An Agent-Based Model of the IL-1 Stimulated Nuclear Factor-kappa B Signalling Pathway

Richard Alun Williams FRSA

PhD

University of York
Computer Science

September 2014

Abstract

The transcription factor NF-κB is a biological component that is central to the regulation of genes involved in the innate immune system. Dysregulation of the pathway is known to be involved in a large number of inflammatory diseases. Although considerable research has been performed since its discovery in 1986, we are still not in a position to control the signalling pathway, and thus limit the effects of NF-κB within promotion of inflammatory diseases.

We believe that computational modelling and simulation of the NF-κB signalling pathway will complement wet-lab experimental approaches, and will facilitate a more comprehensive understanding of this example of a complex biological system. In this study, we have developed an agent-based model of the IL-1 stimulated NF-κB signalling pathway, which has been calibrated to wet-lab data at the single-cell level. Through rigorous software engineering, which followed a principled approach to design and development by adherence to the CoSMoS process, we believe our model provides an abstracted view of the underlying real-world system, and can be used in a predictive capacity through *in silico* experimentation.

A novel approach to domain modelling has been presented, which uses linear and multivariate statistical techniques to complement the Unified Modelling Language. Furthermore, *in silico* experimentation with the newly developed agent-based model, has confirmed the robust yet fragile nature of the signalling pathway. We have discovered that the pathway is robust to perturbations of cell membrane receptor component number, intermediate component number, and the temporal lag between cell membrane receptor activation and subsequent activation of IKK. Conversely however, *in silico* experimentation predicts that the pathway is sensitive to changes in the ratio of *free* IκBα to NF-κB, and fragile to basal dissociation of NF-κB-IκBα outside of a narrow range of probabilities.

Contents

List of Figures

List of Tables

Acknowledgements

This doctoral research was commenced in October 2010 to facilitate a career-change from my previous employment as Project Manager within IT Consulting, to a new life as an academic. Although the journey has been full of challenges, which were interspersed with a number of pitfalls along the way, it has for the most part been strangely enjoyable. In fact, the famous Japanese proverb of resilience *"Nana korobi ya oki"*, comes to mind when I look back on my journey, which means "if you fall down seven times then you should get up eight times".

It must be acknowledged however, that I would not be in the position that I find myself today without the support of a number of key people, to whom I owe my sincere gratitude:

To Jon Timmis for his excellent supervision and unwavering advice and support over the past four years.

To Eva Qwarnstrom, co-supervisor and subject matter expert within this collaboration, who imparted the necessary domain knowledge into a somewhat rusty biochemistry graduate who had taken timeout to work in industry.

To Laura Machin, for her intense belief in me, unwavering patience, and unconditional support when I chose to leave a well paid position in IT Consulting to follow my dream. I couldn't have completed this journey without you. Additionally, writing a thesis is hard, but writing one when in full-time employment, and in a new research discipline, is exponentially harder - you made it as pain-free as possible.

Finally, I wish the best of luck to any doctoral students that come after me and build on this work.

Declaration

This thesis is submitted in part fulfillment for the degree of Doctor of Philosophy. It has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree other than Doctor of Philosophy of the University of York. This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by explicit references.

Some of this work has already been presented in Williams (2011b) and Williams (2012). Additionally, aspects of chapter 3 that focus on the existing computational models of NF-κB have also been published in the review article of Williams et al. (2014b).

I hereby give consent for my thesis, if accepted, to be made available for photocopying and for inter-library loan, and for the title and abstract to be made available to outside organisations.

Publications

A number of journal articles and conference papers have been published during the lifetime of this PhD research. The following relate directly to the work performed during the design, development and experimentation of the agent-based model of the NF-κB signalling pathway:

Williams et al. (2014a) : I am the principal author of this journal article. The paper presents a domain model of the NF-κB signalling pathway, using UML and cartoon diagrammatic notations, along with a number of statistical techniques. This paper has been expanded upon in chapter 4.

Williams et al. (2014b) : I am the principal author of this journal article. The paper reviews the current state of computational modelling of the NF-κB signalling pathway, which has been expanded upon in chapter 3.

Williams (2013) : I am the sole author of this journal article. The paper defines a number of approaches and techniques by which PhD students within computational biology may project manage their doctoral research projects.

Williams (2012) : I am the sole author of this conference extended abstract. The paper discusses how various statistical techniques may be used to complement the Unified Modelling Language when developing domain models of complex biological systems, which has been expanded upon in chapter 4.

Williams (2011b) : I am the sole author of this conference extended abstract. The paper provides an overview to: the IL-1 stimulated NF-κB signalling pathway; agent based modelling of complex biological systems; and the proposed scope of my doctoral research. It has been expanded upon within chapters 2-4. I was awarded **Best Extended Abstract** for this paper.

Williams (2011a) : I am the sole author of this article published in the British Computer Society's professional members magazine (ITNow). The paper discusses how computational techniques are being used in both academic and industrial research for extending our understanding of complex biological systems.

Additionally, the following publications have also resulted from collaborations with the wider York Computational Immunology Lab (YCIL) here at The University of York, of which I am a member:

Read et al. (2013) : I am a co-author of this journal article. This paper presents a computational model of Experimental Autoimmune Encephalomyelitis,

called ARTIMMUS, and discusses the results of associated *in silico* experimentation.

Williams et al. (2013) : I am the principal author of this journal article. The paper presents the results of *in silico* experimentation using the ARTIMMUS computational model (Read et al., 2013).

Williams et al. (2011) : I am the principal author of this conference extended abstract. The paper provides an overview of work-in-progress results using the ARTIMMUS computational model (Read et al., 2013). I was awarded **Best Paper (Computational Immunology Stream)** for this paper.

Part I

Introduction and Literature Review

1 Introduction

This introductory chapter to the doctoral thesis presents: the motivation for the work conducted during the PhD research; an overview to how the thesis is structured; the objectives of the research project; and the contributions that the research has made to the NF-κB modelling community in particular, and the wider computational biology community, more generally. Firstly, section 1.1 provides the motivation behind the use of computational models to increase our understanding of complex biological pathways. Secondly, section 1.2 presents an outline of this doctoral thesis, along with the objectives of the research project, and a summary of the structure that this thesis takes. Finally, section 1.3 presents the contributions that this research project has made to the scientific community.

1.1 Motivation

Cells receive information from their environment through extracellular signals interacting with a class of proteins located on their surfaces that are known as receptors. Signal transduction involves the binding of these extracellular signalling molecules to the cell-surface receptors and the triggering of a cascade of signalling events inside the cell. This intracellular signalling forms part of a complex system of communication, that acts to regulate cell activity and action relative to changes in the external environment. The ability of cells to perceive and correctly respond to their environment is of particular importance for the immune system and the combating of infection. Gene activations and alterations to metabolism are examples of cellular responses to extracellular stimulation that require signal transduction. Errors in the processing of extracellular signals are responsible for the onset of many diseases. The NF-κB signalling pathway is one of the key signal transduction pathways involved in control and regulation of the immune system. It is hoped that through an increased understanding of cell signalling pathways, such as NF-κB, diseases may be treated more effectively, or even eradicated entirely through prevention of pathway dysregulation.

Scientists are designing and creating increasingly complex models and simulations of cells and other biological entities. The modelling and simulation of cell behaviour falls under the wider scope of a relatively new area of science which has been termed *Systems Biology*. Systems biology begins with complex biological phenomena and aims to provide a simpler more abstract framework that explains why these events occur the way they do. Simulation attempts to predict the dynamics of the systems so that the validity of the underlying assumptions can be tested. Within the process of modelling and simulation, detailed behaviours of computer-executable models are compared with experimental observation. Models that survive this validation can then be used for predicting behaviours which can be tested by experiments, as well as to explore questions that are not amenable to experimental inquiry (Kitano, 2002a).

The systems that biologists wish to analyse and model show an increasing number of interdependencies and relationships, meaning that traditional modelling tools are no longer as appropriate as they once were. Computational biologists and computer scientists are beginning to take a more realistic view of these systems through recent advances in computational modelling techniques, such as agent-based modelling and simulation. Furthermore, computational power is advancing rapidly, providing the ability to run large-scale computer simulations of biological systems, which would not have been plausible in the last century (Macal and North, 2005).

Modelling is not the end goal in itself, but is instead a tool to increase our understanding of the complex system in question. Models allow us to develop more directed experiments and facilitate the forming of predictions (Banga, 2008). It is hoped that an increased understanding of the design principles of protein and gene regulatory networks during normal physiology and disease, will pinpoint the underlying causes of diseases and lead to more rationalised and efficacious treatment strategies.

1.2 Thesis Outline

This doctoral research project will use an agent-based modelling approach to explore the IL-1 stimulated NF-κB intracellular signalling pathway. Through an interdisciplinary approach, we aim to use computational modelling and *in silico* experimentation to generate predictions of system dynamics to help drive further *in vitro* wet-lab experimentation. This work makes three major contributions: firstly, to the development of new computational models of the NF-κB signalling pathway, which can be used for hypothesis-generation to further our understanding of recently discovered signalling components; secondly, to the growing knowledge base regarding computational modelling and simulation of complex systems, through reflection on the benefits, constraints and appropriateness of agent-based modelling within computational biology research; and thirdly, to expand the current models (which primarily focus on regulation at the level of the NF-κB-IκB complex), to include events related to cell surface receptor complex formation and activation, which induces the propagation of signal downstream to the gene regulatory events.

We will use a principled approach to design and development of computational models through use of the CoSMoS process (Andrews et al., 2010). The project is organised around two iterations of the computational model, which provide a baseline simulator to allow early experimentation, and an enhanced version later on in the project. The first iteration involves development of a baseline computational model of the NF-κB signalling pathway, which reproduces the functionality from existing differential equation based approaches, but using an agent-based approach. The second iteration involves augmentation of this model for additional granularity of components at the cell membrane receptor complex.

1.2.1 Project Objectives in Relation to Thesis

This thesis concerns the development of an agent-based model of the IL-1 stimulated NF-κB signalling pathway, along with its subsequent use to perform novel *in silico* experimentation. In order to ensure that the computational model is fit for purpose, we will follow a principled approach to design, development and testing of the computational model. Furthermore, in order to ensure that our interpretations of simulation results are valid, we will utilise a number of statistical techniques to qualify the significance of any differences between the results from experimental conditions and those of baseline/control conditions. We believe that statistical underpinnings, such as those used within this thesis, are essential for interpreting the results of *in silico* experimentation, and also believe that without such rigour, it will be hard for the niche area of computational biology to gain credibility with the wider scientific community. This scope is reflected in the thesis aim:

> *To develop an agent-based model of the IL-1 stimulated NF-κB signalling pathway in a quality assured manner, which uses leading practices for software engineering, calibration, verification and validation. Furthermore, once developed, the agent-based model will be used to perform novel in silico experimentation to extend our knowledge of the signalling pathway.*

The following research objectives will guide the work of this thesis towards the overall aim:

Obj 1: Explore the role of diagrammatic and statistical techniques for developing a domain model of the NF-κB case study.

Obj 2: Create an agent-based computational model of the core intracellular components of the IL-1 stimulated NF-κB signalling pathway.

Obj 3: Investigate techniques for calibrating agent-based computational models that have been developed using the FLAME simulation framework.

Obj 4: Perform novel *in silico* experimentation using the agent-based model.

Obj 5: Augment the agent-based model with additional upstream signalling components related to the cell membrane receptor complex.

Obj 6: Perform novel *in silico* experimentation using the augmented simulator.

Obj 7: Investigate the suitability of using the FLAME simulation framework for developing computational models of complex biological systems.

