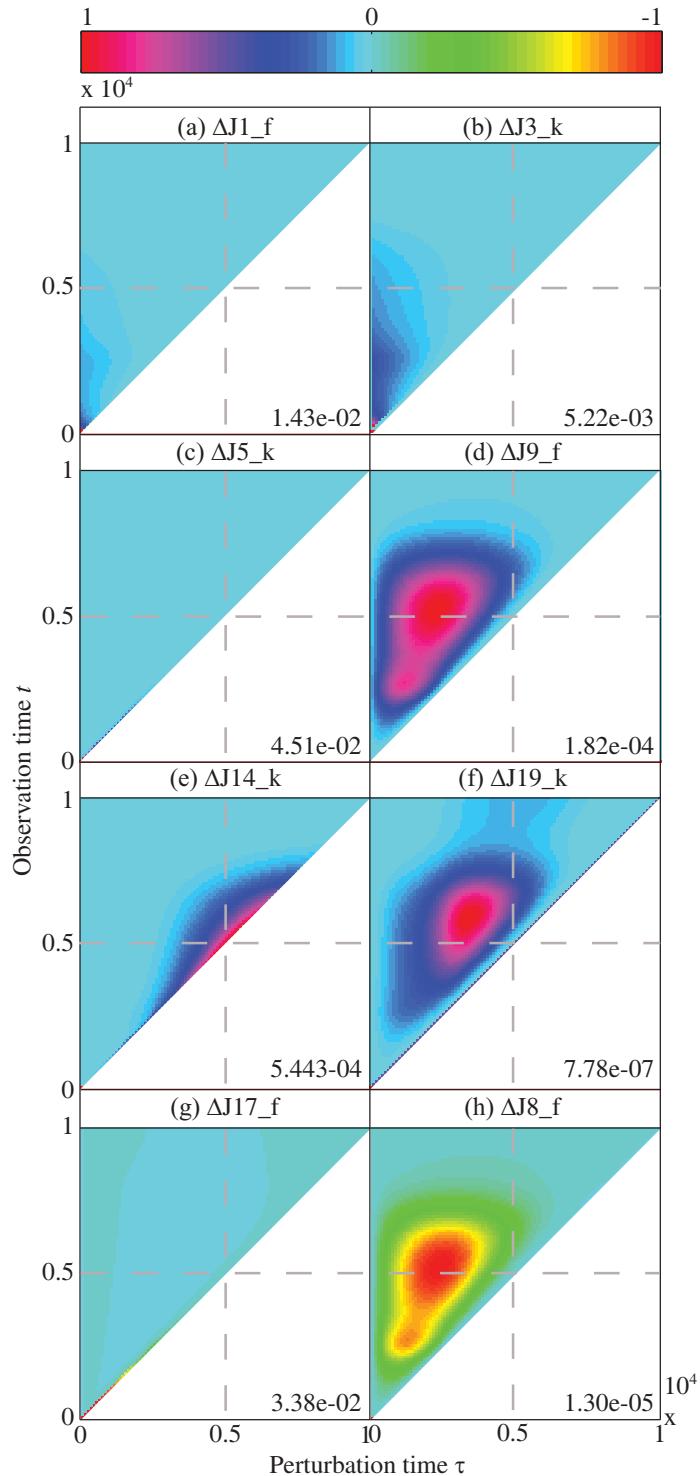


Figure 4.4. The iPSA of caspase-3 activation. (a-h) The heat maps illustrate the iPSA coefficient of active caspase-3 level with respect to perturbations on the specified parameter in the network, indicated in the title. The parameter numbers refer to the reactions shown in Figure 3.5, where the subscripts *f* and *r* denote forward and backward rate constants for reversible reactions and *k* denotes the rate constants for irreversible reactions. The *x*-axis is the time at which impulse perturbation is applied, while the *y*-axis indicates the observation time of caspase-3. Each plot is scaled to have values between ± 1 by the scaling factor in the bottom right corner.

of the important parameters that control caspase-3 according to the PSA are shown in Figure 4.3, using consolidated metrics as in (2.13): infinite norm [120](Figure 4.3(a)), FIM [136] (Figure 4.3(b)), and time integral [99](Figure 4.3(c)), and using the sensitivity magnitudes at switching time (Figure 4.3(d)) (see Figure C.3 for detailed sensitivity rankings). From these rankings, one could not obtain any definitive conclusion regarding the dominance of one pathway over the other.

On the contrary, the iPSA sensitivity of caspase-3 in Figure 4.4 clearly supported a type-II dependent caspase-3 switching with an early type-I dependent activation, in agreement with two previous analyses of this model using the Green's function matrix in Section 3.3.2.2 and model reduction in Section 3.4.3 Figure 4.4(a-h) shows the iPSA coefficient heat maps of selected parameters of the system in caspase-3 activation. By focusing on the observation time between 4000 to 7000 seconds (switching duration) in the y -axis of Figure 4.4, one can identify the sequence of key parameters that control the observed caspase-3 activation. Figure 4.4(a-b) suggests that the upstream reactions r_1 and r_3 (DISC and caspase-8 formation) were among the early significant contributors before switching. Figure 4.4(c) illustrates that the initial caspase-3 activation can be mostly attributed to the type-I mitochondria-independent pathway. Figure 4.4(d-e) suggests that a type-II dominant pathway in the caspase-3 activation. Furthermore, Figure 4.4(c) also points to the lack of type-I role during switching. As seen in Figure 4.4(f) one can also deduce the insignificance of caspase-6 feedback pathway. Also, looking at the impulse sensitivities of forward rate constants of r_{17} and r_8 in Figure 4.4(g-h), one can conclude that XIAP and Bcl-2 are the strong and weak inhibitors of caspase-3 activation, respectively (see Figures C.4 and C.5 for more detail).

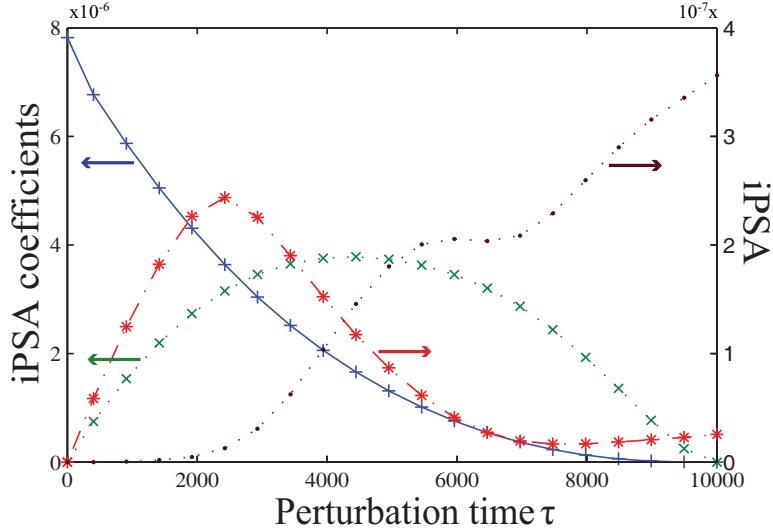


Figure 4.5. The iPSA of caspase-3 at the final output time. iPSA of caspase-3 with respect to impulse parametric perturbations in $J1_f(+)$, $J5_k(*)$ and $J14_k(\bullet)$ inferred at $t = 10000s$ (the final output time). The parameter numbers refer to the reactions shown in Figure 3.5, where the subscripts f and r denote forward and backward rate constants for reversible reactions and k denotes the rate constants for irreversible reactions.

A more detailed understanding can also be obtained by fixing the observation time (t) of interest. The iPSA coefficients of caspase-3 at the final time ($t = 10000s$) in Figure 4.5 confirms the above dynamic regulation of caspase-3 activation, suggesting a type-II dominance in the cell death regulation of human Jurkat T-lymphocyte cell lines (sensitivities to $J5_k$ is larger than that to $J14_k$ only before switching). The same conclusion on caspase-3 activation came from the GFM analysis of this model in Section 3.3.2, which was also in agreement with the previous experimental results [142].

The other dynamical information that can be extracted from the iPSA coefficients relates to how an impulse perturbation given to a rate constant propagates through the network. By setting the perturbation time to zero ($t_0=0$), Figure 4.6 elucidates the magnitude of changes in the molecular concentrations as the impulse signal initiates from the first reaction (i.e., sensitivities to $J1_f$) and travels through the network. In this case, the signal starts from the upstream reactions that are common to both the pathways, followed by type-I and caspase-6 feedback, and later traverses

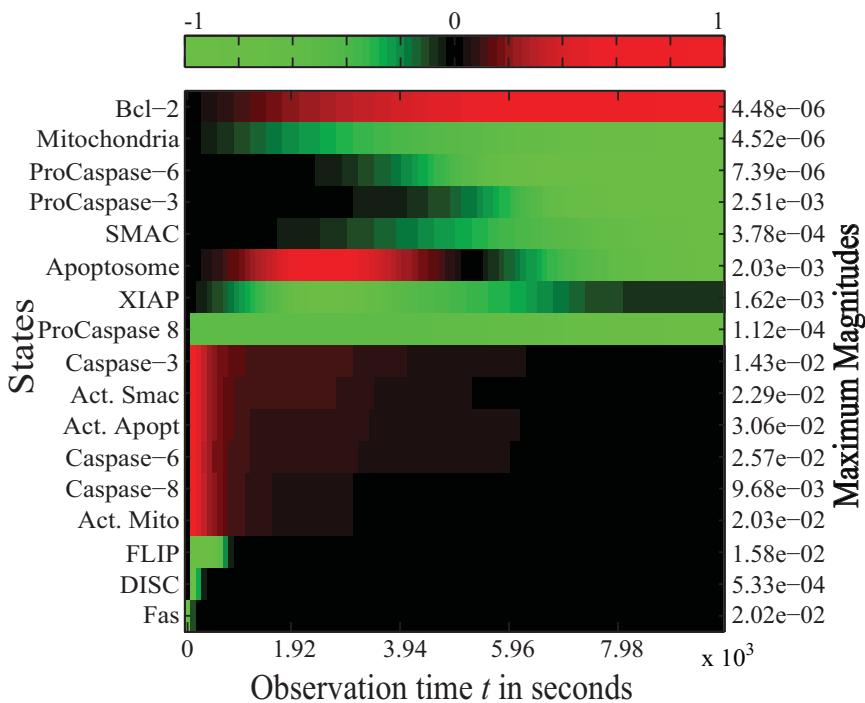


Figure 4.6. Impulse signal progression of J1_f through FasL-induced apoptosis network. Impulse signal here is defined as the sensitivity of system states with respect to the impulse perturbation on the parameter J1_f. The values are scaled between -1 and +1. The scaling factors are shown on the right side of the figure. The coefficients were sorted according to the timing of the signal by way of cosine clustering method.

through the type-II pathway. This finding is consistent with the above finding of caspase-3 activation.

4.4 Discussion

Like the PSA and GFM, the iPSA can be used for the same applications: from model identification to model validation and refinement/reduction, as well as for the purpose to understand the robustness-fragility trade offs and the underlying mechanisms that give rise to the observed system dynamics. The choice of using one over the other depends only on the purpose of the analysis. For example, since system fragility is measured as high sensitivity to parametric perturbations, the iPSA coefficients can reveal system fragilities in terms of key processes and such fragilities (as seen in Section 4.3.2) dynamically move from one process to another. Like the

GFM, one can also obtain signal propagation starting from a specific parameter or process and observe its effect on the molecular concentrations in the network.

The discrepancy between the PSA and iPSA results can again be explained in the context of persistent versus impulse perturbations. As seen in the PSA parameter rankings in Figure 4.3 and following the insights offered by the iPSA in Figure 4.4 and 4.5, the effect of perturbing early response processes, including type-I pathway, was integrated over time in the classical PSA. For example, the highest ranked parameters in the PSA were associated with the first three reactions, r_1 to r_3 , in the model. Such integration masked the dynamical importance of different parameters, while using the PSA metrics. In contrast, the conclusions from the iPSA is in agreement with the simulations of KOs of type-I and type-II pathways in Figure 3.9.

The example above illustrates the problem of using the classical PSA in identifying the controlling mechanisms of a dynamical system. Of course, this does not mean that the PSA of dynamical models is incorrect, but rather as presented in Chapter 2, the interpretation of the sensitivity coefficients should be carefully managed. In particular, a large sensitivity magnitude with respect to a parameter suggests the importance of this parameter in the time period between the perturbation time τ and the state observation time t . In contrast, the iPSA is developed with dynamics in mind, where the impact of a single perturbation on the system is realized only at the perturbation time and subsequently they were delivered at varying perturbation times. By doing so, the iPSA coefficients can elucidate the way system dynamics $x(t)$ is achieved, by indicating which and when parameters or processes are essential. Because of the persistent nature of perturbations used in the PSA, it is still not possible to reproduce the conclusions of the iPSA by varying the time of perturbations (see Figures C.6 and C.7). However, as with the local PSA, impulse

perturbations are also local in nature and thus the impulse sensitivities will depend on the nominal parameter values. The global equivalent of iPSA can be formulated using pulse perturbations and is of future interest.

4.5 Summary

Understanding systems dynamics is of great interest in many fields. To this end, this chapter presented a novel sensitivity analysis, called impulse parametric sensitivity analysis (iPSA) that offers dynamical insights on the functional regulation and signal propagation in a system. In contrast to the state perturbations in the GFM analysis, the iPSA makes use of impulse perturbations introduced at different times on the model parameters to produce the necessary information for understanding system dynamics. In the application to a FasL-induced programmed cell death model of Jurkat cells, the iPSA suggested a type-II regulation in caspase-3 activation in this cell line, which was expectedly in agreement with both the theoretical GFM analysis presented in Chapter 3 and the experimental evidence from literature [142]. But an application of the classical PSA failed in these cases. Since the discrepancy between the PSA and iPSA arises from a fundamental difference in the manner of which parametric perturbations are realized (i.e. persistent vs. impulse), the same caveat and solution can be generalized to the global sensitivity analyses, in which the parameter perturbations are also persistent in nature.

CHAPTER 5

5 Pathway Parametric Sensitivity Analysis^{**}

Synopsis:

Chapters 3 and 4 introduced two novel local dynamic sensitivity analyses for analyzing ODE models of biological systems. In addition to kinetic (dynamic) information, an ODE model also possesses information on the network structure of a biological system. Structural analyses, such as degree distribution or elementary mode analysis, have given insights about the design principles of biological networks. This chapter presents a novel pathway parametric sensitivity analysis (pathPSA), which combines structural and dynamical analysis of cellular network models. The new analysis is based on the identification of pathways in a given ODE model and the iPSA. Consequently, pathPSA can reveal the cause-effect relationship between pathways and the system dynamics. The usefulness of this analysis is demonstrated again through an application to a biological model of FasL-induced apoptotic cell death in human Jurkat T-cell lines.

^{**} Excerpts of this chapter will be part of the following manuscript:

Perumal TM and Gunawan R. pathPSA: A novel dynamic sensitivity analysis tool for analyzing biological systems, *In preparation*.

5.1 Introduction

The structure or topology of biological networks, i.e. the connectivity among molecules, has been shown to be closely related to their functions [156]. Structural analyses have also been used to understand the relationship between network structure and system functionality [157]. Structural analyses methods used in systems biology are typically extended from graph theory, as biological networks are often represented as graphs. Examples of such methods include simple centrality measures [158, 159], communicability measures [160], SigFlux method [161], simple path [162] and more recently, the elementary signaling mode (ESM) analysis [162]. Nevertheless, the application of these methods to a given network reveals how important a particular molecule or a reaction is to network connectivity and system functionality.

Cellular functions rely on coordinated activities of group of molecules and reactions [43]. In this regard, the term “pathway” is often used to define a set of molecules and reactions in the network that connect the input (stimulus) to the output (response). This chapter presents a novel sensitivity analysis that combines structural and dynamical analysis through the identification of pathways and the application of the iPSA. The new analysis, called pathway parametric sensitivity analysis (pathPSA), first obtains pathways connecting the input and output of a given network model using an extension of the Elementary Signaling Mode (ESM) analysis [162], and subsequently applies the iPSA to analyze the sensitivity with respect to perturbation of parameters associated with these pathways. In contrast to the GFM analysis and iPSA, pathPSA can provide direct information on the active pathways which participate dynamically in producing the observed cellular function. The efficacy of the present method is demonstrated again through a biological application

to understand the competing mechanisms of type-I/II apoptosis in Jurkat T-cell lines, which gives rise to the caspase-3 cleavage [98].

5.2 Pathway Parametric Sensitivity Analysis (pathPSA)

5.2.1 Mathematical models

In the pathPSA, ODE models in (1.2), are now written in more detail as:

$$\frac{d\mathbf{x}(t, \mathbf{p})}{dt} = \mathbf{g}(\mathbf{x}, \mathbf{p}) = \mathbf{Nr}(\mathbf{x}, \mathbf{p}) \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (5.1)$$

Here, the vectors $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}_0 \in \mathbb{R}^n$, and $\mathbf{p} \in \mathbb{R}^p$ denote the molecular concentrations, their initial conditions, and the model parameters, respectively as in (1.2). The matrix \mathbf{N} denotes the stoichiometric matrix and the vector function $\mathbf{r}(\mathbf{x}, \mathbf{p}) \in \mathbb{R}^r$ contains the reaction rate equations. This formulation is necessary in order to combine information pertaining to the structure (\mathbf{N}) and dynamics ($\mathbf{r}(\mathbf{x}, \mathbf{p})$).

Like the dynamic sensitivity analysis in previous chapters, the pathPSA is also built on the basis of cause-effect relationship between perturbations and system output changes. The corresponding pathPSA coefficients represent the sensitivity of the system output(s) with respect to simultaneous perturbations to all reactions belonging to each pathway. To introduce such perturbations to pathways, the ODE model in (5.1) has to be re-parametrized (see below). In this case, the calculated sensitivity coefficients can be used to portray the dynamical importance of pathways in regulating the observed output behavior. The pathPSA consists of the following three steps: (1) pathway identification: identification of all potential functional pathways, connecting the input(s) to the output(s) of the system, (2) model re-parametrization:

incorporation of surrogate parameters associated with each pathway in the ODE model in (5.1), called pathway parameters, (3) sensitivity analysis: iPSA of system outputs with respect to perturbations on pathway parameters.

5.2.2 Pathway identification

Pathway identification is done in two steps. First, all possible basic functional units that connect system input(s) to output(s) are identified. Later, associated side processes such as feedback and inhibition and degradation reactions are identified. Finally, the final set of pathways is constructed by combinations of the basic functional units and the side processes.

The explanation of the pathPSA below has been done in the context of a signal transduction network, but the same procedure can be applied to other biological systems, such as metabolic networks. The topology of a signaling network can be represented as a directed graph $G(V, E)$ in which the vertices (or nodes) V denote the signaling components (i.e., molecular concentrations \mathbf{x} in (5.1)) and the edges E denote the biological interactions or transformations (i.e., reactions $\mathbf{r}(\mathbf{x}, \mathbf{p})$ in (5.1)). Given a graph, a variety of methods are available in literature that can identify possible pathways connecting an input (e.g., ligand) and an output (e.g., cellular response) node. Examples of such methods include the identification of classical simple paths [163] and elementary signaling modes (ESMs) [162]. Since the concept of ESM can account for directionality and synergistic interactions among signaling components [162], it is adopted in the present analysis for pathway identification. An ESM is defined as the minimal set of molecules (nodes) that can perform a signal transduction from the input (stimulus) to the output (response). The minimality here signifies that an ESM is composed of non-redundant set of molecules, and in the

pathPSA, the ESMs and the connectivity matrix are used to generate basic signaling units connecting input(s) to output(s). The identification of such ESMs involves rewriting a given network graphs in an enriched representation and a depth-first search of the enriched representation [162].

The basic signaling units do not include side processes and other types of regulatory interactions such as feedback, inhibition and degradation. To identify feedbacks associated with an ESM, the same algorithm for identifying ESMs is used, but applied in the reverse direction. In this case, the ESM algorithm is used to identify reverse connectivity among all pairs of downstream-upstream molecules (except for the stimulus node) in an ESM. In addition, finding inhibition and degradation reactions can be represented as searching for synergistic nodes which are one node away from an ESM. Finally, the pathways considered in the pathPSA are formed by uniquely combining the basic signaling units with all side processes identified above.

5.2.3 Model re-parameterization

The basic idea of pathway perturbation is to scale the rate equations associated with a particular pathway simultaneously by the same relative magnitude. To do so, perturbations of pathways in the pathPSA are carried out by introducing surrogate parameters α . In this case, the rate equations $\mathbf{r}(\mathbf{x}, \mathbf{p})$ in (5.1) that are associated with the i^{th} pathway are premultiplied by a scalar parameter α_i , whose value is set to 1. In other words, the model in (5.1) is rewritten as

$$\frac{d\mathbf{x}(t, \mathbf{p})}{dt} = \mathbf{g}(\mathbf{x}, \mathbf{p}) = \mathbf{N}\mathbf{r}^*(\mathbf{x}, \mathbf{p}, \boldsymbol{\alpha}) \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (5.2)$$

where $\alpha_i = 1$, i : index of pathway

where $\alpha \in \mathbb{R}^{n_{path}}$ and n_{path} is the number of identified pathways. Note that the introduction of the parameters α does not change the model solution. Also, a single rate equation will be pre-multiplied by as many α 's as the number of pathways that involve this particular connection.

5.2.4 Dynamical analysis

The third and final step in the pathPSA is to perform dynamical impulse parametric sensitivity analysis (iPSA) with respect to the pathway parameters α in (5.2). As said earlier, the pathPSA coefficients are obtained by introducing time-varying impulse perturbations on the pathway parameters α at time τ and quantifying the resulting output change at time $t (\geq \tau)$, given by

$$iS_{i,j}^{path}(t, \tau) = \frac{\text{change in the } i\text{-th state at time } t}{\text{impulse perturbation on the } j\text{-th pathway at time } \tau} \quad (5.3)$$

Like the iPSA coefficients, the pathPSA coefficients in (5.3) can be formulated from and directly solved with the model and the GFM sensitivities for the model in (5.2) using the following:

$$iS_{i,j}^{path}(t, \tau) = S_{i,[1..n]}^x(t, \tau) \mathbf{N} \frac{\partial \mathbf{r}}{\partial \alpha_j}(\tau) \quad (5.4)$$

Again, the pathPSA coefficients in (5.4) should also be normalized for comparison between different pathways and outputs, according to

$$i\bar{S}_{i,j}^{path}(t, \tau) = iS_{i,j}^{path}(t, \tau) \frac{\alpha_j}{x_i(t)} \quad (5.5)$$

Since pathPSA also involves two times, i.e., time of perturbation τ and time of observation t , like the GFM in (3.1) and iPSA in (4.1), each $(i,j)^{\text{th}}$ element of the pathPSA coefficients in (5.5) can be visualized as a heat map, similar to that in

Figure 3.2. Again, high (low) sensitivity magnitudes indicate the important (unimportant) pathways in carrying the signal from input to output.

5.3 Case Study: Application to FasL-Induced Apoptotic Cell Death

The efficacy of pathPSA is demonstrated by its application to FasL-induced apoptotic cell death in human Jurkat T-cell lines [98], as presented in Section 3.3.2. Figure 3.5 and Section B.2 summarize the model, which consists of 28 molecular species (nodes) and 32 reactions (edges). The simulation conditions and the output of interest remained the same as that of the GFM and iPSA analysis in Sections 3.3.2.2 and 4.3.2, respectively. As seen in Figure 3.5 (inset), the output executioner caspase-3 followed a switch-like response to a constant FasL stimulus, in which the signaling is effected by two alternative signaling mechanisms: direct type-I and mitochondrial-dependent type-II. The GFM analysis in Section 3.3.2.2 and the iPSA in Section 4.3.2 revealed the importance of type-II mechanism in this activation.

First step of the pathPSA is in the identification of the basic functional units (i.e., main ESMs) and the side (regulatory) processes associated with each of these units. As expected, depth-first search algorithm resulted in two main basic signaling units: type-I and type-II, which can carry out the signal transduction from FasL stimulus to the caspase-3 activation. In addition, reverse application of the same algorithm for each basic unit gave one feedback mechanism (i.e., caspase-6 feedback activation of caspase-8) and four inhibition reactions (i.e., DISC inhibition by FLIP, caspase-3 inhibition by XIAP, mitochondria inhibition by Bcl-2 and the apoptosome inhibition by XIAP). Besides these main ESMs and side process one side or by-product reaction (i.e., Smac inhibition of XIAP), which was not identified by the

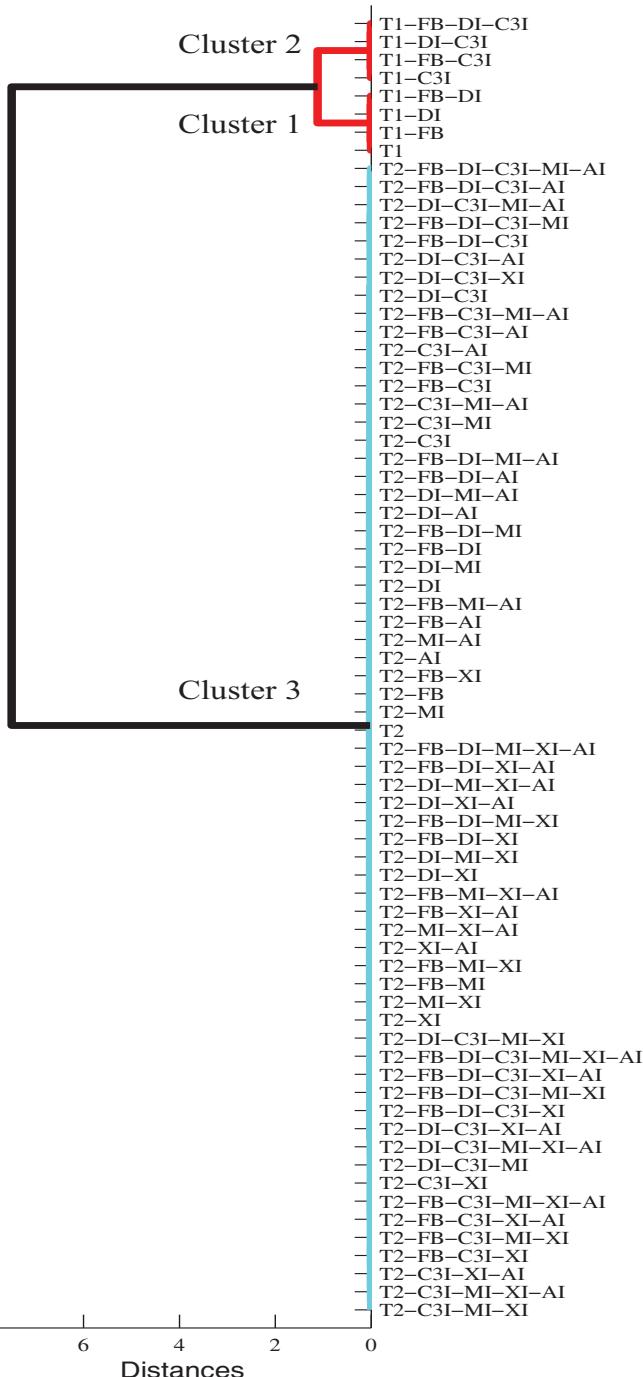


Figure 5.1. Dendrogram of pathPSA coefficients. Binary classification tree for normalized pathPSA coefficients of active caspase-3 to the perturbations on the seventy two pathway parameters specified above. Here, the symbol representations are as follows: T1: type-I, T2: type-II, FB: Caspase-6 feedback, DI: Flip inhibition of DISC, C3I: Caspase-3 inhibition by XIAP, MI: Mitochondria inhibition by Bcl-2, AI: Apoptosome inhibition by XIAP and XI: XIAP inhibition by Smac.

current algorithm, was also considered as a regulatory mechanism for the sake of completeness of the present hypothesis.

Based on the identified basic functional units and their associated regulatory mechanisms, 72 total unique pathway combinations (8 involving type-I and 64

involving type-II) can be formulated (see Table D.1 for the detailed list). The pathPSA analysis was performed again under a constant FasL stimulation (FasL = 2nM) over the time range of 10,000 seconds, like in Sections 3.3.2 and 4.3.2, and for the same purpose: to understand the underlying mechanisms of caspase-3 activation. The pathPSA coefficients of caspase-3 with respect to perturbations on 72 pathways combinations can be analyzed, giving not only the information on which pathway matters but also portray the time (when) it matters (see Figure D.1). However, many of these pathways are overlapping and hence, some of the information obtained is expected to be correlated. Correlation metrics can be computed and clustered, for example based on the Euclidean distances among them and using centroid distance as the linkage class. Figure 5.1 shows the resulting binary classification tree of the pathPSA coefficients of caspase-3. In general, there exist three clusters (with cluster distance greater than 0.1), two for type-I and one for type-II, that have significantly different sensitivity behavior to pathway perturbations. The main difference between the two clusters of type-I is the inclusion of regulatory inhibition of caspase-3 by XIAP. Aside from this inhibition, adding other regulatory mechanisms to the basic signaling units (i.e., type-I/II) did not appreciably change the profile of pathPSA coefficients.

The heat map analysis of the three pathPSA clusters is shown in Figure 5.2. Specifically, the pathPSA analysis suggested the following dynamic picture: (i) early caspase-3 response through type-I pathway is triggered (Figure 5.2(a)), (ii) inhibition by XIAP becomes active following the early caspase-3 activation (Figure 5.2(b)), and (iii) finally, type-II pathway carries the main cell death signaling, giving the switch-like caspase-3 activation (Figure 5.2(c)). Therefore, pathPSA analysis clearly supported a type-II dependent caspase-3 switching with an early type-I dependent

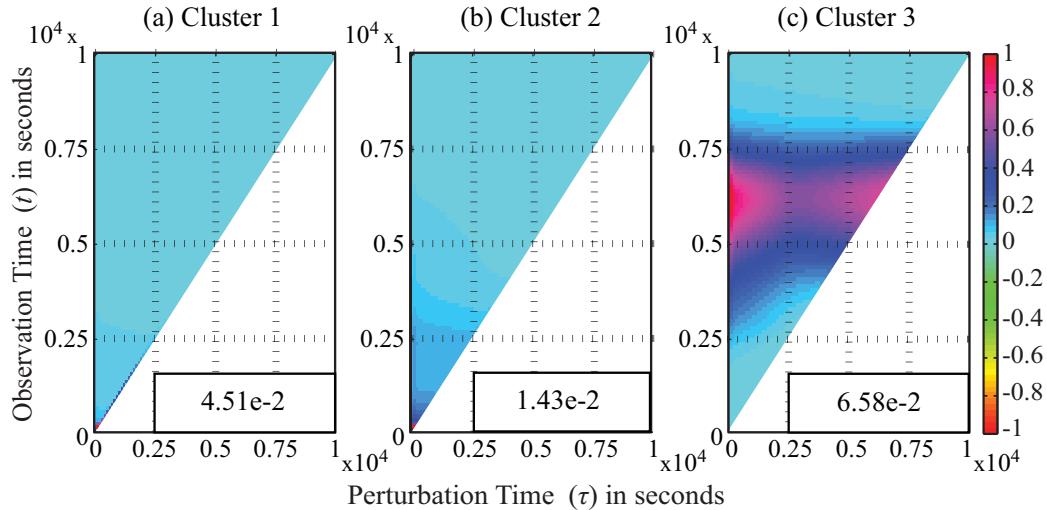


Figure 5.2. PathPSA of caspase-3 activation. (a-c) The heat map plots illustrate the cluster of pathPSA coefficients of active caspase-3 level with respect to perturbations to pathways. Here, clusters 1-3 represents type-I, type-I with caspase-3 inhibition by XIAP and type-II pathways, respectively. The x -axis is the time at which impulse perturbations are applied, while the y -axis indicates the observation time of caspase-3. Each plot is scaled to have values between -1 and +1 by the scaling factor in the bottom right corner.

activation of caspase-3, which is followed by XIAP inhibition for a short period. The inhibition by XIAP explains the transition of importance between type-I and type-II signaling. These observations are in general agreement with the results from the GFM and iPSA of this model in Sections 3.3.2 and 4.3.2. Also, other regulatory pathways such as caspase-6 feedback and DISC, mitochondria, apoptosome inhibitions appeared to have negligible impact on the cell death signaling.

5.4 Discussion

5.4.1 Experimental and biological relevance

Understanding the non-intuitive behavior of biological systems can be done using variety of approaches in systems biology. Existing methods range from static structural analysis to dynamic sensitivity analysis, each having its own advantages and disadvantages. To this end, this chapter presents a novel dynamical sensitivity analysis that considers system topology, called as the pathPSA. This analysis combines both structural and dynamical information, by extracting network

topological information as functional pathways and introducing dynamical impulse perturbations on these functional pathways as parametric perturbations.

Certain cellular network topologies (e.g. scale-free, small world network) are related to (structural) robustness property. In the earlier discussions on dynamical robustness, this property has been defined as insensitivity to perturbations and consequently can be studied using different types of sensitivity analysis. Here, the pathPSA provides a marriage between structural and dynamical robustness using sensitivity analysis with respect to dynamical perturbations of network pathways. Specifically, pathPSA directly gives information about the pathways that are critical in regulating the (robust) system behavior, as well as portrays two dynamical aspects of the cause-effect relationship: the times at which the pathway perturbations become important and at which the system output is significantly affected by these pathways.

Like the GFM and iPSA, such information can again be used in various applications, such as model identification and reduction, drug discovery research in understanding the mechanism of drug action and in identifying the target paths. The choice of using pathPSA over other dynamical sensitivities presented in Chapters 3-4 depends only on the end application. For instance, if the molecules and reactions under query are smaller in number, then the GFM and iPSA can be sufficient enough to infer the system dynamics, but when the system size increases (with increasing number of molecules and even more increasing number of reactions), which is usually the case in biological systems, then the pathPSA provides a viable and efficient tool to analyse the system mechanisms directly by perturbing pathways, rather than system molecules and reactions.

5.4.2 Comparison with classical PSA, GFM and iPSA

While the PSA, GFM and iPSA introduce perturbations either on system parameters or on molecular concentrations, the pathPSA introduce (impulse) perturbations on functional pathways, which represent a set of reactions that can give rise to the observed output behavior. As there are typically much fewer basic functional units than the number of parameters and species in a network (i.e., main ESMs), the interpretation of the results of pathPSA, after clustering, should be easier than that of the PSA, GFM and iPSA. However, like the analyses presented in Chapters 2-4, the results of pathPSA will depend on the choice of the pathway parameters. That is, the pathPSA gives local information about the cellular network based on infinitesimal perturbations around the nominal parameters. The extension of the pathPSA to global (finite) perturbations is of future interest.

The pathway identification considered in this chapter borrows the concept of ESMs and extends it to include other regulatory and side processes, such as feedback, inhibition and degradation reaction. However, the algorithm neglects any side processes, which are two nodes away from the basic functional unit. While such interactions could be identified and included in the pathway identification procedure, this may lead to a combinatorial increase in the number of pathways. Nevertheless, if desired, the pathPSA formulation is general enough to account for such change of pathway definition.

Another limitation of pathPSA is in the number of pathways to be analyzed. Pathways are formulated based on combinations of basic functional units (ESMs) and the associated side processes. This could result in an exponential increase in the number of pathways formed and hence to an increased computational cost. This issue can be avoided by a segregated (step-wise) pathPSA. Initially, pathPSA is performed

only for the basic signaling units (ESMs). Basic units that have “small” sensitivities can be ignored, while pathways can be defined for those that are important. In this case, the analysis need not be done on the complete set of pathways in a given network.

5.5 Summary

Biological complexity arises not only because of large number of components and non linear interactions, but also because of the system topology. Hence, a systems-oriented approach to establish the dynamic cause-effect relationship between the topology of a cellular network model and its experimentally observed dynamic property or function, is needed. To this end, a novel sensitivity analysis, called pathway parametric sensitivity analysis (pathPSA) that offers dynamical insights on the functional regulations of a system output by pathways, is presented here. In contrast to the individual state perturbations in the GFM analysis and parametric perturbation in the iPSA and PSA, the pathPSA makes use of perturbations of pathways that connect system input to output. The analysis is done by first identifying the basic functional units and their associated side processes, followed by re-parametrizing the dynamic ODE model of the system and finally, performing impulse parametric sensitivity analysis on the pathway parameters. In the application to the Fas-induced programmed cell death model for Jurkat T-cells, the pathPSA result suggested an early type-I dominant regulation of caspase-3 activation, which later gets transferred to type-II regulation, after XIAP inhibition. The applications of pathPSA include model diagnosis, model reduction, model refinement and drug discovery, such as for understanding the mechanism of drug action and identifying the target paths.