1.2.2 Thesis Structure

This doctoral thesis addresses the above research objectives over the next eight chapters, the majority of which directly relate to deliverables (e.g. project artefacts) of the CoSMoS process. The structure of this thesis follows below, and focuses on chapter content, relation of the chapter to the CoSMoS process, and the contribution of the chapter in fulfilling the research objectives.

Chapter 2 - Modelling and Simulation of Biological Systems

This literature review chapter provides a summary of computational biology and simulation science, with initial focus on the theoretical underpinnings that stem from: General Systems Theory and Network Theory; Systems Biology and how the advancements in computer technology have led to the disciplines of Computational Biology and Computational Immunology; focus is then applied to two main approaches - mathematical and agent-based - for computational modelling of biological systems. The chapter concludes with an overview of project life-cycles that may be appropriate for computational systems biology research, and specifically the CoSMoS process, which is used within this project. This chapter has no direct relationship to a CoSMoS project deliverable.

Chapter 3 - The Domain: NF-κB Signalling Pathway

This literature review chapter provides a detailed description of the IL-1 stimulated NF-κB intracellular signalling pathway, which is the basis for our computational model. This initial section begins with an introduction to cell signalling and an overview of the NF-κB transcription factor. It then progresses to a detailed description of the NF-κB signalling pathway, consequences of dysregulation of NF-κB, and discussion of the detailed components (receptors, co-receptors, adaptor proteins and kinases) involved in the IL-1 stimulated NF-κB signalling pathway. This initial section then concludes with discussion on recent approaches to study NF-κB.

The chapter ends with a detailed overview of the extent to which the NF-κB signalling pathway, in general, has been modelled, and more specifically the extent to which the IL-1 stimulated NF-κB signalling pathway has been modelled. The majority of these models have used mathematical descriptions (through ordinary differential equations) and have been based on wet-lab biochemical data, thus have utilised averaged dynamics. This final section begins with a discussion on these equation based models, before progressing to agent-based models, which have formed the basis of more recent work, a discussion of recent *minimal models* that have been based on these previous models, and concludes with a discussion around the need for a new computational model of the IL-1 stimulated NF-κB signalling pathway. This chapter represents the domain of the CoSMoS process.

Chapter 4 - Domain Model of IL-1 Stimulated NF-κB Signalling Pathway

This chapter presents the domain model of the IL-1 stimulated NF-κB signalling pathway, which will on the whole be modelled using the Unified Modelling Language (UML) notation, but where UML is deficient will also utilise informal cartoon diagrams. It also utilises a number of statistical techniques to model the heterogeneity and temporal dynamics of wet-lab data, which UML is unable to depict. This chapter represents the domain model from the CoSMoS process, and provides critical reflections on the use of UML and statistical techniques for domain modelling, thereby contributing to research objective 1.

Chapter 5 - The Platform Model

This chapter presents the technical specification of the first iteration of the proposed computational model in the form of a platform model (from the CoSMoS process). As per the domain model, this is predominantly developed using UML, but also includes diagrammatic notations that are specific to communicationg X-Machines, which is the underlying technical architecture for the FLAME simulation framework.

Chapter 6 - Development and Calibration of the Simulation Platform

This chapter focuses on the technical aspects of constructing the first iteration of the computational model. As the objective of this first iteration is to develop a computational model that is calibrated to published wet-lab data, through use of the FLAME simulation framework; any benefits, constraints, or limitations of the FLAME simulation framework are documented here in order to investigate its suitability for developing models of complex biological systems. The computational model associated with this chapter represents the simulation platform from the CoSMoS process. Furthermore, this chapter addresses research objective 2 and 3, and contributes to research objective 7.

Chapter 7 - Experimentation using the Baseline Simulator

This chapter reports the results of novel *in silico* experimentation performed with the baseline simulator, therefore addresses research objective 4. The simulation results and subsequent data analysis form the basis of the results model from the CoSMoS process.

Chapter 8 - The Augmented Simulator

This chapter focuses on extending the baseline computational model to incorporate the additional scope that has been defined for the second iteration. Specific focus is applied to the augmentation of the computational model to incorporate additional receptor complex components at the cell membrane. The chapter then progresses to discuss novel *in silico* experimentation to predict the effects on signal transduction following perturbation of receptor complex formation at the cell membrane. This chapter therefore reflects a second iteration of the CoSMoS process, and defines the updated platform model, simulation platform, and results model, along with further investigations on the suitability of FLAME for developing computational models of complex biological systems. As such, the chapter addresses research objectives 5 and 6, and makes an additional contribution to research objective 7.

Chapter 9 - Discussion, Conclusions and Further Work

This chapter concludes the thesis, and provides critical reflection on the work undertaken, and research performed, within this doctoral research project. Three key sections focus on: contributions to computational biology, including the fitness of the FLAME simulation framework for modelling intracellular signalling

pathways, and the fitness of the CoSMoS process in relation to this project; contributions made to the NF-κB modelling community, and formal evaluation of the research questions identified within the introduction chapter, and the thesis' success in answering them; and future directions of research that result from this thesis. In doing so, this chapter further contributes to research objectives 1, 3 and 7.

1.3 Thesis Contribution

This thesis contributes to the scientific research community in two main ways: those reflecting a computer science bias, which are focused on the application of computational approaches and software engineering principles to cell biology and immunology; and those specific to NF-κB. These are summarised below, and are also expanded upon in chapter 9, which concludes this thesis.

Contributions to the field of computational biology are as follows:

- Assessment of the ability of UML to define the domain model of the IL-1 stimulated NF-κB signalling pathway, and where limitations are found, the ability of statistical techniques to plug these gaps (chapter 4).
- An interdisciplinary and collaborative calibration procedure, which was underpinned by the transformation of single-cell analysis wet-lab data (from the domain model) into desired ranges for output data (from the simulation platform) of individual agent types (chapter 6).
- Build on the discussion by Read et al. (2012) of the necessity to establish a calibrated *baseline* for the simulation platform, which has taken account of the epistemic uncertainty of the underlying real-world biology, and the aleatory uncertainty that arises through the stochasticity of the computational model (chapter 6).
- Development of a novel approach for using the Kolmogorov-Smirnov two-sample test and the Vargha-Delaney A-Test for calculating the minimum number of replicates required to ensure a stable median average of simulation results. This was required because the previously published *consistency analysis* technique by Read et al. (2012), could not be used due to the resource intensive nature of the FLAME simulation framework, making global sensitivity and uncertainty analysis techniques intractable (chapter 6).
- Initiate a debate on the necessity for the computational biology community to ensure a rigourous and robust statistical analysis is performed on simulation results, in order to develop credibility with the wider scientific community (chapters 6, 9 and Appendix B).
- Assessment of the suitability of the FLAME simulation framework to model complex biological systems in general, and the IL-1 stimulated NF-κB signalling pathway in particular (chapters 5-8).
- Assessment of the CoSMoS process as a project lifecycle for computational biology (chapter 9).

In addition, the following contributions have been made to the specific field of NF-κB:

- Review the existing literature around the complexity that is inherent to the NF-κB signalling pathway; the way that computational approaches have facilitated our increased understanding of the pathway; and justify the need for a new computational model to extend our understanding even further, with particular reference to the IL-1 stimulated pathway (chapter 3).
- Design and development of an agent-based model, calibrated to wet-lab data at the single-cell level, through which *in silico* experimentation may be conducted (chapters 5-8).
- Elucidation of the robust, yet fragile[1], nature of the intracellular pathway, with particular emphasis on the probabilistic nature of association and basal-dissociation of the NF-κB-IκBα complex (chapter 7).
- Identification of the substantial robustness within the signalling pathway with respect to the ratios between individual component numbers and their associated interaction dynamics (chapter 7 and 8).

[1]See chapter 2. The term *fragile* in this context uses the systems biology definition, where the concepts of robustness and fragility go hand-in-hand when used to describe characteristics of communicating networks. For example, Albert et al. (2000) argue that biological networks display an unexpected degree of robustness, which allows their individual components to continue to communicate even when under considerable presure from outside perturbations. However, they continue by stating that error tolerance comes at a high price in that these networks are extremely vulnerable to attacks that focus on certain key components within a network. Kitano (2004a) takes this observation further by suggesting that there are trade-offs between robustness, fragility, and overall system performance, and that these trade-offs can explain system behaviour(s) under normal and dysregulated states, including the patterns of system failure.

2 Modelling and Simulation of Biological Systems

During the latter half of the 20th century, biology was dominated by the reductionist approach, resulting in the examination of role and function of increasingly smaller and smaller components within organisms. This approach has successfully generated a large amount of information about individual cellular components and their functions. In fact, the past two decades have seen an acceleration in reductionist approaches through emergence of high-throughput technologies, leading to omic[1] scale data sets. The resultant component-level information has then been used in a 'bottom-up' attempt to infer the properties of the organism(s) as a whole.

Reductionism has served well in the past, but a purely reductionist approach is now openly questioned. A nice introduction to such questioning is Lazebnik's good humoured criticism in *Can a Biologist Fix a Radio* (Lazebnik, 2002). The major drawback of the reductionist approach, which forms a consistent theme through Lazebnik's paper, is that it may promote overly simplified thinking, by focusing on single levels of abstraction. Benoist et al. (2006) argue that this results in the multiplicity of other influences on the chosen pathway being ignored. As such, the emphasis is beginning to move towards a systems approach, with analysis of complex systems using the data gained through reductionist approaches, but maintaining the context of abstraction level and scale. This ensures that interpretations of the data take into account the different hierarchical levels of living systems.

Traditionally, mathematical approaches were used to model the dynamics of such complex biological systems, through for example differential equations, which when solved, generate time-series dynamics at the system level. More recently however, a number of computational approaches taken from the field of computer science have been used that allow modelling at the level of the individual components, and generate *emergent* system-level behaviour through the interactions of the components. With particular reference to the immune system, this interdisciplinary approach has led to the rapidly expanding field of *Computational Immunology*.

2.1 Systems Theory

The idea that a theory could be constructed and used to explain the dynamics of complex systems across multiple disciplines was first advocated by Ludwig

[1] Omic in this context refers to the very large number of measurements or extremely large amounts of data collected from high-throughput experiments that look at genes (genomic), proteins (proteomic), or metabolites (metabolomics).

von Bertalanffy through discussions with his fellow biologists in the late 1920s (von Bertalanffy, 1972). Along with von Bertalanffy, who came from a more theoretical perspective, Norbert Weiner was also one of the early proponents for system-level understanding, and his work led to the birth of cybernetics (Weiner, 1948). Following his early work, von Bertalanffy later went on to propose a *General Systems Theory* (von Bertalanffy, 1950), where he advocates that the reductionist approach should be extended to incorporate relationships between components of a system so that judgements on the system as a whole, and not components in isolation, may be made. He further postulated, that even though the reductionist approach may provide complete explanations of all the individual processes within a complex system, without a more holistic view, we will be unable to understand the integrated nature of the processes, and for example would not be able to explain the total metabolism within a cell. The end goal was to use mathematical terms for developing the general systems theory because he believed mathematics to be an exact language that permitted rigorous deductions and confirmation (or refutal) of theory. von Bertalanffy (1972) states that *"a system may be defined as a set of elements standing in interrelation among themselves and with the environment"*. His general systems theory was conceived as a working hypothesis, where theoretical models could be used to explain and predict system-wide processes and events (von Bertalanffy, 1969). Furthermore, during his latter years as a researcher, von Bertalanffy began to promote the idea that systems theory provided the potential to ultimately 'control' system responses, and highlighted how research into the control of biological systems converges with network theory (von Bertalanffy, 1972).

Boulding (1956) built on von Bertalanffy's initial work by suggesting that general systems theory will become the 'skeleton of science', his emphasis was on biological sciences however, and he highlighted the importance of remembering that functionality (e.g. homeostasis or growth) within these systems are multi-scale in nature, moving from an atom to a cell, up through organs, and culminating with an individual organism. Mesarovic (1968) subsequently built on this biological systems theory when he merged control theory and molecular cell biology, and formally developed the concept of *systems biology*. Furthermore, Barabasi (2003) argues that in order to understand the inherent complexity within biological systems, it is essential that we move beyond the mere structure and topology of the network of components, which only capture the 'skeleton of complexity', and instead focus on the dynamics that take place along the links.

2.2 Network Theory

Investigators from many distinct fields, have over the past two decades discovered that many networks, from the World Wide Web to metabolic pathways, are dominated by a relatively small number of heavily connected nodes. These heavily connected nodes are also known as hubs, and networks containing such hubs, tend to be called 'scale-free', because such nodes have a seemingly unlimited number of links and no node is typical of the others (Barabasi and Bonabeau, 2003; Albert, 2005). A reliance on hubs, depending on the system in question, can have both advantages and disadvantages. For example, hubs provide a degree of robustness to a given system, by providing resilience to random breakdown of

nodes, however they also act as a point of fragility within the system, because if the hub node breaks down, there is a high probability that the whole system will be compromised (Kitano, 2004a, 2010).

Cells use complex networks of interacting molecular components to transfer and process information. It is increasingly appreciated that the robustness of various cellular processes is rooted in the dynamic interactions among its many constituents (Bhalla and Iyengar, 1999), such as proteins, DNA, RNA and small molecules. Barkai and Leibler (1997) advise that the *"computational devices of living cells"* are responsible for many important cellular processes, including metabolic pathways, signalling pathways and regulatory mechanisms. Metabolism is ubiquitous in living cells, and due to the conversion of foodstuffs into energy-rich molecules, is involved in essentially all cellular functions (Voet and Voet, 1995). Signalling pathways allow living cells to sense, measure and evaluate their environment, and to generate appropriate output responses. Signalling influences the cellular fate through initiation of one or more processes, such as differentiation to a more specialised cell-type, replication, apoptosis (programmed cell death), or migration (Hancock, 1997). Signalling pathways have traditionally been portrayed as linear chains of biochemical reactions and protein-protein interactions, starting from receptors, which sense signals outside of the cell, and progressing through intermediate components until they reach their ultimate target inside the cell, which is generally a specific region of DNA (Frank et al., 2009). Finally, regulatory mechanisms ensure that the cellular components involved in expression of genes following activation of signalling pathways are tightly regulated (Palsson, 2011).

Cell biologists use visual representations of these networks, in the form of graphs, to make sense of the myriad signal transduction cascades and metabolic pathways. Graphs in this context, are mathematical diagrams consisting of nodes to represent the system components, and edges to represent the links between components (Diestel, 1997). In a protein interaction graph, for example, two nodes (proteins) are connected by an edge if they interact physically. The interpretation of a directed edge between two nodes (node A to node B) in a graphical model is that an intervention on the first node (A) will alter the second node (B), but that an intervention on the second node will not alter the first node (Blair et al., 2012). Perhaps the most well known of all biological interaction graphs, is the metabolic pathways map developed by Donald Nicholson in the 1960s, which has now reached the 22nd edition (Nicholson, 2001).

Recently, however, it has emerged that more and more components are shared by several pathways, and that these linear networks from individual pathways, merge into large and complex interconnected networks when viewed from a systems perspective. Graph representations of biochemical and signal transduction networks can be extremely useful for gaining systems-level insights into cellular regulation. For example, the interconnected nature of these networks indicate that perturbations of a single gene or protein within a pathway could have seemingly unrelated effects to other pathways within the wider network (Albert, 2005). Oltvai and Barabasi (2002) propose that we build on this approach and believe that by viewing a cell as an interconnected network of genes and proteins, we will be able to develop viable strategies for addressing the complexity of living systems.

Despite the pivotal role that networks play within cell biology, until recently, scientists have had little understanding of their structure and properties. It is through the current drive towards systems biology approaches that we are beginning to fill the gaps in our understanding. For example, recent research has shown that scale-free networks exist among disease states (Frank et al., 2009), and also uncovered a number of causes of diseases, such as cancer, based on the malfunctioning of one or more nodes in complex genetic networks (Barabasi, 2009; Kitano, 2004b, 2007).

2.3 Systems Biology

The concept attributed to Aristotle, that *"the whole is more than the sum of its parts"*, suggests that dissecting an organism, tissue or cell into increasingly smaller subunits and hoping to be able to piece it back together afterwards will not work if the underlying mechanisms and relationships between these subunits are anything but linear and static. While systems theory *per se* is hardly new, having been based on the pioneering, although rather theoretical work of von Bertalanffy (1950), and updated by Boulding (1956) to better reflect biology, the notion of systems biology as the research approach that integrates biology, medicine, computer science and technology to comprehend biological information processing has recently been embraced by the scientific community at large. Understanding the nature of cause and effect is fundamental to all fields of scientific investigation, but the concept of causality can present special difficulties in biology (Mayr, 1961). Biological systems possess a high degree of complexity, with the role of their constituent components not only being defined by their individual function, but also by their interaction with other components (Kitano, 2002b). The field of systems biology is concerned with understanding this complexity by focusing on the interactions between the various components of living cells. There is debate within the scientific community on the exact definition of systems biology (Priami, 2009), however in general, it is the term used for the scientific approaches which discover, model, and understand the dynamic relationships between the molecules that define living organisms at the genetic and molecular levels, and to understand the emergent properties that emanate from these interactions (Banga, 2008; Weston and Hood, 2004). Therefore, a systems biology approach aims to understand how biological systems function at the system level by studying at different abstraction levels the relationships and interactions between various parts (e.g. organelles, cells, metabolic pathways, and cell signalling cycles), and produce results as close as possible to the dynamics and/or behaviour of the biological reality (Corradini et al., 2005; Wolkenhauer, 2001). It uses an integrative approach, combining experimental and computational techniques to identify the molecular mechanisms underlying the properties and behaviours of complex biological system (Mardinoglu and Nielsen, 2012).

2.3.1 Complexity

Biologists study various aspects of cells, and as stated earlier, have traditionally analysed them in greater and greater detail through a reductionist approach. As a convention, there are three standard levels described in biology textbooks: 1)

cellular structure, 2) cellular function through metabolic processes, and 3) genetic processing (Alberts et al., 1994; Voet and Voet, 1995). If we think of systems theory as the study of organisation and behaviour, it may be natural for us to consider systems biology as the application of systems theory to these three levels of cellular function. Through work linking these different abstraction levels together, it is becoming clear that even though the composition of living cells is complex, the number of distinct behaviours (i.e. their phenotypes) that they display is much fewer. Consequently, the principle of simplicity from complexity is emerging from decomposition of gene expression data and from temporal decomposition of complex metabolic models, which show that there are only a few governing dynamic determinants (Palsson, 2000).

Cell function emerges through the control of dynamic intracellular processes that both interact amongst each other (in a given network) and also respond to extracellular signals (Rajapakse et al., 2012). Genetic and metabolic networks are the two main targets of systems biology, because they control the fundamental mechanisms that govern biological systems. In essence, physiological behaviour emerges through the interactions between transcripts, enzymes, and metabolites, which form the constituents of metabolism and regulatory networks (Sauer, 2006). Both genetic and metabolic networks are highly complex, and consist of extensive feedback regulation for the coordination of cell function (Wolkenhauer and Mesarovic, 2005). Feedback loops imply before and after states, which mean that any understanding of cell function should be based on the premise that a cell is a dynamic system. The concept of feedback is one of the most fundamental methods of control in biological systems. An increase in the number of interconnecting cycles in a cascade, or amplification through positive feedback, increases the sensitivity of the target to the input signal, but may also change steady-state responses and therefore facilitate the occurence of instabilities. Negative feedback on the other hand may instil a degree of robustness to parameter variations within the feedback loop (Kholodenko, 2006). An important feature of feedback in biological systems, which has consequences for the stability of signalling processes is that it is often subject to a delay owing to the time taken to translocate molecules between the various cellular compartments (Wolkenhauer and Mesarovic, 2005). This control of signalling pathways aligns well with the physical sciences, as feedback control is a basic engineering strategy for ensuring outputs of a system robustly track the desired value, independently of noise or variations in system parameters (Yi et al., 2000).

Biological cells include numerous examples of constituent components whose behaviours are regulated, often at multiple levels. A prime example is the regulation of gene transcription, and one protein complex that is involved in this regulation of function is the nuclear transcription factor-kappa B (NF-κB), which is itself regulated by a suite of enzyme reactions on associated molecules within its signalling pathway, and indeed promotes the transcription of its own inhibitor (Hoffmann et al., 2002). Cells closely regulate all such enzyme systems, and the NF-κB complex can readily be converted between an inactive complex and an active complex by association and dissociation of inhibitor of kappa B (IκB). Only when the NF-κB complex is in its active form does it actually bind to the cognate binding sites (DNA enhancer and promoter sequences) and promote gene transcription within the nucleus (Wirth and Baltimore, 1988). All of the biochemical

reactions upstream of NF-κB activation involve physical changes at the molecular level. The various enzymes involved in the signalling pathway subtly change the shape of proteins through glycosylation or through addition or removal of phosphate groups, for example. The latter mechanism facilitates the degradation of IκB, which creates 'free' NF-κB. Additionally, complexity becomes evident when it is realised that the individual enzymes (e.g. kinases and phosphatases) involved in the signalling pathway are in turn regulated (at least in part), by the degree of external stimulus to the cell, which activate cell membrane receptors (Karin and Ben-Neriah, 2000).

2.3.2 Modularity

In engineering, a module is defined as a *"functional unit that is capable of maintaining its intrinsic properties irrespective of what it is connected to"* (Sauro, 2008). This is an extremely important abstraction, because it allows engineers to think of these functional units as building blocks, which may be joined together to form systems that display predictable behaviours. With this in mind, modularity can be thought of as either representing the dismantling of a system into its constituent parts (in line with the reductionist thinking of biochemistry and molecular biology), or the recomposition of the entire system from its components (Luttge, 2012). Although quite a general definition, this applies to biology because biological systems often reuse certain components (modules) in multiple, and sometimes very different, applications (Navlahka and Bar-Joseph, 2011).

From a biological perspective, the concept of modularity assumes that functionality within the organism can be divided into groups of components (often proteins) that perform logically separable tasks, but work together to achieve some well-defined overall function (Ravasz et al., 2002; von Dassow and Munro, 1999; Yook et al., 2004). In this way, a module may be thought of (at a high-level) as a biological entity (a structure, a process, or a pathway) that is delineated from other entities with which it interacts. A consequence of this modularity is that the number of interactions between modules (intermodular connections) is low compared to the number of interactions within a given module (intramodular connections) (Bruggeman et al., 2002). The exact definition of a module within systems biology however is context dependent: a computational biologist interested in network theory may view modules as loosely linked clusters of highly connected nodes, whereas a biochemist or cell biologist may see parts of the cell machinery as semi-autonomous modules.

Kitano (2001b) builds on these general definitions by advising that modularity also reflects hierarchical organization of the system. With this in mind, the hierarchical organization of modularity within biological systems can be considered to contain four main levels: individual *components* that can be deemed the elementary units of the system; *devices*, that provide a minimum unit of functional assembly; *modules* that comprise clusters of devices; and the *system*, which is the top-level assembly of modules, and depending on perspective can be a cell or the entire animal. Following this hierarchical organization, we can move from individual genes or proteins (components), to co-expressed genes or multi-protein complexes within signal transduction pathways (devices), to metabolic or signalling pathways (modules), and culminating with a cell, organ, or entire

organism (the system). In this way, the intractability of looking at all aspects of a system simultaneously, may be overcome by the simplification that results from applying modularity, as the system-wide behaviour can be considered as the sum of all modular responses (Bruggeman et al., 2002). Furthermore, the physical partitions from such hierarchical organization provide a degree of robustness to the system. Individual modules may be perturbed, however the effects are generally localised, with limited propagation of perturbations to other modules (Lorenz et al., 2011). Modularity therefore ensures that damage to lower levels of the hierarchy does not spread throughout the entire system, and thus mitigates the risk of system-wide failure resulting from an individual, localised perturbation (Kitano, 2001b).