CHAPTER 6

6 Molecular Density Perturbation Analysis^{††}

Synopsis:

Despite the existence of cell-to-cell variability, the overall function at organ levels typically remains tightly controlled. A quantitative understanding of this cellular phenotype under uncertainty requires modeling frameworks that are able to account for intrinsic and extrinsic cellular variability. To this end, probabilistic models, like the ones presented in Section 1.2.1 have been used. Thus far, three local dynamic sensitivity methods (the GFM, iPSA and pathPSA) have been created for analyzing ODE models of single cells. In contrast this chapter contains a global sensitivity method, called molecular density perturbation (MDP) analysis, to analyze probabilistic models of cellular populations. Like the GFM analysis, this method takes a molecular centric perturbation approach, but under uncertainty. The efficacy of the MDP analysis is demonstrated through an application to a population model of TRAIL-induced apoptosis in Hela cells. The MDP analysis was also performed to compare the sensitivities of apoptotic and non-apoptotic cells to reveal differentially regulated molecules disrupting the apoptotic mechanism.

^{††} Excerpts of this chapter will be part of the following manuscript:

Perumal TM and Gunawan R. Elucidating the dynamics of cellular populations: a dynamic perturbation approach on molecular density functions, *In preparation*

6.1 Introduction

Single cell assays, such as using immunofluorescence-based flow cytometry and high-throughput immunofluorescence microscopy, have revealed the existence of cell-to-cell variability, which can explain phenotypic heterogeneity seen in cellular populations [124, 164]. Such heterogeneity can arise from genetic [165], epigenetic [166], non-genetic/stochastic [66, 68] and/or extracellular environmental variations [167]. Genetic variations are stable and passed on to progenies, and epigenetic modifications are also heritable even for a higher number of cell divisions ($\sim 10\text{-}10^5$) [168]. Since genetic and epigenetic are heritable factors, these determine the mean levels or activities of proteins [124]. Nevertheless, there also exist other stochastic factors such as that arising due to random discrete processes involving low copy numbers of biomolecules (genes and proteins) [68]. In addition, environmental variations (e.g. morphogen gradients during embryogenesis) can further contribute to cell-to-cell differences, and such effect can impact both the mean and variance of gene expressions. Despite the ubiquity of such cell-to-cell variation from microorganisms to higher mammals, the overall population/tissue/organ level phenotype is tightly controlled [164]. Hence understanding the impact of biological variability in cellular information processing and analyzing the population dynamics under uncertainties are important.

Understanding the maintenance of cellular phenotype under uncertainty quantitatively requires modeling frameworks that are different from the ODEs seen in previous chapters. To this end, probabilistic models, such as Fokker-Planck equations (FPE) [126] and chemical master equations [69] (seen in Section 1.2.1), have been used to describe the cell-to-cell variability. These models differ in the manner with which they account for the source of variability. Some examples of these probabilistic

models in biology include the modeling of gene expression in eukaryotes [169], signal transduction [170], and cell death (apoptosis) [66, 171]. The analysis of such models are typically based on global or hybrid sensitivities, as presented in Section 2.1.2 and 2.1.3, such as SOBOL sensitivity [132], DGSM [134], and Glocal analysis [135]. As mentioned in Chapter 2, the dynamical aspect of cellular regulation however may not be immediately apparent from the application of these analyses due to the persistent manner of the underlying perturbations used. Since biological functions are often dynamic in nature, there is a necessity to understand the dynamical regulation as done in the previous chapters, but in the face of cell-to-cell variability.

Like the GFM, iPSA and pathPSA, the novel global analysis presented here can generate dynamical information for a cellular population under uncertainty. This analysis is accomplished through perturbations of marginal density functions of molecular concentrations at different times, hence the name molecular density perturbation (MDP) analysis. The MDP coefficients, analogous to the GFM sensitivity coefficients, are used to reveal the dynamical information about actively participating molecules. The efficacy of this method in comparison to single cell analysis like the GFM (seen in Section 3.2) and iPSA (seen in Section 4.2) and to a common population analysis eFAST [132] (seen in Section 2.1.2), is evaluated through an application to a population (Fokker-Planck) model of TRAIL-induced programmed cell death in Hela cell line [66]. Here, the heterogeneity in population arises only due to the variability in initial concentrations of biomolecules. The MDP analysis is used to reveal the key regulators of cPARP activation in Hela cell population. Also, a separate MDP analysis of apoptotic and non-apoptotic cellular population, predicted by the same model, shows the differentially regulated molecules in cPARP activation.

6.2 Molecular Density Perturbation (MDP) Analysis

The MDP analysis is formulated using the same basis of cause-effect relationship as in the classical PSA in (2.2) [94]. The MDP coefficients are defined as to quantify the change in the output molecular probability density function (effect) due to the change in the density function of molecular input (cause). Similar to the GFM analysis, the MDP takes a molecular centric approach. The first order MDP sensitivity coefficients are obtained by introducing perturbations on marginal densities of input molecules X_j at time τ and quantifying the resulting change in the marginal densities of the output molecules X_i at any time $t \geq \tau$. In this case, the MDP coefficients $gS^x(t, \tau)$, which is an $n \times n$ matrix, is calculated using the following ratio

$$gS_{ij}^x(t, \tau) = sign(\Delta X_i) \frac{\Delta F_{X_i}(t, x_i)}{\Delta F_{X_j}(\tau, x_j)} \quad (6.1)$$

where x_i and x_j are the random variables of molecules X_i and X_j , respectively, and $\Delta F_{X_i}(t, x_i)$ is the distance between positively and negatively perturbed marginal densities of X_i at time t . The positively (negatively) perturbed marginal density of X_i is associated with a positive (negative) mean perturbation of the population of the input molecule X_j . Finally, the function ΔX_i is the difference between the expected values of X_i from the positively and negatively perturbed population. Intuitively, second and higher order MDP coefficients can be obtained by considering the joint density function, $F_x(t, \mathbf{x})$ of inputs and outputs (e.g., second order MDP coefficient of the output molecule X_i with respect to the input molecules X_j and X_k is given by

$$gS_{ijk}^x(t, \tau) = \Delta F_{X_i}(t, x_i) / \Delta F_{X_j, X_k}(\tau, x_j, x_k).$$

This chapter has used both probability density function (PDF) and cumulative density function (CDF) distances to indicate the magnitude. Six density distance metrics from the literature [172] are compared, including four PDF distances: Kullback-Leibler distance (Δ_{KL}), Jeffrey distance (Δ_J), Jenkinson-Shanon (Δ_{JS}) divergence and engineering metric (Δ_e), and two CDF distances: Kolmogorov-Smirnov CDF distance (Δ_{KS}) and Cramer-Von Mise distance (Δ_{CVM}) (see Section E.1 for definitions). Particularly, the Cramer-Von Mise distance (Δ_{CVM}) is given by

$$\Delta_{CVM} F_{X_i}(t, x_i) = \int_{-\infty}^{\infty} \left(F_{X_i^p}(t, x_i^p) - F_{X_i^n}(t, x_i^n) \right)^2 dx_i \quad (6.2)$$

where superscripts p and n denote the densities associated with positive and negative mean perturbation of the input molecular population. In this case, $F_{X_i^p}(t, x_i^p)$ and $F_{X_i^n}(t, x_i^n)$ represent the marginal cumulative distribution functions. Like local sensitivity coefficients in (3.2) and (4.8), the MDP coefficients are formulated in terms of the normalized variables $\bar{x}_i = x_i / x_i^{\max}$; $\bar{x}_j = x_j / x_j^{\max}$ giving the following normalized sensitivity metric

$$g\bar{S}_{ij}^x(t, \tau) = \frac{\Delta F_{\bar{X}_i}(t, \bar{x}_i)}{\Delta F_{\bar{X}_j}(\tau, \bar{x}_j)} \quad (6.3)$$

As in the GFM and iPSA coefficients, the MDP coefficients will have two time axis, time of perturbation τ and time of observation t and again each $(i,j)^{\text{th}}$ element of $g\bar{S}^x(t, \tau)$ can be visualized as heat maps as shown in Figure 6.1. An analogous interpretation of these sensitivities with that of the GFM also exists, where a change in the marginal densities of j^{th} molecular concentration X_j at time τ cause an eventual increase (no change) in the marginal densities of i^{th} molecular concentrations X_i at time t ($t \geq \tau$). Since the system is causal, the lower right half of

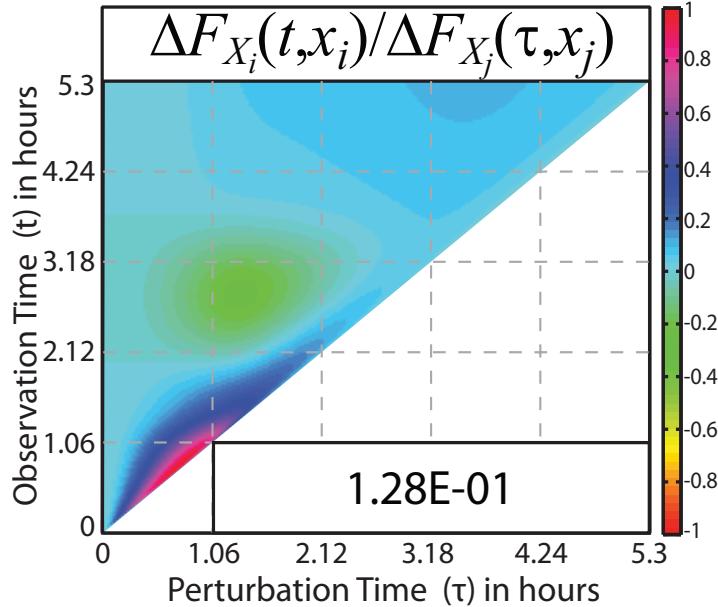


Figure 6.1. A heat map of the MDP coefficients. The heat maps visualize the $(i,j)^{\text{th}}$ element of MDP coefficient, showing the change in molecular density function $F_{X_i}(t, x_i)$ with respect to perturbation on molecular density function $F_{X_j}(\tau, x_j)$. The x-axis of these plots represents the perturbation time τ at which perturbations are introduced on random variable X_j while the y-axis represents the observation time t at which changes in random variable X_i are measured. MDP coefficient in the heat map is scaled accordingly to have magnitudes bounded by ± 1 , by the scaling factor reported in the lower right corner of the plot.

the plot is again null. The sign of the MDP sensitivities may reflect the potential action of the input molecule, i.e. positive sensitivities indicate activators (or substrates) and negative sensitivities indicate inhibitors.

As noted above, the proposed MDP analysis differs from the existing global sensitivity analyses in the way that the perturbations are realized and the resulting changes are quantified. This is in contrast to the use of moments of the density function than quantifying the change of the density functions, e.g. by variance decomposition in eFAST [132] or by mean change in DGSM [134]. Therefore, the MDP analysis avoids the shortcomings of eFAST in handling co-linearity between input perturbations and of DGSM in handling multi-modal distributions. Also, the perturbations in MDP analysis are dynamic in nature, i.e., introduced at varying times τ . Consequently, the result of MDP analysis can reveal network fragility in the form

of key biological molecules, where such fragilities may dynamically change from one molecule to the next during cellular regulation.

Similar to the GFM analysis of single cell models presented in Chapter 3, the MDP analysis of cellular population model can also reveal two major biological insights: either by choosing a particular output x_i and observing the sensitivities of its variability (density function) to changes in the uncertainty of other molecules x_j 's (a row of $\mathbf{g}\bar{\mathbf{S}}^x(t, \tau)$), or by choosing a particular perturbation in the uncertainty of x_j and observing the resulting effects on the variability of the other molecules x_i 's in the network (a column of $\mathbf{g}\bar{\mathbf{S}}^x(t, \tau)$). The information gained from MDP analysis can be used to understand the consequences of cell-to-cell variability [124], such as in drug discovery research, to identify potential drug targets, to understand drug efficacy and specificity, to optimize drug dosing and timing and also in analyzing the robustness-fragility of systems under uncertainty [164].

6.3 Application to TRAIL-induced Cell Death Model of HeLa cell Population

Figure 6.2 depicts the extrinsic cell death (apoptosis) signaling network triggered by the TNF-related apoptosis-inducing ligand (TRAIL) in Hela cells. A mathematical model was constructed on the basis of mass-action kinetics using ordinary differential equations (ODEs) [66]. The model has 58 species, 28 reactions, 18 non-zero initial conditions and 70 parameters (see Section E.2 for initial conditions, parameter values rate equations, ode solver type and its settings). The model parameters and initial conditions were acquired from a parameter fitting to experimental data, consisting of both single-cells and population-based experiments such as imaging, flow cytometry, and immunoblotting [59, 66, 173]. Although the model lacks detailed descriptions of

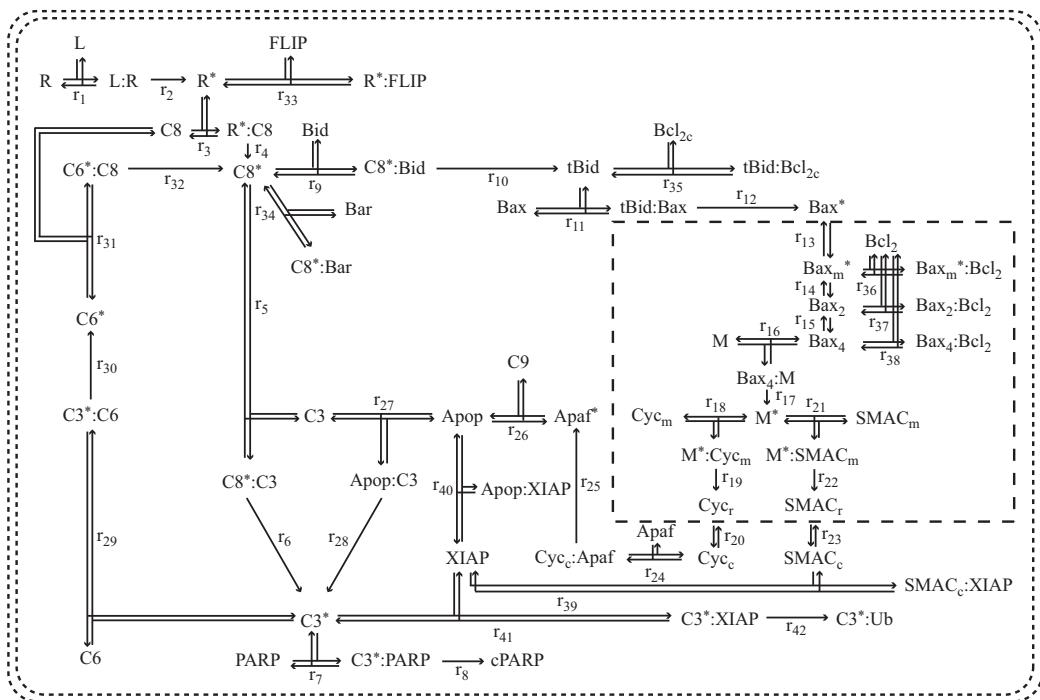


Figure 6.2. Network model of TRAIL-induced cell death in HeLa cell lines. Type-I pathway involves a direct cleavage of pro-caspase-3 by caspase-8 to form an active caspase-3 which cleaves the substrate PARP to cPARP, while the type-II pathway describes a mitochondria-dependent activation of caspase-3 in turn activating the substrate.

few key mechanisms in the extrinsic cell death signaling (e.g., details of DISC and mitochondrial pore formation), it has extensive details on most of the significant mechanisms for the activation of endogenous executioner caspase (C3) and its downstream substrate (PARP) cleavage [59]. Specifically the model has four major mechanisms: (i) upstream pathway, representing TRAIL activation of pro-caspase-8 (to caspase-8), (ii) mitochondrial independent pathway (type-I), representing an enzyme cascade in which caspase-8 cleaves inactive pro-caspase-3 (to active caspase-3) and inhibition of active caspase-3 by XIAP, (iii) a mitochondrial pathway (type-II), in which caspase-8 promotes the formation of mitochondrial pores, the consequent release of CyC (cytochrome-C) into the cytosol, and finally the activation of caspase-3, and (iv) pro-caspase-6 positive feedback-loop, cleaving pro-caspase-8 (to caspase-8).

The output of interest in this model is the cleavage of the substrate PARP (to cPARP) by activated executioner caspase-3. Here, single cell model analysis using the

GFM analysis and iPSA and population analysis using the eFAST and MDP analysis are applied to reveal the important mechanism(s) of cPARP activation, and to identify key differences between cells that respond positively to TRAIL (apoptotic) and those that escape cell death (non-apoptotic).

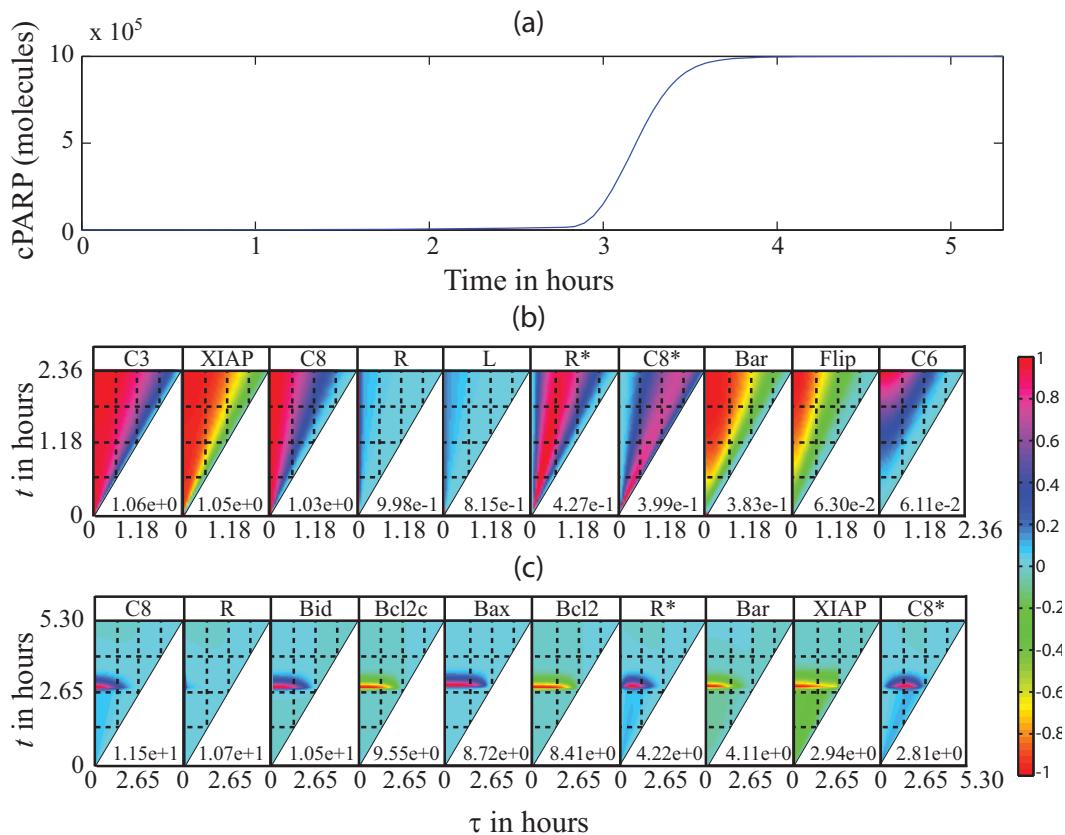


Figure 6.3. Green's Function Matrix (GFM) analysis of cPARP activation by a constant TRAIL stimulus in a single cell. (a) cPARP activation follows a delayed snap-action behaviour in response to a constant TRAIL stimulus. (b-c) The heat maps represent the ten largest GFM coefficients of the cPARP active level with respect to perturbations on different molecules in the network. The x -axis of these plots represents the time τ at which perturbation is introduced on each molecule, while the y -axis represents the time t at which the active cPARP level changes. Each heat map is scaled accordingly to have values bounded by ± 1 , using the scaling factor reported in the lower right corner of the plot. (b) GFM analysis during pre-MOMP period (before 2.36 hours). The analysis showed that the common upstream ((in) active receptor, pro- and caspase-8) and downstream molecules (C3 and XIAP) constituted the early responders to TRAIL stimulus. (c) Post-MOMP (after 5.3 hours), both upstream molecules (receptor, caspase-8), XIAP and specific type-II molecules (Bid and Bax) are directly implicated to be the key regulators of cPARP switch by the location and magnitude of the peak sensitivities.

6.3.1 GFM and iPSA analysis of TRAIL induced apoptosis of a single cell

GFM analysis of apoptosis in a single Hela cell (a representative mean of population) reveals pro-caspase-8 and XIAP to be the key components in cPARP activation:

The GFM analysis was performed under a constant TRAIL stimulation on the single cell ODE model using population mean concentration as the initial condition (see Section E.2) [173]. The analysis was carried out over a time range of 0 to 5.3 hours to allow the system to reach a new steady state. During this time period, the single cell model simulated an apoptotic cell, in which the activation of cPARP in response to TRAIL follows a delayed snap action switch like profile (Figure 6.3(a)). Each subplot in Figure 6.3(b-c) here represents an element in the row of GFM matrix corresponding to active cPARP.

To study the activation dynamics of cPARP in greater detail, the GFM analysis was split into two phases: pre-MOMP (before 2.36 hours) and post-MOMP (after 2.36 hours). Figure 6.3(b) and 6.3(c) portray the ten most important molecules in the cPARP activation according to the GFM sensitivities during the pre- and post-MOMP, respectively (see Section E.3 for detailed analysis). The sensitivities with respect to molecular complexes are omitted, as these molecules are either reaction intermediates or mathematical approximations and carry little physical significance (see Video E.1 for detailed ranking and analysis). According to the results of the GFM analysis, the delayed snap-action activation is majorly governed by the mitochondrial dependent pathway (type-II), but the most important molecules are pro-caspase-8 and XIAP. The same conclusion is also obtained through an application of the impulse parametric sensitivity analysis (iPSA) (see Section E.4).

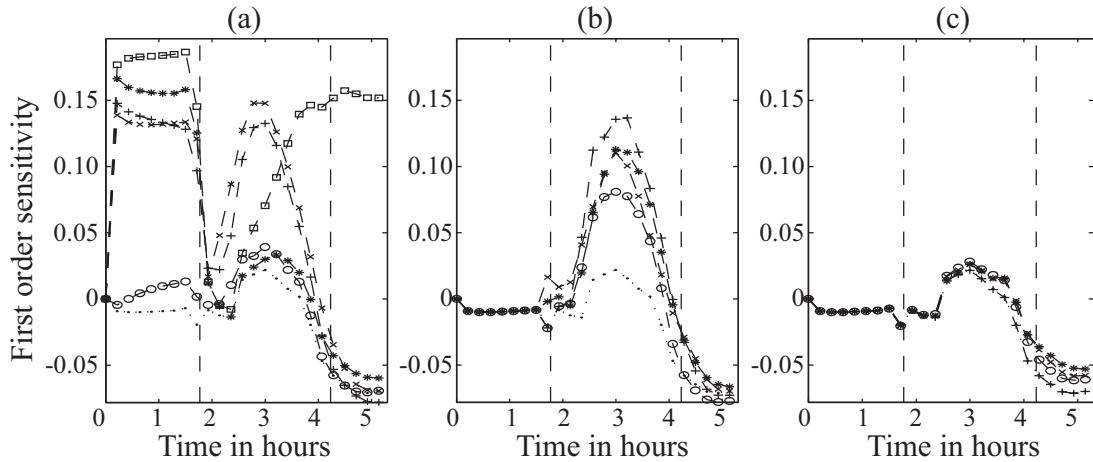


Figure 6.4. Extended Fourier Amplitude Sensitivity Test (eFAST) of cPARP activation in a cellular population. (a-c) represents the first order sensitivity of cPARP to the molecules indicated below. The dashed lines (--) at 1.77 and 4.23 hours indicates the start and end of MOMP, respectively. (a) upstream molecules, receptor (+), pro-caspase-8 (x), Bar (o), pro-caspase-3 (*), XIAP (□), pseudo (●), (b) type – II molecules (upstream of MOMP), Bid (+), Bcl2c (x), Bax (o), Bcl2 (*), pseudo (●), (c) type – II molecules (downstream of MOMP), membrane bound cytochrome-c (+), membrane bound Smac (x), Apaf (o), pro-caspase-9 (*), pseudo (●). eFAST analysis points to the initial variability in XIAP and pro-caspase-8 to be the important regulators of cPARP activation.

6.3.2 eFAST analysis of TRAIL induced apoptosis

eFAST analysis of apoptosis in Hela cell population reveal XIAP and pro-caspase-8 to be the key components in cPARP activation: A global sensitivity analysis of the cellular population version of the same apoptosis model was performed using the eFAST algorithm [131], as presented in Section 2.1.2. The analysis was done under the same TRAIL stimulation over the time range between 0 and 5.3 hours, using a cell population size of 1000 and with five replicates. The initial conditions of the cells were sampled from a log-normal distribution using Latin Hypercube Sampling (LHS) algorithm (see Section E.2 for mean values and coefficients of variations). Note that the sensitivity coefficients were calculated only with respect to the 18 molecules whose nominal initial conditions were non-zero. The cell population simulation revealed that 94.69% of the cells undergo apoptosis. Figure 6.4 shows the time-

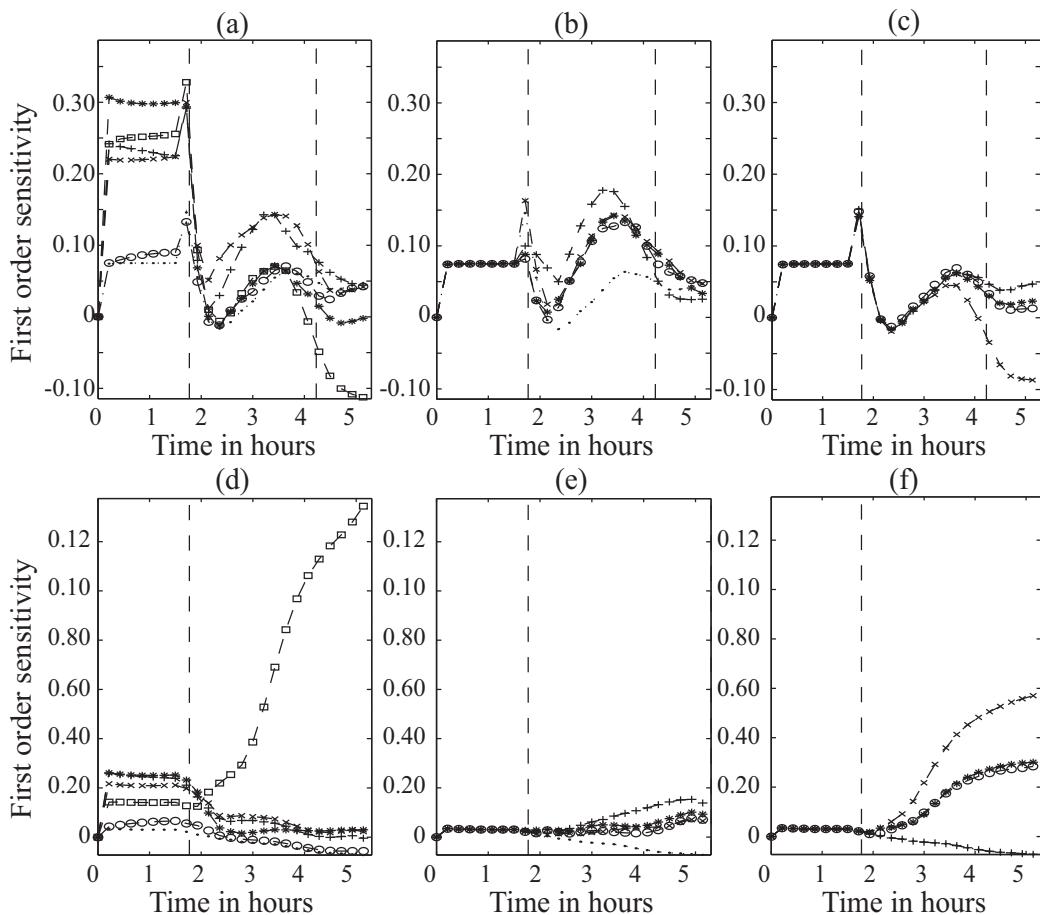


Figure 6.5. Extended Fourier Amplitude Sensitivity Test (eFAST) of cPARP activation in apoptotic and non-apoptotic cell population. (a-f) represents the first order sensitivity of cPARP to the molecules indicated below and the dotted lines (--) at 1.77 hours indicates the start of MOMP. (a-c) the dotted lines at 4.23 hours indicate the end of MOMP in apoptotic cell population; (d-f) the dotted lines at 5.3 hours indicate the end of MOMP in non-apoptotic cell population. (a,d) upstream molecules, receptor (+), pro-caspase-8 (\times), Bar (\circ), pro-caspase-3 (*), XIAP (\square), pseudo (\bullet), (b,e) type – II molecules (before MOMP), Bid (+), Bcl2c (\times), Bax (\circ), Bcl2 (*), pseudo (\bullet), (c,f) type – II molecules (after MOMP), membrane bound cytochrome-c (+), membrane bound Smac (\times), Apaf (\circ), pro-caspase-9 (*), pseudo (\bullet). A comparative eFAST analysis of apoptotic and non-apoptotic cell population points to the initial variability in XIAP and membrane bound Smac to be the differentially regulated molecules responsible for cPARP activation.

varying first-order sensitivity coefficients of cPARP with respect to 13 molecules with the largest cPARP sensitivities (see Section E.5 for more detailed results). A pseudo molecule with no dynamics was artificially introduced into the system to quantify for the variability in the calculated sensitivity coefficients due to finite sampling.

The global analysis was split into three dynamic regimens: (i) pre-MOMP (before 1.77 hours), defined as the period of which at least 1% of the population

possess one mitochondrial pore, (ii) MOMP (between 1.77 and 4.23 hours), defined as the period between the end of pre-MOMP and the time at which 99% of the population have accomplished 99% of CyC/Smac release, and (iii) post-MOMP (after 4.23 hours). From Figure 6.4(a), it is clear that variability in the upstream molecules: pro-caspase-8 and pro-caspase-3, and the inhibitor XIAP significantly contributes to the final cPARP variability. During MOMP, Figure 6.4(b) further suggests that in addition to these molecules, a few type-II related molecules, such as Bid, Bax and Bcl2, also regulate cPARP activation, as indicated by the sensitivities. Interestingly, Figure 6.4(c) indicates that the variability of downstream type-II molecules, e.g. membrane bound cytochrome-c, Smac, Apaf and pro-caspase-9, does not carry much effect on the MOMP. In summary, eFAST analysis shows that the variability in the initial concentrations of pro-caspase-8, XIAP and Bid are the determinant of the cell death decision (by way of cPARP activation), which is in much agreement with the single cell GFM analysis above.

Comparative eFAST analysis of apoptotic and non-apoptotic cell population reveals initial XIAP and membrane bound Smac to be the differentially regulated molecules in cPARP activation: The eFAST analysis was repeated to analyze two subpopulations with cells that responded positively to TRAIL (committing to apoptosis) and those that did not. The purpose of the analysis is to elucidate the key components (molecular species) that contribute to the differences in the final cPARP concentration. Figure 6.5(a-f) shows the time-varying global (first order) sensitivities with respect to thirteen molecules with the largest sensitivities (in magnitude) and to the pseudo molecule (see Section E.6 for more detailed results). As expected, the results of the eFAST analysis of apoptotic subpopulation, as shown in Figure 6.5(a-c),

were equivalent to that done for the whole population, as the majority of cells in the population (~95%) were apoptotic. The comparison between the eFAST results of apoptotic and non-apoptotic subpopulations suggests that the initial XIAP and membrane bound Smac are the determining factors for apoptotic decision.

6.3.3 MDP analysis of TRAIL induced apoptosis

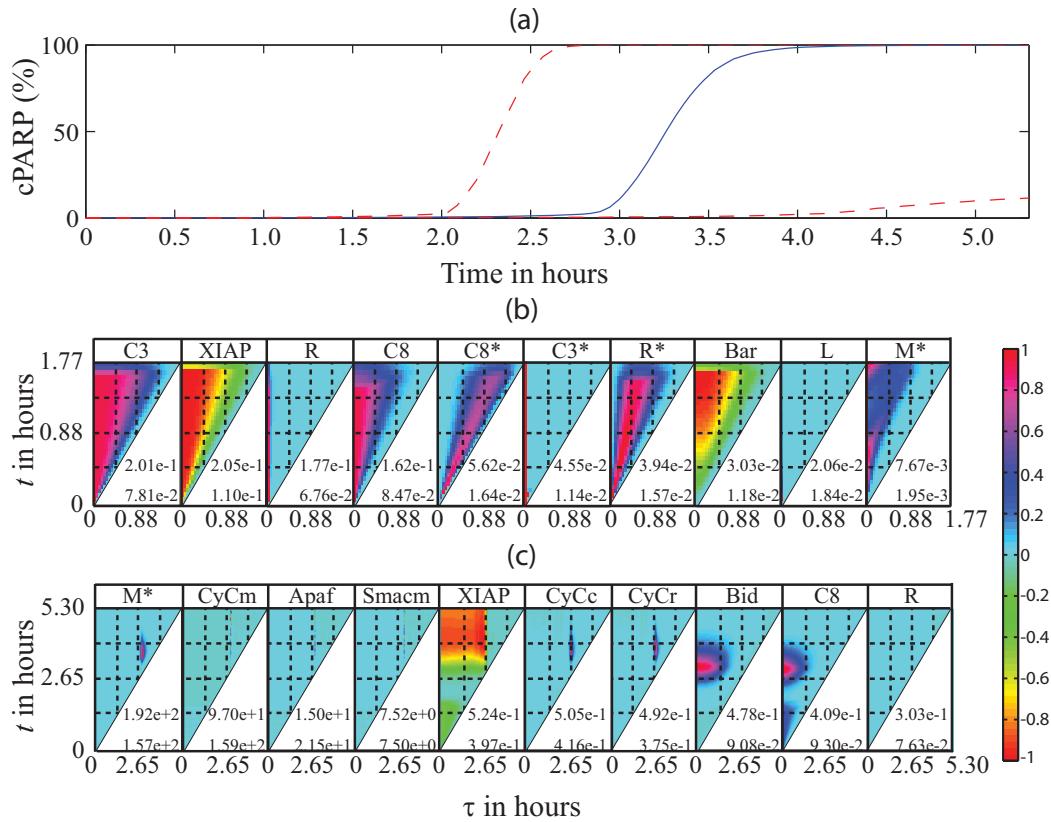


Figure 6.6. Molecular Density based Perturbation (MDP) analysis of cPARP activation by a constant TRAIL stimulus in cellular population. (a) cPARP activation follows a switch-like behaviour in response to a constant TRAIL stimulus. Dotted lines (--) indicate the 1% and 99% bounds and lines (-) represents the median evolution, respectively. (b-c) The heat maps represent the average values of top ten MDP coefficients of cPARP active level with respect to the perturbations on different molecules in the network. The x-axis of these plots represents the time τ at which perturbation is introduced on each molecule, while the y-axis represents the time t at which the active cPARP level changes. Each heat map is scaled accordingly to have magnitudes bounded by ± 1 , by the scaling factor reported in the lower right corner of the plot along with the variance of MDP coefficients over five runs. (b) MDP analysis performed till MOMP (till $t = \tau = 1.76$ hours) delay. As expected, in agreement with the single cell analysis, MDP analysis showed that the common upstream ((in) active receptor, pro- and caspase-8) and downstream molecules (C3 and XIAP) constituted the early responders to TRAIL stimulus. (c) After MOMP (till $t = \tau = 5.3$ hours), contradicting to the single cell analysis, MDP analysis type-II specific molecules (number of mitochondrial pore opening, membrane bound cytochrome-c, Smac and initial available Apaf) are directly implicated to be the key regulators by the location and magnitude of the peak sensitivities.

MDP analysis of HeLa cell population reveals mitochondrial pore opening, initial membrane bound cytochrome-c, Smac, and Apaf in cytosol to be the key components in cPARP activation:

The MDP analysis was done in the same spirit as the GFM and eFAST analyses above and under the same conditions. The MDP sensitivities were calculated using the Δ_{CVM} density distances, in which active cPARP was selected as the output molecule. Figure 6.6(b-c) shows the heat map of the MDP sensitivities, analogous to those of the GFM analysis (see Video E.5 for more detailed results). Following the eFAST analysis, the MDP analysis was split into three dynamic regimens (before, during and after MOMP).

Figure 6.6(b-c) portrays the ten molecules with the largest cPARP MDP sensitivities in magnitude, before and during MOMP, respectively. The sensitivities with respect to molecular complexes are again omitted (included in Section E.7). In this case, the MDP sensitivities reveal that: (i) caspase-8 and caspase-3 activation and caspase-3 inhibition by XIAP are the important global mechanisms during pre-MOMP, (ii) mitochondrial dependent (type-II) pathway molecules, including Bid, Bax, Apaf, CyC and Smac gain importance over direct caspase-8 pathway during MOMP, (iii) post-MOMP cPARP activation is majorly determined by the mitochondrial dependent (type-II) pathway, such as active mitochondrial pore opening, and by initial available concentrations of Apaf and membrane bound cytochrome-c/ Smac. In comparison to the single cell GFM analysis and the global eFAST analysis, the MDP analysis suggest the same initial regulator of cPARP activation to be caspase-8, -3 and XIAP, but unlike the GFM and eFAST analyses, the MDP sensitivities indicate active mitochondrial pores and the initial available membrane bound CyC/Smac and inactive Apaf as the most important factors during and after MOMP.