2.3.3 Robustness and Fragility

Kitano (2007) advises that complex biological systems, which have evolved to be optimal, have a *robust yet fragile* nature. Robustness in the context of systems biology is the ability of a biological process or system to withstand or overcome perturbations, and thus maintain stable functionality of the system (Kitano, 2004b). For a system to initiate a response to a perturbation, information about the perturbation is required to spread within its network. The short path lengths of biological networks (metabolic, signalling, and protein interaction) is an extremely important feature, which facilitates fast and efficient reactions to perturbations (Albert, 2005).

A system is robust as long as it maintains functionality, which within biology commonly entails a state transition through a new steady-state when facing stress conditions (Kitano, 2007). The mechanisms and components that promote robustness for a system, often provide the ability to maintain functionality against a large variety of common perturbations, such as mutation, toxins and environmental changes, but unfortunately may also introduce fragility of the system to other, less common perturbations (Carlson and Doyle, 2002; Csete and Doyle, 2002). This has been termed *highly optimized tolerance* by Carlson and Doyle (1999), who propose that optimally robust systems are those that provide a high degree of robustness towards common perturbations, but unfortunately incur catastrophic failures when some rare variations of events occur (Carlson and Doyle, 2000).

Complex biological systems, in particular the immune system, are successful if they are robust against a wide range of external and internal perturbations. It would however be both impractical and indeed undesirable for systems to be robust to every change in external and internal conditions. For example, as Morohashi et al. (2002) state, *"a system should be sensitive to particular types of variation in inputs, otherwise it would not respond to anything!"*. Examples of system properties used in biological systems to maintain robustness include: tolerance of stochastic fluctuations in the dynamics of protein-protein interactions, resilience to fluctuations in protein concentrations, tolerance of stochastic noise, and adaptation to external signals and stimuli (Kitano, 2004b). These features of robustness are enabled through feedback control systems (whether positive or negative) in order to facilitate homeostasis; functional redundancy, via cellular heterogeneity and the use of functionally equivalent modules that can substitute

for each other; modularity, which prevents propagation or amplification of local perturbations (within a module) to other parts of the system (Kitano, 2007); and structural stability, to physically contain physiological effects (Kitano, 2004b).

Along with these features of robustness within complex biological systems, there is a specific network architecture commonly observed in biological interaction networks, called a *bow-tie* motif (Csete and Doyle, 2004). The bow-tie motif has been found in various subsystems of living organisms, including metablic networks, signalling networks and the immune system (Kitano, 2007). It is composed of a highly conserved, and often robust core (the *hub*), connected by diverse and redundant input and output subnetworks, potentially including a variety of feedback control loops (see figure 2.1). Csete and Doyle (2004) and Barabasi and Oltvai (2004) argue that such network architectures provide the biological system with robust and flexible responses to various effector molecules (both stimulatory and inhibitory) due to the high degree of redundancy in input and output modules. However, Albert et al. (2000) and Kitano (2007) also argue that the architecture is also inherently fragile in response to alterations in the network hub, as these molecules may not be redundant, or may not be controlled through regulatory feedback.

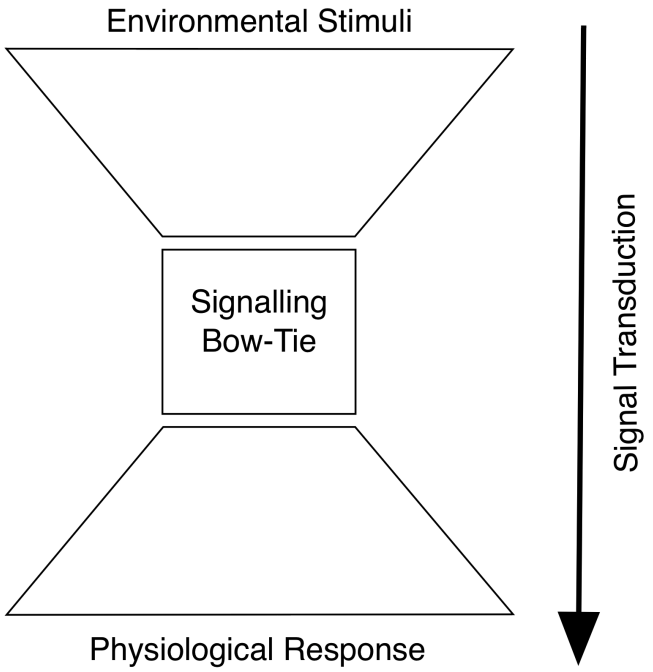

Figure 2.1: Example architecture of the bow-tie motif, which is commonly seen within complex dynamical systems. Here specific emphasis has been given to signalling networks, with a diverse range of environmental stimuli being identified outside of the cell, a centralised signalling pathway being initiated within the cell, and finally a plethora of physiological responses enacted. After Csete and Doyle (2004).

2.3.4 The Systems Biology Project Lifecycle

The merits of following a systems approach to biological research have been discussed in depth by Kitano (2001a, 2002b, 2004a), with his project lifecycle diagram having the potential to become the classic diagrammatic representation of how systems biology is underpinned by a hypothesis-driven research cycle (figure 2.2). Here, research begins through selection of a complex biological system with underlying questions that need answering, and the development of a data-driven, computational model that represents the system. In this respect, a model is a partial representation of the system, that can be used to identify which features are essential. Noble (2002) advises that partial representations of systems are essential, because if a model provided a complete representation, we would be left *"just as wise, or ignorant as before"*. Following development of the model, computational *in silico* experiments (termed *"dry"* experiments by Kitano, as they occur in a computer and not the wet-lab), such as simulation, are then performed. Simulation attempts to predict the dynamics of systems, so that the validity of the underlying assumptions behind the models can be tested. Any necessary refinements to the model are made iteratively until the model is believed to be consistent with the biological system in question (and at the appropriate level of abstraction). Once the computational model is deemed *fit for purpose*, having survived the initial validation, it can be used to generate predictions for further wet-lab based experiments, and to explore questions that are not yet amenable to wet-lab experimental inquiry[2]. The cycle is completed through comparison of the results of these wet-lab experiments with the predictions from *in silico* experimentation.

Over the past twenty years, a systems biology approach to research has become more widespread, with researchers in the biological sciences increasingly using computer models and simulations to better understand intercellular (between cells) and intracellular (within cells) processes of living organisms (Giersch, 2000; Tieri et al., 2005; Young et al., 2008). One of the main strengths of the systems biology approach is that it focuses on three key properties of complex systems: 1) system structures, 2) system dynamics, and 3) system control (Ideker et al., 2001). As discussed above, before a computational model can be developed, an initial degree of understanding is required of the system's structure and dynamics. Such understanding can be based around the network of gene and metabolic interactions inherent in the system, as well as the mechanisms by which these interactions modulate the physical properties of intracellular and multicellular structures; or indeed how the system behaves over time under various conditions, and through identification of essential mechanisms underlying specific behaviours. Once the model is developed, and confirmed through testing to be a reasonable approximation to the real system, simulations may be performed that apply various perturbations to the system so that an understanding

[2]An example of predictions made through *in silico* experimentation that are not yet amenable to wet-lab experimental inquiry is the work of Williams et al. (2013). Here they used *in silico* experimentation to develop hypotheses of dendritic cell mediated regulation of CD8 Treg cells within murine Experimental Autoimmune Encephalomyelitis. The wet-lab technologies have not yet developed sufficiently for these hypotheses to be tested, however this in no way reduces the quality of the work, or its benefits towards increasing our understanding of the intercellular dynamics leading to disease propagation and ultimately spontaneous recovery.

of the mechanisms controlling the state of the system may be ascertained. Once this latter goal is achieved, we may harness this knowledge to control the system through modulation of system states. It is hoped that through such an ability to control system states, we may be able to minimise malfunction and provide therapeutic targets for treatment of disease (Weston and Hood, 2004). Kitano (2000) believes that the future will transform biology and medicine into precision engineering through the use of systems biology and computer engineering to design novel drugs as therapeutic treatments.

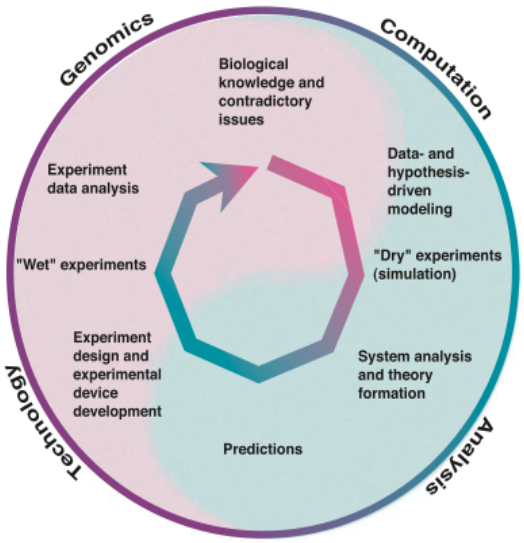

Figure 2.2: The hypothesis-driven research cycle in systems biology from Kitano (2002b). Here Kitano advocates that systems biology research begins with the development of a computational model of a complex biological system of interest, which has questions that need answering. The model is tested through simulation against biological data and underpinning assumptions, updated as necessary, and then used to perform *in silico* experiments. Results of which are then used for hypothesis generation, and the design of new wet-lab based experiments. Following wet-lab experimentation, the results of the two approaches are then compared to test the predictions from *in silico* experimentation. Additionally, these wet-lab experiments may yield new facts/data for use in augmenting the model, with the cycle continuing on an iterative basis.

2.4 Computational Biology

Advances in experimental approaches are expected to continue, however insights into the functioning of biological systems will not result from purely intuitive assaults alone, due to their intrinsic complexity (Kitano, 2002a). Navlahka and Bar-Joseph (2011) argue that biological processes can be considered algorithms that nature uses to solve problems, and that we can better understand their properties by viewing them as information processing units. Recently, the systems biology approach, through harnessing advances in computer software and computational power, has enabled the creation and analysis of reasonably realistic computational models of biology. Models in this context are partial representations of biology, whose purpose is to facilitate a deeper understanding of the

underlying biological system, through for example, identifying which features of the system are necessary and sufficient to understand it (Noble, 2002). Kholodenko (2006) discusses how computational models based on networks, aim to create *in silico* replicas that provide a framework for estimating the temporal dynamics of causal relationships. An example where such computational models are proving essential for advancing our knowledge of system dynamics is regulatory feedback (either positive or negative) within biological networks, along with the phenomena of *combinatorial complexity*[3], which often makes it difficult to manually predict the various implications of hypothetical perturbations to the network (Chakraborty and Das, 2010). Furthermore, Blair et al. (2012) advise that computational models of biological networks can be used to perform *in silico* experimentation to predict responses to genetic and environmental perturbations. This approach of using computational models to further our understanding of the underlying biology has been termed *Computational Biology*.

There are many analogous features between biological systems and engineered computational systems. Firstly, like large-scale computing infrastructures, biological systems are distributed in nature, consisting of molecules, cells or organisms that interact, coordinate and make decisions without central control (Seeley, 2002; Babaoglu et al., 2006). Secondly, just as computing systems are designed with fault tolerance in mind, biological processes also need to be robust, in order to successfully manage external threats and internal failures (Jeong et al., 2000; Kitano, 2004a). Thirdly, both biological and computational networks are necessary structures for the propagation of information (Alon, 2006) throughout the systems to elicit relevant responses. Fourthly, in a similar way to the design of object-oriented computer systems, biological systems are often modular, having the capability to reuse certain components on multiple, and sometimes very different applications (Navlahka and Bar-Joseph, 2011). Finally, as with computational systems that employ pseudo-random number generators, biological systems often comprise a number of stochastic processes (Kaern et al., 2005).

In general, the term computational biology can be defined as the study of how computers and computational techniques are used to address issues relating to biological structure and function. The precise definition however is a matter of some debate, with the most narrow usage refering to the creation and management of biological data, and the most broad usage covering all applications of computers for solving problems in biology (Altman, 1998). In practice however, computational biology can be thought of as having two distinct branches: knowledge discovery or data-mining, which is also known as bioinformatics, and tries to generate hypotheses by extracting the deep underlying patterns from massive experimental datasets; and simulation-based analysis, which tests hypotheses using *in silico* experimental approaches, to generate predictions for testing through additional *in vivo* (within a living organism) and *in vitro* (outside of a living organism) experimentation (see Kitano's hypothesis-driven research cycle in figure 2.2). These theoretical investigations use programming languages to encode algo-

[3]Combinatorial complexity relates to network topologies where individual components are linked to more than two other components. In such instances an individual component may be connected to one or more input components and one or more output components. Any increase in the total number of components, results in an even greater increase in the total number of connections within the network. Within biological networks, this increase in total connections often follows non-linear dynamics.

rithms[4] that can be used to further our understanding of the temporal, spatial and causal aspects of biological systems (Priami, 2009). Wada (2000) believes that through the use of computational models, the biological sciences will be transformed from a *data-driven* to a *model-driven* enterprise. Furthermore, Kitano (2004b) and Slepchenko et al. (2002) believe that the theoretical investigations (through *in silico* experimentation) into the underlying cellular dynamics of biological systems, followed up by verification in actual biological systems, needs to be promoted as a new aspect of *the scientific method*.

Modern computer architectures and modelling frameworks can monitor and record every possible event that could occur within a hypothetical network model and uncover how particular network perturbations may influence overall network dynamics. Computational models can therefore be useful tools for exploring the behaviours and dynamics of biological systems, especially the networks that underpin these systems (Andrews and Bray, 2004). Through *in silico* investigations, these models provide a relatively easy mechanism for testing complex hypotheses of how complex low-level system dynamics result in the myriad of system-wide behaviours. For example, Noble (2006) describes a multi-scale computational model of the heart, that utilises the low-level interactions of protein biochemistry and through aggregation up the component hierarchy (protein to cell to tissue to organ) simulates organ-level physiology. The Physiome Project[5] is taking this approach further by developing a framework for modelling the entire human body using computational approaches (Bassingthwaighte, 2000). The range of spatial and temporal scales is too great for a single model, therefore the project aims to develop multiple computational models incorporating biochemical, biophysical and anatomical data relating to cells, tissues and organs, and it is expected that these will be linked (via aggregation of lower-level sub-system dynamics) to allow simulation of the entire human body (Hunter and Borg, 2003; Hunter et al., 2006).

2.5 Computational Immunology

The discipline of computational biology is vast, encompassing all fields of biology which have synergistically joined forces with computer science. Just as biology can be broken down into sub-disciplines, such as cell biology, biochemistry, immunology, etc, a growing speciality within the discipline of computational biology relates to the immune system, and has been termed *computational immunology*. Like all scientific disciplines, breakthroughs in immunology are driven by experimental observations. These observations provide the ability to record facts that can be integrated into hypotheses and theoretical models, which are amenable to further experimental tests (Chakraborty et al., 2003).

The immune system provides our primary protection mechanism against invading organisms (e.g. bacteria, viruses, fungi, or other microorganisms) and cells that have become uncontrollably changed (Perelson and Weisbuch, 1997). It is a highly concurrent system, with changes in the surrounding environment having the potential to trigger multiple parallel processes. An example is that a

[4]These algorithms can be solved if encoded within a mathematical equation, or simulated if encoded within a simulator

[5]http://www.physiome.org

cell carries out multiple tasks in parallel, which may relate to the immunological state, the phase of the cell within its cell-cycle, and the cells anatomical location (Kam et al., 2001). Furthermore, the immune system functions over multiple hierarchies, locations and timescales, which involve feedback to fine-tune the appropriate defensive response (Murphy et al., 2008).

Defects in the detection of pathogens, signal transduction following pathogenic attack, or cessation of the resulting immune response, form a number of the underlying causes of many diseases, and a greater understanding of these defects may facilitate a therapeutic treatment. With this in mind, diseases can be viewed as the malfunction of system robustness under normal physiological dynamics, and re-establishment of robust and progressive disease states (Kitano, 2004b). One reason for understanding these immune response (or disease state) network dynamics is to adapt system-wide control (i.e. alter the systems robustness under a diseased state) for the design and development of therapeutic drug interventions (Kitano, 2004b). Indeed, faulty receptors controlling signalling pathways can be targeted with antibody drugs, whilst the individual components of signalling pathways inside the cell are often targeted with small-molecule drugs (Kumar et al., 2006). Finding such fragilities within systems that demonstrate a diseased state, requires an in-depth understanding of the dynamics of the gene regulatory and biochemical networks of the cells. It must be stated however, that most diseases are far too complex to be determined by the activity of a single rogue component (Aksenov et al., 2005). Due to the complexity of the immune system containing parallel tasks, multiple locations and diversity of cells and signals; computational immunology is becoming an essential component of modern immunological research and drug discovery. It aims to study the complex cellular and subcellular interactions and networks, to elucidate a better understanding of immune responses and their role during normal, diseased, and reconstituted states (Tong and Ren, 2009), and ultimately to capture systematically the effect of a given intervention strategy on complex molecular networks.

The complexity of the immune system due to it spanning multiple levels (genes up to whole organisms), time scales (subsecond through lifetime), the large number of components (genes, proteins, cells, organs), and stochastic events, makes it difficult to build simple models. Computational approaches to modelling have a lot to offer towards this end, as techniques such as agent-based modelling and simulation allow models to be developed at varying levels of abstraction, and are able to generate bottom-up system dynamics over a number of hierarchies. Computational models of immunological responses and disease take a systems biology approach to understand the underlying complexity of the *normal* and *abnormal* system dynamics (Dancik et al., 2010). Furthermore, computational modelling provides an opportunity to integrate data generated from multiple types of experimentation, to perform a kind of meta-analysis targeted to the immune response in healthy and diseased states (Kleinstein, 2008).

Models and simulations are generally based on the tools available at the time, and early examples used cellular automata for the simulation of the humoral immune system (Seiden and Celada, 1992) and dynamics of HIV infection (dos Santos and Coutinho, 2001), differential equations to model immune responses to HIV (Perelson and Nelson, 1999), and development of a network of immune system cells using petri nets (Ootsuki and Sekiguchi, 1999). With the advances over

time to computing power, hardware (particularly graphics), and the widespread use of object-oriented programming languages, immune system models evolved to utilise parallel processing through advanced cellular automata (Bernaschi and Castiglione, 2001), and more recently agent-based modelling and simulation. Examples of the latter include a model of immune system interactions (Kleinstein and Seiden, 2000), a simulation depicting the dynamics of thymocyte development (Efroni et al., 2007), and the use of a simulator to perform *in silico* experimentation with an animal model of Multiple Sclerosis (Williams et al., 2013).

Although it is still very early in the lifetime of computational immunology, the examples above indicate that computational models can be useful complements to genetic and biochemical experiments. We expect these successes will encourage further collaborations between wet-lab experimentalists and computer scientists, aimed towards elucidation of the complex biology of cells. The development of computational models of infection and disease, and their response to computer generated perturbations allow the initial testing of hypotheses to be performed *in silico*, which may then be validated through subsequent wet-lab experimentation. Petrovsky and Brusic (2002) suggest that this offers the ability to accelerate the drug discovery process relating to therapeutic treatment of diseases, by reducing the number of potential hypotheses of drug targets through *in silico* experimentation, to yield only those hypotheses with the greatest probability of success during laboratory-based research and development, and resultant clinical trials.

The examples presented here illustrate the power of computational approaches to complex immunological problems, however we believe there to be difficulties in modelling the inherent scale of the immune system. To the best of our knowledge, we are not aware of a model that has seemlessly passed information from one level to another, building up a hierarchy analogous to the human body. Gary An developed a multiscale model of acute inflammation[6] through the NF-κB signalling pathway (An, 2008), however upon close inspection it appears that the model was in fact a proof of concept, which positioned an architecture based on a set of separate models, developed at different levels of abstraction and linked together in a hierarchical fashion. The first level was that of individual cells within an immune response, notably endothelial cells, blood-borne inflammatory cells, and epithelial cells. The second level was that of an organ, the examples used being the gut, lung and the pulmonary system. Finally, the third level was deemed the whole human body, although this was not visualised in the model, as it consisted of the cumulative effects of the organs. A detailed inspection of the model design shows that the organs are linked in a pair-wise manner, and resulting cell dynamics are confined to movement between these two organs. As such, the model does not represent a true multiscale hierarchy, as it does not conform to the pyramid of life proposed by Oltvai and Barabasi (2002), but instead operates at the level of a set of boundaries between pairs of organs, and therefore could be thought of as more of a *sandwich* than a pyramid. We believe that this work provides a useful incremental step towards the goal of multiscale modelling of the immune system, but conclude that there is still plenty of work to do in this

[6]Inflammation is a hallmark of many human diseases, and is a localised physical reaction to injury or infection. Elucidating the mechanisms of uncontrolled inflammation has long been a key aim of immunological research (Chen et al., 2008).

area. Indeed, a workshop held in Tokyo and consisting of systems biologists and representatives of the pharmaceutical industry laid down a grand challenge to create a 'virtual human' within the next 30 years (Jones, 2008).

2.6 Computational Modelling

The last half of the twentieth century saw phenomenal use of, and indeed success with, the reductionist approach. This success comes at a cost however, and has resulted in the prodigious challenge, which the first branch of computational biology (knowledge generation and data-mining) may be able to help with. Indeed, the challenge from the huge amount of data generated was neatly summed up by Denis Noble in 2002 when he asked *what do all those data mean?*. With regards to the second branch of computational biology (modelling and simulation), it appears that the notion of modelling a biological process computationally is almost as old as the electronic computer itself. Alan Turing, in 1952, modelled the movement of chemical substances (which he termed morphogens) within a cell, along with a number of basic chemical reactions associated with these morphogens, and their diffusion across the cell membrane into other cells. Turing defined the state of the system and how the state of an individual cell is determined by using the state at a moment very shortly before. He modelled the chemical reactions using linear differential equations, and when certain conditions were met regarding a cell's morphogen composition and concentrations, the cell's state changed.

The complexity of systems now under consideration, along with the technicalities of the experiments and the non-linear nature of biochemical and molecular interactions make it necessary to use computational modelling in the pursuit of furthering our understanding of biology (Wolkenhauer et al., 2005). Computational models allow us to gain insights into the complex relationships between the extracellular stimuli, the intracellular reactions, and the overall cellular responses. They also provide an avenue for us to predict the mechanisms that are responsible for signal amplification, noise reduction and discontinuous bistable dynamics, through for example feedback inhibition. It is believed that an increased understanding of protein, gene regulatory, and signalling networks will help identify the critical controlling factors of human diseases (Kholodenko, 2006), in particular complex diseases that cannot be mapped to a single gene or component, and that this information will improve drug development efforts and ultimately lead to preventive drugs (Weston and Hood, 2004). Sontag (2005) argues that cells can be thought of as comprising a large number of subsystems that are involved in various processes such as cell growth and maintenance, division, and death, and that the study of these simpler subsystems is a first step towards understanding the emergent properties of the cell as a whole. Along with a focus at the intra- and inter-cellular levels to identify the controlling factors of diseases, modelling efforts for drug discovery and development must also simulate responses at the tissue or organ level, and as such are required to cover multiple hierarchies of biological function. Multiscale models are however difficult to develop as they require the efficient integration of molecular, cellular and organ levels within a single model, which can require very large amounts of computational resources using current software platforms. Furthermore, there are limitations in the extent of our bottom-up knowledge, hence our use of computational models as tools

to extend our understanding; but also a lack of standardised experimental data (i.e. using similar wet-lab methodologies and reagents) to validate these models (Butcher et al., 2004). Agreement between the computational model and the true underlying biology is a central goal of systems biology (Kitano, 2002a).