Comparative MDP analysis of apoptotic and non-apoptotic Hela cell population reveals XIAP and membrane bound Smac to be the differentially regulated molecules in cPARP activation: Similar to the global eFAST analysis, the MDP analysis was repeated to analyze subpopulation of apoptotic and non-apoptotic cells. In this case, the final cPARP concentration was taken to be the biomarker of apoptosis (see Figure 6.7(a-b)). The MDP sensitivities of the final cPARP activation were further compared according to the infinite norm over the perturbation times τ . Figure

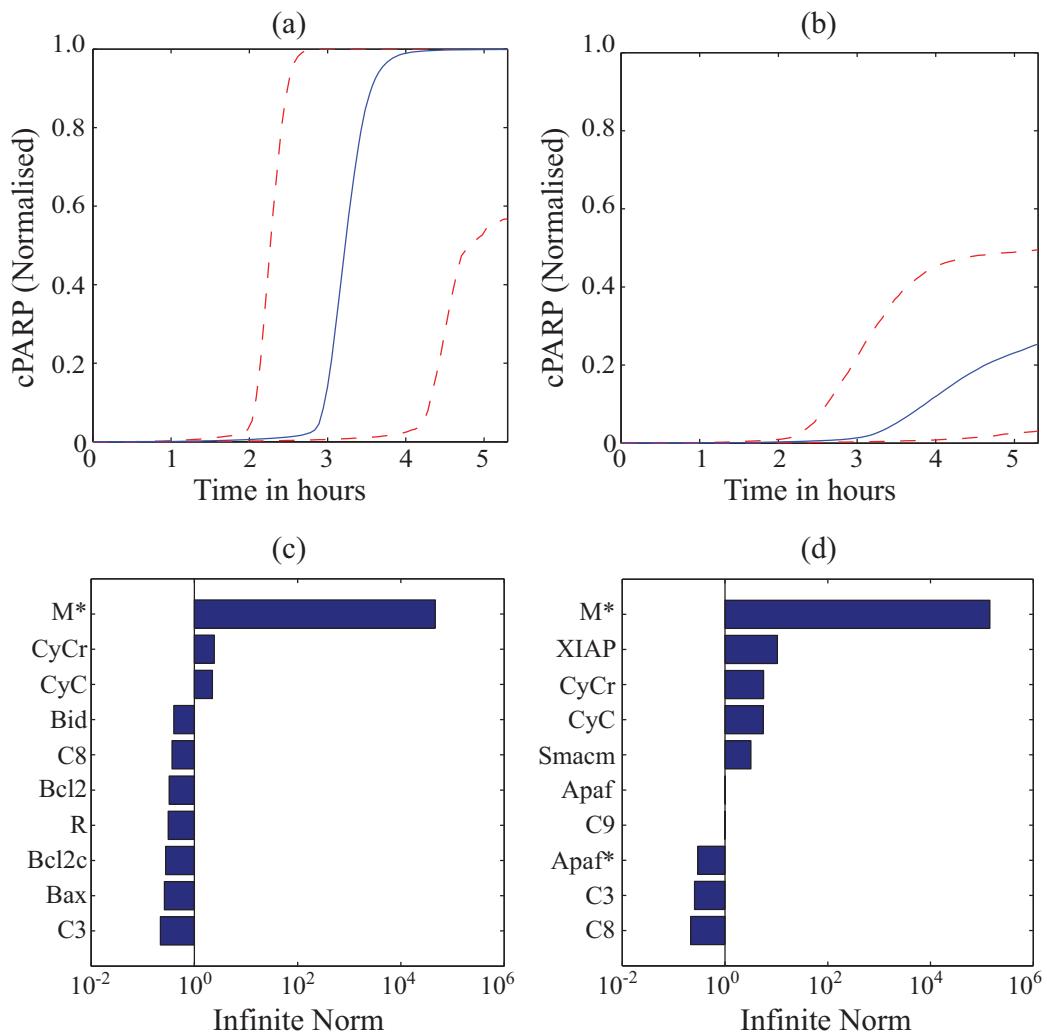


Figure 6.7. Comparative MDP analysis of cPARP activation by a constant TRAIL stimulus in (a,c) apoptotic and (b,d) non-apoptotic cell population. (a-b) Activation levels of cPARP to a constant TRAIL stimulus normalized with respect to the initial available PARP levels. Dotted lines (--) indicate the 1% and 99% bounds and lines (-) represents the median evolution, respectively. (c-d) Bar graphs represent the top ten sensitive molecules at steady state. As seen from the graph XIAP, Smac, Apaf(s) and C9 are the differentially regulated molecules.

6.7(c-d) shows the ranking of the top ten molecules according to the magnitude of the infinite norm of MDP sensitivities above (see Section E.8 for detailed analysis). The MDP ranking of apoptotic subpopulation was in much agreement with the analysis for the whole population, as most cells in the complete population were apoptotic. The analysis of the non-apoptotic subpopulation again differ from that of the apoptotic cells in the importance of XIAP and membrane bound Smac, as also found previously by the global eFAST analysis.

6.4 Discussion

6.4.1 Variability in the conclusion of MDP analysis

Since the MDP (as well as the eFAST) coefficients are calculated from Monte Carlo simulations, there is a variability associated with finite sampling. This variability can arise from the different stages in the density distance calculation, ranging from the random sample generation (e.g. from the choice of random number generator used) to the normalization technique and the reconstruction of density functions. Partly, the magnitude of the error (bias) in the MDP coefficients can be estimated using a pseudo

Molecules	$E\left[\max_{t,\tau}\left(\left g\bar{S}_{i,j}^x(t,\tau)\right \right)\right]$	$CV\left[\max_{t,\tau}\left(\left g\bar{S}_{i,j}^x(t,\tau)\right \right)\right]$	$E[\text{Rank}]$	$CV[\text{Rank}]$
Ms	210.552	0.651	1.200	0.333
CyCm	162.247	0.833	1.800	0.222
Apaf	21.445	0.889	4.200	0.350
Smacm	10.486	0.370	3.600	0.136
CyC	0.647	0.421	10.800	0.170
CyCr	0.601	0.442	12.200	0.256
XIAP	0.588	0.546	13.200	0.297
Bid	0.492	0.156	11.400	0.043
C8	0.429	0.144	12.400	0.040
R	0.342	0.053	14.400	0.034

Table 6.1. Represent the variability in the top ten MDP coefficients, calculated with respect to five replicated experiments. First and second column represents the mean and coefficients of variations of the infinite norms of sensitivity magnitudes calculated till the end time ($t = \tau = 5.3$ hours). Third and fourth column represents the corresponding mean and coefficient of variation of the sensitivity rankings considered.

state in the system with a nominal value of one and no dynamics. In theory, the pseudo variable should have no effect nor it should be affected by other system states and thus, is the MDP coefficients associated with this state should be zero. In this case, non-zero MDP coefficient of the pseudo state provides an estimate of the finite sampling error. Another way of accounting the variability in MDP coefficients is to use repeated experiments. In the current work, five random samples of size 1000 cell were analyzed. The coefficient of variation in the infinite norm of the top ten MDP sensitivities and the associated ranking are presented in Table 6.1. While the coefficient of variation can reach as high as 0.889, the relative ranking of the molecules is more consistent among replicates. In other words, the use of the MDP sensitivities in a comparative study to rank the importance of molecules, such as those done in the case study above, is reliable.

Also, the results of the MDP analysis depend on the metrics and parameters used to perform the analysis, including: (i) the density distance (i.e., Δ_{CVM} , Δ_e , Δ_J , Δ_{JS} , Δ_{KL} and Δ_{KS}), (ii) the type of perturbation (e.g., mean, variance or both), and (iii) the magnitude of perturbation. The MDP analysis in the case study above has used Δ_{CVM} with 10% mean perturbation of the input density function, while the variance of the log-normal distribution was kept the same. The effect of using a different density

Table 6.2(a): Ranking correlation between distance metrics of MDP analysis (mean, 10%)						
Measure	Δ_{CVM}	Δ_e	Δ_J	Δ_{JS}	Δ_{KL}	Δ_{KS}
Δ_{CVM}	1.000	0.871	0.793	0.816	0.361	0.935
Δ_e	0.871	1.000	0.785	0.811	0.461	0.906
Δ_J	0.793	0.785	1.000	0.963	0.511	0.801
Δ_{JS}	0.816	0.811	0.963	1.000	0.533	0.863
Δ_{KL}	0.361	0.461	0.511	0.533	1.000	0.369
Δ_{KS}	0.935	0.906	0.801	0.863	0.369	1.000

Table 6.2(b): Ranking correlation between different perturbation magnitudes (Δ_{CVM} , mean)

Magnitude (%)	1	10	100	Type	Mean	Variance
1	1.000	0.998	0.965	Mean	1.000	0.786
10	0.998	1.000	0.973	Variance	0.786	1.000
100	0.965	0.973	1.000			

Table 6.2(c): Ranking correlation between different perturbation types (Δ_{CVM} , 10%)

distance and different type and magnitude of perturbations on the results of the MDP analysis (i.e. the ranking of importance of molecules) of the case study above is summarized in Table 6.2. Here, the correlations between MDP sensitivity rankings produced by six different density distances (Table 6.2(a)) or by three different perturbation magnitudes (Table 6.2(b)) or by two different types of perturbation (Table 6.2(c)) show that with the exception of the Kullback-Leibler (Δ_{KL}) divergence, the rankings of sensitivities are relatively similar, i.e. molecules that are found to be important by one calculation of MDP analysis are also relatively important by MDP sensitivities computed using different metric or perturbation. The correlation between two CDF distance metrics Δ_{CVM} and Δ_{KS} is expectedly high. Also, PDF-based metrics are generally more sensitive to finite sampling error, as shown in Table 6.3, and for this reason, CDF distance metrics are preferred. In the case study above, Δ_{CVM} was used as the distance metric, since it has better sample-to-sample reproducibility than Δ_{KS} .

6.4.2 Relevance and Application

Predicting and understanding biological behavior are further complicated by cellular stochasticity that arises from genetic and non-genetic factors. As demonstrated in the application to the population model of cell death, the MDP analysis can give insights into how a particular behavior of a cell population is dynamically regulated. Also, similar to the signal tracking analysis using the GFM coefficients, a complementary analysis of the MDP coefficients can reveal how a cellular population responds to changes in the variability of one of its molecules.

Some parts of the MDP analysis can be validated through experiments that perturb the variability of external molecules, for example using a bolus injection. In such an example, the experiment reproduces the mean perturbation in the calculation of MDP coefficients. But, to validate the conclusions obtained from the MDP analysis of the TRAIL induced cell death case study, *in silico* perturbation experiments were performed. In this *in silico* experiment, the mean level of pro-caspase-8 in the cellular population is increased by 100%, at two different times: $\tau = 0$ and $\tau = 2.14$ hours. As predicted by the MDP analysis in Figure 6.6, the perturbation at $\tau = 0$ led to a significant change in the cPARP activation among cells (measured by the change in

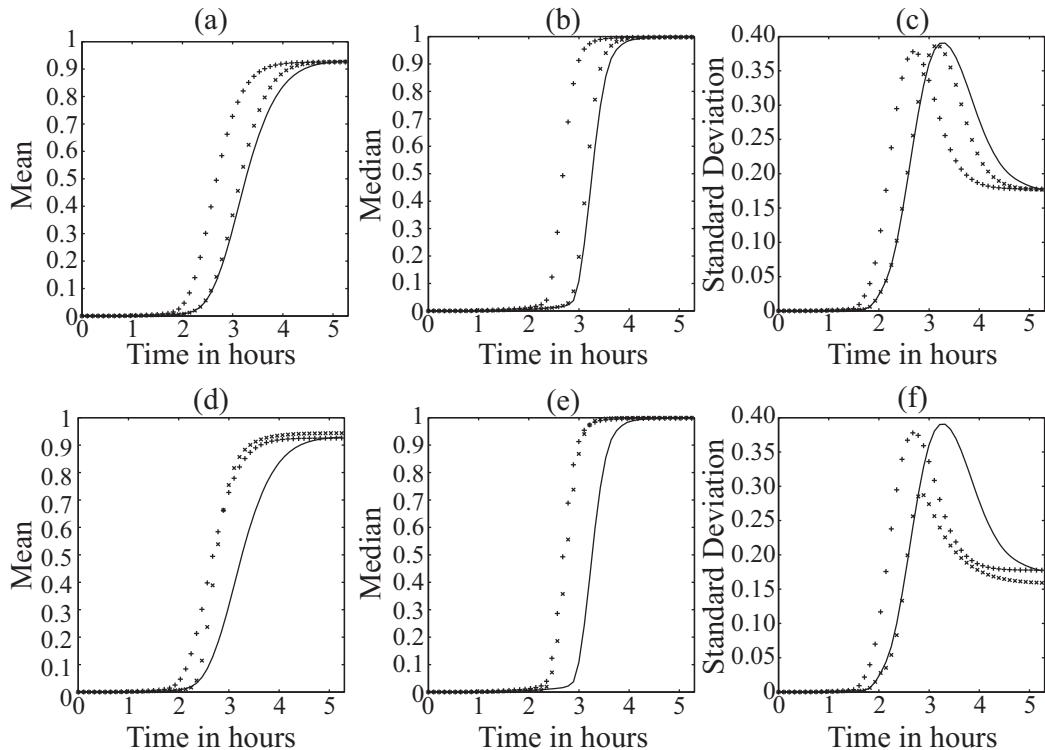


Figure 6.8. *In silico* validation of and hypothesis generation using the MDP Analysis of cPARP activation under constant TRAIL Stimulus. Different characteristic (first column - mean, second column - median and third column - standard deviation) profiles of normalised active cPARP levels for nominal (line) and perturbed profiles (symbols) are compared. (a-c) *In silico* validation experiments for MDP analysis, mean levels of molecular density function of pro-caspase-8 are increased by a fold at both the initial time, $\tau = 0$ and $\tau = 2.14$ hours. The characteristics of normalised active levels of cPARP molecular density functions are plotted for nominal (—), change in pro-caspase-8 at $\tau = 0$ (+), and at $\tau = 2.14$ hours (x). (d-f) Hypothesis generation for the whole cell population, mean levels of initial procaspase-8 ($\tau = 0$ hours) or the available mitochondrial pore opening during switching ($\tau = 2.14$ hours) are increased by a fold. The characteristics of normalised active levels of cPARP molecular density functions are plotted for nominal (—), change in pro-caspase-8 at $\tau = 0$ (+) and change in available mitochondrial pore opening during switching at $\tau = 2.14$ hours (x).

mean, median and standard deviation), while the perturbation at $\tau = 2.14$ hours rendered little change (see Figure 6.8(a-c)).

The consideration of variability of the cell population in sensitivity analysis is important as single cell analysis can be misleading. However, the type of global analysis used also matters. For example, the eFAST and MDP analysis gave different conclusions about the molecules regulating cPARP activation during and post-MOMP (see Figures 6.4 and 6.6). The MDP analysis predicted that mitochondrial pore opening and its subsequent cytochrome-c release to be the most important factors during this period, while the eFAST analysis indicated the initial pro-caspase-8 to be the most important. As shown in Figure 6.8(d-f), an increase of pro-caspase-8 initially (at $\tau = 0$ hour) by one fold and a similar increase in mitochondrial pores during switching ($\tau = 2.14$ hours) lead to equally higher cPARP (mean and median) activation levels in the population. The effect of having higher initial pro-caspase-8 level decreases the switching time in the cPARP activation, but this render a small change in the cPARP variability among the cells and no change in the fraction of apoptotic cells in the population. But, higher number of mitochondrial pores around cPARP switching period leads to much lower variability among cells and importantly, increases the percentage of apoptotic cells from 94.36% to 96.82%. Hence, the eFAST analysis is unable to provide the dynamical importance of molecules due to the timing of the perturbations. This issue is similar to the one raised in Section 2.2, where an analysis based on fixed time perturbation will not be able to describe the dynamical importance of the system components. Extending the eFAST analysis using perturbations at different times will violates the orthogonal pair assumption (i.e., input distributions must be non-correlated), that is needed in the calculation of the SOBOL's sensitivity indices in this analysis (see [132] for details).

6.4.3 Comparison with GFM, iPSA and eFAST analyses

Since MDP analysis is based on perturbations with respect to molecular density functions introduced at multiple time points and on different molecules, it directly provides a molecule-by-molecule illustration of events happening during the functional regulation of system output behaviour under uncertainty. Even though local dynamic sensitivity analysis, like GFM presented in Chapter 3, on a representative single cell model of the cellular population can give such molecule-by-molecule interpretation, it fails to predict the truthful mechanism, as seen in its application to the cell death model in Section 6.3.1. This is because the nature of the perturbation considered here is local and on a single moment (i.e., mean change) of the distribution. Hence, it does not characterize the whole change in distribution.

On the other hand, a more common global parametric sensitivity analysis, like eFAST also fails to predict the population dynamics (as exemplified by the cell death analysis in Section 6.3.2). The reason stems from the fact that the eFAST analysis only quantifies or decomposes the output variability in terms of initial input variability alone. The contribution of intermediate complexes and activated species are neglected, since they are not present during the initial time. Extending eFAST analysis to times other than initial time violates the orthogonal pair assumption (i.e., input distributions must be non-correlated), the basis by which SOBOL's sensitivity indices have been formulated (see [132] for details). Therefore, other than indicating the contribution of input variability to that of the output variability, eFAST analysis cannot give mechanistic insights on the system output behaviour. Due to this, existing global sensitivity analysis may be misleading or not even give any conclusions about the key regulatory molecules and pathways.

6.4.4 Applications of MDP analysis

The results from the MDP analysis can be useful in many applications, some of which have been mentioned previously. For example, the agreement or disagreement between model analysis and experimental observations or biological knowledge can be used for model validation and design of experiments. Furthermore, despite the importance of cell-to-cell variability is increasingly recognized, the extent to which this affects the cell population phenotypic response to drug resistance/sensitivity remains unanswered [124]. Hence, analysis tools for models under uncertainty, like the MDP analysis presented here, can be used in drug discovery research to identify potential molecules for drug targets or drug cocktail to treat cancer cells, in which population variability is an important factor. In this case, the failure to guarantee 100% positive response of the cancer cell population to the drugs can lead to cancer recurrence. By building cell population models and applying population-based analysis, like the MDP, the decision of drug combinations to use and the appropriate dose to use will no longer need to be done by trial and error.

6.5 Summary

In summary, cellular heterogeneity and biological complexity necessitate the use of systems-oriented approach to establish the cause-effect relationship between molecules and their interactions and population phenotype. Based on cell population models and perturbations on the molecular density functions, the MDP analysis offers dynamical insights on the functional regulation in the cellular network, for which either the single cell analysis (like PSA, GFM and iPSA) or standard global sensitivity analysis (like eFAST) may not. In the application to the TRAIL induced cell death model, the MDP analysis predicted that mitochondrial pore opening and cytochrome-

c release to be the important deciding factors for cells in a population to commit to cell death, while the single cell GFM analysis and the eFAST global analysis both indicated other molecules (XIAP and pro-caspase 8). In this case, *in silico* perturbation experiments agreed with the findings from the MDP analysis. Finally, the MDP analysis can guide drug discovery efforts in the identification of potential drug targets, the understanding of drug efficacy and specificity, and the optimization of drug dosing and timing, in which cell-to-cell variability can affect the efficacy of drugs, such as in cancer.

CHAPTER 7

7 Dimensionality Reduction of Large Kinetic Models^{‡‡}

Synopsis:

This chapter presents novel automated model reduction techniques based on dynamic sensitivities of the GFM and iPSA presented in Chapters 3 and 4. The efficacies of the proposed methods are compared with a model reduction algorithm based on the classical PSA, through applications of these methods to produce reduced-order models of the reaction kinetics in alkane pyrolysis, natural gas combustion (GRI Mech 3.0) and industrial steam cracking of ethane.

^{‡‡} Excerpts of this chapter will be part of the following manuscript:

Perumal TM, Krishna SM, Tallam SS and Gunawan R. Reduction of kinetic models using dynamic sensitivities, *Comp. and Chem. Engg.*, (In submission)

7.1 Introduction

Many contemporary world problems, such as the ones discussed in this thesis (understanding biological complexity) or elsewhere (global climate change and energy issues), are in one way or another related to complex chemical processes, ranging from enzyme kinetics to atmospheric reactions to combustion and hydrocarbon processing. The understanding of these processes often necessitates the creation and use of detailed kinetic models that describe the intermediates and reactions in the system. Such mechanistic knowledge is important for the control, design or manipulation of these systems. Nevertheless, a reduced-order model that is valid under certain conditions is often desired for computational reasons (e.g., in model-based optimization and control applications).

A variety of methods exist that can provide reduced order models from the full ordinary differential equation (ODE) kinetic models, which are the most common formalism used to describe chemically reacting systems. The general aim of these methods is to obtain the simplest model with the fewest intermediates and/or reactions, while still retaining the essential features of the detailed model [174]. For example, reduced-order linear ODE models could be obtained using methods from linear systems theory, such as balanced truncation or generalized Grammians [175], for which the upper bound of the H_∞ norm of reduced model error could be computed. However, the applications of linear system methods to complex chemical processes have been limited as the kinetic models typically involve nonlinear rate equations.

For nonlinear models, dimensionality reductions have conventionally been done using the quasi-steady species (QSS) and partial equilibrium assumptions. The reduction is achieved by assuming the radicals and fast reversible reactions to be in equilibrium and by ignoring the slow (unimportant) reactions [176, 177]. Such

reduction techniques however require inputs from an experienced and knowledgeable user. Another common method relies on model sensitivity coefficients to rank the importance of reactions [93, 178-180] as these coefficients reflect the ratio of changes in the system outputs with respect to perturbations in kinetic parameters [94]. In this case, model reduction was performed by eliminating species and reactions that were not affecting the system outputs, as indicated by small sensitivity magnitudes. While this method could potentially be automated, its good performance often requires knowledgeable user guidance as to which reactions to remove [178]. More advanced model reduction methods also exist, including computational singular perturbation (CSP) [181], chemical lumping [182], intrinsic low dimensional manifolds (ILDM) [183], directed relation graph with (DRGEP) or without (DRG) error propagation [184], target factor analysis [185], and flux based methods [186]. Each of the methods has its own advantages and disadvantages, some of which have been discussed elsewhere [187].

In this chapter, five new dynamic sensitivity-based model reduction methods are proposed. The crucial difference between these and existing sensitivity coefficient based methods is the use of time-dependent perturbations in the sensitivity analysis, i.e. the use of Green's function matrix (GFM) analysis from Chapter 3 and the impulse parametric sensitivity analysis (iPSA) from Chapter 4. It was clearly understood from Chapters 2-4 that the dynamical importance of reactions cannot be inferred from the traditional parametric sensitivity coefficients, but is immediately apparent from the iPSA and GFM analysis. Since very large scale curated kinetic models of signaling networks are seldom available and since the reduction analyses presented here are applicable for general ODE models, the efficacies of the proposed methods are compared through the reduction of chemical kinetic models, including

alkane pyrolysis [188], natural gas combustion (GRI Mech 3.0) [189] and a recently published mechanistic model for industrial steam cracking of ethane [60].

7.2 Methods

The most common formalism in the detailed kinetic modeling of chemically reacting systems is the ordinary differential equations (ODEs), as in (1.2). For chemically reacting systems, the right hand side of (1.2), $\mathbf{g}(\mathbf{x}, \mathbf{p})$ can be written as $\mathbf{Nr}(\mathbf{x}, \mathbf{p})$, where the matrix \mathbf{N} denotes the stoichiometric matrix, and the vector-valued function $\mathbf{r}(\mathbf{x}, \mathbf{p}) \in \mathbb{R}^r$ contains the reaction rate equations. The solution to the ODE model gives the time trajectory of the chemical species concentrations $\mathbf{x}(t)$ in the system.

7.2.1 Model reduction using parametric sensitivity analysis (PSA)

One of the many applications of local PSA is model reduction, which is done using a simple algorithm: remove reactions whose rate constants have low sensitivity

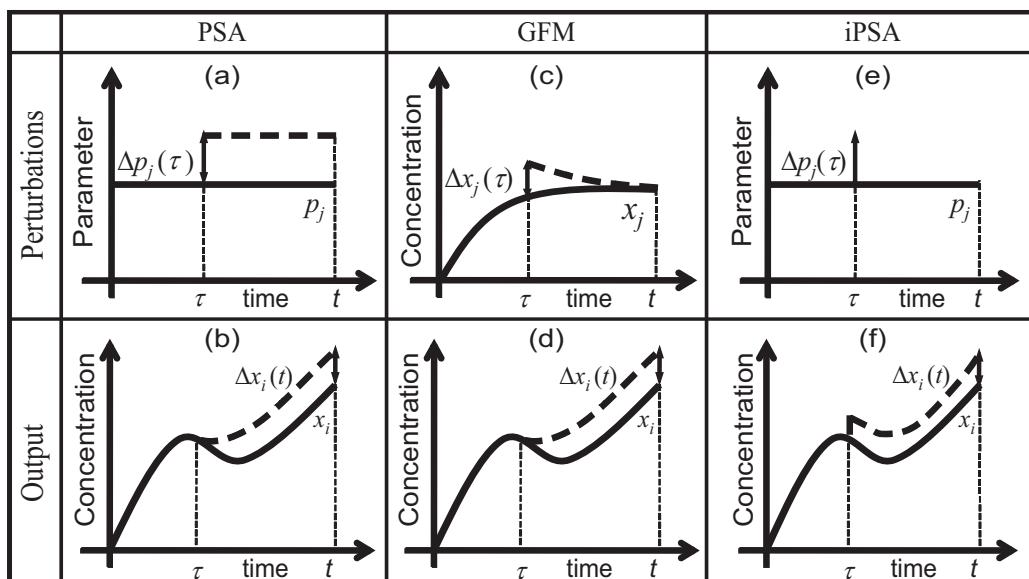


Figure 7.1. Sensitivity analysis. Illustrations of different types of perturbations and their effect on system dynamics in (a-b) the traditional PSA, (c-d) the GFM and (e-f) the iPSA. Solid lines represent the nominal trajectory and the dashed lines show the perturbed trajectory, respectively. Figures are only meant to illustrate the basic concepts and thus are not drawn to scale.

magnitudes, as these are deemed unimportant. When using local PSA sensitivities, parameters are first sorted based on some metrics of importance, such as norm of the sensitivity coefficients in (2.2) (e.g. infinite norm: $\|S_{i,j}\|_\infty = \max_t (\|S_{i,j}(t)\|)$ [99]) or eigenvalues of the Fisher information matrix FIM (i.e., $\text{FIM} = \mathbf{S}^T \mathbf{V}^{-1} \mathbf{S}$, where \mathbf{V} is a weighting matrix [178, 190]). Subsequently, reactions are removed one at a time, starting from the lowest importance metric, until a certain criteria based on model reduction error (i.e. difference between the full and reduced model), model size or some combinations of the two, is optimized.

It has been discussed in detail in Section 2.2 that the parameter perturbations used in the local PSA are persistent in nature. Hence, the information obtained from the sensitivity coefficients above reflects the integrated changes in system output that resulted due to a step change in the parameter values at the initial time, as depicted in Figure 7.1(a-b). As a result, any parametric perturbations which cause large changes in the species concentrations at early times may generate artificially high sensitivity ranking throughout the simulation period, while key reactions participating only at later times, maybe incorrectly ranked low in sensitivity (see Section 2.2.2 for detail). Consequently, a model reduction procedure based on such analysis will favor towards preserving reactions of early importance in the model, while mistakenly removing reactions with later significance.

7.2.2 Model reduction using dynamic sensitivity analysis (iPSA and GFM)

As the caveat of PSA is rooted from the persistent nature of the parameter perturbations, Chapters 3 and 4 have presented two (local) sensitivity analyses, namely the Green's function matrix (GFM) analysis and impulse parametric

sensitivity analysis (iPSA). In the iPSA, perturbations are introduced as impulses on system parameters at different initial times (see Figure 7.1(c-d)). Meanwhile, the GFM relies on time-dependent perturbations on species concentrations (see Figure 7.1(e-f)). By using time-dependent perturbations, the associated sensitivity coefficients can directly reveal the dynamical importance of chemical species and reactions, without the integrated effect as in the traditional PSA. Thus, these sensitivity analyses not only give answers to which species and parameters are important, but also when they matter.

Following the model reduction procedure in the traditional PSA, the dynamical importance of species and parameters are first sorted based on some metrics of the sensitivity coefficients in (3.2) and (4.8). In the model reduction using the GFM and iPSA below, infinite norm has been used as the metric for ranking, i.e.

for the GFM: $\|S_{i,j}^x\|_\infty = \max_{t,\tau} (|S_{i,j}^x(t, \tau)|)$ and for the iPSA: $\|iS_{i,j}\|_\infty = \max_{t,\tau} (|iS_{i,j}(t, \tau)|)$.

Species (parameters) having the same sensitivity metrics are further sorted based on the number of reactions (species) that they are associated with. Starting from the lowest sensitivities, the model reduction is carried out by sequentially removing species (in the GFM based method) or reactions (in the iPSA based method) of the model. In the GFM based reduction, deleting a species from the model means removing the corresponding differential equation from the model and all rate equations associated with reactions in which this species acts as a substrate.

There exists a trade-off between reducing models reaction-wise and species-wise. When the number of reactions far exceeds the number of species, as typically encountered in complex chemical processes, the elimination of reactions guided by parametric sensitivity ranking generally produces a finer gradation of reduced order models and of model reduction error, in comparison with the elimination of species by

the GFM reduction. On the other hand, species-wise removal will outpace reaction-wise method in terms of model dimensionality reduction, either measured by the number of remaining species or reactions. Naturally, if the number of species exceeds that of reactions, then the reverse circumstances apply.

7.2.3 Combined methods for model reduction

The iPSA and GFM analyses offer different, but complementary, information regarding the dynamical importance of reaction pathways. By combining the two analyses, three combined model reduction strategies have been formulated in this chapter. The proposed methods follow reaction-wise reduction approach as this will provide more refined control over model reduction error in a typical complex chemical reaction network, as discussed above. The strategies are designed to be fully automated, but at the same time, to allow some flexibility to accept user inputs, such as maximum tolerable error or model size. Existing model reduction methods that combine the sensitivities with respect to parameters and initial concentration of species typically follow a two-stage procedure. The first stage involves the removal of species using information from the initial condition sensitivities (i.e. the GFM coefficients with $\tau = 0$) or other methods such as the DRGEP, which is then followed by the elimination of reactions using the traditional PSA, PCA (principal component analysis), or CSP methods [178, 191]. The transition between the first and second stage is usually not known *a priori*, and a user specification, for example on the reduced model error, is required (e.g. remove species until before the model reduction error exceeds the specification, then continue with the removal of reactions) [178]. To become fully automated, the proposed algorithms here rely on the Pareto optimization of model reduction, using criteria such as the extent and the error of model reduction.

In these algorithms, the parameter and species perturbation sensitivities are combined to produce new ways to rank reactions. In particular, new rankings are generated by first creating a 2-dimensional plot of the permuted stoichiometric matrix, where the rows and columns are sorted according to the GFM and iPSA sensitivity metrics, respectively, from low to high (see Figure 7.2). For each reaction (or parameter), the maximum distance to the origin (lower left corner) or the minimum distance to the highest sensitivity rank (upper right corner) are computed, from which combined reaction rankings and algorithms for obtaining reduced order models are developed (see Algorithms 1 and 2 in Section 7.2.3.1 and 7.2.3.2). Alternatively, the reaction-wise removal of parameters could also be done sequentially according to the

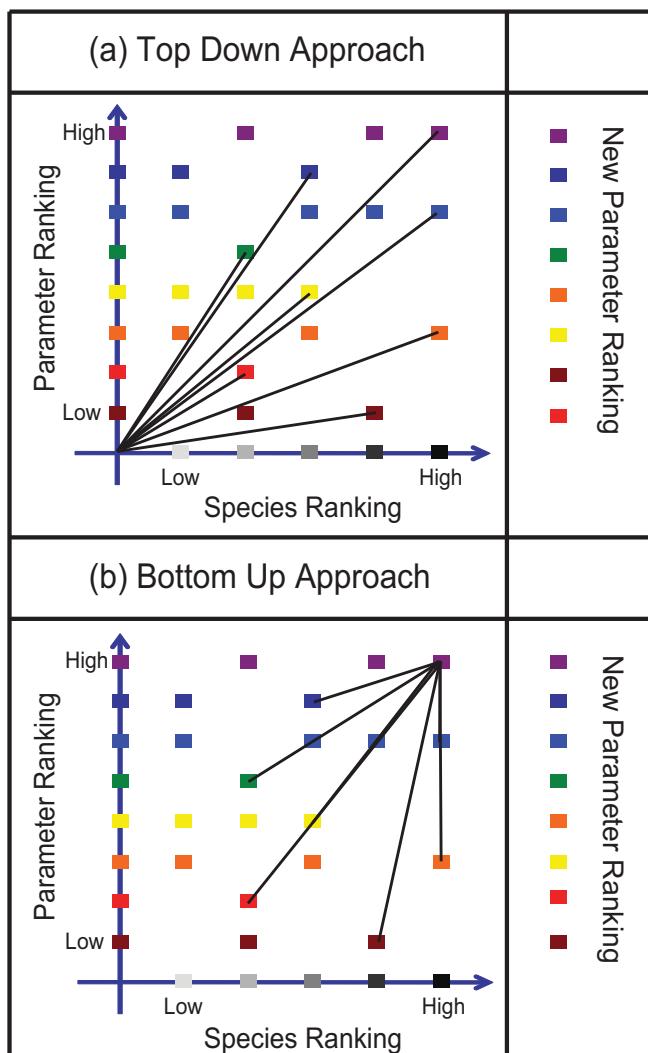


Figure 7.2. Top – Down (TD) and Bottom – Up (BU) algorithms. Abscissa represents species sensitivity rankings (low to high) and ordinate represents parameter rankings (low to high).

species ranking by the GFM sensitivities (see Algorithm 3 in Section 7.2.3.3). If the number of species becomes larger than the number of reactions, it is also possible to generate rankings for the importance of chemical species from Figure 7.2. In such a case, model reduction could be done in a similar fashion as the three algorithms below, except using a species-wise removal.

7.2.3.1 Algorithm 1 (Top – Down Approach, iPSA-GFM TD)

In essence, the first algorithm makes use of the ranking obtained from the maximum distance to the origin in Figure 7.2(a) to order reactions based on the potential direct and indirect error of removing a single reaction. The direct error here is gauged by the parametric sensitivities (iPSA), while indirect error is in relation to changes in the associated species concentrations and is incorporated by considering the GFM sensitivities. In the first algorithm, a top-down model reduction approach was taken by removing one reaction at a time, starting from the parameter with the lowest maximum distance. In this case, a species can be removed indirectly from the model when the respective ODE has a zero right hand side. The best reduced model is determined from the Pareto optimal front between the root mean square relative error (RMSRE) of and the fraction of remaining reactions in the reduced model.

7.2.3.2 Algorithm 2 (Bottom – Up Approach, iPSA-GFM BU)

The second combined sensitivity rankings using the minimum distance to the highest sensitivity rank reflects the potential benefit of including a particular reaction. Consequently, the second algorithm takes a bottom-up approach, in which reduced order models are created by including reactions (and indirectly chemical species) into the model. Again, the best reduced order model is determined using the same Pareto optimal front as in the first algorithm.

7.2.3.3 Algorithm 3 (Sequential Approach, iPSA-GFM SQ)

In the third algorithm, the reduction is done using sequential Pareto optimizations in a sequential manner. Starting from the species with the lowest GFM sensitivity metric, a set of reduced order models are generated by removing reactions sequentially according to the iPSA sensitivity metrics. The Pareto optimal reduced-order model is then determined and this model becomes the starting point for the subsequent reduction step using the next species of the GFM ranking. The procedure is repeated and the last Pareto optimal model is taken to be the best reduced model.

7.3 Examples

The efficacy of the new methods (iPSA, GFM, iPSA-GFM TD, iPSA-GFM BU, and iPSA-GFM SQ) are first compared with that of the traditional PSA based model reduction by applying these reduction methods to the models of : (i) alkane pyrolysis [188], (ii) natural gas combustion (GRI Mech 3.0) [189], and (iii) an industrial ethane pyrolysis [60]. The performance criteria of the reduction techniques are based on (i) the model reduction error, calculated as the root mean square of the relative error (RMSRE) between the reduced and the full model simulation, and (ii) the number of reactions remaining, calculated as the fraction of reactions that remain in the reduced model. Particularly, the Pareto optimal point of both the performance metrics, i.e. the minimum Euclidean distance to the origin, is used in the final comparison.