A variety of modelling approaches have been used within the systems biology community to study complex biological systems. The quality of such models is invariably related to the quality, completeness and detail of the data available (Giersch, 2000), and indeed the data will usually drive the approach to be taken for modelling, e.g. a diagrammatic versus mathematic versus agent-based approach. Traditionally, diagrammatic approaches have been used for modelling biological systems, as they allow an easy and intuitive method for conceptualising our understanding of component interactions and hyothesising the underlying mechanisms of system behaviour. Diagrammatic approaches suffer however from the fact that they depict static representations of the system, and are therefore unable to suitably convey dynamic representations of the system in time and space. Other computational modelling approaches can be used to overcome these limitations, and are providing increasingly powerful approaches that allow us to elucidate underlying functional principles relating to the dynamics of biology.

In the context of systems biology, computational modelling can be divided into two broad categories based on the way in which they describe system dynamics - continuous or discrete event. The first category aims to convert a detailed conceptual model of the system in question (e.g. signalling pathway) and the measured values of parameters (e.g. protein concentrations and chemical reaction rates) to a computational model that may be used to evaluate quantitative hypotheses, and generally use mathematical models (Chakraborty and Das, 2010). The second category is more qualitative in nature and aims to study the mechanistic behaviour of systems, and to generate new hypotheses for currently unexplained phenomena. This latter category uses simulation-based approaches, and focuses on the dynamics and interplay between biological components (such as cells, tissues and organs) using a holistic approach rather than reductionism, which typically excludes information regarding time and space (Ahn et al., 2006). This rule-based approach provides the advantage of being able to distinguish individual components (e.g. cells and molecules) through their location, developmental states, and specificities, whereas equations limit you to a homogenous population (Woelke et al., 2010). Through the translation of biological knowledge into equations or rules, a framework becomes available upon which descriptions, and ultimately understanding of the system can be developed (Kirschner and Linderman, 2009). Furthermore, by extending and integrating the effects of assumptions made in these mathematical and rule-based models at various levels of biological organisation (e.g. cell, organ, organism, or population), these approaches allow us to visualise hypothetical scenarios across time, space and hierarchies (Wildermuth, 2000).

2.6.1 Diagrammatic Modelling

Perhaps the simplest models of complex biology are diagrammatic in nature. Following on from early work visualising bicohemical pathways by joining individual enzyme-catalyzed reaction components, such as the citric acid cycle (Krebs,

1940), and the consolidated network diagram of Nicholson (2001), a number of new approaches have been used since the 1990s which have borrowed standardised diagrammatic languages from computer science. One such standard is the Unified Modelling Language (UML)[7], which although originally developed to document technical requirements for the analysis and design of computer systems (Booch et al., 1998), has recently been used to model complex biological systems. The UML specification (version 2.4) (Object Management Group, 2011) states that *"The objective of UML is to provide system architects, software engineers, and software developers with the tools for analysis, design, and implementation of software based systems as well as for modelling business and similar processes"*. The UML specification defines 14 separate diagramming notations, split across three main groups: structure diagrams, which show the static structure of components within a system; behaviour diagrams, which show the dynamic behaviour(s) of components within a system; and implementation diagrams, which show the hardware and software infrastructures within a system (see figure 2.3).

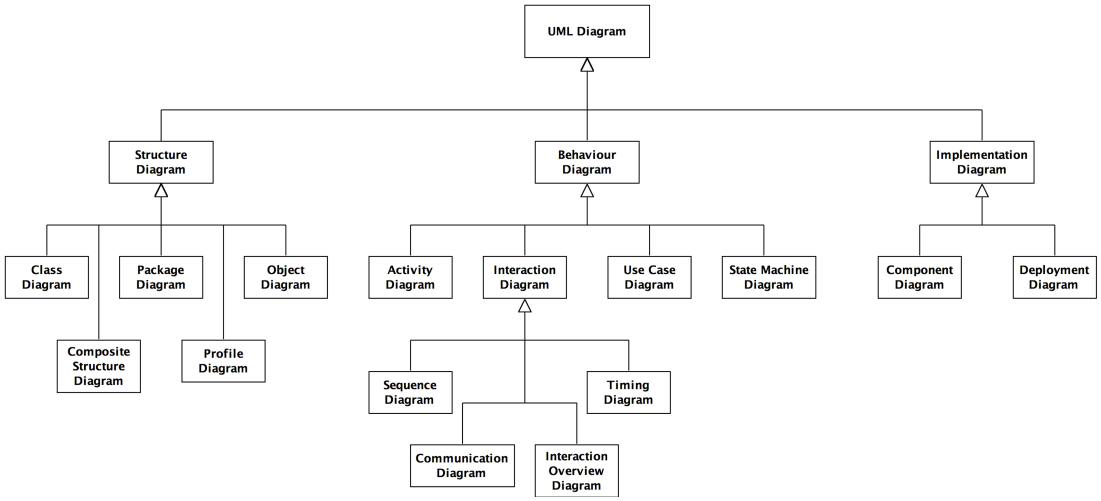

Figure 2.3: Taxonomy of UML diagramming notations (after Object Management Group (2011)). Structure diagrams show the static structure of the components within a system, and comprise: Class, Composite Structure, Package, Profile and Object diagrams. Behaviour diagrams show the dynamic behaviour of the components within a system, and comprise: Activity, Sequence, Communication, State Machine, Use Case, Interaction Overview, and Timing diagrams. Finally, implementation diagrams comprise Component and Deployment diagrams.

Webb and White (2005) and Bersini et al. (2012) argue that the principles of object-oriented analysis and design inherent in UML can be directly applied to the top-down modelling of cells, and bottom-up modelling of metabolic pathways and cell signalling cycles. Examples of models that use UML include Kam et al. (2001) who modelled T-cell activation using statecharts, which are the forerunner to UML's state machine diagram notation, and Webb and White (2004) who extensively used UML to model the structure of a cell, by representing the relationships/hierarchies and composition of various biological components, along

[7]UML is maintained by the Object Management Group, and is available from www.uml.org Two useful books on the UML notations are (Fowler, 2004) and (Arlow and Neustadt, 2009).

with a generic enzyme-catalysed biochemical reaction. Similarly, they later used UML to graphically design an executable model of a simple biological control system in which an enzyme (glycogen phosphorylase) continuously transforms glycogen into a more readily usable sugar (Webb, 2007). More recently, UML has been used to model the intercellular interaction network of Experimental Autoimmune Encephalomyelitis (an animal form of Multiple Sclerosis) (Read et al., 2009a, 2014), the differentiation of T-cells during their normal lifecycle (Bersini et al., 2012), and as part of the design process for an agent-based model of tissue formation in the lymphoid organ (Alden et al., 2012; Patel et al., 2012).

Following on from the early days of visualising biological systems using UML, there has recently been considerable activity around diagrammatic notations in biology. A number of key advances have been: molecular interaction maps, which have been used to model a generic G-protein coupled reaction at the plasma membrane (Kohn et al., 2006); Biocharts, which have been used to represent bacterial movement through chemotaxis (Kugler et al., 2009); and the Systems Biology Graphical Notation (SBGN, (le Novere et al., 2009)), which extends the principles inherent in UML and has been used to diagrammatically model the Toll-like receptor network (Oda and Kitano, 2006), the mTOR signalling network (Caron et al., 2010), and NF-κB related interactions (in supplementary figure 3 of le Novere et al. (2009)).

Of these newer diagrammatic notations, we believe that the SBGN has the most potential to rival UML as the language of choice for diagrammatically modelling complex biological systems. Kitano (2003) advises that most authors of biological papers relating to biochemical networks use a standard notation of arrow-headed lines and bar-headed lines to indicate activation and inhibition, but then go on to use mixed and often inconsistent semantics for the low-level details of the reactions. The SBGN was developed by an international collaboration of biochemists, computational biologists and computer scientists, with the overriding objective to allow scientists to diagrammatically represent networks of biochemical interactions using standardised terminology and notation. Whereas UML contains 14 different notations, the SBGN currently only contains 3, being: the process diagram, the entity-relationship diagram[8], and the activity diagram (Kitano et al., 2005; le Novere et al., 2009; van Iersel et al., 2012).

Taking these 3 notations in turn, firstly, the process diagrams are used for modelling the interactions that take place between biomolecules and the various state-transitions that occur as part of the biochemical reaction. They are able to convey the temporal aspects of molecular events occuring in biochemical reactions, and are analogous to UML sequence and communication diagrams. The main drawback with process diagrams appears to be that a given component must appear multiple times on the same diagram if it exists under several states, whereas in UML you can have one object with several activities coming off, that through the use of *guard conditions*, can specify which activity occur under specific circumstances. Indeed, the requirement for SBGN process diagrams to diagrammatically define all states that a component can take, can become problematic. For example, a biological component that acts as a *hub* in a network will have a large number of connections and therefore possible network states. These

[8]The SBGN entity-relationship diagrams should not be confused with the Entity-Relationship (ER) diagrams relating to database schema, i.e. normal form or Boyce-Codd normal form

all have to be defined separately in process diagrams, which leads to the issue of *combinatorial explosion* identified by Tiger et al. (2012). Secondly, the entity-relationship diagrams are based on Kohn's molecular interaction maps and are used for modelling the relationships between biomolecules. These focus on the influences that entities have on each other, but not the state transformations that occur following interactions; they are akin to UML class diagrams and activity diagrams. Unlike the process diagrams, an entity appears only once, which is closer to the approach of UML. An enhancement over UML with respect to modelling biology is that these diagrams have specific notations for low-level biochemical reactions such as phosphorylation, which can be displayed on specific amino acid residues of protein entities. Finally, the activity flow diagrams, are used for modelling the activities of biomolecules at a high-level of abstraction. They can be used to convey component-level interactions (e.g. protein-protein), without the need to show the detail of specific chemical reactions at the level of individual amino acids (e.g. phosphorylation events). As the activity diagram ignores the specific biochemical processes that entities are involved in, and their associated state transitions, they are quite compact in nature, and can be thought of as the typical network diagram found in traditional biochemical textbooks. As per UML, these 3 notations complement each other and are used to diagrammatically model different aspects/views of the biological system.

One of the most comprehensive examples of diagrammatic modelling using SBGN is that of Mizuno et al. (2012), who constructed a comprehensive map of the intra-, inter- and extra-cellular pathways involved in Alzheimer's disease. They manually curated over 100 review articles, and the resulting diagrammatic network map consisted of 1,347 molecules and 1,070 biochemical reactions. These diagrammatic models developed using UML and SBGN are useful for providing a static picture of the biological system, but unfortunately describe to a lesser extent the relationships between components that lead to system behaviours. Therefore, in order to model the full behaviour of a system, we require more dynamic approaches to computational modelling, that are capable of modelling the systems associated component-level interactions in time and space.

2.6.2 Mathematical Modelling

Intercellular networks and the various metabolic and signalling pathways within cells are complex and often display non-linear dynamics. Systems biologists often model these biological processes mathematically through equations, be they reaction kinetics, which look at the rates of chemical processes, or differential equations, which look at unknown functions within processes or systems, and link these to a number of known input and output variables. Mathematical modelling is interested in the quantitative dynamics of complex systems at the population level of components, which for the immune system would be the population of individual cells or biochemical/molecular components. It allows the investigation and analysis of aspects of complex systems that we are unable to observe or understand directly. This may be due to difficulties in observing system dynamics due to extremely short or indeed large timescales, to the magnitude of the system to be observed (e.g. whole population), or to the location or complexity of the system (Cho and Wolkenhauer, 2003).

Fisher and Henzinger (2007) advise that *"a mathematical model is a formal model whose primary semantics is denotational; that is, the model describes by equations a relationship between quantities and how they change over time"*. Equations are typically used to quantitatively describe patterns in data, and extrapolate to predict dynamics outside of the range of the data from wet-lab experimentation. As described earlier, the immune system is a dynamic network of components (e.g. cell type, protein species, etc), and these become the variables of mathematical models, allowing the generation of time-dependent behaviour of the component concentrations (if intracellular, e.g. protein) or the total number of components within the population (if a cell type). Three interesting examples of how mathematical modelling techniques can be applied to the immune system have focused on the evolutionary dynamics of the Human Immunodeficiency Virus (HIV). The first, by Sguanci et al. (2006), modelled the within-patient evolutionary process of HIV infection, using differential equations that contained variables for mutant virus strains, along with cell surface receptors and co-receptors; they hypothesise that accumulated mutations in HIV gained through molecular evolutionary processes may enable interaction with other cell surface receptors over time. The second, by Bagnoli et al. (2006), modelled the coexistence of different HIV viral strains and the competition between them under different immune system conditions. The third, Sguanci et al. (2007), presented a model of HIV early infection, which describes the infection dynamics of different HIV viral species.

There are two main approaches to mathematically model the dynamics of biological systems: the *deterministic* approach or the *stochastic* approach (Gillespie, 1977). Deterministic reactions are assumed to take place in uniform (i.e. homogenous) biochemical environments, such as the internal cellular compartments that focus on energy production, e.g. the mitochondria, where large-scale biochemical reactions take place to harness the energy stored within various sugars. Alternatively, the stochastic approach to modelling is necessary for environments where system components are not in such abundance, and where there is a degree of random chance involved in a given reaction occuring (Sreenath et al., 2008).

Most metabolic pathways can be assumed to have components in large numbers that are well mixed. In these circumstances, the dynamics of each component population can be described mathematically by an Ordinary Differential Equation (ODE, see below). ODEs simulate behaviour in a deterministic manner, and are good for modelling population/system-level dynamics using continuous time (Khan et al., 2003).

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum f(x) - \sum g(x)$$

The generic ODE above can be used to describe the kinetics of a system over continuous time. When solved it provides the total number of system components (x) at a given time interval (t). This is made up of the cumulative effects of all increases in the population through function f(x) and all decreases in the population through function g(x).

Situations where the components require separation into discrete locations (such as organs or sub-cellular structures) cannot be effectively modelled using ODEs, and in this case require Partial Differential Equation (PDE) based

models, where the dynamics of the system are situated within physical space. Individual differential equations (whether ODE or PDE) can be coupled together to form linked systems of differential equations, and this is where their power lies to model complex biological behaviours.

Differential equations assume that each dynamical *species* (e.g. protein, RNA, gene) of a system is present in large numbers, so that their spatial concentrations can be treated as continuous variables. This assumption is usually fine for certain types of proteins, which may have an abundance in the tens (or hundreds) of thousands, however genes and some mRNA may be present at far lower levels, for example one or two molecules per cell (Behn, 2007; Ferrell, 2009). At such low numbers, a mean average description is no longer justified because each individual event becomes pivotal to cell dynamics. In this situation, the modelling approach needs to take account of the probabilistic interactions within the system, with one suitable mathematical approach being the use of Stochastic Differential Equations (SDE), which are able to describe systems where the number of molecules involved is so small that microscopic fluctuations can produce macroscopic effects (Alves et al., 2006). Furthermore, these SDEs are able to approximate the dynamic behaviour of the system propagating through space, for example subcellular compartmentalisation, or more specifically in relation to the signalling pathways, the translocation of a molecule from the cytoplasm to the nucleus. Stochastic modelling approaches are therefore more representative of biological behaviour because they consider each molecule in the system as a discrete entity and each productive collision of molecules as a discrete reaction event. There are drawbacks however, in that they require each process within the system to be described in terms of elementary reactions, therefore requiring a much greater understanding of the *in vitro* or *in vivo* biology; they rely on the assumption that system parameters do not change over time; they do not explicitly account for spatial issues (i.e. actual cartesian co-ordinates); and more importantly from a computational perspective, they tend to be more computationally intensive, especially when the models have large numbers (in the hundreds of thousands) of molecules.

2.6.3 Agent-Based Modelling and Simulation

Although quantitative mathematical models are well established tools for modelling complex biological phenomena, they require an exhaustive set of precise parameters to be specified for each variable. This is fine for small models that align to dynamics of a few components, however when the scale is increased to capture a more realistic scope at the system-level (Acerbi et al., 2012), they begin to suffer from limitations in accuracy as data for use in analysing the effectiveness of differential equation models is often unavailable from wet-lab immunological experiments (Perelson and Weisbuch, 1997). Furthermore, as they rely on population-level averaged dynamics, they are unable to model dynamics at the individual component-level and therefore suffer considerably from their inability to capture the natural variation inherent to all biological processes. An alternative to these equation based approaches is to model molecules or cells individually and assign probabilities to each possible interaction or state change through rule-based techniques. The individual component behaviours may then

be aggregated up to system-level dynamics, which are then extrapolated in order to make predictions of the system-level behaviours in real biology (Cohn and Mata, 2007; Stark et al., 2007).

Cohen (2007) suggests that a computational model should comprise seven key characteristics in order to make it suitable for simulation and generation of predictions for biological system-level behaviours. First, the model must be data-driven and therefore designed and developed from the bottom-up. Second, biologists experiment at an abstracted level, focusing on molecules, cells and organisms; the model should therefore take a rule-based approach based on these abstracted 'objects'. Third, the model must be dynamic and not a static diagrammatic representation. Fourth, in order to allow the emergence of system-level behaviour from individual components, the model must be multi-scalar. Fifth, the model must be designed and developed in a modular fashion, so that new functionality can be added following advancements in wet-lab technologies and new knowledge or information regarding the biological system. Sixth, in order to be useful the computational model should be interactive, with the ability to peform novel *in silico* experiments as a precursor to doing the real experiments *in vivo* or *in vitro*. Finally, the model needs to be realistic, in that its design should be based on real data (further emphasising the first key characteristic) and accepted knowledge from the wet-lab, with the bare minimum of *artistic license* used for modelling mechanistic behaviours, and where this is necessary it should be fully documented as design assumptions.

Interestingly, it would appear that a large amount of data generated experimentally in biology actually accumulates in an object-oriented manner. The reductionist approach, which endeavours to look at systems using the smallest indivisible unit, is analogous to looking for 'objects' within nature. Object-oriented approaches to modelling, therefore provide a useful formalism for constructing computational models by designing systems from a bottom-up perspective and organising information around individual objects (Kam et al., 2001). Agent-Based Modelling and Simulation (ABMS) builds on the object-oriented paradigm, with the key enhancement being that an *agent* is active rather than passive, and that ABMS has multiple threads of control. Macal and North (2005) nicely describe this by stating that the *"fundamental feature of an agent is the capacity of the component to make independent decisions"*. For example, a given protein within a biochemical pathway may be defined as an agent; as there are potentially tens (if not hundreds) of thousands of such protein molecules within a cell, an agent-based approach would allow each of these molecules to have their own distinct *life*. The concept of an agent has been further defined by Wooldridge (1997), who states that an agent is *"an encapsulated computer system that is situated in some environment and that is capable of flexible, autonomous action in that environment, in order to meet its design objectives"*. Jennings (2000) and Bauer et al. (2009) extend the idea of an agent being situated (or embedded) within a particular environment, by advising that they are asynchronous, thus do not evolve at constant time-steps, but are instead interactive with the environment - they will receive inputs related to the state of their environment through sensors, and may act on the environment through effectors. Therefore, to formally define agents, a model is needed which not only represents them as individuals, but also provides the ability for individual agents to process information from the

environment in which they are situated and to communicate with each other.

Fisher and Henzinger (2007) advise that an agent-based model *"is a formal model whose primary semantics is operational; that is, the model prescribes a sequence of steps or instructions that can be executed by an abstract machine, which can be implemented on a real computer"*. The main roots of ABMS are in modelling social networks of humans and the dynamics of their decision making (Bonabeau, 2002). More specifically, it was developed to investigate complex adaptive systems from a reductionist perspective, by using the underlying notion that systems are built from the ground up, in contrast to the top-down system view taken by systems theory (Macal and North, 2005). Since its inception, ABMS has rapidly expanded out of the realm of the social sciences, and has received major interest from other fields, including biology and medicine. A good example of research within medicine is the work of Zhang et al. (2007) who developed a 3D multi-scale agent-based model of a brain tumour. This model integrated data and yielded experimentally testable hypotheses regarding switching behaviour of tumour cells, which are able to proliferate, but do not do both at the same time. They also showed that over time, proliferative and migratory cell populations oscillate, suggesting a dynamic relationship, and indeed provided insight into the growth of tumours through migration-proliferation oscillations. Agent-based modelling at the cellular and subcellular levels is also becoming an area of increasing research interest, due to the desire to understand cellular processes at increasing levels of detail. Good examples include carbohydrate oxidation through glycolysis and the TCA cycle (Corradini et al., 2005), the immune response to atherogenesis due to low density lipoproteins (Pappalardo et al., 2008), and the intracellular NF-κB signalling pathway (Pogson et al., 2006), which was also used to predict that actin filaments of the cytoskeleton sequestered excess IκB, which affected the control of NF-κB (Pogson et al., 2008).

Examples of the use of ABMS within computational biology include: Pappalardo et al. (2011) who used an agent-based model as a predictive tool for novel wet-lab experimentation into tumour formation in skin cancer; Ray et al. (2009) who investigated the role of Tumour Necrosis Factor (TNF) in the control of tuberculosis; and Patel et al. (2012) who predicted computationally that lymphoid tissue initiator cells were not required for the initiation phase of cell aggregation to form Peyers Patches, and subsequently confirmed this prediction through wet-lab experimentation.

Communicating X-Machines

Agents can be represented through a number of different computational techniques. The one thing these techniques have in common however is that they portray the agent as having a defined *state* at any particular moment in time. Representations of the agent may therefore be viewed as a *state machine*, where the current input in combination with all past inputs (previous states) determines the output (next state) of the agent. A *finite state machine* is an abstract mathematical model that represents the computation of system dynamics as transitions through a finite set of states (Kehris et al., 2000). As such, they provide an intuitive means to describe the dynamical behaviour of systems, through use of a formalised 8-tuple notation, as defined by Ipate and Holcombe (1998):

$$X = (\Sigma,\Gamma,Q,M,\Phi,F,q_0,m_0)$$

where:
- Σ and Γ are the input and output alphabets respectively.
- Q is the finite set of states.
- M is the (possibly) infinite set called memory.
- Φ, the *type* of the X-Machine X, is a set of partial functions φ that map an input and a memory state to an output and a possibly different memory state, $\varphi : \Sigma \text{ x } M \rightarrow \Gamma \text{ x } M$.
- F is the next state partial function, $F : Q \text{ x } \Phi \rightarrow Q$, which given a state and a function from the type Φ determines the next state. F is often described as a state transition diagram.
- q_0 and m_0 are the initial state and initial memory respectively.