7.3.1 Alkane pyrolysis model

The low-temperature alkane pyrolysis model consists of 38 species and 98 irreversible reactions [188]. In the model reduction below, only six species C₃H₈, H₂, CH₄, C₃H₆, C₂H₄ and C₂H₆ trajectories, as shown in Figure 7.3, were considered as the main

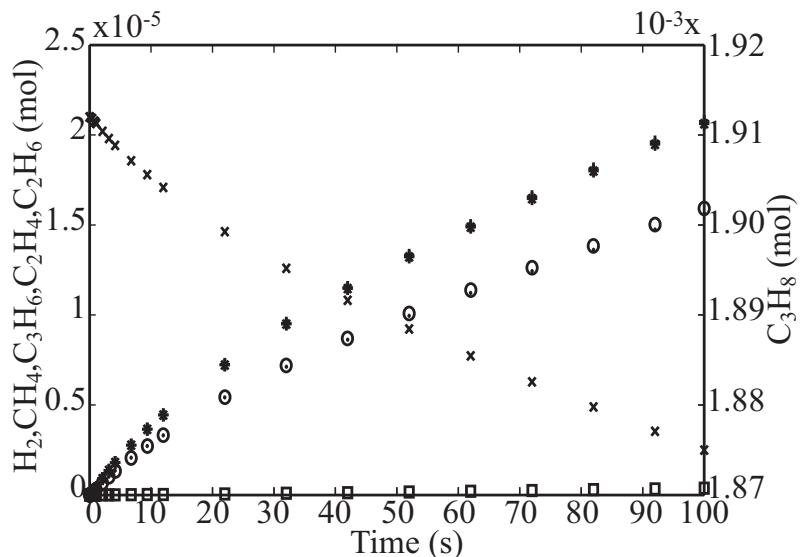


Figure 7.3. Low temperature alkane pyrolysis model. Time profiles of C_3H_8 (\times) in the right ordinates and H_2 ($+$), CH_4 (\circ), C_3H_6 ($*$), C_2H_4 (\bullet) and C_2H_6 (\square) in the left ordinates. The reactions are simulated at 817.16K and initiated only with 1.912×10^{-3} mol dm $^{-3}$ of propane. The simulation was carried out for 100 seconds to reach 2% propane conversion, for which the original model was validated.

model outputs and used in the calculation of sensitivities and model reduction errors.

The performance of individual model reduction techniques based on the PSA, iPSA and GFM are compared in Figure 7.4(a), while Table 7.1(a) summarizes the Pareto optimal reduced model from each of these methods (see Table F.1 for more detailed results). The PSA and iPSA coefficients were computed with respect to the rate constants. By comparing the Pareto optimal distances to the origin in Table 7.1(a), the

Table 7.1: Reduction of Alkane Pyrolysis Model using PSA, iPSA and GFM

Methods*	Species Remaining	Reactions Remaining	RMSRE (%)	Pareto Optimal Distance
(a): Pareto Optimal Model				
iPSA	15	23	11.860	0.263
PSA	18	33	7.558	0.345
GFM	29	63	10.822	0.652
(b): Fixed Reactions (38)				
iPSA	19	35	1.613	0.358
PSA	21	38	3.181	0.389
GFM	24	34	76.940	0.844
(c): Fixed Error (0.5%)				
iPSA	21	43	0.381	0.439
PSA	24	55	0.469	0.561
GFM	30	81	0.175	0.827

* Sorted based on the optimal Pareto distance

model reduction based on the iPSA provided the best balance between model reduction error and model dimensionality. Meanwhile, the GFM based reduction outperformed the other two methods when the desired number of species was specified (see Table F.1(d) for results). Such an advantage is not surprising considering that the GFM sensitivities directly relate to the dynamical importance of species and thus reflect the impact of removing a chemical species from the model. Taken together, the results of the performance comparison above demonstrate the benefit of using time-dependent sensitivities over the traditional PSA in model reduction.

The automated reduction algorithms in this example were also compared with a manual reduction of the same model, which resulted in a reduced model with 38 reactions and a relative error of 0.5% [178]. In this case, new Pareto optimal reduced-order models were obtained from the Pareto frontiers in Figure 7.4(a) by setting an upper bound on either the reduction error (0.5%) or the total number of remaining reactions (38), as summarized in Table 7.1(b-c). The iPSA model reduction still gave a better Pareto optimal distance than the PSA and GFM. Importantly, without any user guidance, the iPSA model reduction was able to produce reduced order models that are comparable to the reduction done by an expert user [178].

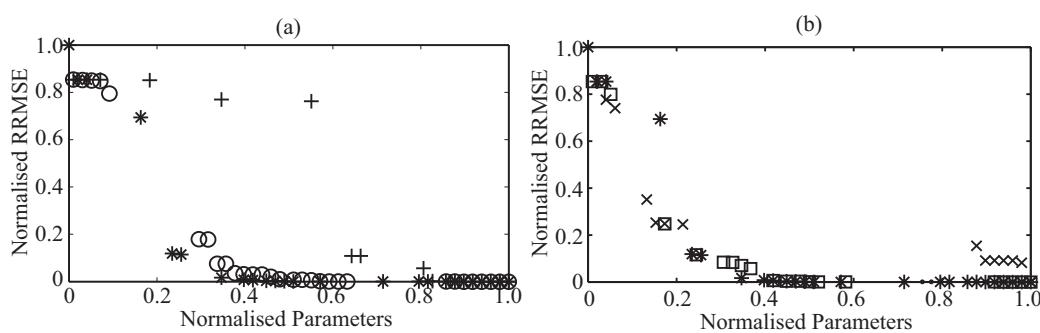


Figure 7.4. Reduction of low temperature alkane pyrolysis model. (a-b) the Pareto frontiers in terms of normalized relative root mean square error and normalized model dimension (number of reactions). (a) Pareto frontiers of the PSA (\circ), iPSA (*) and GFM (+) based individual methods. (b) Pareto frontiers of the iPSA (*), iPSA-GFM BU (\bullet), iPSA-GFM TD (\square) and iPSA-GFM SW (\times) based combined methods.

Table 7.2:Reduction of Alkane Pyrolysis Model using Combined iPSA and GFM

Methods*	Species Remaining	Reactions Remaining	RMSRE (%)	Pareto Optimal Distance
(a): Pareto Optimal Model				
iPSA-GFM (SQ)	14	16	15.820	0.227
iPSA	15	23	11.860	0.263
iPSA-GFM (BU)	15	23	11.860	0.263
iPSA-GFM (TD)	15	23	11.860	0.263
(b): Fixed Reactions (38)				
iPSA	19	35	1.613	0.358
iPSA-GFM (BU)	19	35	1.613	0.358
iPSA-GFM (TD)	19	36	1.777	0.368
iPSA-GFM (SQ)	17	37	10.160	0.391
(c): Fixed Error (0.5%)				
iPSA	21	43	0.381	0.439
iPSA-GFM (BU)	21	43	0.381	0.439
iPSA-GFM (TD)	21	43	0.381	0.439
iPSA-GFM (SQ)	20	54	0.402	0.551

* Sorted based on the optimal Pareto distance

The performances of the three combined reduction methods were subsequently evaluated against that of the iPSA reduction, as illustrated by the Pareto frontiers in Figure 7.4(b). Table 7.2 suggests that the combined methods generally perform as well as or better than the iPSA model reduction and that the sequential approach (iPSA-GFM SQ) approach gives the best Pareto optimal model reduction (also see Table F.1 for more detailed comparison). Thus, the use of combined dynamical information from the iPSA and GFM sensitivities can provide some advantage over using individual sensitivities, but off course at the cost of increased computational effort.

7.3.2 Natural gas combustion – GRI-Mech 3.0

The GRI-Mech 3.0 is a detailed kinetic model of isobaric-isothermal natural gas combustion with NO formation and reburn chemistry [189]. The model has 53 species and 650 irreversible (or 325 reversible) reactions. In this case, the reduction was done to obtain a reduced order kinetic model for NO emission during methane oxidation at

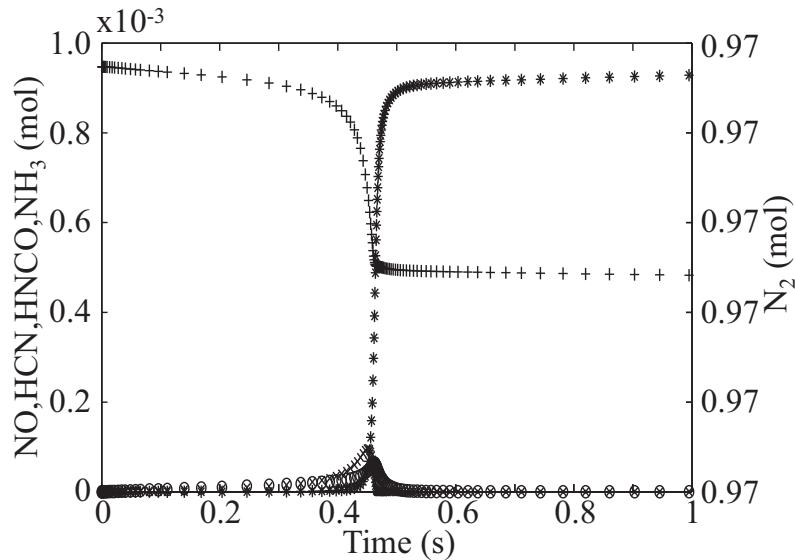


Figure 7.5. Isobaric-Isothermal natural gas combustion (GRI Mech 3.0) model. Simulation of Nitrogen – N_2 (\times), NO (+), HCN (\circ), HNCO (*) and NH_3 (\bullet) profiles at 1300K and 1atm with a feed mixture of CH_4 , C_2H_6 , O_2 , NO, H_2O and N_2 , of 2864 ppm, 5090 ppm, 947 ppm, 2.16%, 96.92%, respectively. The simulation was run until CH_4 was exhausted (0.8 seconds).

a constant pressure and temperature. Nitrogen containing species, including NO, HNCO, HCN, N_2 and NH_3 were here considered as the main products, whose time trajectories are shown in Figure 7.5. The PSA and iPSA sensitivities were computed with respect to the rate constants.

Figure 7.6 shows the Pareto frontiers from each model reduction algorithm based on the sensitivities of the five nitrogen containing chemical species mentioned above, while Table 7.3 summarizes the Pareto optimal reduced order model from each method (see Table F.2 for more detailed results). Again, the iPSA reduction

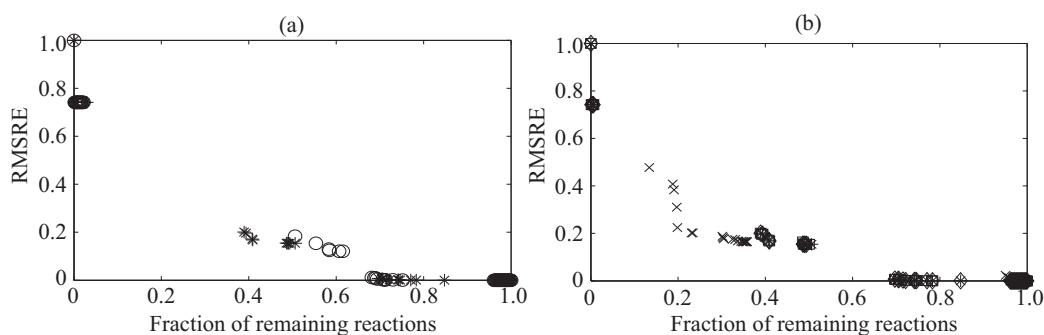


Figure 7.6. Reduction of isobaric-isothermal natural gas combustion (GRI Mech 3.0) model. (a-b) the Pareto frontiers in terms of normalized relative root mean square error and normalized model dimension (number of reactions). (a) Pareto frontiers of the PSA (\circ), iPSA (*) and GFM (+) based individual methods. (b) Pareto frontiers of the iPSA (*), iPSA-GFM BU (\bullet), iPSA-GFM TD (\square) and iPSA-GFM SQ (\times) based combined methods.

Table 7.3: Model Reduction of Natural Gas Combustion - GRI Mech 3.0 Model

Methods*	Species Remaining	Reactions Remaining	RMSRE (%)	Pareto Optimal Distance
iPSA-GFM (SQ)	39	129	22.462	0.300
iPSA	49	253	20.074	0.438
iPSA-GFM (BU)	49	253	20.074	0.438
iPSA-GFM (TD)	49	253	20.074	0.438
PSA	51	329	18.332	0.538
GFM	3	1	74.161	0.742

* Sorted based on the optimal Pareto distance

performed better than the PSA and GFM based methods and the combined methods had a slight advantage over the iPSA alone. Likewise, the combined sequential approach (iPSA-GFM SQ) gave the best overall reduced order model (i.e., the one with the smallest Pareto optimal distance).

7.3.3 *Ab initio* kinetic model of industrial ethane pyrolysis

In a recent publication, Sun and Sayes proposed a first-principle kinetic model of an industrial steam cracking of ethane [60]. The reaction network consists of 20 chemical species with 10 intermediate radicals and 150 irreversible elementary reactions. In this case, heavy hydrocarbons larger than C₄ have been neglected. Thermodynamic and kinetic parameters of the reactions were obtained from first principle calculations. The

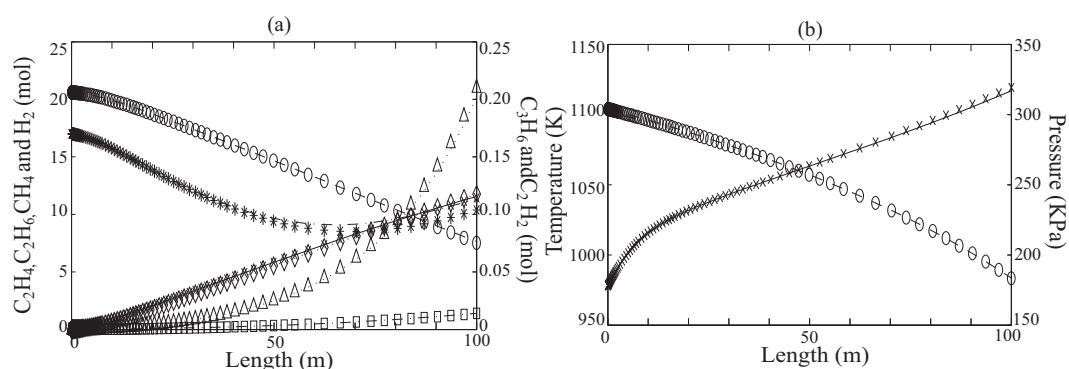


Figure 7.7. Industrial steam cracking of ethane at steady state in a PFR. (a) Concentration profiles of C₂H₄ (—/×), C₂H₆ (---/○), CH₄ (—/□) and H₂ (···/◊) in the left y-axis and C₃H₆ (—/*) and C₂H₂ (···/Δ) in the right y-axis, (b) Temperature (—/×) profile in left y-axis and the pressure (---/○) profile in the right y-axis, as predicted by the full model (lines) and the reduced model (symbols) at steady state along the length of the PFR. Model simulations were performed at steady state in a PFR of length 100 m and cross-section 0.009161 m², with an inlet feed mixture consisting of 21.53 moles of C₂H₆, 0.22 moles of C₂H₄ and 0.18 moles of C₃H₆, at 1000K and 3atm. The end of reactor length corresponds to 60% conversion of C₂H₆.

Table 7.4: Reduction of Ethane Cracking Model

Methods*	Species Remaining	Reactions Remaining	RMSRE (%)	Pareto Optimal Distance
(a): Pareto Optimal Model				
iPSA-GFM (SW)	11	15	17.532	0.202
iPSA	19	36	10.140	0.261
iPSA-GFM (BU)	19	36	10.140	0.261
iPSA-GFM (TD)	19	36	10.140	0.261
PSA	19	28	18.308	0.261
GFM	18	46	40.411	0.507
(b): Fixed Error (1%)				
iPSA	20	67	0.838	0.447
iPSA-GFM (BU)	20	67	0.838	0.447
iPSA-GFM (TD)	20	67	0.838	0.447
PSA	20	79	0.413	0.527
iPSA-GFM (SW)	20	134	0.887	0.893
GFM	20	150	0.000	1.000

* Sorted based on optimum Pareto distance

model prediction of C₂H₆, C₂H₄ and H₂ yields were within 5% deviation of the experimental data generated from a plug flow reactor (PFR) at steady state [60].

In this example, reduced order kinetic models of the ethane pyrolysis process were obtained based on the sensitivities of six chemical species of interest: C₂H₆, C₂H₄, C₃H₆, CH₄, H₂ and C₂H₂, at steady state along the length of the PFR, whose concentrations are shown in Figure 7.7(a). The iPSA and PSA coefficients were computed with respect to perturbations in the pre-exponential factors of the rate constants. The iPSA-GFM SQ algorithm again gave the best Pareto optimal reduced order model, as shown in Table 7.4a. However, if the RMSRE among the six species mentioned above was desired to be less than 1%, the iPSA method gave the best Pareto optimal model (see Table 7.4b), reducing the full model from 150 reactions to 67 reactions, with a RMSRE of only 0.838%. Table 7.5 further shows the percent removal of reactions associated with each species in the reduced model. Importantly, the model reduction reveals that reactions involving higher carbons, such as C₄H₁₀, C₃H₈, C₄H₉ radicals can be removed without much change in the model prediction (see Table F.3 for more detailed list). As demonstrated in Figure 7.7, the iPSA

Table 7.5: Reduction of Ethane Cracking Model

Species Name*	Before Reduction	After Reduction	% Reduction
C₄H₁₀	40	5	87.50
C₃H₈	38	7	81.58
·C₄H₉	24	5	79.17
CH₄	20	5	75.00
·C₄H₇	24	6	75.00
C₄H₈	28	7	75.00
·C₃H₇	26	7	73.08
·2-C₃H₇	22	8	63.64
·2-C₄H₉	24	9	62.50
·CH₃	36	16	55.56
H₂	20	9	55.00
C₄H₆	4	2	50.00
·H	52	27	48.08
·C₂H₃	30	17	43.33
C₂H₄	28	18	35.71
·C₂H₅	30	20	33.33
·C₃H₅	24	16	33.33
C₃H₆	28	19	32.14
C₂H₆	22	15	31.82
C₂H₂	4	4	0.00

* Sorted based on percentage change

reduced model is able to recapitulate the concentration profiles of the product species as well as the temperature and pressure profiles along the reactor length, at steady state.

7.4 Summary

Reduced order models of complex chemical kinetics are often desired for many reasons. While the traditional sensitivity analysis has been commonly used for this purpose, the reduction often requires supervision from experts to produce a good reduced order model. In this chapter, five automated model reduction algorithms are introduced based on previously developed time-varying species or impulse parametric perturbation sensitivities, named as the GFM and the iPSA sensitivity coefficients, as seen in Chapters 3 and 4, respectively. The use of impulse and species perturbations provided direct dynamical information regarding the degree and timing of the

importance of reactions and species, which was not straightforwardly available from the traditional PSA. Applications to three large kinetic models, including the GRI Mech 3.0 with 650 reactions, demonstrated the advantage of using impulse and species sensitivities for model reduction over the classical PSA. In general, the Pareto optimal reduction performance can be achieved by combining both parameter and species perturbation sensitivities using the iPSA-GFM SQ method. However, when a small relative reduction error is desired, the iPSA based reduction method usually produces the smallest model dimension. For example, the application of the iPSA reduction on a model of industrial ethane pyrolysis led to a reduction from 150 to 67 reactions with a relative error of <1%.

CHAPTER 8

8 Thesis Outlook

Synopsis:

This chapter summarizes the major outcomes of the present dissertation and discusses the implications for future research. In particular, the needs for dynamic sensitivity analyses are explained and the specific contributions of this dissertation to meet this need are presented. In addition, this chapter also presents the major applications and discusses the limitations of the sensitivity analyses presented here. Finally, future directions, in both theoretical and application areas, for the improvement of the current dynamic sensitivity analyses are presented.

8.1 Dynamical Sensitivity Analyses of Kinetic Models in Biology

8.1.1 Motivation

Through millennia of evolution, cells have developed intricate networks of signaling, gene regulation, and metabolism to accomplish their myriad functions under significant intrinsic and extrinsic uncertainties. Reductionist approach in the genomic era has uncovered significant amount of information about the cellular components and their interactions in these networks. However, as seen in Chapter 1, post genomic era has given rise to new set of challenges, such as in understanding how a cellular behaviour or function is accomplished by large complex networks. These challenges are addressed using system level understanding of biology, in which quantitative (mathematical) representations of cellular networks are constructed and then analyzed to discover the basis of cellular phenotypes, to predict the effect of network changes on cellular behavior, and to ultimately design networks with new and desired properties and functions.

As shown in Section 1.2.1, there exist many modeling paradigms and as many, if not more, systems analysis tools that can be used to understand the underlying cellular mechanisms. Each of these methods has their own advantages and limitations. In order to ascertain the importance and role of various cellular processes, most of these methods investigate the effects of perturbations on model parameters and/or on system inputs in the system output behaviour. As these methods focus solely on model parameters, including initial conditions, its suitability to understand the dynamical aspects of cellular regulation has not been addressed to date. As seen in

Section 2.2, with a rigorous theoretical formalism and a synthetic network model, this dissertation shows the caveat of using classical sensitivity analysis in obtaining the mechanistic understanding of a dynamical system. The reason mainly stems from the fact that the perturbations effected are on system parameters that have no dynamics and are static, i.e., introduced only at the initial time point. Hence, providing the time integrated effect of perturbation-output relationship. Therefore, a dynamic perturbation framework is needed to understand the underlying mechanisms, which gives rise to the observed system behavior.

8.1.2 Sensitivity analyses

To avoid the above said caveat of the classical PSA in understanding system dynamics, this dissertation presents four novel dynamic sensitivity analyses. These analyses are the Green's function matrix (GFM) analysis, the impulse parametric sensitivity analysis (iPSA), the pathway parametric sensitivity analysis (pathPSA) and for cellular population models, the molecular density perturbation (MDP) analysis. While the GFM analysis, iPSA and pathPSA are used to study deterministic ODE models, MDP analysis is applied to understand the dynamics of probabilistic models in systems biology. However, irrespective of the modeling paradigms, these analyses are built on a cause-effect relationship, in which time-varying perturbations are introduced to a system parameter and the resulting change in output behavior is measured. Based on the resulting perturbation-effect ratio, these sensitivity coefficients offers step-by-step dynamical insights on the functional regulations and in some cases signal propagation in the cellular network.

Specifically, the GFM analysis (Chapter 3) and iPSA (Chapter 4) dynamically perturb the molecular concentrations and system parameters, respectively. In contrast to the perturbations in the GFM analysis, the iPSA makes use of impulse perturbations to produce the necessary information. Based on the resulting sensitivity coefficients, the dynamical state-by-state and parameter-by-parameter dependence of system behaviour are mapped out. On the other hand, the pathPSA (Chapter 5) decomposes system structure into functional pathways and introduces time-varying dynamic perturbations on the pathways. Analogous to the GFM analysis for ODE models, the MDP analysis (seen in Chapter 6) introduces time-varying perturbations on the molecular density functions of probabilistic models, revealing the dynamical state dependence of system behavior under uncertainty. Not only the analyses reveal which states, parameters or pathways are critical, the dynamic sensitivities presented in this dissertation also indicate when they become important.

The efficacies of these analyses are demonstrated through ODE and population models of synthetic networks and apoptotic cell death. The analyses presented above can guide the drug discovery efforts in the identification of potential drug targets, the understanding of drug efficacy and specificity, in understanding the mechanism of drug action and in the optimization of drug dosing and timing. The development of these methods represents a concrete step towards robustness-based drug design through systems biology.

8.1.3 Model reduction

In the above said examples, dynamic sensitivities are used to obtain the step-by-step mechanistic insights and to identify the important states, parameters or pathways, which give rise to the observed system behavior. Other than understanding system

dynamics and robustness-fragility, dynamic sensitivities are also used in many applications in systems biology, ranging from model reduction/refinement to model validation/analysis and the design of experiments. One application of the dynamic sensitivities for dimensionality reduction of large reaction kinetic models is presented in Chapter 7. Here, five automated model reduction algorithms are introduced based on previously developed the GFM and iPSA sensitivities. The use of species and impulse perturbations provides direct dynamical information regarding the degree and timing of the importance of species and reactions, which are not available from the classical PSA. In addition, a better reduction performance could be achieved by combining both parameter and species perturbation sensitivities.

8.1.4 Contributions

The specific contributions of this dissertation include:

1. an illustration of the caveat of using classical PSA to understand the system dynamics and to obtain the mechanistic details of dynamic ODE models of biological systems.
2. three novel dynamic sensitivity analysis methods based on perturbations on molecular concentrations (GFM), system parameters (iPSA), and system structure (pathPSA), for analyzing the dynamics of deterministic ODE models in systems biology.
3. a novel dynamic sensitivity analysis method based on perturbations on molecular density functions to analyze the dynamics of cellular population models in systems biology.
4. five new reduction algorithms based on dynamic GFM and iPSA sensitivities to reduce models of detailed chemical kinetics.

8.1.5 Advantages and limitations

8.1.5.1 Sensitivity analyses for deterministic models

Results of the dynamic sensitivities used to analyse ODE models, i.e., GFM, iPSA and pathPSA, depend on the choice of the model parameters as well as the initial conditions. That is, these analyses introduce infinitesimal perturbations around a fixed (nominal value) local point of the system. Hence provide only with the local linear information about the cellular network at a given operating conditions. To account for cellular and parametric uncertainty, multiple local dynamic sensitivity analysis can be performed along the system trajectory. Such global analysis should consider a region in the parameter or initial state space and improves upon the local analysis that focuses only at one choice of parameter set or initial conditions.

Computational Efficiency: As seen from Sections 4.2 and 5.2, both iPSA and pathPSA coefficients are calculated from the GFM coefficients by simple matrix multiplication with the non-homogeneous terms, which is computationally inexpensive. Therefore, the major computational efforts of all the local methods are based on the calculation of GFM. From Section 3.2, calculation of GFM coefficients for different τ 's is computationally inexpensive by taking advantage of the semigroup property. In this case, one only needs to solve for the GFM coefficients for one time step $\Delta\tau$ from each τ_k , while the remaining GFM for different (t, τ) pairs are calculated using simple, inexpensive matrix multiplications. Therefore, all the local methods are computationally efficient. For example, the application of dynamic sensitivities to the Fas-induced apoptosis model with 28 species and 32 reversible reactions took less

than 10 minutes using a computer workstation with dual-core CPU (Intel 6300 @ 1.86GHz) and 3GB RAM.

Scalability: Most of the systems in biology are modularized and hence for such modular systems, all the local dynamic sensitivity analyses presented in this thesis are scalable. Since all the local dynamic sensitivities are based on GFM and in turn as seen from Section 3.2 that the GFM analysis of the overall network can be reconstructed from the analysis of individual subsystems, the local dynamic sensitivities of the systems are scalable. Unfortunately, if there exist any feedback mechanisms from downstream molecules, forming loops, such scalability are no longer possible.

8.1.5.2 Sensitivity analysis for probabilistic models

In comparison to the local methods, the MDP analysis presented in Chapter 6 is a global analysis method. The result from such global analysis should not depend upon the nominal values of the system as the states or parameters of the system are considered as random variables, described by probability density functions (e.g. uniform or Gaussian).

Variability of Metrics: The sensitivity metrics in the MDP analysis, like any other Monte Carlo methods, are affected by noise due to the use of finite samples in approximating the probability density function. As seen in Section 6.4.1, even though this variability is prominently seen in the absolute values of the global sensitivity coefficients, the importance ranking is preserved and hence the end conclusion from such ranking is still reliable.

Computational efficiency: As seen in Section 6.2, solving global methods requires simulating the ODE model in (1.2) a large number of times. For example, for an uncertain system of n states with s sample size and t time points, the MDP analysis for a single output state requires $2 \times n \times t \times s$ evaluations of the nominal model. Hence, the numerical calculations of global sensitivity coefficients are computationally expensive. For the TRAIL-induced apoptosis model presented in Section 6.3.3 with 52 molecular species and 70 parameters, the MDP analysis with 1000 samples and for 100 time points took roughly 52 hours using MATLAB distributed computing engine with 16 computing nodes, each having a single core CPU (Intel 6300 @ 1.86GHz) and 1GB of RAM.

Higher order sensitivities: Both the local and global analyses presented in this dissertation are first order sensitivities. i.e., the perturbations on system states, parameters or pathways are effected one at a time (OAT). Second and higher order sensitivities will involve simultaneous perturbations on states or parameters. Even though the higher order extensions of the current dynamic sensitivities are possible, they are less commonly computed and hence the focus of the current work is only on the first-order sensitivities and the higher order sensitivities are of future interest. Nevertheless, the pathPSA can analogously be compared to multi-perturbation approach, since perturbing a single pathway means perturbing multiple reactions simultaneously.

8.2 Recommendations for Future Work

Following the developments and observations made during the course of this research work, recommendations for future work, both in the area of theoretical development

and application of the existing methods, are shown in Figure 8.1 and are outlined below.

Theoretical development: The immediate development in the current theoretical framework is to extend the local iPSA and pathPSA to account for cellular variability, i.e. translating these into global analysis. The resulting global analysis can mimic the MDP analysis in the way perturbations are introduced on the system (i.e. instead of molecular perturbations, the analyses will introduce pulse perturbations on the parametric distributions).

Though the dynamic sensitivities presented here investigates only the first order sensitivities (one at a time), second and higher order sensitivity (combinatorial perturbation) approach is of future interest. Such an approach is definitely expected to shed more light on the underlying system mechanisms which gives rise to the observed behaviour. The only problem that can be immediately foreseen is the high computational cost in doing such calculations.

Existence of summation theorem: As seen in Chapter 3, there exist similarities between the local dynamic sensitivities (the GFM, iPSA and pathPSA) and the metabolic control coefficients. Mathematically, the metabolic control coefficients are equivalent to the classical PSA. Hence, the difference between control coefficients and local dynamic sensitivities are the same as with the classical PSA coefficients. It has already been shown in Section 3.4.2 that the summation theorem of control analysis does not generalize to the GFM analysis, whereas, the possibility of such a summation theorem for iPSA and pathPSA is of greater interest to pursue in the near

future. One such summation theorem is expected to shed lights on the underlying biological principles.

Mathematical accuracy and Computational efficiency: Even though the calculations of local dynamic sensitivities (the GFM, iPSA and pathPSA) presented in Chapters 3-5 are computationally efficient and scalable, they are mathematically inaccurate. The reason stems from the fact that the local dynamic sensitivities are now solved as ODEs with continuous observation time t and a discrete perturbation time τ . To avoid this discretization a partial differential equation (PDE) implementation of local dynamic sensitivities is of future interest.

The present calculation of MDP sensitivities shows lot of variability and also computationally expensive to calculate. Hence, a future development in this area could aim to reduce both variability and computational cost. The possible approach ranges from formulating rigorous sensitivity metrics to choosing a good sampling strategy and normalisation procedure.

Applications of dynamic sensitivities: As seen in Chapter 7, one of the applications of local dynamic sensitivities is for model reduction. Hence, using the pathPSA and MDP analysis to reduce kinetic models is an obvious extension. In addition, another potential application of these dynamic analyses is in the area of design of experiments and in identifiability analysis. Using dynamic sensitivities is expected to give more information for experimental design and parameter identifiability.

Dynamical Sensitivity Analysis for Biological Systems

Complexity, Robustness and Dynamics of Biological Systems [Chapter 1]

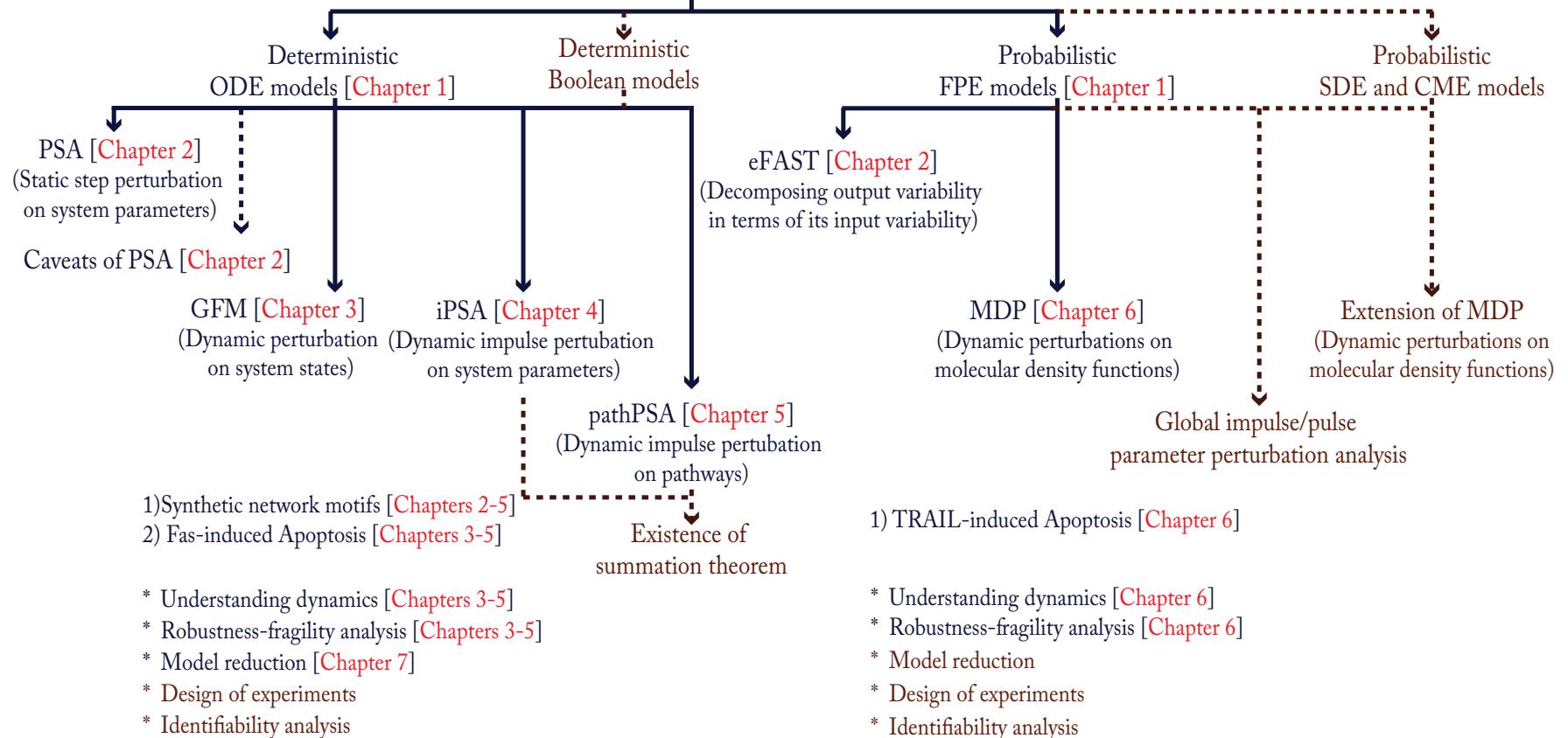


Figure 8.1. Thesis outlook and future work. This figure highlights different methods developed in the current dissertation and some of their applications. It also summarizes the breath and depth of the current research. In addition, possible extension of the current works are mentioned (highlighted in brown color)

9 Bibliography

1. Konopka, A.K., 'Systems Biology Principles, Methods, and Concepts', (CRC Press/Taylor & Francis, 2007)
2. Porter, R., Bynum, W.F., and Browne, E.J., *Macmillan Dictionary of the History of Science*, (: Macmillan, 1983)
3. Solé, R.V. and Goodwin, B.C., *Signs of Life : How Complexity Pervades Biology*, (Basic Books, 2000)
4. Ross, J. and Arkin, A.P., 'Complex Systems: From Chemistry to Systems Biology', *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106, (16), pp. 6433-6434.
5. Tyson, J.J., Chen, K.C., and Novak, B., 'Sniffers, Buzzers, Toggles and Blinkers: Dynamics of Regulatory and Signaling Pathways in the Cell', *Current Opinion in Cell Biology*, 2003, 15, (2), pp. 221-231.
6. Szallasi, Z., Stelling, J., and Periwal, V., *System Modeling in Cell Biology from Concepts to Nuts and Bolts*, (The MIT Press, 2006)
7. Rihani, S., *Complex Systems Theory and Development Practice : Understanding Non-Linear Realities*, (Zed Books, 2002)
8. Weng, G., Bhalla, U.S., and Iyengar, R., 'Complexity in Biological Signaling Systems', *Science*, 1999, 284, (5411), pp. 92-96.
9. Kitano, H., 'A Robustness-Based Approach to Systems-Oriented Drug Design', *Nature Reviews Drug Discovery*, 2007, 6, (3), pp. 202--210.
10. Bonchev, D., Rouvray, D.H., and SpringerLink (Online service), *Complexity in Chemistry, Biology, and Ecology*, (Springer, 2005)
11. Barkai, N. and Leibler, S., 'Robustness in Simple Biochemical Networks', *Nature*, 1997, 387, (6636), pp. 913--917.
12. Kitano, H., 'Biological Robustness', *Nature Reviews Genetics*, 2004, 5, (11), pp. 826--837.
13. Stelling, J.r., Sauer, U., Szallasi, Z., Doyle, F.J., and Doyle, J., 'Robustness of Cellular Functions', *Cell*, 2004, 118, (6), pp. 675--685.
14. Little, J.W., Shepley, D.P., and Wert, D.W., 'Robustness of a Gene Regulatory Circuit', *The EMBO Journal*, 1999, 18, (15), pp. 4299--4307.
15. Alon, U., Surette, M.G., Barkai, N., and Leibler, S., 'Robustness in Bacterial Chemotaxis', *Nature*, 1999, 397, (6715), pp. 168--171.
16. Stelling, J.r., Gilles, E.D., and Doyle, F.J., 'Robustness Properties of Circadian Clock Architectures', *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101, (36), pp. 13210--13215.
17. Stelling, J., Sauer, U., Szallasi, Z., Doyle, F.J., and Doyle, J., 'Robustness of Cellular Functions', *Cell*, 2004, 118, (6), pp. 675--685.
18. Kitano, H., 'Systems Biology: A Brief Overview', *Science*, 2002, 295, (5560), pp. 1662--1664.
19. Kitano, H., 'Computational Systems Biology', *Nature*, 2002, 420, (6912), pp. 206--210.

20. Yi, T.M., Huang, Y., Simon, M.I., and Doyle, J., 'Robust Perfect Adaptation in Bacterial Chemotaxis through Integral Feedback Control', *Proceedings of the National Academy of Sciences of the United States of America*, 2000, 97, (9), pp. 4649--4653.
21. Aldridge, B.B., Haller, G., Sorger, P.K., and Lauffenburger, D.A., 'Direct Lyapunov Exponent Analysis Enables Parametric Study of Transient Signalling Governing Cell Behaviour', *Systems Biology (Stevenage)*, 2006, 153, (6), pp. 425--432.
22. Batt, G., Yordanov, B., Weiss, R., and Belta, C., 'Robustness Analysis and Tuning of Synthetic Gene Networks', *Bioinformatics*, 2007, 23, (18), pp. 2415--2422.
23. Perumal, T.M. and Gunawan, R., 'Understanding Dynamics Using Sensitivity Analysis: Caveat and Solution', *BMC Systems Biology*, 2011, 5, (1), p. 41.
24. Perumal, T.M., Wu, Y., and Gunawan, R., 'Dynamical Analysis of Cellular Networks Based on the Green's Function Matrix', *Journal of Theoretical Biology*, 2009, 261, (2), pp. 248-259.
25. Chaves, M., Sontag, E.D., and Albert, R., 'Methods of Robustness Analysis for Boolean Models of Gene Control Networks', *Systems Biology (Stevenage)*, 2006, 153, (4), pp. 154--167.
26. Kim, J., Bates, D.G., Postlethwaite, I., Ma, L., and Iglesias, P.A., 'Robustness Analysis of Biochemical Network Models', *Systems Biology (Stevenage)*, 2006, 153, (3), pp. 96--104.
27. Cho, K.H. and Wolkenhauer, O., 'Analysis and Modelling of Signal Transduction Pathways in Systems Biology', *Biochemical Society Transactions*, 2003, 31, (Pt 6), pp. 1503-1509.
28. Krakauer, D.C. and Plotkin, J.B., 'Redundancy, Antiredundancy, and the Robustness of Genomes', *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99, (3), pp. 1405--1409.
29. Kitano, H., 'Towards a Theory of Biological Robustness', *Molecular Systems Biology*, 2007, 3, p. 137.
30. Csete, M.E. and Doyle, J.C., 'Reverse Engineering of Biological Complexity', *Science*, 2002, 295, (5560), pp. 1664--1669.
31. Qi, Y. and Ge, H., 'Modularity and Dynamics of Cellular Networks', *PLoS Computational Biology*, 2006, 2, (12), p. e174.
32. Kitano, H., 'Cancer Robustness: Tumour Tactics', *Nature*, 2003, 426, (6963), p. 125.
33. Jin, Y., Peng, X., Liang, Y., and Ma, J., 'Uniform Design-Based Sensitivity Analysis of Circadian Rhythm Model in Neurospora', *Computers and Chemical Engineering*, 2008, 32, (8), pp. 1956-1962.
34. Lovrics, A., Zsély, I.G., Csikász-Nagy, A., Zádor, J., Turányi, T., and Novák, B., 'Analysis of a Budding Yeast Cell Cycle Model Using the Shapes of Local Sensitivity Functions', *International Journal of Chemical Kinetics*, 2008, 40, (11), pp. 710-720.
35. Zhang, T., Brazhnik, P., and Tyson, J.J., 'Computational Analysis of Dynamical Responses to the Intrinsic Pathway of Programmed Cell Death', *Biophysical Journal*, 2009, 97, (2), pp. 415-434.
36. Hansen, C.H., Endres, R.G., and Wingreen, N.S., 'Chemotaxis in Escherichia Coli: A Molecular Model for Robust Precise Adaptation', *PLoS Computational Biology*, 2008, 4, (1), p. e1.
37. Reed, J.C., Doctor, K.S., and Godzik, A., 'The Domains of Apoptosis: A Genomics Perspective', *Science Signaling: STKE*, 2004, 2004, (239), p. re9.

38. Chu, Y., Jayaraman, A., and Hahn, J., 'Parameter Sensitivity Analysis of IL-6 Signalling Pathways', *IET Systems Biology*, 2007, 1, (6), pp. 342-352.
39. Ihekweaba, A.E., Broomhead, D.S., Grimley, R.L., Benson, N., and Kell, D.B., 'Sensitivity Analysis of Parameters Controlling Oscillatory Signalling in the NF-Kappab Pathway: The Roles of Ikk and Ikappabalpha', *Systems Biology (Stevenage)*, 2004, 1, (1), pp. 93-103.
40. Ihekweaba, A.E., Broomhead, D.S., Grimley, R., Benson, N., White, M.R., and Kell, D.B., 'Synergistic Control of Oscillations in the NF-Kappab Signalling Pathway', *Systems Biology (Stevenage)*, 2005, 152, (3), pp. 153-160.
41. Adler, P., Peterson, H., Agius, P., Reimand, J., and Vilo, J., 'Ranking Genes by Their Co-Expression to Subsets of Pathway Members', *Annals of the New York Academy of Sciences*, 2009, 1158, (1), pp. 1-13.
42. Iwamoto, K., Tashima, Y., Hamada, H., Eguchi, Y., and Okamoto, M., 'Mathematical Modeling and Sensitivity Analysis of G1/S Phase in the Cell Cycle Including the DNA-Damage Signal Transduction Pathway', *Biosystems*, 2008, 94, (1-2), pp. 109-117.
43. Bhalla, U.S. and Iyengar, R., 'Emergent Properties of Networks of Biological Signaling Pathways', *Science*, 1999, 283, (5400), pp. 381--387.
44. Auffray, C., Imbeaud, S., Roux-Rouquie, M., and Hood, L., 'From Functional Genomics to Systems Biology: Concepts and Practices', *Comptes Rendus Biologies*, 2003, 326, (10-11), pp. 879--892.
45. Bhalla, U.S., Ram, P.T., and Iyengar, R., 'Map Kinase Phosphatase as a Locus of Flexibility in a Mitogen-Activated Protein Kinase Signaling Network', *Science*, 2002, 297, (5583), pp. 1018--1023.
46. Novak, B. and Tyson, J.J., 'Modeling the Control of DNA Replication in Fission Yeast', *Proceedings of the National Academy of Sciences of the United States of America*, 1997, 94, (17), pp. 9147--9152.
47. Forger, D.B. and Peskin, C.S., 'A Detailed Predictive Model of the Mammalian Circadian Clock', *Proceedings of the National Academy of Sciences of the United States of America*, 2003, 100, (25), pp. 14806-14811.
48. Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srinivas, R., and Palsson, B.Ø., 'Global Reconstruction of the Human Metabolic Network Based on Genomic and Bibliomic Data', *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104, (6), pp. 1777--1782.
49. Kitano, H., 'Biological Robustness in Complex Host-Pathogen Systems', *Progress in Drug Research*, 2007, 64, pp. 239, 241--239, 263.
50. Kitano, H., 'The Theory of Biological Robustness and Its Implication in Cancer', *Ernst Schering Research Foundation Workshop*, 2007, (61), pp. 69--88.
51. Bailey, J.E., 'Toward a Science of Metabolic Engineering', *Science*, 1991, 252, (5013), pp. 1668--1675.
52. Stephanopoulos, G., 'Challenges in Engineering Microbes for Biofuels Production', *Science*, 2007, 315, (5813), pp. 801--804.
53. Ideker, T. and Lauffenburger, D., 'Building with a Scaffold: Emerging Strategies for High- to Low-Level Cellular Modeling', *Trends in Biotechnology*, 2003, 21, (6), pp. 255--262.
54. Wilkinson, D.J., *Stochastic Modelling for Systems Biology*, (Taylor & Francis, 2006)
55. Huang, S., 'Non-Genetic Heterogeneity of Cells in Development: More Than Just Noise', *Development*, 2009, 136, (23), pp. 3853-3862.

56. Faure, A., Naldi, A., Chaouiya, C., and Thieffry, D., 'Dynamical Analysis of a Generic Boolean Model for the Control of the Mammalian Cell Cycle', *Bioinformatics*, 2006, 22, (14), pp. e124-131.
57. Sanchez, L. and Thieffry, D., 'A Logical Analysis of the Drosophila Gap-Gene System', *Journal of Theoretical Biology*, 2001, 211, (2), pp. 115--141.
58. Mendoza, L., Thieffry, D., and Alvarez-Buylla, E.R., 'Genetic Control of Flower Morphogenesis in Arabidopsis Thaliana: A Logical Analysis', *Bioinformatics*, 1999, 15, (7-8), pp. 593--606.
59. Albeck, J.G., Burke, J.M., Aldridge, B.B., Zhang, M., Lauffenburger, D.A., and Sorger, P.K., 'Quantitative Analysis of Pathways Controlling Extrinsic Apoptosis in Single Cells', *Molecular Cell*, 2008, 30, (1), pp. 11-25.
60. Sun, W. and Saeys, M., 'Construction of an Ab Initio Kinetic Model for Industrial Ethane Pyrolysis', *AIChE Journal*, 2011, 57, (9), pp. 2458-2471.
61. Fell, D.A., 'Metabolic Control Analysis: A Survey of Its Theoretical and Experimental Development', *Biochemical Journal*, 1992, 286 (Pt 2), pp. 313--330.
62. Conrad, E.D. and Tyson, J.J., 'Systems Modeling in Cellular Biology: From Concepts to Nuts and Bolts', in Szallasi, Z., Stelling, J., and V, P. (eds.), (MIT Press, Cambridge, MA, 2006)
63. Fusenegger, M., Bailey, J.E., and Varner, J., 'A Mathematical Model of Caspase Function in Apoptosis', *Nature Biotechnology*, 2000, 18, (7), pp. 768--774.
64. Novák, B. and Tyson, J.J., 'A Model for Restriction Point Control of the Mammalian Cell Cycle', *Journal of Theoretical Biology*, 2004, 230, (4), pp. 563--579.
65. Sjöberg, P., Lötstedt, P., and Elf, J., 'Fokker-Planck Approximation of the Master Equation in Molecular Biology', *Computing and Visualization in Science*, 2009, 12, (1), pp. 37-50.
66. Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K., 'Non-Genetic Origins of Cell-to-Cell Variability in Trail-Induced Apoptosis', *Nature*, 2009, 459, (7245), pp. 428-432.
67. Risken, H., *The Fokker-Planck Equation : Methods of Solution and Applications*, (: Springer-Verlag, 1989, 2nd edn)
68. McAdams, H.H. and Arkin, A., 'It's a Noisy Business! Genetic Regulation at the Nanomolar Scale', *Trends in Genetics*, 1999, 15, (2), pp. 65--69.
69. Gillespie, D.T., 'A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions', *Journal of Computational Physics*, 1976, 22, (4), pp. 403--434.
70. Gillespie, D.T., 'Exact Stochastic Simulation of Coupled Chemical Reactions', *Journal of Physical Chemistry*, 1977, 81, (25), pp. 2340--2361.
71. Gillespie, D.T., 'Stochastic Simulation of Chemical Kinetics', *Annual Review of Physical Chemistry*, 2007, 58, pp. 35-55.
72. Gillespie, D.T., 'The Chemical Langevin Equation', *The Journal of Chemical Physics*, 2000, 113, (1), pp. 297-306.
73. Arkin, A., Ross, J., and McAdams, H.H., 'Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage Lambda-Infected Escherichia Coli Cells', *Genetics*, 1998, 149, (4), pp. 1633--1648.
74. Forger, D.B. and Peskin, C.S., 'Stochastic Simulation of the Mammalian Circadian Clock', *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102, (2), pp. 321--324.

75. Morohashi, M., Winn, A.E., Borisuk, M.T., Bolouri, H., Doyle, J., and Kitano, H., 'Robustness as a Measure of Plausibility in Models of Biochemical Networks', *Journal of Theoretical Biology*, 2002, 216, (1), pp. 19--30.
76. Schmidt, H. and Jacobsen, E.W., 'Linear Systems Approach to Analysis of Complex Dynamic Behaviours in Biochemical Networks', *Systems Biology (Stevenage)*, 2004, 1, (1), pp. 149-158.
77. Gunawan, R., Cao, Y., Petzold, L., and Doyle, F.J., 'Sensitivity Analysis of Discrete Stochastic Systems', *Biophysical Journal*, 2005, 88, (4), pp. 2530--2540.
78. Gunawan, R. and Doyle, F.J., 'Isochron-Based Phase Response Analysis of Circadian Rhythms', *Biophysical Journal*, 2006, 91, (6), pp. 2131--2141.
79. Ingalls, B.P., 'Autonomously Oscillating Biochemical Systems: Parametric Sensitivity of Extrema and Period', *Systems Biology (Stevenage)*, 2004, 1, (1), pp. 62--70.
80. Kim, P.-J., Lee, D.-Y., Kim, T.Y., Lee, K.H., Jeong, H., Lee, S.Y., and Park, S., 'Metabolite Essentiality Elucidates Robustness of Escherichia Coli Metabolism', *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104, (34), pp. 13638--13642.
81. Ma, H.-W. and Zeng, A.-P., 'The Connectivity Structure, Giant Strong Component and Centrality of Metabolic Networks', *Bioinformatics*, 2003, 19, (11), pp. 1423--1430.
82. Skogestad, S. and Postlewaite, I., *Multivariate Feedback Control - Analysis and Design*, (John Wiley \& Sons Ltd, 1996, Second edn)
83. Shoemaker, J.E. and Doyle III, F.J., 'Identifying Fragilities in Biochemical Networks: Robust Performance Analysis of Fas Signaling-Induced Apoptosis', *Biophysical Journal*, 2008.
84. Strogatz, S.H., *Nonlinear Dynamics and Chaos : With Applications to Physics, Biology, Chemistry, and Engineering*, (: Addison-Wesley Pub., 1994)
85. Lu, J., Engl, H.W., and Schuster, P., 'Inverse Bifurcation Analysis: Application to Simple Gene Systems', *Algorithms for Molecular Biology*, 2006, 1, p. 11.
86. Chickarmane, V., Paladugu, S.R., Bergmann, F., and Sauro, H.M., 'Bifurcation Discovery Tool', *Bioinformatics*, 2005, 21, (18), pp. 3688-3690.
87. Borisuk, M.T. and Tyson, J.J., 'Bifurcation Analysis of a Model of Mitotic Control in Frog Eggs', *Journal of Theoretical Biology*, 1998, 195, (1), pp. 69-85.
88. Battogtokh, D. and Tyson, J.J., 'Bifurcation Analysis of a Model of the Budding Yeast Cell Cycle', *Chaos*, 2004, 14, (3), pp. 653-661.
89. Saltelli, A., Chan, K., and Scott, E.M., *Sensitivity Analysis: Gauging the Worth of Scientific Models*, (John Wiley \& Sons, Ltd., 2000)
90. Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F., 'Sensitivity Analysis for Chemical Models', *Chemical Reviews*, 2005, 105, (7), pp. 2811--2827.
91. Saltelli, A., Tarantola, S., and Campolongo, F., 'Sensitivity Analysis as an Ingredient of Modeling', *Statistical Science*, 2000, 15, (4), pp. 377--395.
92. Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M., *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, (John Wiley \& Sons, Ltd., 2004)
93. Turányi, T., 'Sensitivity Analysis of Complex Kinetic Systems. Tools and Applications', *Journal of Mathematical Chemistry*, 1990, 5, (3), pp. 203-248.
94. Varma, A., Morbidelli, M., and Wu, H., *Parametric Sensitivity in Chemical Systems*, (Cambridge University Press, Cambridge, UK, 1999)
95. Frey, H.C. and Patil, S.R., 'Identification and Review of Sensitivity Analysis Methods', *Risk Analysis*, 2002, 22, (3), pp. 553-578.

96. Eissing, T., Allgower, F., and Bullinger, E., 'Robustness Properties of Apoptosis Models with Respect to Parameter Variations and Intrinsic Noise', *Systems Biology (Stevenage)*, 2005, 152, (4), pp. 221-228.
97. Hua, F., Cornejo, M.G., Cardone, M.H., Stokes, C.L., and Lauffenburger, D.A., 'Effects of Bcl-2 Levels on Fas Signaling-Induced Caspase-3 Activation: Molecular Genetic Tests of Computational Model Predictions', *The Journal of Immunology*, 2005, 175, (2), pp. 985--995.
98. Hua, F., Hautaniemi, S., Yokoo, R., and Lauffenburger, D.A., 'Integrated Mechanistic and Data-Driven Modelling for Multivariate Analysis of Signalling Pathways', *Journal of the Royal Society Interface*, 2006, 3, (9), pp. 515--526.
99. Bentele, M., Lavrik, I., Ulrich, M., Stober, S., Heermann, D.W., Kalthoff, H., Krammer, P.H., and Eils, R., 'Mathematical Modeling Reveals Threshold Mechanism in Cd95-Induced Apoptosis', *The Journal of Cell Biology*, 2004, 166, (6), pp. 839--851.
100. Hu, D. and Yuan, J.M., 'Time-Dependent Sensitivity Analysis of Biological Networks: Coupled Mapk and Pi3k Signal Transduction Pathways', *Journal of Physical Chemistry A*, 2006, 110, (16), pp. 5361-5370.
101. Hornberg, J.J., Binder, B., Bruggeman, F.J., Schoeberl, B., Heinrich, R., and Westerhoff, H.V., 'Control of Mapk Signalling: From Complexity to What Really Matters', *Oncogene*, 2005, 24, (36), pp. 5533--5542.
102. Friboulet, A. and Thomas, D., 'Systems Biology-an Interdisciplinary Approach', *Biosensors & Bioelectronics*, 2005, 20, (12), pp. 2404--2407.
103. Gunawan, R. and Doyle, F.J., 'Phase Sensitivity Analysis of Circadian Rhythm Entrainment', *Journal of Biological Rhythms*, 2007, 22, (2), pp. 180--194.
104. Aguda, B.D. and Tang, Y., 'The Kinetic Origins of the Restriction Point in the Mammalian Cell Cycle', *Cell Proliferation*, 1999, 32, (5), pp. 321--335.
105. Ingalls, B., 'Sensitivity Analysis: From Model Parameters to System Behaviour', *Essays in Biochemistry*, 2008, 45, pp. 177-193.
106. <http://www.mathworks.com/>, accessed Date Accessed
107. <http://www.math.pitt.edu/~bard/xpp/xpp.html>, accessed Date Accessed
108. Schmidt, H. and Jirstrand, M., 'Systems Biology Toolbox for Matlab: A Computational Platform for Research in Systems Biology', *Bioinformatics*, 2006, 22, (4), pp. 514-515.
109. Maiwald, T. and Timmer, J., 'Dynamical Modeling and Multi-Experiment Fitting with Potterswheel', *Bioinformatics*, 2008, 24, (18), pp. 2037-2043.
110. Mendes, P., 'Gepasi: A Software Package for Modelling the Dynamics, Steady States and Control of Biochemical and Other Systems', *Computer Application in Biosciences*, 1993, 9, (5), pp. 563-571.
111. Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., and Kummer, U., 'Copasi--a Complex Pathway Simulator', *Bioinformatics*, 2006, 22, (24), pp. 3067-3074.
112. <http://sbw.kgi.edu/>, accessed Date Accessed
113. <http://www.physiome.org/jsim/>, accessed Date Accessed
114. <http://www.chemengr.ucsb.edu/~ceweb/faculty/doyle/biosens/BioSens.htm>, accessed Date Accessed
115. Zi, Z., Zheng, Y., Rundell, A.E., and Klipp, E., 'Sbml-Sat: A Systems Biology Markup Language (Sbml) Based Sensitivity Analysis Tool', *BMC Bioinformatics*, 2008, 9, p. 342.

116. Rodriguez-Fernandez, M. and Banga, J.R., 'Senssb: A Software Toolbox for the Development and Sensitivity Analysis of Systems Biology Models', *Bioinformatics*, 26, (13), pp. 1675-1676.
117. Alves, R., Antunes, F., and Salvador, A., 'Tools for Kinetic Modeling of Biochemical Networks', *Nature Biotechnology*, 2006, 24, (6), pp. 667-672.
118. Klipp, E., Liebermeister, W., Helbig, A., Kowald, A., and Schaber, J., 'Systems Biology Standards—the Community Speaks', *Nature Biotechnology*, 2007, 25, (4), pp. 390-391.
119. Yamada, S., Shiono, S., Joo, A., and Yoshimura, A., 'Control Mechanism of Jak/Stat Signal Transduction Pathway', *FEBS Letters*, 2003, 534, (1-3), pp. 190-196.
120. Yue, H., Brown, M., Knowles, J., Wang, H., Broomhead, D.S., and Kell, D.B., 'Insights into the Behaviour of Systems Biology Models from Dynamic Sensitivity and Identifiability Analysis: A Case Study of an Nf-Kappab Signalling Pathway', *Molecular Biosystems*, 2006, 2, (12), pp. 640-649.
121. Zi, Z., Cho, K.H., Sung, M.H., Xia, X., Zheng, J., and Sun, Z., 'In Silico Identification of the Key Components and Steps in Ifn-Gamma Induced Jak-Stat Signaling Pathway', *FEBS Letters*, 2005, 579, (5), pp. 1101-1108.
122. Hwang, J.T., Dougherty, E.P., Rabitz, S., and Rabitz, H., 'The Green's Function Method of Sensitivity Analysis in Chemical Kinetics', *The Journal of Chemical Physics*, 1978, 69, (11), pp. 5180-5191.
123. Sandu, A., Daescu, D.N., and Carmichael, G.R., 'Direct and Adjoint Sensitivity Analysis of Chemical Kinetic Systems with Kpp: Part I—Theory and Software Tools', *Atmospheric Environment*, 2003, 37, (36), pp. 5083-5096.
124. Niepel, M., Spencer, S.L., and Sorger, P.K., 'Non-Genetic Cell-to-Cell Variability and the Consequences for Pharmacology', *Current Opinion in Chemical Biology*, 2009, 13, (5-6), pp. 556-561.
125. Marino, S., Hogue, I.B., Ray, C.J., and Kirschner, D.E., 'A Methodology for Performing Global Uncertainty and Sensitivity Analysis in Systems Biology', *Journal of Theoretical Biology*, 2008, 254, (1), pp. 178-196.
126. Costanza, V. and Seinfeld, J.H., 'Stochastic Sensitivity Analysis in Chemical Kinetics', *Journal of Chemical Physics*, 1981, 74, (7).
127. Bieniasz, L.K. and Speiser, B., 'Use of Sensitivity Analysis Methods in the Modelling of Electrochemical Transients Part 1. Gaining More Insight into the Behaviour of Kinetic Models', *Journal of Electroanalytical Chemistry*, 1998, 441, (1-2), pp. 271-285.
128. Sargsyan, K., Debusschere, B., Najm, H., and Marzouk, Y., 'Bayesian Inference of Spectral Expansions for Predictability Assessment in Stochastic Reaction Networks', *Journal of Computational and Theoretical Nanoscience*, 2009, 6, (10), pp. 2283-2297.
129. Zhang, Y., Vijayaraghavan, K., and Seigneur, C., 'Evaluation of Three Probing Techniques in a Three-Dimensional Air Quality Model', *Journal of Geophysical Research D: Atmospheres*, 2005, 110, (2), pp. 1-21.
130. Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G., and Schaibly, J.H., 'Study of the Sensitivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients. I Theory', *The Journal of Chemical Physics*, 1970, pp. 3873-3878.
131. Saltelli, A., Tarantola, S., and Chan, K.P.S., 'A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output', *Technometrics*, 1999, 41, (1), pp. 39-56.
132. Saltelli, A., *Global Sensitivity Analysis : The Primer*, (John Wiley, 2008)

133. Saltelli, A. and Tarantola, S., 'On the Relative Importance of Input Factors in Mathematical Models: Safety Assessment for Nuclear Waste Disposal', *Journal of the American Statistical Association*, 2002, 97, (459), pp. 702-709.
134. Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C., and Shah, N., 'Monte Carlo Evaluation of Derivative-Based Global Sensitivity Measures', *Reliability Engineering & System Safety*, 2009, 94, (7), pp. 1135-1148.
135. Hafner, M., Koepll, H., Hasler, M., and Wagner, A., "Glocal" Robustness Analysis and Model Discrimination for Circadian Oscillators', *PLoS Computational Biology*, 2009, 5, (10), p. e1000534.
136. Yue, H., Brown, M., He, F., Jia, J., and Kell, D.B., 'Sensitivity Analysis and Robust Experimental Design of a Signal Transduction Pathway System', *International Journal of Chemical Kinetics*, 2008, 40, (11), pp. 730-741.
137. Arfken, G.B. and Weber, H.-J., *Mathematical Methods for Physicists*, (Elsevier, 2005, 6th edn)
138. Yetter, R.A., Dryer, F.L., and Rabitz, H., 'Some Interpretive Aspects of Elementary Sensitivity Gradients in Combustion Kinetics Modeling', *Combustion and Flame*, 1985, 59, (2), pp. 107-133.
139. Chen, C.-T., *Linear System Theory and Design*, (Oxford University Press, 1999, Third edition edn)
140. Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W., 'From Molecular to Modular Cell Biology', *Nature*, 1999, 402, (6761 Suppl), pp. C47--C52.
141. Kremling, A., Fischer, S., Gadkar, K., Doyle, F.J., Sauter, T., Bullinger, E., Allgower, F., and Gilles, E.D., 'A Benchmark for Methods in Reverse Engineering and Model Discrimination: Problem Formulation and Solutions', *Genome Research*, 2004, 14, (9), pp. 1773-1785.
142. Scaffidi, C., Fulda, S., Srinivasan, A., Friesen, C., Li, F., Tomaselli, K.J., Debatin, K.M., Krammer, P.H., and Peter, M.E., 'Two Cd95 (Apo-1/Fas) Signaling Pathways', *The EMBO Journal*, 1998, 17, (6), pp. 1675--1687.
143. Chen, B.-S., Wang, Y.-C., Wu, W.-S., and Li, W.-H., 'A New Measure of the Robustness of Biochemical Networks', *Bioinformatics*, 2005, 21, (11), pp. 2698--2705.
144. Bhalla, U.S. and Iyengar, R., 'Robustness of the Bistable Behavior of a Biological Signaling Feedback Loop', *Chaos*, 2001, 11, (1), pp. 221--226.
145. Wagner, A., 'Robustness, Evolvability, and Neutrality', *FEBS Letters*, 2005, 579, (8), pp. 1772--1778.
146. Carlson, J.M. and Doyle, J., 'Complexity and Robustness', *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99 Suppl 1, pp. 2538--2545.
147. Westerhoff, H.V. and Chen, Y.D., 'How Do Enzyme Activities Control Metabolite Concentrations? An Additional Theorem in the Theory of Metabolic Control', *European Journal of Biochemistry*, 1984, 142, (2), pp. 425-430.
148. Okazaki, N., Asano, R., Kinoshita, T., and Chuman, H., 'Simple Computational Models of Type I/Type II Cells in Fas Signaling-Induced Apoptosis', *Journal of Theoretical Biology*, 2008, 250, (4), pp. 621--633.
149. Fesik, S.W., 'Promoting Apoptosis as a Strategy for Cancer Drug Discovery', *Nature Reviews Cancer*, 2005, 5, (11), pp. 876--885.
150. Breckenridge, D.G., Nguyen, M., Kuppig, S., Reth, M., and Shore, G.C., 'The Procaspsase-8 Isoform, Procaspsase-8I, Recruited to the Bap31 Complex at the Endoplasmic

Reticulum', *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99, (7), pp. 4331--4336.

151. Nishimura, S., Adachi, M., Ishida, T., Matsunaga, T., Uchida, H., Hamada, H., and Imai, K., 'Adenovirus-Mediated Transfection of Caspase-8 Augments Anoikis and Inhibits Peritoneal Dissemination of Human Gastric Carcinoma Cells', *Cancer Research*, 2001, 61, (19), pp. 7009--7014.
152. Nelson, H.S., 'Advair: Combination Treatment with Fluticasone Propionate/Salmeterol in the Treatment of Asthma', *Journal of Allergy and Clinical Immunology*, 2001, 107, (2), pp. 398--416.
153. Larder, B.A., Kemp, S.D., and Harrigan, P.R., 'Potential Mechanism for Sustained Antiretroviral Efficacy of Azt-3tc Combination Therapy', *Science*, 1995, 269, (5224), pp. 696--699.
154. Borisy, A.A., Elliott, P.J., Hurst, N.W., Lee, M.S., Lehar, J., Price, E.R., Serbedzija, G., Zimmermann, G.R., Foley, M.A., Stockwell, B.R., and Keith, C.T., 'Systematic Discovery of Multicomponent Therapeutics', *Proceedings of the National Academy of Sciences of the United States of America*, 2003, 100, (13), pp. 7977--7982.
155. Pollack, I.F., Erff, M., and Ashkenazi, A., 'Direct Stimulation of Apoptotic Signaling by Soluble Apo2l/Tumor Necrosis Factor-Related Apoptosis-Inducing Ligand Leads to Selective Killing of Glioma Cells', *Clinical Cancer Research*, 2001, 7, (5), pp. 1362--1369.
156. Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E.D., 'Metabolic Network Structure Determines Key Aspects of Functionality and Regulation', *Nature*, 2002, 420, (6912), pp. 190-193.
157. Klamt, S., Saez-Rodriguez, J., and Gilles, E.D., 'Structural and Functional Analysis of Cellular Networks with Cellnetanalyzer', *BMC Systems Biology*, 2007, 1, p. 2.
158. Wasserman, S. and Faust, K., *Social Network Analysis : Methods and Applications*, (: Cambridge University Press, 1994)
159. Perra, N. and Fortunato, S., 'Spectral Centrality Measures in Complex Networks', *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 2008, 78, (3 Pt 2), p. 036107.
160. Crofts, J.J. and Higham, D.J., 'A Weighted Communicability Measure Applied to Complex Brain Networks', *Journal of the Royal Society Interface*, 2009, 6, (33), pp. 411-414.
161. Liu, W., Li, D., Zhang, J., Zhu, Y., and He, F., 'Sigflux: A Novel Network Feature to Evaluate the Importance of Proteins in Signal Transduction Networks', *BMC Bioinformatics*, 2006, 7, p. 515.
162. Wang, R.S. and Albert, R., 'Elementary Signaling Modes Predict the Essentiality of Signal Transduction Network Components', *BMC Systems Biology*, 2011, 5, p. 44.
163. Rubin, F., 'Enumerating All Simple Paths in a Graph', *IEEE Transactions on Circuits and Systems*, 1978, (8), pp. 641-642.
164. Paszek, P., Ryan, S., Ashall, L., Sillitoe, K., Harper, C.V., Spiller, D.G., Rand, D.A., and White, M.R., 'Population Robustness Arising from Cellular Heterogeneity', *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107, (25), pp. 11644-11649.
165. Alberts, B., *Molecular Biology of the Cell*, (Garland Science, 2002, 4th edn)
166. Goldberg, A.D., Allis, C.D., and Bernstein, E., 'Epigenetics: A Landscape Takes Shape', *Cell*, 2007, 128, (4), pp. 635-638.
167. Elowitz, M.B. and Leibler, S., 'A Synthetic Oscillatory Network of Transcriptional Regulators', *Nature*, 2000, 403, (6767), pp. 335--338.

168. Rando, O.J. and Verstrepen, K.J., 'Timescales of Genetic and Epigenetic Inheritance', *Cell*, 2007, 128, (4), pp. 655-668.
169. Blake, W.J., KAErn, M., Cantor, C.R., and Collins, J.J., 'Noise in Eukaryotic Gene Expression', *Nature*, 2003, 422, (6932), pp. 633-637.
170. Manninen, T., Linne, M.-L., and Ruohonen, K., 'A Novel Approach to Model Neuronal Signal Transduction Using Stochastic Differential Equations', *Neurocomputing*, 2006, 69, (10-12), pp. 1066-1069.
171. Fennell, D.A., Pallaska, A., Corbo, M., and Cotter, F.E., 'Stochastic Modelling of Apoptosis Kinetics', *Apoptosis*, 2005, 10, (2), pp. 447-452.
172. Deza, M. and Deza, E., *Encyclopedia of Distances*, (Springer, 2009)
173. Albeck, J.G., Burke, J.M., Spencer, S.L., Lauffenburger, D.A., and Sorger, P.K., 'Modeling a Snap-Action, Variable-Delay Switch Controlling Extrinsic Cell Death', *PLoS Biology*, 2008, 6, (12), pp. 2831-2852.
174. Petzold, L. and Zhu, W., 'Model Reduction for Chemical Kinetics: An Optimization Approach', *AICHE Journal*, 1999, 45, (4), pp. 869-886.
175. Dullerud, G.E. and Paganini, F.G., *A Course in Robust Control Theory : A Convex Approach*, (Springer, 2000)
176. Peters, N., Numerical and Asymptotic Analysis of Systematically Reduced Reaction Schemes for Hydrocarbon Flames', in Glowinski, R., Larrouтуrou, B., and Temam, R. (eds.), *Numerical Simulation of Combustion Phenomena*, (Springer Berlin / Heidelberg, 1985)
177. Williams, F.A., *Combustion Theory : The Fundamental Theory of Chemically Reacting Flow Systems*, (Westview, 1985, 2nd edn)
178. Turanyi, T., 'Reduction of Large Reaction Mechanisms', *New Journal of Chemistry*, 1990, 14, pp. 795-803.
179. Androulakis, I.P., 'Kinetic Mechanism Reduction Based on an Integer Programming Approach', *AICHE Journal*, 2000, 46, (2), pp. 361-371.
180. Bhattacharjee, B., Schwer, D.A., Barton, P.I., and Green, W.H., 'Optimally-Reduced Kinetic Models: Reaction Elimination in Large-Scale Kinetic Mechanisms', *Combustion and Flame*, 2003, 135, (3), pp. 191-208.
181. Lam, S.H. and Goussis, D.A., 'Csp Method for Simplifying Kinetics', *International Journal of Chemical Kinetics*, 1994, 26, (4), pp. 461-486.
182. Weekman Jr, V.W., 'Lumps, Models, and Kinetics in Practice', *AICHE Monograph Series*, 1979, 75, (11).
183. Maas, U. and Pope, S.B., 'Simplifying Chemical Kinetics: Intrinsic Low-Dimensional Manifolds in Composition Space', *Combustion and Flame*, 1992, 88, (3-4), pp. 239-264.
184. Lu, T. and Law, C.K., 'A Directed Relation Graph Method for Mechanism Reduction', in, (2005, 1 edn.)
185. Bonvin, D. and Rippin, D.W.T., 'Target Factor Analysis for the Identification of Stoichiometric Models', *Chemical Engineering Science*, 1990, 45, (12), pp. 3417-3426.
186. He, K., Ierapetritou, M.G., and Androulakis, I.P., 'Integration of on-the-Fly Kinetic Reduction with Multidimensional Cfd', *AICHE Journal*, 2010, 56, (5), pp. 1305-1314.
187. Tomlin, A.S., Turányi, T., and Pilling, M.J., Chapter 4 Mathematical Tools for the Construction, Investigation and Reduction of Combustion Mechanisms', (1997)
188. Edelson, D. and Allara, D.L., 'A Computational Analysis of the Alkane Pyrolysis Mechanism: Sensitivity Analysis of Individual Reaction Steps', *International Journal of Chemical Kinetics*, 1980, 12, (9), pp. 605-621.
189. http://www.me.berkeley.edu/gri_mech/, accessed Date Accessed

190. Degenring, D., 'Sensitivity Analysis for the Reduction of Complex Metabolism Models', *Journal of Process Control*, 2004, 14, (7), pp. 729-745.
191. Shi, Y., Ge, H.W., Brakora, J.L., and Reitz, R.D., 'Automatic Chemistry Mechanism Reduction of Hydrocarbon Fuels for Hcci Engines Based on Drgep and Pca Methods with Error Control', *Energy and Fuels*, 2010, 24, (3), pp. 1646-1654.

Appendix A

A.1 Simple Network Model

Table A.1 - Simple Network Model			
Solvers		ODE15S of MATLAB/CVODE	
Solver Settings		AbsTol = RelTol = 1e-12	
Initial Conditions		Rate Constants	
x_1	1	k_{f1}	0.06
x_2	0	k_{v2}	1
x_3	1	k_{k2}	2
x_4	1	k_{v3}	1
x_5	0	k_{k3}	2
x_6	0	k_{v4}	1
		k_{k4}	2
Reaction Number		Rate Equation	
r_1		$k_{f1} * x_1$	
r_2		$k_{v2} * x_2 * x_4 / (k_{k2} + x_4)$	
r_3		$k_{v3} * x_2 * x_3 / (k_{k3} + x_3)$	
r_4		$k_{v4} * x_5 * x_4 / (k_{k4} + x_4)$	
Differential Equations			
dx_1/dt		$-r_1$	
dx_2/dt		$+r_1$	
dx_3/dt		$-r_3$	
dx_4/dt		$-r_2 - r_4$	
dx_5/dt		$+r_3$	
dx_6/dt		$+r_2 + r_4$	

Appendix B

B.1 Common Network Motifs

B.1.1 Perfectly Adapted Signal-Response Model

Table B.1 – Perfectly adapted signal-response model	
Solvers	ODE15S of MATLAB
Solver Settings	AbsTol = RelTol = 1e-12
Initial Conditions	Value (arbitrary units)
R	1.00
X	0.00
Parameter Name	Value (arbitrary units)
k_1	2.00
k_2	2.00
k_3	1.00
k_4	1.00
Differential Equations	
$d[R]/dt$	$k_1*S - k_2*X*R;$
$d[X]/dt$	$k_3*S - k_4*X;$

B.1.2 Mutual Activation Model

Table B.2 – Mutual activation model	
Solvers	ODE15S of MATLAB
Solver Settings	AbsTol = RelTol = 1e-12
Initial Conditions	Value (arbitrary units)
R	0.50
EP	0.50
Parameter Name	Value (arbitrary units)
k_0	0.40
k_1	0.01
k_2	1.00
k_3	1.00
k_4	0.20
J_3	0.05
J_4	0.05
Differential Equations	
$d[R]/dt$	$k_0*EP + k_1*S - k_2*X*R$
$[EP]$	$G(k_3*R, k_4, J_3, J_4)$
$G(u, v, J, K)$	$2*u*K/(v-u+v^*J+u^*K+sqrt((v-u+v^*J+u^*K)^2-4*(v-u)*u^*K))$

B.1.3 Negative Feedback Oscillator Model

Table B.3 – Negative Feedback Oscillator model	
Solvers	ODE15S of MATLAB
Solver Settings	AbsTol = RelTol = 1e-12
Initial Conditions	Value (arbitrary units)
X	5.4504
YP	0.7990
RP	0.0435
Parameter Name	Value (arbitrary units)
k_0	0.00
k_1	1.00
k_2	0.01
k_2P	10.00
k_3	0.10
k_4	0.20
k_5	0.10
k_6	0.05
K_{m3}	0.01
K_{m4}	0.01
K_{m5}	0.01
K_{m6}	0.01
Differential Equations	
$d[X]/dt$	$k_0+k_1*S-k_2*X-k_2P*RP*X;$
$d[YP]/dt$	$(k_3*X*(1-YP)/(K_{m3}+1-YP))-(k_4*YP/(K_{m4}+YP));$
$d[RP]/dt$	$(k_5*YP*(1-RP)/(K_{m5}+1-RP))-(k_6*RP/(K_{m6}+RP));$

B.2 FasL-induced Cell Death Model of Human Jurkat Cells

Table B.4 –FasL-induced cell death model	
Solvers	ODE15S of MATLAB/CVODE
Solver Settings	AbsTol = RelTol = 1e-12
State Name	Initial Condition (nM)
Fas	16.67
ProCaspase 8	33.33
DISC	0.00
DISC:Casp8	0.00
Caspase-8	0.00
ProCaspase-3	200.00
Casp-8:Casp-3	0.00
Caspase-3	0.00
Bcl-2	75.00

Bcl-2:Mito	0.00
Mitochondria	83.06
Caspase-8:Mitochondria	0.00
Activated Mitochondria	0.00
Apoptosome	16.61
Activated Apoptosome	0.00
Activated Apoptosome:Caspase-3	0.00
Smac	100.00
Activated Smac	0.00
XIAP	30.00
Apoptosome:XIAP	0.00
Activated Smac:XIAP	0.00
Caspase-3:XIAP	0.00
Caspase-3:Caspase-6	0.00
Caspase-6	0.00
Caspase-6:Caspase-8	0.00
ProCaspase-6	10.00
FLIP	8.00
DISC:Flip	0.00
FasLigand	2.00
Parameter Name	Value
P1_f	0.0101
P1_r	0.896
P2_f	0.00889
P2_r	0.785
P3_k	0.0294
P4_f	0.000438
P4_r	0.788
P5_k	0.0225
P8_f	0.0662
P8_r	0.00526
P9_f	1.00E-05
P9_r	0.767
P11_k	0.0168
P12_k	0.00182
P13_f	0.00844
P13_r	0.591
P14_k	0.0125
P15_f	0.0903
P15_r	0.0102
P16_r	0.00132
P18_f	1.08E-05
P18_r	0.537
P19_k	0.00197
P20_f	0.0268
P20_r	0.467

P21_k	0.00531
P22_f	0.0799
P22_r	1.00E-06
P16_f	0.007
P17_f	0.0025
P17_r	0.0024
P10_k	0.0225
Rate Equations	
r1_f	P1_f*FasL*FADD;
r1_r	P1_r*DISC;
r2_f	P2_f*Casp8*DISC;
r2_r	P2_r*Casp8_DISC;
r3_k	P3_k*Casp8_DISC;
r4_f	P4_f*Casp8_act*Casp3;
r4_r	P4_r*Casp3_Casp8_act;
r5_k	P5_k*Casp3_Casp8_act;
r8_f	P8_f*Bcl2*Mit;
r8_r	P8_r*Bcl2_Mit;
r9_f	P9_f*Casp8_act*Mit;
r9_r	P9_r*Mit_Casp8_act;
r10_k	P10_k*Mit_Casp8_act;
r11_k	P11_k*Apopt*Mit_act;
r12_k	P12_k*Smac*Mit_act;
r13_f	P13_f*Apopt_act*Casp3;
r13_r	P13_r*Casp3_Apopt_act;
r14_k	P14_k*Casp3_Apopt_act;
r15_f	P15_f*Apopt*XIAP;
r15_r	P15_r*XIAP_Apopt;
r16_f	P16_f*XIAP*Smac_act;
r16_r	P16_r*XIAP_Smac;
r17_f	P17_f*XIAP*Casp3_act;
r17_r	P17_r*XIAP_Casp3_act;
r18_f	P18_f*Casp3_act*Casp6;
r18_r	P18_r*Casp6_Casp3_act;
r19_k	P19_k*Casp6_Casp3_act;
r20_f	P20_f*Casp8*Casp6_act;
r20_r	P20_r*Casp8_Casp6_act;
r21_k	P21_k*Casp8_Casp6_act;
r22_f	P22_f*DISC*Flip;
r22_r	P22_r*Flip_DISC;
Differential Equations	
d[FADD]/dt	r1_r - r1_f
d[Casp8]/dt	r20_r - r20_f - r2_f + r2_r
d[DISC]/dt	r1_f - r1_r - r22_f + r22_r - r2_f + r2_r + r3_k
d[Casp8_DISC]/dt	r2_f - r2_r - r3_k
d[Casp8_act]/dt	r10_k + r21_k + r3_k - r4_f + r4_r + r5_k - r9_f + r9_r

d[Casp3]/dt	r13_r - r13_f - r4_f + r4_r
d[Casp3_Casp8_act]/dt	r4_f - r4_r - r5_k
d[Casp3_act]/dt	r14_k - r17_f + r17_r - r18_f + r18_r + r19_k + r5_k
d[Bcl2]/dt	r8_r - r8_f
d[Bcl2_Mit]/dt	r8_f - r8_r
d[Mit]/dt	r8_r - r8_f - r9_f + r9_r
d[Mit_Casp8_act]/dt	r9_f - r10_k - r9_r
d[Mit_act]/dt	r10_k
d[Apopt]/dt	r15_r - r15_f - r11_k
d[Apopt_act]/dt	r11_k - r13_f + r13_r + r14_k
d[Casp3_Apopt_act]/dt	r13_f - r13_r - r14_k
d[Smac]/dt	0 - r12_k
d[Smac_act]/dt	r12_k - r16_f + r16_r
d[XIAP]/dt	r15_r - r15_f - r16_f + r16_r - r17_f + r17_r
d[XIAP_Apopt]/dt	r15_f - r15_r
d[XIAP_Smac]/dt	r16_f - r16_r
d[XIAP_Casp3_act]/dt	r17_f - r17_r
d[Casp6_Casp3_act]/dt	r18_f - r18_r - r19_k
d[Casp6_act]/dt	r19_k - r20_f + r20_r + r21_k
d[Casp8_Casp6_act]/dt	r20_f - r20_r - r21_k
d[Casp6]/dt	r18_r - r18_f
d[Flip]/dt	r22_r - r22_f
d[Flip_DISC]/dt	r22_f - r22_r

B.3 Clustering Method

To better track the signal propagation through the network, the nodes are clustered according to the cosine distance d_{rs} between two time-vectors \dot{S}_m and \dot{S}_n , as follows

$$d_{rs} = \sqrt{1 - \frac{\dot{S}_m^T \dot{S}_n}{(\dot{S}_m^T \dot{S}_m)^{1/2} (\dot{S}_n^T \dot{S}_n)^{1/2}}} \quad (H.1)$$

where $\dot{S}_m = [dS_{m,j}^x(t_1, \tau)/dt \quad dS_{m,j}^x(t_2, \tau)/dt \quad \dots \quad dS_{m,j}^x(t_N, \tau)/dt]^T$ for an impulse signal starting from x_j at time τ and N is the total number of time points. As the name suggests, the cosine distance measures the cosine of angle between the vectors, and hence vectors with similar distance values appear together. Finally, the clusters are then sorted according to the timing of the first measurable signal.

B.4 Appendix Figures

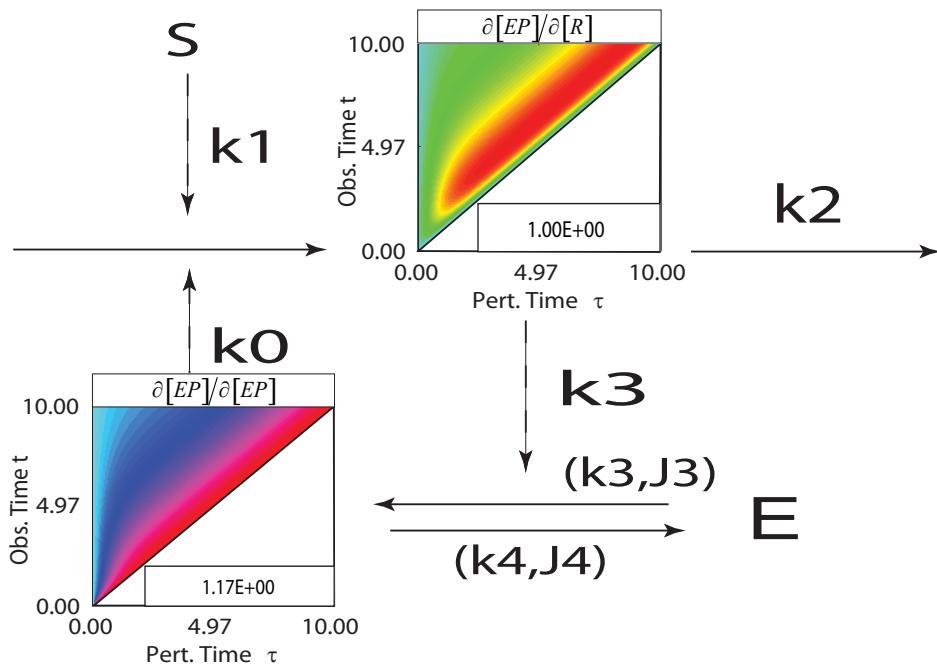


Figure B.1. The heat maps represent the GFM coefficients of output response EP with respect to the perturbations on different molecules in the network, $\partial[EP]/\partial x_j$. For comparison, each heat map is scaled accordingly to have values between -1 and 1 by the scaling factor reported in the lower right corner of the plot.

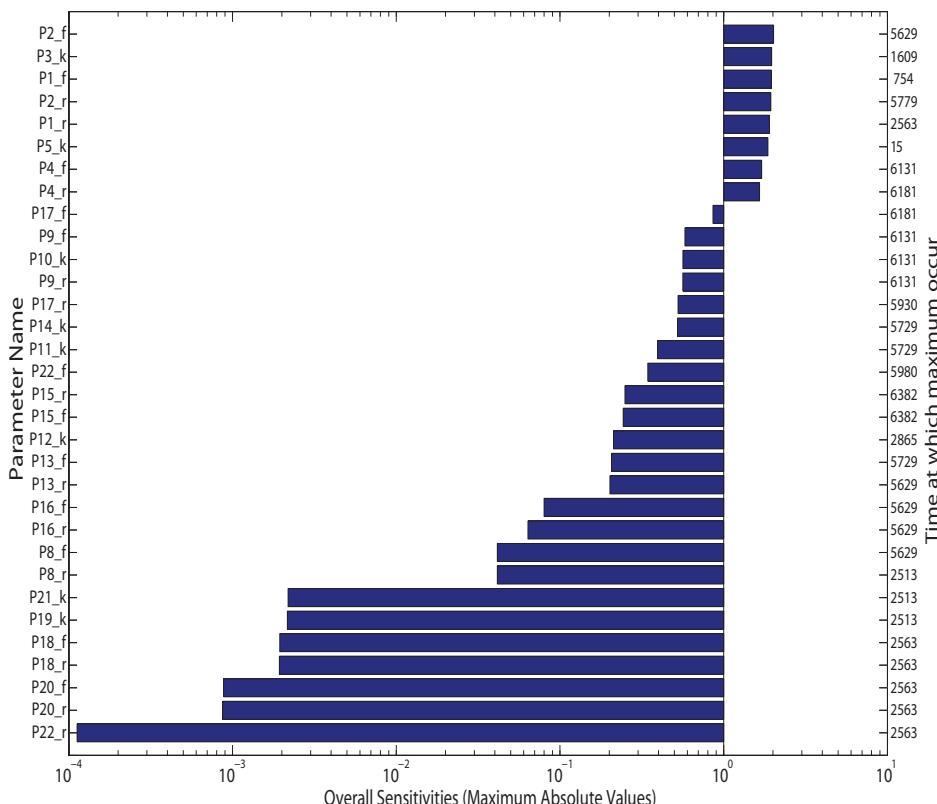


Figure B.2. Bar graph showing parametric sensitivity indices of activated caspase 3 with respect to all kinetic parameters sorted according to their maximum values over the entire simulation time (0 –

10000s). Parameter numbers corresponds to the reaction numbers as shown in Figure 3.5. The right ordinate represents the time at which these maximum occurs.

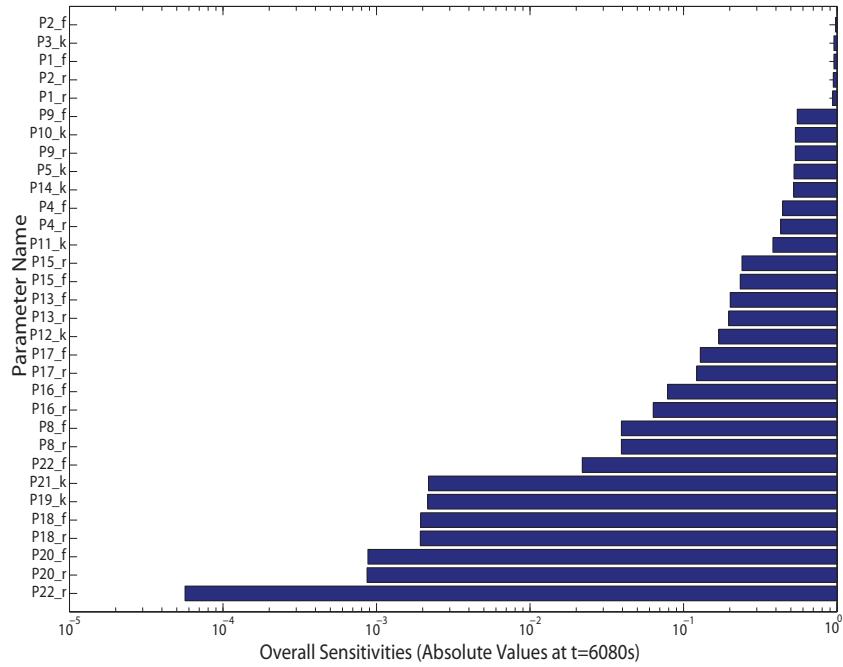


Figure B.3. Bargraph showing absolute value of parametric sensitivity indices of activated caspase3 with respect to kinetic parameters at the switching time ($t = 6080.4$ s). Parameter numbers corresponds to the reaction numbers as shown in Figure 3.5.

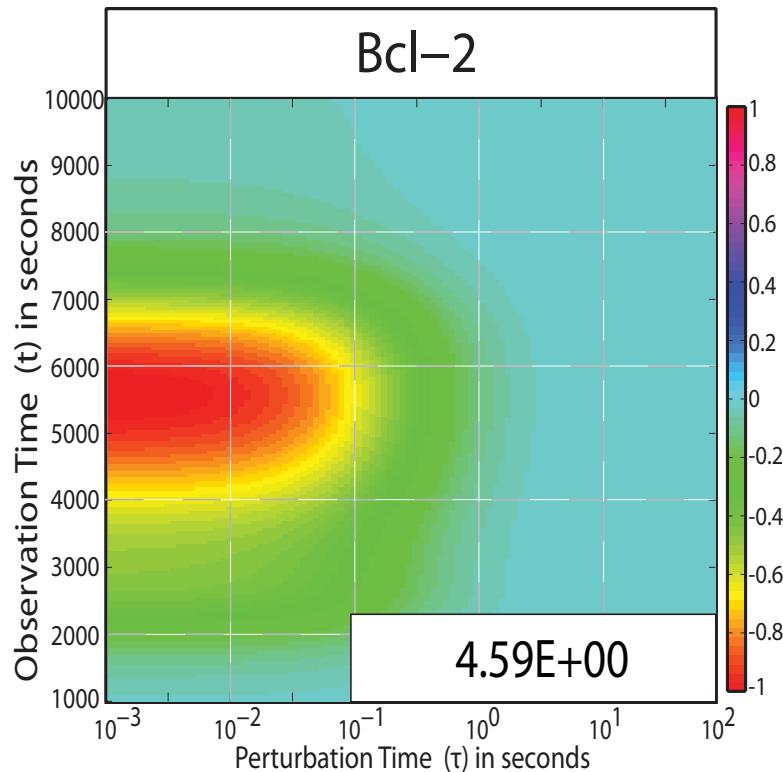


Figure B.4. The heat map represents the GFM coefficient of caspase-3 active level with respect to perturbations on Bcl-2 in the network.

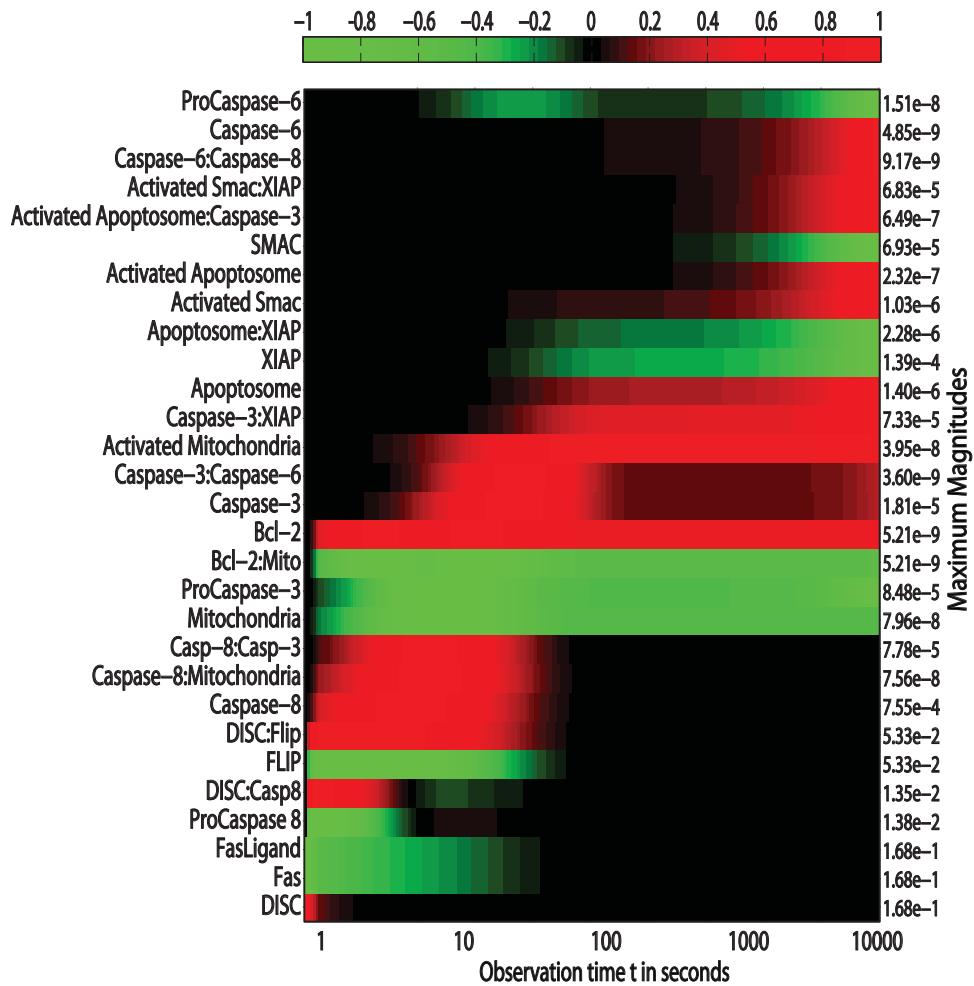


Figure B.5. Complete Signal Progression of FasL Impulse through the Programmed Cell Death Network (detailed version along with the complexes). The signal progression is quantified under zero FasL background. The values are again scaled between -1 and 1 and the scaling factors are listed as the right ordinate.

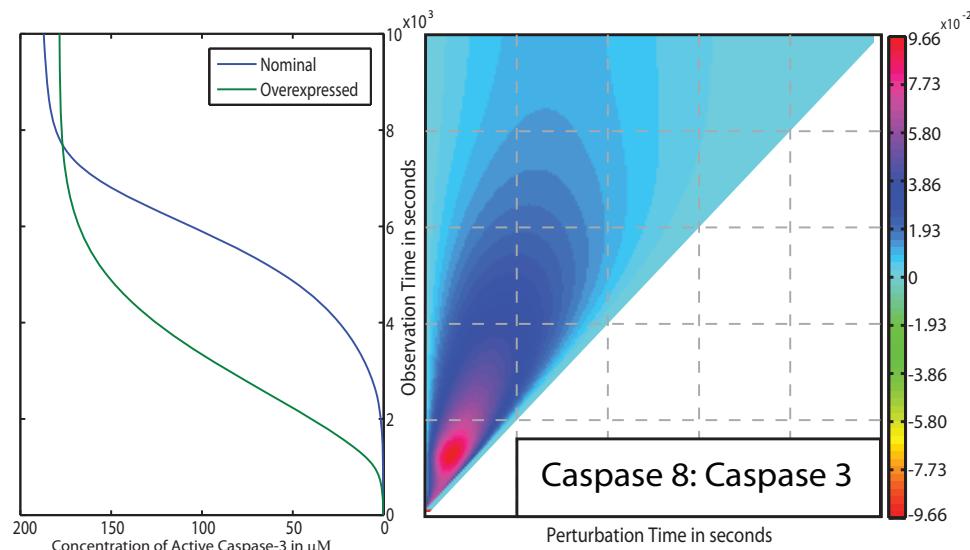


Figure B.6. Overexpression of procaspase-8 and Bcl-2 reroutes the apoptosis regulation through the type-I pathway. The procaspase-8 and Bcl-2 level is increased by 6 and 10 folds, respectively. The shift to the type-I regulation is evident from the heat map of caspase-8:caspase-3 complex, a type-I

molecule. The peak sensitivity of the complex now coincides with the activation time of caspase-3. A complete pathway analysis is given in Figure B.7.

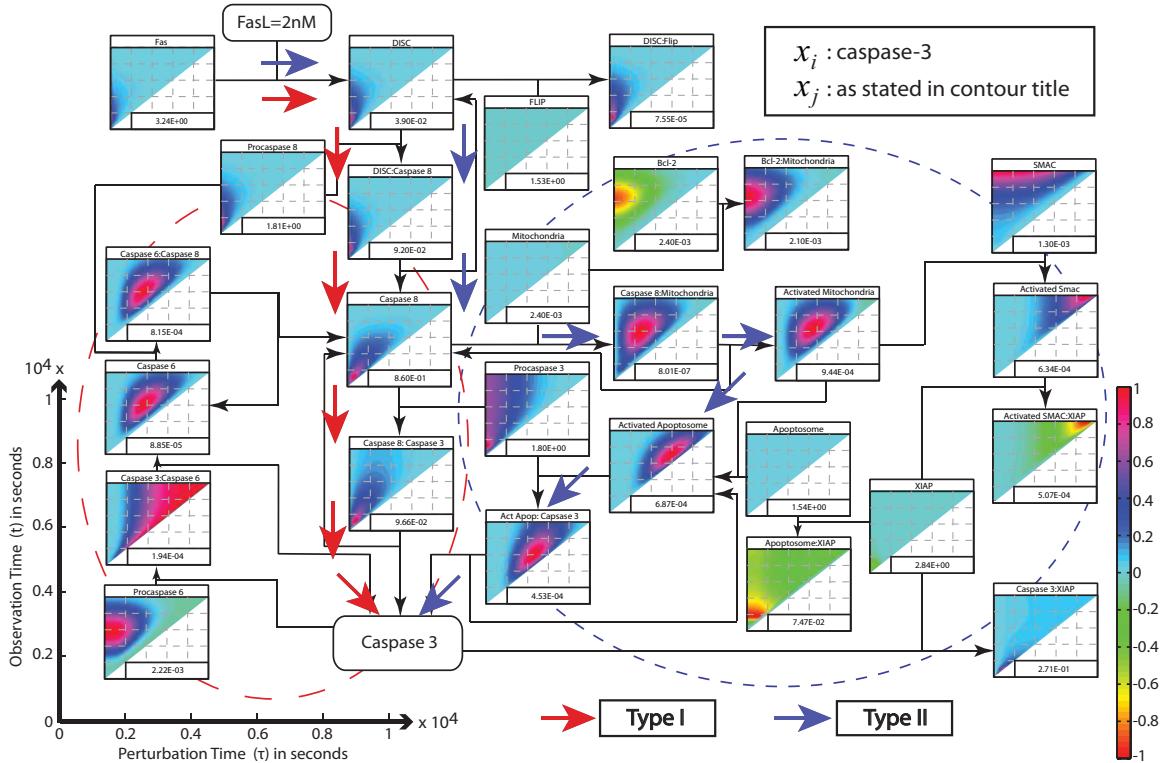


Figure B.7. Greens Function Matrix (GFM) Analysis of Caspase-3 Activation by FasL Stimulus for upregulated pathway. The heat maps represent the GFM coefficients of caspase-3 active level with respect to the perturbations on different molecules in the network, $\partial[\text{caspase-3}]/\partial x_j$. Each heat map is scaled accordingly to have values between -1 and 1 by the scaling factor reported in the lower right corner of the plot. The red arrows illustrate the type-I regulation of caspase-3 activation, while the blue arrows show the type-II mitochondrial-pathway. The analysis showed that the upstream molecules (Fas, DISC, pro- and caspase-8) constituted the early responders to FasL stimulus, as seen in the localization of high sensitivities around low τ values. During the caspase-3 switch (1000 – 4000 seconds), the type-I molecule (caspase8:caspase3 complex) were directly implicated to be the key regulators by the (diagonal) location and magnitude of the peak sensitivities. On the other hand, the type-II pathway has very less contribution to the late activation of caspase-3, which was again illustrated by the location of the peak sensitivities and their maximum values.

Appendix C

C.1 Appendix Figures

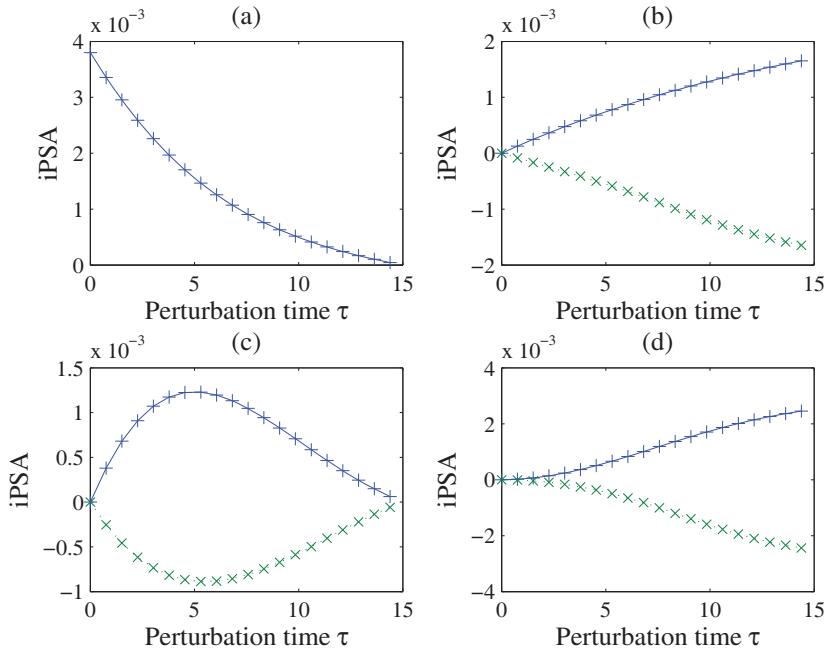


Figure C.1. Impulse Parametric Sensitivity Analysis (iPSA) of x_6 activation in response to x_1 stimulus at switching time ($t = 7.12$). (a) kf1(+), (b) kv2(+) and kk2(x), (c) kv3(+) and kk3(x), (d) kv4(+) and kk4(x)

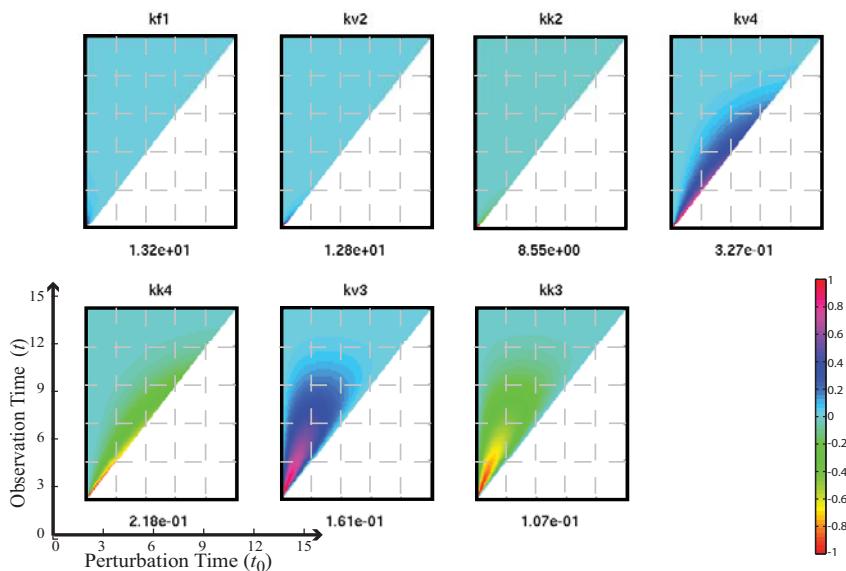


Figure C.2. Impulse Parametric Sensitivity Analysis (iPSA) of x_6 activation in response to x_1 stimulus. Each heatmap illustrates the iPSA coefficient of active x_6 level with respect to perturbations on one parameter in the network, indicated in the title. The x-axis gives the time at which impulse perturbation is applied to the parameter, while the y-axis indicates the observation time of $x_6(t)$. For comparison purpose, each plot is scaled to have values between -1 and +1 by the scaling factor reported in the abscissa. The plots are arranged in the decreasing order of their scaling factors (left to right, top

to bottom). The heat map suggests that: (a) the early response to the stimulus depends on k_{f1}, as indicated by the large sensitivity peak near the *y*-axis (early perturbation is important and the impact is delayed); (b) the early activation of x_6 depends on k_{v2} and k_{k2}, as indicated by the sensitivity peaks along the diagonal (little or no delay in between perturbation and impact) during initial times; (c) the activation depends mainly on k_{v4} and k_{k4}, as indicated by the peaks along the diagonal during the switching time; and finally (d) k_{v3} and k_{k3} control the intermediates, as the peaks lie in between the *y*-axis and the diagonal.

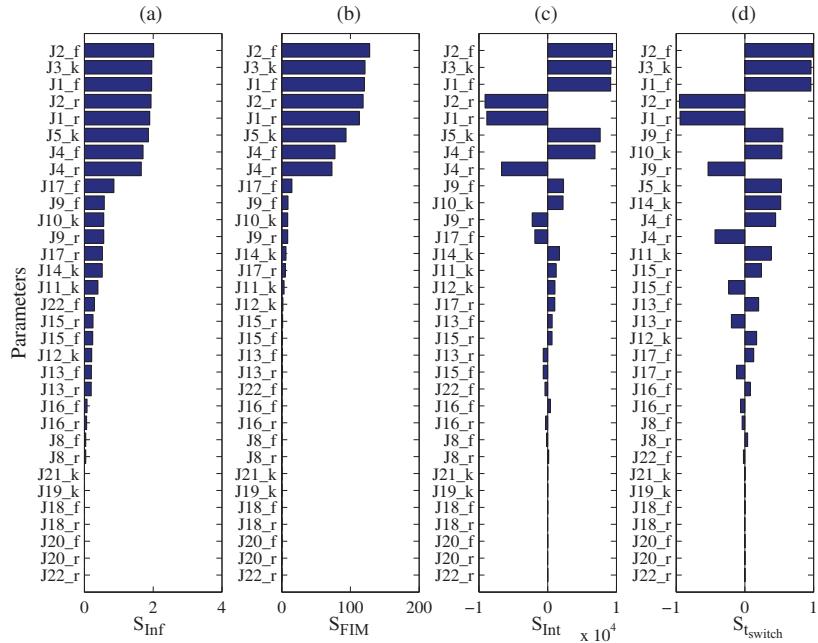


Figure C.3. Local Parametric Sensitivity Analysis of caspase-3 activation under a constant FasL stimulus. (a-c) The bar graphs represent consolidated sensitivity metrics of caspase-3: infinite norm, Fisher Information Matrix (FIM), time integrated sensitivity coefficients, respectively, with respect to the parameters indicated on the *y*-axis. (d) The PSA coefficients of caspase-3 magnitudes with respect to the same parameters at switching time ($t=6060$ s). The parameter numbers refer to the reactions as shown in Figure 3.6, where the subscripts *f* and *r* denote forward and backward rate constants for reversible reactions and *k* denotes the rate constants for irreversible reactions.

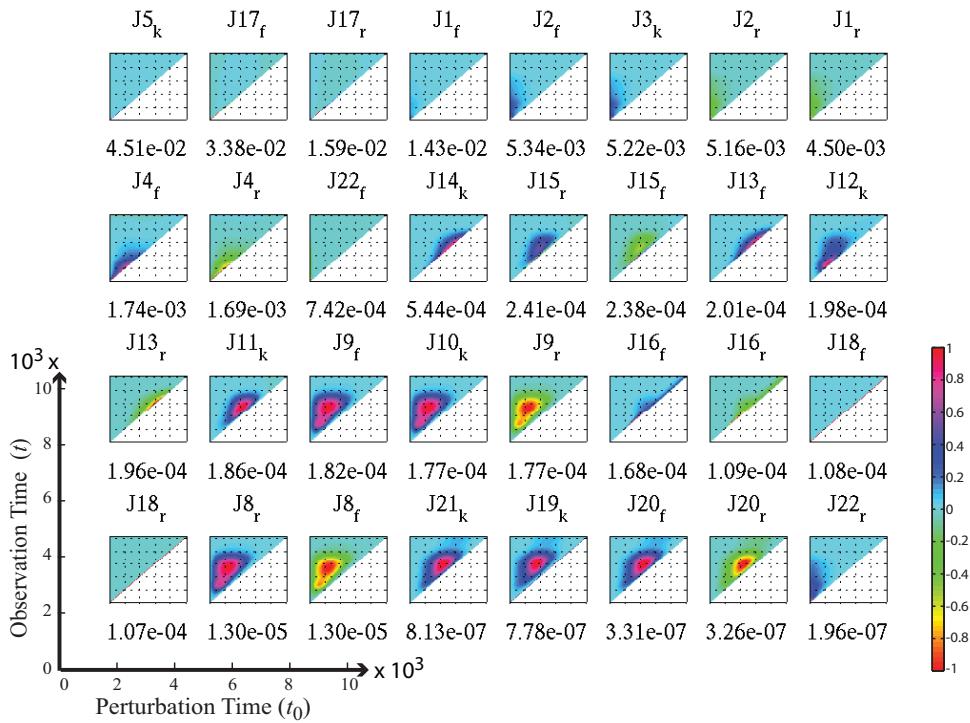


Figure C.4. Impulse Parametric Sensitivity Analysis (iPSA) of caspase-3 activation by a constant FasL (2nM) stimulus. Each heatmap illustrates the iPSA coefficient of active caspase-3 level with respect to perturbations on one parameter in the network, indicated in the title. The x -axis gives the time at which impulse perturbation is applied to the parameter, while the y -axis indicates the observation time of caspase-3(t). For comparison purpose, each plot is scaled to have values between -1 and +1 by the scaling factor reported in the abscissa. The plots are arranged in the decreasing order of their scaling factors (left to right, top to bottom). The heatmap mainly suggests that: (a) the early activation of caspase-3 is due to the type-I pathway, and (b) the procaspase-3 cleaving by type -II pathway is active during the switching. Therefore, the iPSA suggests a type-II dependent activation of caspase-3.

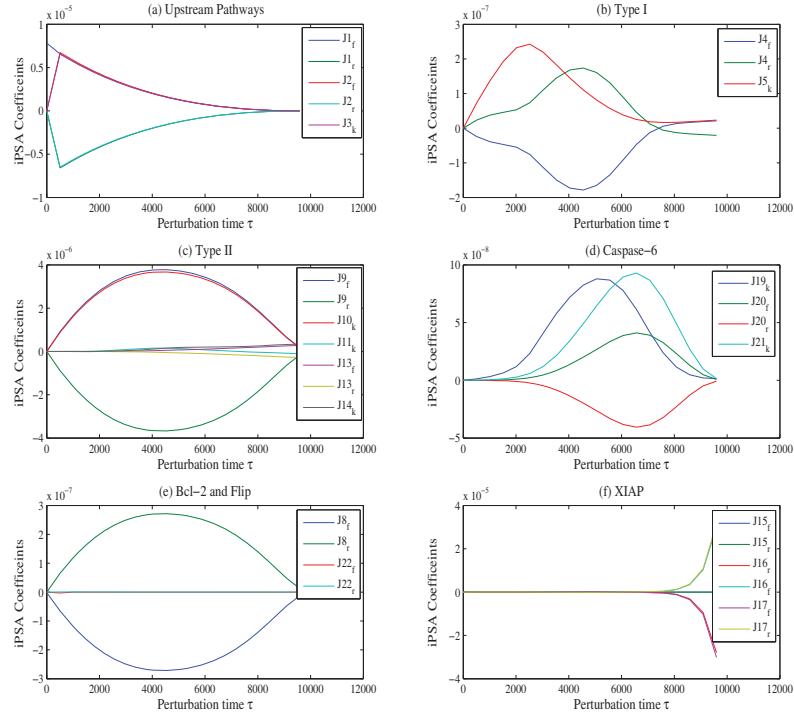


Figure C.5. Impulse Parametric Sensitivity Analysis (iPSA) of caspase-3 at $t = 10,000$ seconds by a constant FasL (2nM) stimulus

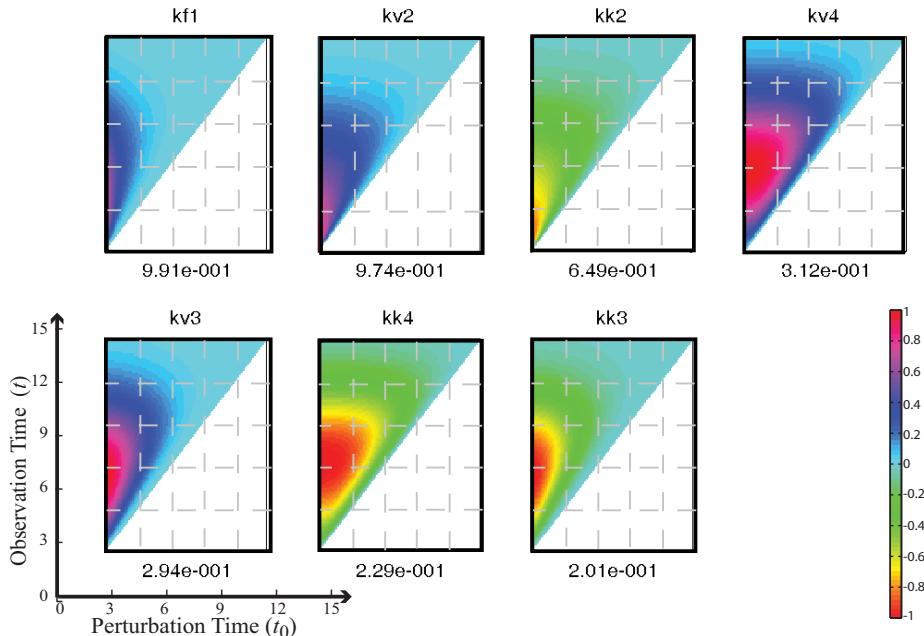


Figure C.6. Local Parametric Sensitivity Analysis (PSA) of x_6 activation in response to x_1 stimulus. Each heatmap illustrates the local PSA coefficient of active x_6 level with respect to perturbations on one parameter in the network, indicated in the title. The x -axis gives the time at which perturbation is applied to the parameter, while the y -axis indicates the observation time of $x_6(t)$. For comparison purpose, each plot is scaled to have values between -1 and +1 by the scaling factor reported in the abscissa. The plots are arranged in the decreasing order of their scaling factors (left to right, top to bottom).

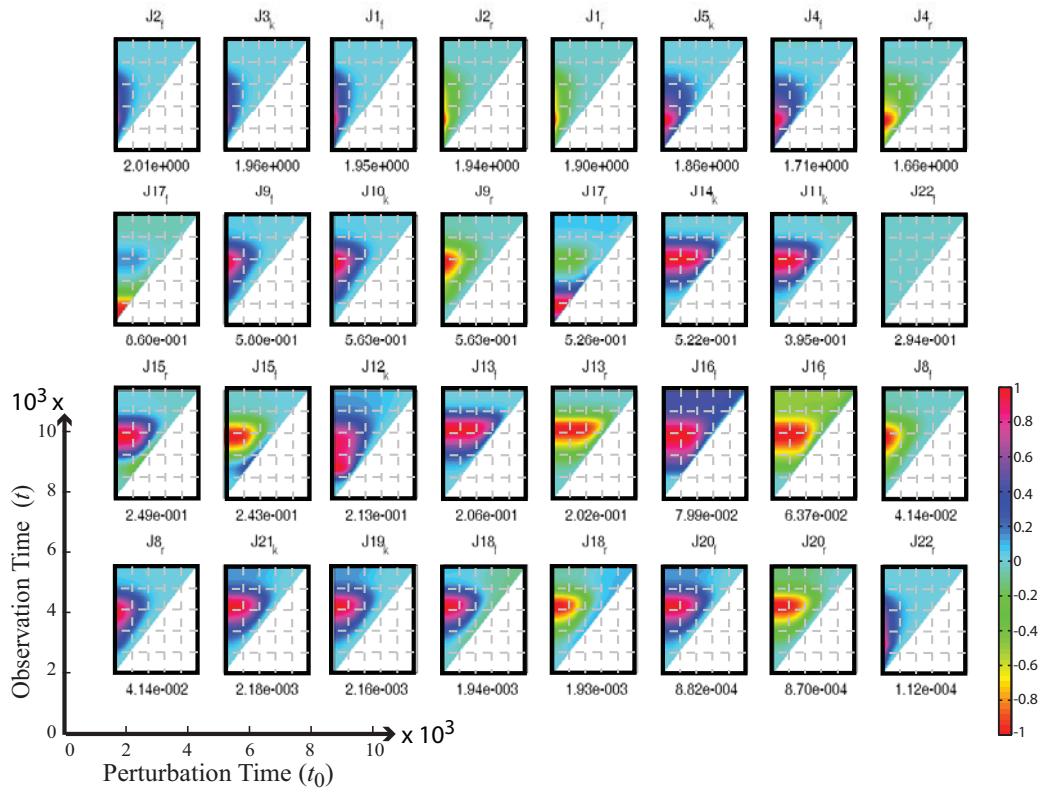


Figure C.7. Local Parametric Sensitivity Analysis (PSA) of caspase-3 activation by a constant FasL (2nM) stimulus. Each heatmap illustrates the local PSA coefficient of active caspase-3 level with respect to perturbations on one parameter in the network, indicated in the title. The x -axis gives the time at which impulse perturbation is applied to the parameter, while the y -axis indicates the observation time of caspase-3(t). For comparison purpose, each plot is scaled to have values between -1 and +1 by the scaling factor reported in the abscissa. The plots are arranged in the decreasing order of their scaling factors (left to right, top to bottom).

Reaction Rates																
	T2-DI-C3I-MI-XI															
	T2-DI-C3I-MI-AI															
	T2-DI-C3I-XI-AI															
	T2-DI-MI-XI-AI															
r1_f	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r1_r			
r1_r	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r1_f			
r2_f	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r2_r			
r2_r	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r2_f			
r3_k	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r3_r			
r4_f													r4_r			
r4_r													r5_k			
r5_k													r8_f			
r8_f	α_{61}	α_{62}		α_{64}	α_{65}	α_{66}	α_{67}		α_{69}	α_{70}	α_{71}	α_{72}	r8_r			
r8_r	α_{61}	α_{62}		α_{64}	α_{65}	α_{66}	α_{67}		α_{69}	α_{70}	α_{71}	α_{72}	r8_f			
r9_f	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r9_r			
r9_r	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r10_k			
r10_k	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r11_k			
r11_k	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r12_k			
r12_k	α_{61}		α_{63}	α_{64}	α_{65}	α_{66}		α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r13_f			
r13_f	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r13_r			
r13_r	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r14_k			
r14_k	α_{61}	α_{62}	α_{63}	α_{64}	α_{65}	α_{66}	α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r15_f			
r15_f		α_{62}	α_{63}	α_{64}	α_{65}		α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r15_r			
r15_r		α_{62}	α_{63}	α_{64}	α_{65}		α_{67}	α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r16_f			
r16_f	α_{61}		α_{63}	α_{64}	α_{65}	α_{66}		α_{68}	α_{69}	α_{70}	α_{71}	α_{72}	r16_r			

r17_f	α61	α62	α63		α65	α66	α67	α68		α70	α71	α72	r17_f
r17_r	α61	α62	α63		α65	α66	α67	α68		α70	α71	α72	r17_r
r18_f					α66	α67	α68	α69	α70		α72		r18_f
r18_r					α66	α67	α68	α69	α70		α72		r18_r
r19_k					α66	α67	α68	α69	α70		α72		r19_k
r20_f					α66	α67	α68	α69	α70		α72		r20_f
r20_r					α66	α67	α68	α69	α70		α72		r20_r
r21_k					α66	α67	α68	α69	α70		α72		r21_k
r22_f	α61	α62	α63	α64		α66	α67	α68	α69		α71	α72	r22_f
r22_r	α61	α62	α63	α64		α66	α67	α68	α69		α71	α72	r22_r

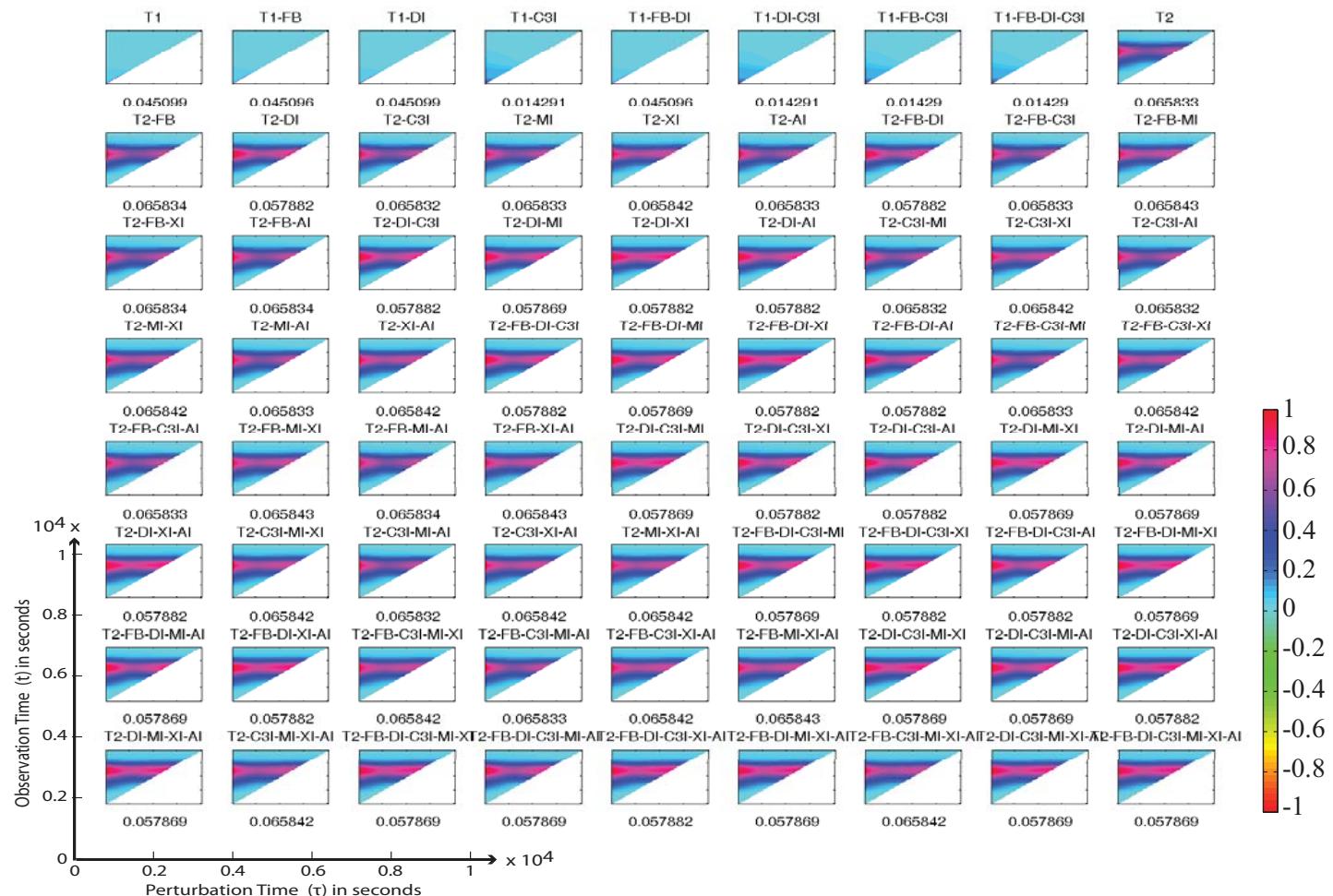


Figure D.1. The pathPSA of caspase-3 activation. Each heat maps illustrate the pathPSA coefficient of active caspase-3 level with respect to perturbations on the pathways, whose names are indicated in the title. The x-axis is the time at which impulse perturbations are applied, while the y-axis indicates the observation time of caspase-3. Each plot is scaled to have values between 0 and ± 1 by the scaling factor given below the heat maps as x-axis label.

Appendix E

E.1 Probability distance metrics

Engineering Metric:

$$\Delta_e f_{\bar{X}_i}(t, \bar{x}_i) = \left| \int_{-\infty}^{\infty} \bar{x}_i^p f_{\bar{X}_i^p}(t, \bar{x}_i^p) - \bar{x}_i^n f_{\bar{X}_i^n}(t, \bar{x}_i^n) d\bar{x}_i \right| \quad (\text{H.2})$$

Jeffrey Divergence:

$$\Delta_J f_{\bar{X}_i}(t, \bar{x}_i) = \int_{-\infty}^{\infty} \left(f_{\bar{X}_i^p}(t, \bar{x}_i^p) \ln f_{\bar{X}_i^p}(t, \bar{x}_i^p) - f_{\bar{X}_i^n}(t, \bar{x}_i^n) \ln f_{\bar{X}_i^n}(t, \bar{x}_i^n) \right) d\bar{x}_i \quad (\text{H.3})$$

Jensen-Shannon Divergence:

$$\Delta_{JS} f_{\bar{X}_i}(t, \bar{x}_i) = \left(\frac{\Delta_{KL} f_{\bar{X}_i^p}(t, \bar{x}_i^p) + \Delta_{KL} f_{\bar{X}_i^n}(t, \bar{x}_i^n)}{2} \right)$$

where

$$\Delta_{KL} f_{\bar{X}_i^p}(t, \bar{x}_i^p) = \int_{-\infty}^{\infty} \left(f_{\bar{X}_i^p}(t, \bar{x}_i^p) \ln \frac{f_{\bar{X}_i^p}(t, \bar{x}_i^p)}{f_{\bar{X}_i^m}(t, \bar{x}_i^m)} \right) d\bar{x}_i \quad (\text{H.4})$$

$$\Delta_{KL} f_{\bar{X}_i^n}(t, \bar{x}_i^n) = \int_{-\infty}^{\infty} \left(f_{\bar{X}_i^n}(t, \bar{x}_i^n) \ln \frac{f_{\bar{X}_i^n}(t, \bar{x}_i^n)}{f_{\bar{X}_i^m}(t, \bar{x}_i^m)} \right) d\bar{x}_i$$

$$\text{and } f_{\bar{X}_i^m}(t, \bar{x}_i^m) = \frac{f_{\bar{X}_i^p}(t, \bar{x}_i^p) + f_{\bar{X}_i^n}(t, \bar{x}_i^n)}{2}$$

Kullback-Leibler Distance:

$$\Delta_{KL} f_{\bar{X}_i}(t, \bar{x}_i) = \int_{-\infty}^{\infty} \left(f_{\bar{X}_i^n}(t, \bar{x}_i^n) \ln \frac{f_{\bar{X}_i^n}(t, \bar{x}_i^n)}{f_{\bar{X}_i^p}(t, \bar{x}_i^p)} \right) d\bar{x}_i \quad (\text{H.5})$$

Kolmogorov-Smirnov Metric:

$$\Delta_{KS} F_{\bar{X}_i}(t, \bar{x}_i) = \sup_{\bar{x}_i \in \mathbb{R}} |F_{\bar{X}_i^p}(t, \bar{x}_i^p) - F_{\bar{X}_i^n}(t, \bar{x}_i^n)| \quad (\text{H.6})$$

E.2 TRAIL induced extrinsic cell death (apoptosis) model of Hela cells:

Solvers		ODE15S of MATLAB
Solver Settings		AbsTol = RelTol = 1e-6
Initial Conditions:	#/cell	Coefficient of variation
L	15000	0
R	200	0.25
L:R	0	0
Rs	0	0
flip	100	0.25
Rs:flip	0	0
C8	20000	0.25
Rs:C8	0	0
C8s	0	0
Bar	1000	0.25
C8s:Bar	0	0
C3	10000	0.282
C8s:C3	0	0
C3s	0	0
C6	10000	0.25
C3s:C6	0	0
C6s	0	0
C6s:C8	0	0
XIAP	100000	0.288
C3s:XIAP	0	0
PARP	1000000	0.25
C3s:PARP	0	0
cPARP	0	0
Bid	40000	0.288
C8s:Bid	0	0
tBid	0	0
Bcl2c	20000	0.25
Bid:Bcl2c	0	0
Bax	100000	0.271
tBid:Bax	0	0
Baxs	0	0
Baxms	0	0
Bcl2	20000	0.294
Baxms:Bcl2	0	0
Bax2	0	0
Bax2:Bcl2	0	0

Bax4	0	0
Bax4:Bcl2	0	0
M	500000	0.25
Bax4:M	0	0
Ms	0	0
CyCm	500000	0.25
Ms:CyCm	0	0
CyCr	0	0
Smacm	100000	0.25
Ms:Smacm	0	0
Smacr	0	0
CyC	0	0
Apaf	100000	0.25
CyC:Apaf	0	0
Apafs	0	0
C9	100000	0.25
Apop	0	0
Apop:C3	0	0
Smac	0	0
Apop:XIAP	0	0
Smac:XIAP	0	0
C3s:Ub	0	0
Pseudo	1	0

Rate Constants:	(#/CC) ⁻¹ .sec ⁻¹	Rate Constants:	sec ⁻¹	Rate Constants:	sec ⁻¹
Kf1	4.00E-07	kr1	1.00E-03	kf2	1.00E-05
Kf3	1.00E-06	kr3	1.00E-03	kf4	1.00E+00
Kf5	1.00E-07	kr5	1.00E-03	kf6	1.00E+00
Kf7	1.00E-06	kr7	1.00E-02	kf8	1.00E+00
Kf9	1.00E-07	kr9	1.00E-03	kf10	1.00E+00
Kf11	1.00E-07	kr11	1.00E-03	kf12	1.00E+00
Kf13	1.00E-02	kr13	1.00E-02	kf17	1.00E+00
Kf14	1.00E-06	kr14	1.00E-03	kf19	1.00E+01
Kf15	1.00E-06	kr15	1.00E-03	kf22	1.00E+01
Kf16	1.00E-06	kr16	1.00E-03	kf25	1.00E+00
Kf18	2.00E-06	kr18	1.00E-03	kf28	1.00E+00
Kf20	1.00E-02	kr20	1.00E-02	kf30	1.00E+00
Kf21	2.00E-06	kr21	1.00E-03	kf32	1.00E+00
Kf23	1.00E-02	kr23	1.00E-02	kf42	1.00E-01
Kf24	5.00E-07	kr24	1.00E-03		
Kf26	5.00E-08	kr26	1.00E-03		
Kf27	5.00E-09	kr27	1.00E-03		
Kf29	1.00E-06	kr29	1.00E-03		
Kf31	3.00E-08	kr31	1.00E-03		

Kf33	1.00E-06	kr33	1.00E-03		
Kf34	1.00E-06	kr34	1.00E-03		
Kf35	1.00E-06	kr35	1.00E-03		
Kf36	1.00E-06	kr36	1.00E-03		
Kf37	1.00E-06	kr37	1.00E-03		
Kf38	1.00E-06	kr38	1.00E-03		
Kf39	7.00E-06	kr39	1.00E-03		
Kf40	2.00E-06	kr40	1.00E-03		
Kf41	2.00E-06	kr41	1.00E-03		

Ratio of mitochondrial volume to cell volume (v) is taken as 0.07

Rate Equations:
r1 = kf1*L*R - kr1*L:R
r2 = kf2*L:R
r3 = kf3*C8*Rs - kr3*Rs:C8
r4 = kf4*Rs:C8
r5 = kf5*C3*C8s - kr5*C8s:C3
r6 = kf6*C8s:C3
r7 = kf7*C3s*PARP - kr7*C3s:PARP
r8 = kf8*C3s:PARP
r9 = kf9*Bid*C8s - kr9*C8s:Bid
r10 = kf10*C8s:Bid
r11 = kf11*Bax*tBid - kr11*tBid:Bax
r12 = kf12*tBid:Bax
r13 = kf13*Baxs - kr13*Baxms
r14 = (kf14*(Baxms*Baxms))/v - kr14*Bax2
r15 = (kf15*(Bax2*Bax2))/v - kr15*Bax4
r16 = (kf16*Bax4*M)/v - kr16*Bax4:M
r17 = kf17*Bax4:M
r18 = (kf18*CyCm*Ms)/v - kr18*Ms:CyCm
r19 = kf19*Ms:CyCm
r20 = kf20*CyCr - kr20*CyCc
r21 = (kf21*Ms*Smacm)/v - kr21*Ms:Smacm
r22 = kf22*Ms:Smacm
r23 = kf23*Smacr - kr23*Smacc
r24 = kf24*Apaf*CyCc - kr24*CyCc:Apaf
r25 = kf25*CyCc:Apaf
r26 = kf26*Apafs*C9 - kr26*Apop
r27 = kf27*Apop*C3 - kr27*Apop:C3
r28 = kf28*Apop:C3
r29 = kf29*C3s*C6 - kr29*C3s:C6
r30 = kf30*C3s:C6
r31 = kf31*C6s*C8 - kr31*C6s:C8

r32 = kf32*C6s:C8
r33 = kf33*Rs*flip - kr33*Rs:flip
r34 = kf34*Bar*C8s - kr34*C8s:Bar
r35 = kf35*Bcl2c*tBid - kr35*tBid:Bcl2c
r36 = (kf36*Baxms*Bcl2)/v - kr36*Baxms:Bcl2
r37 = (kf37*Bax2*Bcl2)/v - kr37*Bax2:Bcl2
r38 = (kf38*Bax4*Bcl2)/v - kr38*Bax4:Bcl2
r39 = kf39*Smacc*XIAP - kr39*Smacc:XIAP
r40 = kf40*Apop*XIAP - kr40*Apop:XIAP
r41 = kf41*C3s*XIAP - kr41*C3s:XIAP
r42 = kf42*C3s:XIAP

RHS Functions/Equations:
d[L]/dt = -r1
d[R]/dt = -r1
d[L:R]/dt = r1 -r2
d[Rs]/dt = r2 -r3 +r4 -r33
d[flip]/dt = -r33
d[Rs:flip]/dt = r33
d[C8]/dt = -r3 -r31
d[Rs:C8]/dt = r3 -r4
d[C8s]/dt = r4 -r5 +r6 -r9 +r10 +r32 -r34
d[Bar]/dt = -r34
d[C8s:Bar]/dt = r34
d[C3]/dt = -r5 -r27
d[C8s:C3]/dt = r5 -r6
d[C3s]/dt = r6 -r7 +r8 +r28 -r29 +r30 -r41
d[C6]/dt = -r29
d[C3s:C6]/dt = r29 -r30
d[C6s]/dt = r30 -r31 +r32
d[C6s:C8]/dt = r31 -r32
d[XIAP]/dt = -r39 -r40 -r41 +r42
d[C3s:XIAP]/dt = r41 -r42
d[PARP]/dt = -r7
d[C3s:PARP]/dt = r7 -r8
d[cPARP]/dt = r8
d[Bid]/dt = -r9
d[C8s:Bid]/dt = r9 -r10
d[tBid]/dt = r10 -r11 +r12 -r35
d[Bcl2c]/dt = -r35
d[tBid:Bcl2c]/dt = r35
d[Bax]/dt = -r11

$d[tBid:Bax]/dt = r11 - r12$
$d[Baxs]/dt = r12 - r13$
$d[Baxms]/dt = r13 - 2*r14 - r36$
$d[Bcl2]/dt = -r36 - r37 - r38$
$d[Baxms:Bcl2]/dt = r36$
$d[Bax2]/dt = r14 - 2*r15 - r37$
$d[Bax2:Bcl2]/dt = r37$
$d[Bax4]/dt = r15 - r16 - r38$
$d[Bax4:Bcl2]/dt = r38$
$d[M]/dt = -r16$
$d[Bax4:M]/dt = r16 - r17$
$d[Ms]/dt = r17 - r18 + r19 - r21 + r22$
$d[CyCm]/dt = -r18$
$d[Ms:CyCm]/dt = r18 - r19$
$d[CyCr]/dt = r19 - r20$
$d[Smacm]/dt = -r21$
$d[Ms:Smacm]/dt = r21 - r22$
$d[Smacr]/dt = r22 - r23$
$d[CyCc]/dt = r20 - r24 + r25$
$d[Apaf]/dt = -r24$
$d[CyCc:Apaf]/dt = r24 - r25$
$d[Apafs]/dt = r25 - r26$
$d[C9]/dt = -r26$
$d[Apop]/dt = r26 - r27 + r28 - r40$
$d[Apop:C3]/dt = r27 - r28$
$d[Smacc]/dt = r23 - r39$
$d[Apop:XIAP]/dt = r40$
$d[Smacc:XIAP]/dt = r39$
$d[C3s:Ub]/dt = r42$
$d[Pseudo:Variable]/dt = 0$

E.3 GFM analysis of TRAIL induced apoptosis

Time lapsed GFM coefficients of cPARP levels with respect to perturbations in all the molecules (rows of GFM matrix) are given in the attached video EV.1. The heat maps are arranged in the order of their infinite matrix norms at a given perturbation and output time. The activation of cPARP, whose mechanistic insight is at query, is given in the first row and first column as a reference. A careful observation of these coefficients paints the following dynamic picture in cPARP activation: (i) Ligand-Receptor complex formation followed by pro-caspase-8 cleavage (to caspase-8) and XIAP inhibition of active caspase-3 are found to be the most important molecules regulating cPARP activation during early phase of pre-MOMP (till 1 hours), (ii) Later transiently (1-2.57 hours), caspase-6 feedback and Bar inhibition of caspase-8 are also found to be important, but not as significant as type-I apoptotic molecules and C3

inhibition-degradation of XIAP, (iii) during MOMP (2.57-2.89 hours), caspase-8 activation of Bid (to tBid) and its competitive inhibition by Bcl2 (both in cytosol and mitochondria), followed by activation of Bax (to Bax*) are found to gain importance than type-I apoptotic molecules, and (iv) finally in post-MOMP (after 2.89 hours), there is no appreciable change in the molecular rankings, mitochondria dependent (type-II) apoptotic molecules, especially competitive tBid activation-inhibition, are found to be the regulatory molecules.

E.4 iPSA analysis of TRAIL induced apoptosis

Impulse Parametric Sensitivity Analysis (iPSA) is applied to the extrinsic cell death model. As in GFM analysis, iPSA also uses sensitivity metrics, to reveal the dynamic information about the actively participating reactions that gives rise to the output behavior. Simulation conditions remain the same as that of GFM analysis. The model simulates an apoptotic cell, and the purpose of the analysis is to elucidate the key reactions in the regulation of cPARP activation during extrinsic cell death. By choosing cPARP activation as the output of interest, the iPSA sensitivities of cPARP level to perturbations on different parameters of the network at different perturbation times (τ) are illustrated in the video EV.2. Each subplot here represents an element in the row of iPSA matrix, corresponding to the active cPARP. The heat maps are arranged in the order of their infinite matrix norms over a given perturbation and output time. As in GFM analysis, even iPSA analysis is split into two phases: pre-MOMP (before 2.57 hours) and post-MOMP (after 2.57 hours). A careful observation of iPSA coefficients reveals the following dynamic picture in cPARP activation: (i) Activation of receptor, pro-caspase-8 (to C8*), and pro-caspase-3 (to C3*) followed by the inhibition of C3* by XIAP and C8* by Bar are found to be the important reactions in pre-MOMP (till 2.57 hours), (ii) other than most of the reactions of pre-MOMP, post-MOMP is characterized by the activation of Bid (to tBid), Bax (to Bax*) and the Bcl2 inhibition of tBid, (iii) in both the phases, C6 feed-back loop is found to be insignificant. The major difference between GFM and iPSA analysis lies in the proteasome mediated degradation of XIAP, which cannot be inferred from GFM coefficients but found not so important from the iPSA coefficients. Hence the XIAP inhibition and not the degradation of C3* which controls the pre-MOMP phase.

E.5 eFAST analysis of TRAIL induced apoptosis (Whole population)

Time lapsed first and total order eFAST sensitivity coefficients of cPARP activation levels to the eighteen non-zero initial conditions and a pseudo variable are given in the attached video EV.3. Bar graphs (a-b) and (c-d) are arranged in the decreasing order of eFAST sensitivities at a given time and integrated till the given time, respectively, where the given time is specified in the top of the chart. Investigation of these bar graphs paints the following dynamic picture in cPARP activation: (i) first order and total contributions of pro-caspase-8 activation to caspase-8 by receptor and caspase-3 activation by caspase-8 and its subsequent inhibition by XIAP are the important molecules which contributes to the output variability during pre-MOMP (till 1.77 hours), (ii) Later transiently during MOMP (1.77 hours to 4.23 hours), total order sensitivities of bid activation to tbid and its subsequent inhibition by Bcl2 and

Bax are found to be important, (iii) finally post-MOMP (after 4.23 hours), XIAP and membrane bound Smac are found to be important for the cells to avoid apoptosis. There is no appreciable change in the eFAST sensitivities of pro-caspase-6 and hence the corresponding feedback loop renders no appreciable change to the output variability.

E.6 eFAST analysis of TRAIL induced apoptosis (apoptotic and non-apoptotic cell population)

Similar to eFAST analysis of the whole population, a comparative eFAST analysis was performed for a segregated apoptotic and non-apoptotic cell population. Time lapsed first and total order eFAST sensitivity coefficients of cPARP activation levels to the eighteen non-zero initial conditions and a pseudo variable are given in the attached video EV.4. Bar graphs (a-b) and (c-d) indicates the apoptotic and non-apoptotic cell sensitivities respectively. Investigation of these bar graphs gives the following dynamic picture in cPARP activation: pre-MOMP contributions remained the same in both the cells. Whereas post-MOMP in non-apoptotic cells are characterized by a significant change in XIAP and membrane bound Smac.

E.7 MDP analysis of TRAIL induced apoptosis (Whole population)

Like GFM and iPSA analysis, time lapse heat maps of MDP coefficients of cPARP with respect to 10% (mean increase) perturbation in all the molecules (rows of MDP matrix) are given in the attached video EV.5. The heat maps are normalized using their absolute maximum over perturbation and observation time. They are arranged in the order of decreasing importance. The activation dynamics of cPARP (i.e., 1% and 99%) in the Hela cell population, whose mechanistic insight is at query, is given in the first row and first column as a reference. The analysis is divided into three major dynamic regions: (i) pre-MOMP (till 1.77 hours), which is characterized by at least 1% of mitochondrial pore opening, (ii) MOMP (from 1.77 hours to 4.28 hours), from 1% pore opening till 99% of CyC/Smac release, and (iii) post-MOMP (from 4.28 hours to steady state). A careful observation of these coefficients paints the following dynamic picture in cPARP activation in Hela cells population: (i) Pre-MOMP period is characterized by ligand-receptor complex formation followed by pro-caspase-8 cleavage (to caspase-8) and XIAP inhibition of active caspase-3, of which XIAP inhibition is found to be most important, (ii) Later transiently, for a very short period, caspase-6 feedback and Bar inhibition of caspase-8 are also found to be important, but not as significant as type-I apoptotic molecules and C3 inhibition-degradation of XIAP, (iii) during MOMP (1.77-4.23 hours), mitochondrial pore opening and the subsequent release of cytochrome-c and Smac into cytosol are found to gain importance than type-I apoptotic molecules, and (iv) finally in post-MOMP (after 4.23 hours), there is no appreciable change in the molecular ranking. Therefore, MDP analysis points to mitochondrial pore opening and subsequent downstream events to be highly important in cPARP activation.

E.8 MDP analysis of TRAIL induced apoptosis (apoptotic and non-apoptotic cells)

MDP analysis of both the apoptotic and non-apoptotic cell population is performed separately and compared. Time lapse comparison of the infinite norms of MDP coefficients of cPARP activation with respect to 10% (mean increase-decrease) perturbation in all the molecules (rows of MDP matrix) are given in the attached video EV.6. They are arranged in the order of decreasing importance. Since 95% of the whole population has undergone apoptosis, MDP analysis of apoptotic cell population is same as that of the whole population. Whereas, MDP analysis of non-apoptotic cell population paints a different picture with respect to cPAPR activation. Pre-MOMP regimen of non-apoptotic cells are characterized by the same molecules as that of apoptotic cells, which is also evident from their rank order correlation. On the other hand, MOMP is significantly delayed in non-apoptotic cell population. This is due to the differential regulation of XIAP, Smac and Apaf, as seen in EV.6.

Appendix F

Table F.1: Low Temperature Alkane Pyrolysis [188]

Methods*	Species Remaining	Reactions Remaining	% Error	Pareto Optimal Distance
Whole Model	38	98	0.000	-
(a): Based on Pareto Distance (Parameter vs Error)				
iPSA-GFM (SQ)	14	16	15.820	0.227
PSA-GFM (SQ)	15	22	8.203	0.239
iPSA	15	23	11.860	0.263
iPSA-GFM (BU)	15	23	11.860	0.263
iPSA-GFM (TD)	15	23	11.860	0.263
PSA	18	33	7.558	0.345
PSA-GFM (BU)	18	33	7.558	0.345
PSA-GFM (TD)	16	29	17.867	0.346
GFM	29	63	10.822	0.652
(b): Fixed Reactions (38)				
iPSA	19	35	1.613	0.358
iPSA-GFM (BU)	19	35	1.613	0.358
iPSA-GFM (TD)	19	36	1.777	0.368
PSA	21	38	3.181	0.389
PSA-GFM (BU)	21	38	3.181	0.389
iPSA-GFM (SQ)	17	37	10.160	0.391
PSA-GFM (SQ)	17	38	6.934	0.394
PSA-GFM (TD)	21	38	7.063	0.394
GFM	24	34	76.940	0.844
(c): Fixed Error (0.5%)				
iPSA	21	43	0.381	0.439
iPSA-GFM (BU)	21	43	0.381	0.439
iPSA-GFM (TD)	21	43	0.381	0.439
iPSA-GFM (SQ)	20	54	0.402	0.551
PSA-GFM (TD)	24	55	0.273	0.561
PSA-GFM (SQ)	20	55	0.406	0.561
PSA	24	55	0.469	0.561
PSA-GFM (BU)	24	55	0.469	0.561
GFM	30	81	0.175	0.827
(d): Fixed States (34)				
GFM	31	87	0.000	0.888
PSA	34	92	0.000	0.939
iPSA	34	92	0.000	0.939
PSA-GFM (BU)	34	92	0.000	0.939
PSA-GFM (TD)	34	92	0.000	0.939
iPSA-GFM (BU)	34	92	0.000	0.939
iPSA-GFM (TD)	34	92	0.000	0.939

PSA-GFM (SQ)	34	94	9.131	0.964
iPSA-GFM (SQ)	34	94	9.131	0.964

* Sorted based on the optimal Pareto distance

Table F.2: Natural Gas Combustion - GRI Mech 3.0 [189]				
Methods*	Species Remaining	Reactions Remaining	% Error	Pareto Optimal Distance
Whole Model	53	650	0.000	-
(a): Based on Pareto Distance (Parameter vs Error)				
iPSA-GFM (SQ)	39	129	22.462	0.300
PSA-GFM (SQ)	45	206	20.804	0.379
iPSA	49	253	20.074	0.438
iPSA-GFM (BU)	49	253	20.074	0.438
iPSA-GFM (TD)	49	253	20.074	0.438
PSA	51	329	18.332	0.538
PSA-GFM (BU)	51	330	21.062	0.550
PSA-GFM (TD)	51	330	21.062	0.550
GFM	3	1	74.161	0.742
(b): Fixed Reactions (450)				
iPSA-GFM (SQ)	43	233	16.462	0.394
iPSA-GFM (BU)	52	321	15.426	0.517
iPSA-GFM (TD)	52	325	15.406	0.523
iPSA	52	329	15.417	0.529
PSA-GFM (SQ)	49	360	11.061	0.565
PSA-GFM (TD)	52	444	0.244	0.683
PSA	52	449	0.979	0.691
PSA-GFM (BU)	52	449	0.979	0.691
GFM	18	19	74.159	0.742
(c): Fixed Error (0.5%)				
PSA-GFM (TD)	52	444	0.244	0.683
PSA	52	447	0.980	0.688
PSA-GFM (BU)	52	447	0.980	0.688
iPSA	52	453	0.736	0.697
iPSA-GFM (BU)	52	453	0.736	0.697
iPSA-GFM (TD)	52	453	0.736	0.697
PSA-GFM (SQ)	52	638	0.512	0.982
iPSA-GFM (SQ)	52	638	0.512	0.982
GFM	53	650	0.000	1.000

* Sorted based on the optimal Pareto distance

Table F.3: Ethane Pyrolysis Model [60]		
Reaction Number	Reactions	Status in Reduced Model
Radical recombination and scission reactions		
1	$\cdot\text{CH}_3 + \cdot\text{CH}_3 \rightarrow \text{C}_2\text{H}_6$	Retained
2	$\text{C}_2\text{H}_6 \rightarrow \cdot\text{CH}_3 + \cdot\text{CH}_3$	Retained

3	$\cdot\text{C}_2\text{H}_5 + \cdot\text{C}_2\text{H}_5 \rightarrow \text{C}_4\text{H}_{10}$	Retained
4	$\text{C}_4\text{H}_{10} \rightarrow \cdot\text{C}_2\text{H}_5 + \cdot\text{C}_2\text{H}_5$	Removed
5	$\cdot\text{CH}_3 + \cdot\text{C}_2\text{H}_5 \rightarrow \text{C}_3\text{H}_8$	Retained
6	$\text{C}_3\text{H}_8 \rightarrow \cdot\text{CH}_3 + \cdot\text{C}_2\text{H}_5$	Removed
7	$\cdot\text{C}_2\text{H}_5 + \cdot\text{H} \rightarrow \text{C}_2\text{H}_6$	Retained
8	$\text{C}_2\text{H}_6 \rightarrow \cdot\text{C}_2\text{H}_5 + \cdot\text{H}$	Retained
9	$\cdot\text{H} + \cdot\text{C}_2\text{H}_3 \rightarrow \text{C}_2\text{H}_4$	Removed
10	$\text{C}_2\text{H}_4 \rightarrow \cdot\text{H} + \cdot\text{C}_2\text{H}_3$	Retained
11	$\cdot\text{C}_3\text{H}_5 + \cdot\text{H} \rightarrow \text{C}_3\text{H}_6$	Retained
12	$\text{C}_3\text{H}_6 \rightarrow \cdot\text{C}_3\text{H}_5 + \cdot\text{H}$	Retained
13	$\cdot\text{H} + \cdot\text{C}_3\text{H}_7 \rightarrow \text{C}_3\text{H}_8$	Removed
14	$\text{C}_3\text{H}_8 \rightarrow \cdot\text{H} + \cdot\text{C}_3\text{H}_7$	Removed
15	$\cdot\text{CH}_3 + \cdot\text{C}_2\text{H}_3 \rightarrow \text{C}_3\text{H}_6$	Removed
16	$\text{C}_3\text{H}_6 \rightarrow \cdot\text{CH}_3 + \cdot\text{C}_2\text{H}_3$	Retained
17	$\cdot\text{H} + \cdot\text{H} + \text{M} \rightarrow \text{H}_2 + \text{M}$	Removed
18	$\text{H}_2 \rightarrow \cdot\text{H} + \cdot\text{H}$	Removed
19	$\cdot\text{C}_2\text{H}_3 + \cdot\text{C}_2\text{H}_3 \rightarrow \text{C}_4\text{H}_6$	Removed
20	$\text{C}_4\text{H}_6 \rightarrow \cdot\text{C}_2\text{H}_3 + \cdot\text{C}_2\text{H}_3$	Removed
21	$\cdot\text{C}_2\text{H}_3 + \cdot\text{C}_2\text{H}_5 \rightarrow \text{C}_4\text{H}_8$	Retained
22	$\text{C}_4\text{H}_8 \rightarrow \cdot\text{C}_2\text{H}_3 + \cdot\text{C}_2\text{H}_5$	Removed
23	$\cdot\text{CH}_3 + \cdot\text{H} \rightarrow \text{CH}_4$	Retained
24	$\text{CH}_4 \rightarrow \cdot\text{CH}_3 + \cdot\text{H}$	Removed
25	$\cdot\text{H} + \cdot\text{C}_4\text{H}_9 \rightarrow \text{C}_4\text{H}_{10}$	Removed
26	$\text{C}_4\text{H}_{10} \rightarrow \cdot\text{H} + \cdot\text{C}_4\text{H}_9$	Removed
27	$\cdot\text{H} + \cdot\text{2-C}_4\text{H}_9 \rightarrow \text{C}_4\text{H}_{10}$	Removed
28	$\text{C}_4\text{H}_{10} \rightarrow \cdot\text{H} + \cdot\text{2-C}_4\text{H}_9$	Removed
29	$\cdot\text{H} + \cdot\text{2-C}_3\text{H}_7 \rightarrow \text{C}_3\text{H}_8$	Removed
30	$\text{C}_3\text{H}_8 \rightarrow \cdot\text{H} + \cdot\text{2-C}_3\text{H}_7$	Removed
31	$\cdot\text{H} + \cdot\text{C}_4\text{H}_7 \rightarrow \text{C}_4\text{H}_8$	Removed
32	$\text{C}_4\text{H}_8 \rightarrow \cdot\text{H} + \cdot\text{C}_4\text{H}_7$	Removed
33	$\cdot\text{CH}_3 + \cdot\text{C}_3\text{H}_7 + \text{C}_4\text{H}_{10}$	Removed
34	$\text{C}_4\text{H}_{10} \rightarrow \cdot\text{CH}_3 + \cdot\text{C}_3\text{H}_7$	Removed
35	$\cdot\text{CH}_3 + \cdot\text{C}_3\text{H}_5 \rightarrow \text{C}_4\text{H}_8$	Retained
36	$\text{C}_4\text{H}_8 \rightarrow \cdot\text{CH}_3 + \cdot\text{C}_3\text{H}_5$	Retained

Hydrogen abstraction reactions

37	$\text{C}_2\text{H}_6 + \cdot\text{CH}_3 \rightarrow \cdot\text{C}_2\text{H}_5 + \text{CH}_4$	Retained
38	$\text{CH}_4 + \cdot\text{C}_2\text{H}_5 \rightarrow \text{C}_2\text{H}_6 + \cdot\text{CH}_3$	Removed
39	$\text{CH}_4 + \cdot\text{C}_2\text{H}_3 \rightarrow \text{C}_2\text{H}_4 + \cdot\text{CH}_3$	Removed
40	$\cdot\text{CH}_3 + \text{C}_2\text{H}_4 \rightarrow \text{CH}_4 + \cdot\text{C}_2\text{H}_3$	Retained
41	$\text{C}_2\text{H}_6 + \cdot\text{H} \rightarrow \cdot\text{C}_2\text{H}_5 + \text{H}_2$	Retained
42	$\cdot\text{C}_2\text{H}_5 + \text{H}_2 \rightarrow \text{C}_2\text{H}_6 + \cdot\text{H}$	Retained
43	$\cdot\text{C}_2\text{H}_3 + \text{H}_2 \rightarrow \text{C}_2\text{H}_4 + \cdot\text{H}$	Retained
44	$\text{C}_2\text{H}_4 + \cdot\text{H} \rightarrow \text{H}_2 + \cdot\text{C}_2\text{H}_3$	Retained
45	$\cdot\text{CH}_3 + \text{H}_2 \rightarrow \text{CH}_4 + \cdot\text{H}$	Retained

46	$\text{CH}_4 + \cdot\text{H} \rightarrow \cdot\text{CH}_3 + \text{H}_2$	Removed
47	$\text{C}_4\text{H}_{10} + \cdot\text{H} \rightarrow \text{H}_2 + \cdot\text{C}_4\text{H}_9$	Removed
48	$\cdot\text{C}_4\text{H}_9 + \text{H}_2 \rightarrow \text{C}_4\text{H}_{10} + \cdot\text{H}$	Removed
49	$\text{C}_4\text{H}_{10} + \cdot\text{H} \rightarrow \text{H}_2 + \cdot 2\text{-C}_4\text{H}_9$	Retained
50	$\cdot 2\text{-C}_4\text{H}_9 + \text{H}_2 \rightarrow \text{C}_4\text{H}_{10} + \cdot\text{H}$	Removed
51	$\text{C}_3\text{H}_8 + \cdot\text{H} \rightarrow \text{H}_2 + \cdot\text{C}_3\text{H}_7$	Removed
52	$\cdot\text{C}_3\text{H}_7 + \text{H}_2 \rightarrow \text{C}_3\text{H}_8 + \cdot\text{H}$	Removed
53	$\text{C}_3\text{H}_8 + \cdot\text{H} \rightarrow \text{H}_2 + \cdot 2\text{-C}_3\text{H}_7$	Retained
54	$\cdot 2\text{-C}_3\text{H}_7 + \text{H}_2 \rightarrow \text{C}_3\text{H}_8 + \cdot\text{H}$	Removed
55	$\text{C}_3\text{H}_6 + \cdot\text{H} \rightarrow \text{H}_2 + \cdot\text{C}_3\text{H}_5$	Retained
56	$\cdot\text{C}_3\text{H}_5 + \text{H}_2 \rightarrow \text{C}_3\text{H}_6 + \cdot\text{H}$	Retained
57	$\text{C}_4\text{H}_8 + \cdot\text{H} \rightarrow \text{H}_2 + \cdot\text{C}_4\text{H}_7$	Removed
58	$\cdot\text{C}_4\text{H}_7 + \text{H}_2 \rightarrow \text{C}_4\text{H}_8 + \cdot\text{H}$	Removed
59	$\cdot\text{CH}_3 + \text{C}_4\text{H}_{10} \rightarrow \text{CH}_4 + \cdot\text{C}_4\text{H}_9$	Removed
60	$\text{CH}_4 + \cdot\text{C}_4\text{H}_9 \rightarrow \text{C}_4\text{H}_{10} + \cdot\text{CH}_3$	Removed
61	$\cdot\text{CH}_3 + \text{C}_4\text{H}_{10} \rightarrow \text{CH}_4 + \cdot 2\text{-C}_4\text{H}_9$	Removed
62	$\text{CH}_4 + \cdot 2\text{-C}_4\text{H}_9 \rightarrow \text{C}_4\text{H}_{10} + \cdot\text{CH}_3$	Removed
63	$\cdot\text{CH}_3 + \text{C}_3\text{H}_8 \rightarrow \text{CH}_4 + \cdot\text{C}_3\text{H}_7$	Removed
64	$\text{CH}_4 + \cdot\text{C}_3\text{H}_7 \rightarrow \text{C}_3\text{H}_8 + \cdot\text{CH}_3$	Removed
65	$\cdot\text{CH}_3 + \text{C}_3\text{H}_8 \rightarrow \text{CH}_4 + \cdot 2\text{-C}_3\text{H}_7$	Removed
66	$\text{CH}_4 + \cdot 2\text{-C}_3\text{H}_7 \rightarrow \text{C}_3\text{H}_8 + \cdot\text{CH}_3$	Removed
67	$\cdot\text{CH}_3 + \text{C}_3\text{H}_6 \rightarrow \text{CH}_4 + \cdot\text{C}_3\text{H}_5$	Retained
68	$\text{CH}_4 + \cdot\text{C}_3\text{H}_5 \rightarrow \text{C}_3\text{H}_6 + \cdot\text{CH}_3$	Removed
69	$\cdot\text{CH}_3 + \text{C}_4\text{H}_8 \rightarrow \text{CH}_4 + \cdot\text{C}_4\text{H}_7$	Removed
70	$\text{CH}_4 + \cdot\text{C}_4\text{H}_7 \rightarrow \text{C}_4\text{H}_8 + \cdot\text{CH}_3$	Removed
71	$\text{C}_2\text{H}_6 + \cdot\text{C}_2\text{H}_3 \rightarrow \text{C}_2\text{H}_4 + \cdot\text{C}_2\text{H}_5$	Retained
72	$\text{C}_2\text{H}_4 + \cdot\text{C}_2\text{H}_5 \rightarrow \text{C}_2\text{H}_6 + \cdot\text{C}_2\text{H}_3$	Retained
73	$\text{C}_2\text{H}_6 + \cdot\text{C}_4\text{H}_9 \rightarrow \text{C}_4\text{H}_{10} + \cdot\text{C}_2\text{H}_5$	Removed
74	$\text{C}_4\text{H}_{10} + \cdot\text{C}_2\text{H}_5 \rightarrow \text{C}_2\text{H}_6 + \cdot\text{C}_4\text{H}_9$	Removed
75	$\text{C}_4\text{H}_{10} + \cdot\text{C}_2\text{H}_5 \rightarrow \text{C}_2\text{H}_6 + \cdot 2\text{-C}_4\text{H}_9$	Removed
76	$\text{C}_2\text{H}_6 + \cdot 2\text{-C}_4\text{H}_9 \rightarrow \text{C}_4\text{H}_{10} + \cdot\text{C}_2\text{H}_5$	Retained
77	$\text{C}_2\text{H}_6 + \cdot\text{C}_3\text{H}_7 \rightarrow \text{C}_3\text{H}_8 + \cdot\text{C}_2\text{H}_5$	Retained
78	$\text{C}_3\text{H}_8 + \cdot\text{C}_2\text{H}_5 \rightarrow \text{C}_2\text{H}_6 + \cdot\text{C}_3\text{H}_7$	Removed
79	$\text{C}_3\text{H}_8 + \cdot\text{C}_2\text{H}_5 \rightarrow \text{C}_2\text{H}_6 + \cdot 2\text{-C}_3\text{H}_7$	Retained
80	$\text{C}_2\text{H}_6 + \cdot 2\text{-C}_3\text{H}_7 \rightarrow \text{C}_3\text{H}_8 + \cdot\text{C}_2\text{H}_5$	Removed
81	$\text{C}_3\text{H}_6 + \cdot\text{C}_2\text{H}_5 \rightarrow \text{C}_2\text{H}_6 + \cdot\text{C}_3\text{H}_5$	Retained
82	$\text{C}_2\text{H}_6 + \cdot\text{C}_3\text{H}_5 \rightarrow \text{C}_3\text{H}_6 + \cdot\text{C}_2\text{H}_5$	Retained
83	$\text{C}_2\text{H}_6 + \cdot\text{C}_4\text{H}_7 \rightarrow \text{C}_4\text{H}_8 + \cdot\text{C}_2\text{H}_5$	Retained
84	$\text{C}_4\text{H}_8 + \cdot\text{C}_2\text{H}_5 \rightarrow \text{C}_2\text{H}_6 + \cdot\text{C}_4\text{H}_7$	Removed
85	$\text{C}_4\text{H}_{10} + \cdot\text{C}_2\text{H}_3 \rightarrow \text{C}_2\text{H}_4 + \cdot\text{C}_4\text{H}_9$	Removed
86	$\text{C}_2\text{H}_4 + \cdot\text{C}_4\text{H}_9 \rightarrow \text{C}_4\text{H}_{10} + \cdot\text{C}_2\text{H}_3$	Removed
87	$\text{C}_4\text{H}_{10} + \cdot\text{C}_2\text{H}_3 \rightarrow \text{C}_2\text{H}_4 + \cdot 2\text{-C}_4\text{H}_9$	Removed
88	$\text{C}_2\text{H}_4 + \cdot 2\text{-C}_4\text{H}_9 \rightarrow \text{C}_4\text{H}_{10} + \cdot\text{C}_2\text{H}_3$	Retained
89	$\text{C}_3\text{H}_8 + \cdot\text{C}_2\text{H}_3 \rightarrow \text{C}_2\text{H}_4 + \cdot\text{C}_3\text{H}_7$	Removed

90	$C_2H_4 + \cdot C_3H_7 \rightarrow C_3H_8 + \cdot C_2H_3$	Retained
91	$C_3H_8 + \cdot C_2H_3 \rightarrow C_2H_4 + \cdot 2-C_3H_7$	Removed
92	$C_2H_4 + \cdot 2-C_3H_7 \rightarrow C_3H_8 + \cdot C_2H_3$	Retained
93	$C_3H_6 + \cdot C_2H_3 \rightarrow C_2H_4 + \cdot C_3H_5$	Retained
94	$2H_4 + \cdot C_3H_5 \rightarrow C_3H_6 + \cdot C_2H_3$	Retained
95	$C_4H_8 + \cdot C_2H_3 \rightarrow C_2H_4 + \cdot C_4H_7$	Removed
96	$C_2H_4 + \cdot C_4H_7 \rightarrow C_4H_8 + \cdot C_2H_3$	Removed
97	$C_4H_{10} + \cdot C_3H_7 \rightarrow C_3H_8 + \cdot C_4H_9$	Removed
98	$C_3H_8 + \cdot C_4H_9 \rightarrow C_4H_{10} + \cdot C_3H_7$	Removed
99	$C_4H_{10} + \cdot 2-C_3H_7 \rightarrow C_3H_8 + \cdot C_4H_9$	Removed
100	$C_3H_8 + \cdot C_4H_9 \rightarrow C_4H_{10} + \cdot 2-C_3H_7$	Removed
101	$C_3H_6 + \cdot C_4H_9 \rightarrow C_4H_{10} + \cdot C_3H_5$	Removed
102	$C_4H_{10} + \cdot C_3H_5 \rightarrow C_3H_6 + \cdot C_4H_9$	Removed
103	$C_4H_{10} + \cdot C_4H_7 \rightarrow C_4H_8 + \cdot C_4H_9$	Removed
104	$C_4H_8 + \cdot C_4H_9 \rightarrow C_4H_{10} + \cdot C_4H_7$	Removed
105	$C_4H_{10} + \cdot C_3H_7 \rightarrow C_3H_8 + \cdot 2-C_4H_9$	Removed
106	$C_3H_8 + \cdot 2-C_4H_9 \rightarrow C_4H_{10} + \cdot C_3H_7$	Removed
107	$C_4H_{10} + \cdot 2-C_3H_7 \rightarrow C_3H_8 + \cdot 2-C_4H_9$	Removed
108	$C_3H_8 + \cdot 2-C_4H_9 \rightarrow C_4H_{10} + \cdot 2-C_3H_7$	Removed
109	$C_3H_6 + \cdot 2-C_4H_9 \rightarrow C_4H_{10} + \cdot C_3H_5$	Retained
110	$C_4H_{10} + \cdot C_3H_5 \rightarrow C_3H_6 + \cdot 2-C_4H_9$	Removed
111	$C_4H_{10} + \cdot C_4H_7 \rightarrow C_4H_8 + \cdot 2-C_4H_9$	Removed
112	$C_4H_8 + \cdot 2-C_4H_9 \rightarrow C_4H_{10} + \cdot C_4H_7$	Removed
113	$C_3H_6 + \cdot C_3H_7 \rightarrow C_3H_8 + \cdot C_3H_5$	Removed
114	$C_3H_8 + \cdot C_3H_5 \rightarrow C_3H_6 + \cdot C_3H_7$	Removed
115	$C_3H_8 + \cdot C_4H_7 \rightarrow C_4H_8 + \cdot C_3H_7$	Removed
116	$C_4H_8 + \cdot C_3H_7 \rightarrow C_3H_8 + \cdot C_4H_7$	Removed
117	$C_3H_6 + \cdot 2-C_3H_7 \rightarrow C_3H_8 + \cdot C_3H_5$	Retained
118	$C_3H_8 + \cdot C_3H_5 \rightarrow C_3H_6 + \cdot 2-C_3H_7$	Removed
119	$C_3H_6 + \cdot C_4H_7 \rightarrow C_4H_8 + \cdot C_3H_5$	Retained
120	$C_4H_8 + \cdot C_3H_5 \rightarrow C_3H_6 + \cdot C_4H_7$	Removed
121	$C_3H_8 + \cdot C_4H_7 \rightarrow C_4H_8 + \cdot 2-C_3H_7$	Removed
122	$C_4H_8 + \cdot 2-C_3H_7 \rightarrow C_3H_8 + \cdot C_4H_7$	Removed

Radical addition and β -scission reactions

123	$C_2H_4 + \cdot H \rightarrow \cdot C_2H_5$	Retained
124	$\cdot C_2H_5 \rightarrow C_2H_4 + \cdot H$	Retained
125	$\cdot C_2H_5 + C_2H_4 \rightarrow \cdot C_4H_9$	Retained
126	$\cdot C_4H_9 \rightarrow \cdot C_2H_5 + C_2H_4$	Retained
127	$\cdot CH_3 + C_2H_4 \rightarrow \cdot C_3H_7$	Removed
128	$\cdot C_3H_7 \rightarrow \cdot CH_3 + C_2H_4$	Retained
129	$\cdot C_2H_3 + C_2H_4 \rightarrow \cdot C_4H_7$	Retained
130	$\cdot C_4H_7 \rightarrow \cdot C_2H_3 + C_2H_4$	Retained
131	$C_2H_2 + \cdot H \rightarrow \cdot C_2H_3$	Retained
132	$\cdot C_2H_3 \rightarrow C_2H_2 + \cdot H$	Retained

133	$\cdot\text{CH}_3 + \text{C}_2\text{H}_2 \rightarrow \cdot\text{C}_3\text{H}_5$	Retained
134	$\cdot\text{C}_3\text{H}_5 \rightarrow \text{C}_2\text{H}_2 + \cdot\text{CH}_3$	Retained
135	$\text{C}_3\text{H}_6 + \cdot\text{H} \rightarrow \cdot\text{C}_3\text{H}_7$	Retained
136	$\cdot\text{C}_3\text{H}_7 \rightarrow \text{C}_3\text{H}_6 + \cdot\text{H}$	Retained
137	$\text{C}_4\text{H}_6 + \cdot\text{H} \rightarrow \cdot\text{C}_4\text{H}_7$	Retained
138	$\cdot\text{C}_4\text{H}_7 \rightarrow \text{C}_4\text{H}_6 + \cdot\text{H}$	Retained
139	$\text{C}_4\text{H}_8 + \cdot\text{H} \rightarrow \cdot\text{C}_4\text{H}_9$	Removed
140	$\cdot\text{C}_4\text{H}_9 \rightarrow \text{C}_4\text{H}_8 + \cdot\text{H}$	Retained
141	$\text{C}_3\text{H}_6 + \cdot\text{H} \rightarrow \cdot\text{2-C}_3\text{H}_7$	Retained
142	$\cdot\text{2-C}_3\text{H}_7 \rightarrow \text{C}_3\text{H}_6 + \cdot\text{H}$	Retained
143	$\text{C}_4\text{H}_8 + \cdot\text{H} \rightarrow \cdot\text{2-C}_4\text{H}_9$	Removed
144	$\cdot\text{2-C}_4\text{H}_9 \rightarrow \text{C}_4\text{H}_8 + \cdot\text{H}$	Retained
145	$\text{CH}_3 + \text{C}_3\text{H}_6 \rightarrow \cdot\text{2-C}_4\text{H}_9$	Retained
146	$\cdot\text{2-C}_4\text{H}_9 \rightarrow \text{C}_3\text{H}_6 + \cdot\text{CH}_3$	Retained

Isomerization reactions

147	$\cdot\text{C}_3\text{H}_7 \rightarrow \cdot\text{2-C}_3\text{H}_7$	Retained
148	$\cdot\text{2-C}_3\text{H}_7 \rightarrow \cdot\text{C}_3\text{H}_7$	Retained
149	$\cdot\text{C}_4\text{H}_9 \rightarrow \cdot\text{2-C}_4\text{H}_9$	Retained
150	$\cdot\text{2-C}_4\text{H}_9 \rightarrow \cdot\text{C}_4\text{H}_9$	Retained

Curriculum Vita

Thanneer Malai Perumal

Permanent Address: 51/1525, Kandi Dawoodshah Lane,
West Main Street, Thanjavur – 613 009,
Tamil Nadu, India, Ph: (H): +91-4362-223653

Email: (P) th.vairam@gmail.com, (P) th_vairam@yahoo.co.in

Linkedin: <http://sg.linkedin.com/pub/thanneermalai-perumal/1a/865/279>

Website: http://gunawan.openwetware.org/Thanneer_Malai.html

Career Interest	Mathematical modeling and analysis, Computational systems biology, System identification and data mining, Statistical analysis, Robust control theory, Information Theory, Bioremediation and Biodegradation	
Education	<p>Ph. D. Jan, 2007- Dec, 2011</p> <p>Major: Chemical and Biomolecular Engineering</p> <p>Inst: National University of Singapore, Singapore.</p> <p>Thesis: <i>Dynamical Sensitivity Analyses of Kinetic models in Biology</i></p> <p>Emphasis: Computational Systems Biology, Mathematical Modeling and Simulation, Robust Control Theory, Information theory and Statistical Analysis</p> <p>Advisor: Dr. Rudiyanto Gunawan CGPA: 4.83 / 5</p> <p>B.E. Aug, 2002 – May, 2006</p> <p>Major: Chemical Engineering</p> <p>Inst: National Institute of Technology Karnataka, Surathkal, India.</p> <p>Thesis: <i>Recirculation studies on solid waste leachate using lysimeter</i></p> <p>Emphasis: Biodegradation, Bioremediation</p> <p>Advisor: Dr. Gopal Mugeraya CGPA: 3.75 / 4</p>	
Positions	<p>Collaborateur scientifique sur projet Feb, 2012 – Jan, 2013</p> <p>Luxembourg Center for Systems Biomedicine, University of Luxembourg, Luxembourg</p> <p>Academic Visitor Jun, 2011 – Sep, 2011</p> <p>Institute for Chemical and Bioengineering, D-CHAB, ETH Zurich, Switzerland</p> <p>Research Assistant May, 2006 – Nov, 2006</p> <p>Center for Environmental Science & Engineering, Indian Institute of Technology Bombay, Mumbai, India.</p>	

	Summer Intern	May, 2004 – Jul, 2004
	<u>Oil and Natural Gas corporation Limited, Uran, Mumbai, India.</u>	
Awards	2008-2012	<u>Singapore Millennium Foundation (SMF) Scholarship</u>
	2007-2008	<u>NUS Research Scholarship</u>
Software Skills	<p>Platforms: Windows, Linux, Unix</p> <p>Languages: C, C++, VB, HTML</p> <p>Packages: MATLAB (with major toolbox), COMSOL, Gensym G2</p> <p>HPC: MATLAB (Parallel Computing Toolbox), OpenMP, MPI</p> <p>Clusters: MATLAB(Distributed Computing Servers), SGE</p>	
Research Experience	<p>Dynamical sensitivity analyses of kinetic models in biology</p> <p>Developing novel sensitivity based model analysis tools to elucidate the system dynamics and to identify robustness-fragility trade-offs of both deterministic and stochastic models of chemical and biochemical systems. Specifically,</p> <ul style="list-style-type: none"> ▪ Identified the caveats of classical Parametric Sensitivity Analysis (PSA) ▪ Developed a Green's Function Matrix (GFM) and an Impulse Parametric Sensitivity Analysis (iPSA) tools for analyzing deterministic ODE models ▪ Developed an information theoretic sensitivity analysis based on distribution perturbations of stochastic ODE and CME models ▪ Developed a novel pathway based sensitivity analysis tool, which can identify the pathways that are responsible for the system dynamics and identify the robustness causing mechanisms ▪ Used these methods to determine the fragility and understand the system dynamics of cellular signaling networks such as Apoptosis, Cell cycle, MAPK, p53-Mdm2 and JAK-STAT pathways ▪ Used the above deterministic tools for understanding and reducing the large complex kinetic mechanisms of chemical and biochemical engineering <p>This work provides the technology for the use of mathematical models to benefit human health and medicine through robustness-based systems-oriented drug design.</p> <p>Reverse engineering of biological networks</p> <p>Developing algorithms to efficiently infer network topography from heterogeneous multivariate biological data. Specifically,</p> <ul style="list-style-type: none"> ▪ Identified the state-of-art methods, opportunities and challenges in reverse engineering biological networks ▪ Generated identifiable synthetic data sets with multivariate perturbations for topography identification, based on Design of Experiments (DOE) ▪ Developed a standard to evaluate the performance of parametric methods for network identification ▪ Based on the standards developed, in parametric methods, 	

- proved that topography prediction is not equivalent to data fitting problem
- At present, developing performance metrics and benchmark examples to evaluate the non-parametric or information theory based methods

Application of oil biodegrading cultures for the treatment of oil water emulsion wastes

Biodegradation of both aliphatic and aromatic components of Non-Aqueous Phase Liquids (Ex., Diesel) using a predesigned consortia on an attached growth bioreactor (RBC). Specifically,

- Involved in the startup of medium scaled rotating biological contactor (RBC) for diesel degradation using predesigned consortia
- Studied oxygen mass transfer characteristics in the reactor
- Studied nitrogen carryover effects during degradation
- Optimized the buffer concentration of media
- Trained on equipments such as GC-MS, HPLC, ICP, AAS and TOC analyzer

Recirculation studies on solid waste leachate using lysimeter

Testing the effect of recirculation of the leachate as a simple and economical treatment of waste water from landfills. Specifically

- Designed and constructed the lysimeter (a model landfill)
- Conducted leachate recirculation studies
- Measured most of the physiochemical parameters of the leachate (i.e., pH, temperature, BOD, COD, TOC, NH₃-N₂, organic N₂, chloride, sulfate, sodium, magnesium, calcium, potassium)

Journal Publications:

1. [Perumal TM, Wu Y, and Gunawan R. Dynamical analysis of cellular networks based on the Green's function matrix. J Theor Biol 2009, 261\(2\) 248-59. doi:10.1016/j.jtbi.2009.07.037 PMID:19660478](#)
2. [Perumal TM and Gunawan R. Understanding dynamics using sensitivity analysis: caveat and solution. BMC Syst Biol 2011, 5\(1\): 41. PMID: 214060955](#)
3. Perumal TM, Krishna SM, Tallam SS and Gunawan R. Reduction of Kinetic Models using Dynamic Sensitivities (In Submission)
4. Perumal TM and Gunawan R. Robustness-Fragility Analysis of Cellular Population Models. (In Preparation)
5. Perumal TM and Gunawan R. A pathway based robustness mechanism identification. (In Preparation)
6. Perumal TM and Gunawan R. GRN identification: Caveat and solution. (In Preparation)

Peer-Reviewed Conference Proceedings:

1. Perumal TM and Gunawan R. Impulse Parametric Sensitivity Analysis, In Proceedings of the 18th World Congress - The International Federation of Automatic Control (IFAC), Milano, Italy, 2011, August 28- September 2, PP 9686-9690.

2. Perumal TM and **Gunawan R**. Dynamical Analysis and Model Reduction of Complex Systems, In Proceedings of the 13th Asia Pacific Confederation of Chemical Engineering Congress (APCChE), Taipei, 2010, October 5-8
3. **Perumal TM** and Gunawan R. Information-theoretic global robustness analysis of cellular systems: A molecular perturbation approach, In Proceedings of the 3rd International Conference on Foundations of Systems Biology in Engineering (FOSBE), Denver, Colorado, USA, 2009, August 9 - 12, PP: 52 – 55
4. Perumal TM, Yan W and **Gunawan R**. Robustness Analysis of Cellular Systems for In Silico Drug Discovery, In Proceedings of the 17th World Congress - The International Federation of Automatic Control (IFAC), Seoul, Korea, 2008, July 6-11, PP: 12607-12612

Oral & Poster Presentations:

1. Perumal TM and **Gunawan R**. Dynamical Pathway Sensitivity Analysis for Biological Systems. Oral presented in the AIChE Annual Meeting, Minneapolis, MN, USA, 2011, October 28 - November 2 [Oral]
2. Perumal TM and **Gunawan R**. Caveats of Parametric Sensitivity Analysis (PSA): In analyzing the dynamics of biological systems. Oral presented in the AIChE Annual Meeting, Salt Lake City, UT, USA, 2010, November 7-12 [Oral]
3. Perumal TM and **Gunawan R**. Dynamical Model Reduction of Large Reaction Mechanisms: A Green's Function Matrix (GFM) Based Approach. Oral presented in the AIChE Annual Meeting, Salt Lake City, UT, USA, 2010, November 7-12 [Oral]
4. **Perumal TM** and Gunawan R. In Analyzing the Complex Dynamics of Biochemical Pathways. Oral presented in the satellite conference on Application of Control Theory and Optimization Techniques in Biochemical Pathways, HICC, Hyderabad, India, 2010, August 16-18 [Oral]
5. **Perumal TM** and Gunawan R. Information Transfer in Biological Network Motifs, Poster presented in the 10th International Conference on Systems Biology (ICSB), Standford, California, USA, 2009, August 30 - September 4 [Poster]
6. **Perumal TM**, Yan W and Gunawan R. In Silico Dynamical Analysis of Cellular Systems: A Molecular Perturbation Approach, Poster presented in the 12th Annual International Conference on Research in Computational Molecular Biology (RECOMB), Singapore, 2008, March 30 - April 2 [Poster]
7. Perumal TM, Yan W and **Gunawan R**. In Silico Molecular Analysis: Application to Fas-Induced Apoptotic Pathway, Poster presented in the 10th International Conference on Molecular Systems Biology (ICMSB), Manila, Philippines, 2008, February 25-28 [Poster]
8. Perumal TM, Yan W and **Gunawan R**. In Silico Molecular Analysis: Application to the Fas-induced Apoptotic Pathway, Poster presented in the 8th International Conference on Systems Biology (ICSB), California, USA, 2007, October 1-6 [Poster]

Training

Summer school on “Cell and Molecular Mechanics in Biomedicine” in NUS from 25th June to 6th July 2007

Professional Societies

[American Institute of Chemical Engineering](#)
[Asia Pacific BioInformatics Network](#)
[The Indus Entrepreneurs \(TiE\), Singapore](#)

Leadership Activities

One among the three brand campaigner for promoting [Singapore Education](#) in India in the year 2009

Guided 4 Undergraduate and 4 Masters students to work towards their research thesis

Worked as a teaching assistant for 2 undergraduate and 2 graduate courses

Executive Committee member, of Graduate Students Association (GSA), Dept. of ChBE, NUS

National Cadet Corps (NCC), India, 'A' and 'B' certificate holder.

NCC is a unit of Indian Army which imparts leadership and military training to students

Black belt holder in Indian Martial Arts from Indian Martial Arts Federation.

Detailed research portfolio, references and career summary available on request

4th Jul 2012

Thanneer Malai Perumal

Note

Note