Although X-Machines provide the ability to formally define a mathematical model of an individual state machine, they do not have a mechanism for communication between state machines. As such, they are only able to model a single component, and are therefore not suitable for agent-based modelling, as they could only be used to model the system as a whole. An extended version of the finite state machine has been proposed that allows communication between individual state machines, which has been termed the communicating X-Machine. This is a formalised mathematical model that can be used to compute the functional behaviour of smaller components (i.e. individual agents), whose dynamics may be aggregated to generate the emergent behaviour within the entire system. Like the finite state machine, the communicating X-Machine is based upon a simple set of rules that describe what the agent must/could do under different circumstances. The enhancement over finite state machine is that the state transitions are not simply labelled by inputs, but may include functions that operate on these inputs, along with another attribute storing the memory state, thus utilising a 10-tuple formal notation. Here C_i^X represents the ith communicating X-Machine component, and comprises the 8-tuple as per finite state machines, along with an operational function that utilises the input messages and also produces output messages (Stamatopoulou et al., 2007):

$$C_i^X = (\Sigma_i,\Gamma_i,Q_i,M_i,\Phi_i,F_i,q_{0i},m_{0i},I\Phi_i,O\Phi_i)$$

where:
- $\Sigma_i,\Gamma_i,Q_i,M_i,\Phi_i,F_i,q_{0i},m_{0i}$ are the same as in a finite state machine.
- $I\Phi_i$ is the communication interface for the input messages.
- $O\Phi_i$ is the communication interface for the output messages.

Holcombe (1988) defines an X-Machine as a system that has an internal *computational* state and an internal memory, which can transition to another state dependent on environmental input and their current internal state. The communicating X-Machine is therefore able to encapsulate both the dynamic and functional behaviour of an agent, as well as the underlying data that it is modelled on, in a single process specification (Barnard et al., 1996). Communication between individual X-Machines occurs through a 'communication matrix', which is essentially a *message board* that facilitates the reading and writing of information between every X-Machine, allowing communication and interaction between

the machines. Individual agents start with an initial computational state, and upon receiving an external input (e.g. communication from another agent), they update this state, based on the rule regarding their current state and the particular external input received (Balanescu et al., 1999; Kefalas et al., 2003). Following this, it will change internal and/or memory state, potentially generate a message for communication, or continue its current behaviour.

Flexible Large-scale Agent-based Modelling Environment

Agent-based modelling has proven itself to be a powerful technique for simulating the emergent behaviour of groups, through the use of communication between the individual autonomous agents. As defined previously, these individual agents behave according to predefined rules, and in most cases may be defined as communicating X-Machines. A consequence of designing and developing models using such a bottom-up approach however, is that simulating many individual agents is computationally expensive (Richmond et al., 2009), especially when the overall scale of the model is increased to generate results consistent with system-level dynamics from the real world domain, and furthermore when hundreds (if not thousands) of replicates are required in order to provide a stable average, due to the underlying stochasticity within individual simulation runs.

The Flexible Large-scale Agent-based Modelling Environment (FLAME[9]) was developed by Simon Coakley during his PhD (Coakley, 2007) as a framework for the simulation of large-scale agent-based models over high-performance computing clusters. Kiran et al. (2008) and Coakley et al. (2012) advise that FLAME was designed and developed from the outset to be able to deal with massive simulations, incorporating a large scope with respect to the underlying real world domain, and very large numbers of individual agents - in the order of hundreds of thousands to millions. Furthermore, it appears that Coakley (2007) was mindful that researchers were often hindered by complexities of ensuring simulations can run across multiple operating systems or porting models on parallel platforms. With this in mind, FLAME was developed to work across the various scales of computer hardware, allowing initial development and testing on individual laptops or desktop computers (using Linux, UNIX, MAC OS or Windows), and full-scale testing and simulation on large supercomputers utilising parallel hardware platforms.

FLAME itself is not a true modelling platform in the purest sense, instead it requires the modeller to use templates to define the agent-based model, and then parse and compile these using the packaged parser and associated APIs (Coakley et al., 2006a; Holcombe et al., 2006). Due to FLAME utilising the concept of communicating X-Machines, the agents are modelled using XML templates to define the attributes and internal states of the agent, and C code is used to define the rule-based functions of the agent behaviour (see figure 2.4). Through the use of messages to communicate changes within the system's environment, and transition functions, which define the rule-based logic of the computational model, the agents are able to transition to a new state and update their internal memory as the simulation progresses. Once a model is specified using the XML and C templates, the FLAME modelling framework is able to automatically

[9]http://www.flame.ac.uk/joomla/index.php

generate simulation code (via the *xparser* programme) that allows communication between X-Machines (agents) through its own communication library, called *Message Board*. Through interfacing to the Message Passing Interface (MPI) communication framework, the simulation code is also fully compliant with parallel hardware platforms, enabling efficient communication between the individual agents, and ensuring that concurrently executing agents remain in sync with each other (Foster, 1995).

To date, FLAME has been used across a number of disciplines, ranging from Biology, to Economics, to transport and logistics of Utilities. A number of examples where FLAME has been used for modelling the immune system are: the simulation of wounded epithelial cells (Walker et al., 2004), intercellular signalling via epithelial growth factor (Walker et al., 2006), and a hypothetical study of Transforming Growth Factor-β1 in epidermal wound healing (Sun et al., 2009).

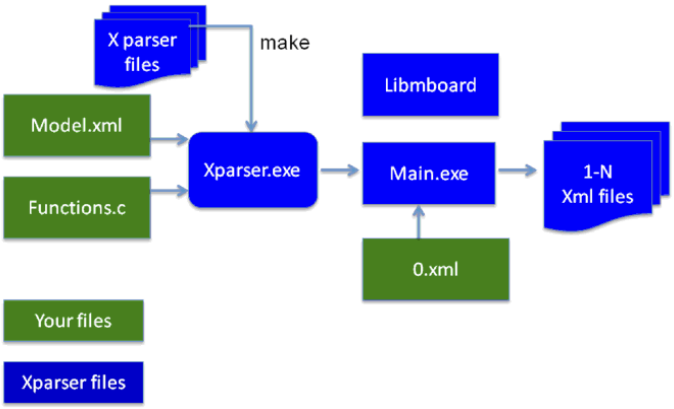

Figure 2.4: The underlying mechanics of FLAME rely on the parsing and subsequent compilation into C files of the computational model. Simulations are run through pointing an initial starting parameters file (0.xml) towards the main C executable file. As FLAME is a discrete-event simulator, it generates an individual output file for each time-step within the simulation, which act as input parameter file for the subsequent time-step. The blocks in blue are automatically generated through parsing and compilation, whereas the blocks in green are modeller files (from Coakley and Kiran (2009)).

2.7 Project Lifecycles

As defined by The Project Management Institute (2004), a project is a *"temporary endeavour undertaken to create a unique product, service or result"*. This means that every project has a defined beginning and end, and creates unique deliverables, potentially through progressive elaboration, by developing in steps and continung by increments. Projects are carried out in an environment broader than that of the project itself, and therefore require rigorous management to ensure that they complete: within the allocated timescales; to budget; having developed the agreed deliverables (scope); to an acceptable quality. Such management is termed *Project Management*, and one of the most important aspects of project

management is the adherence to a *Project Lifecycle*, which defines the phases that facilitate progression of the project from its beginning to its end.

Until the turn of the millenia, implementation of software projects traditionally used the *waterfall* model as the project lifecycle. The waterfall model is based on a manufacturing approach for development, is sequential in nature, and requires the completion of each phase before progression to the next, using a single pass through all phases (Royce, 1970). Contemporary approaches for software development however, especially within the rapidly growing areas of academic research within computational biology, use a more iterative project lifecycle. Indeed, Kitano's hypothesis-driven research cycle within systems biology (see figure 2.2) forms the basis of such a project lifecycle and can be broken down into four main phases. The first phase investigates the underlying biology and poses research questions of interest. The second phase develops computational models, which are then used for *in silico* experimentation, analysis and hypothesis generation for additional wet-lab experimentation. The third phase is where these additional wet-lab experiments are performed, and finally the fourth phase is where the results of these wet-lab experiments are then analysed in relation to the original research questions, with potential augmentation of the computational model, for continuation of the cycle.

There are numerous software development project lifecycles in the commercial world, indeed, most large software organisations will have their own *in house* standards, but these will usually follow a generalised pattern. A good example is Oracle Corporation's Unified Method (OUM) which is used by their Consulting[10] division. Whereas Kitano's hypothesis-driven lifecycle has four phases, the OUM (Oracle Corporation, 2012) has five, relating to: Inception, Elaboration, Construction, Transition and Production. The inception phase is where the project is planned, the functional and technical requirements are captured, and a prototype is built, which is demonstrated to the end users through a Conference Room Pilot (CRP). The elaboration phase is where the requirements are refined following CRP1, the prototype is enhanced to transform into a full draft software release (along with necessary peripheral code objects), before demonstrating through another CRP iteration. The construction phase is where the updates and refinements from CRP2 are performed, and demonstrated within the final CRP. Following CRP3, the individual software components are deemed to have satisfied the functional requirements, and are then linked together for systems integration testing, and subsequent user acceptance testing. The software is then deemed to have passed all functional requirements tests, and the transition phase is where the software is validated against full-volume data loads, and if being developed to replace an existing system, is where a parallel run is performed for verification purposes. Finally, the new system is switched on for end users (and any corresponding old system turned off) within the production phase.

Recently, the Complex Systems Modelling and Simulation (CoSMoS)[11] process has been developed through a collaboration between the Universities of York, Kent, Abertay and the University of the West of England (Andrews et al., 2010). This process provides a framework of leading practice for developing and using simulations to explore complex systems, and is comparable to project lifecycle

[10] www.oracle.com/uk/products/consulting/index.html
[11] www.cosmos-research.org

methodologies used in industry. Like OUM, the CoSMoS process is organised around phases, which contain a set of products (deliverables), and associated activities. The CoSMoS process has three phases: the Discovery phase, which establishes the scientific basis of the project, identifies and models the domain of interest, and formulates scientific questions; the Development phase, which produces the simulator; and the Exploration phase, which uses the simulator for *in silico* experimentation, the results of which are used to explore the scientific questions defined previously. Along with these phases, there are key products associated with CoSMoS projects: Domain Model, Platform Model, Simulation Platform, and Results Model (figure 2.5).

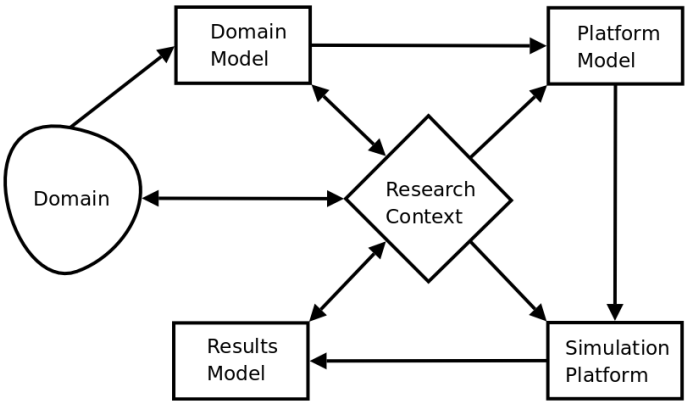

Figure 2.5: CoSMoS products and their relationships to each other and the biological domain of interest. For the first iteration of phases within a CoSMoS project, there is a distinct sequence for creation of the products, which are always developed with the research context and the underlying domain of interest, in mind (figure 2.1 of Andrews et al. (2010)).

The domain in this context represents the real-world system, or part of the system due to an appropriate abstraction level. Similarly, the research context in the CoSMoS process relates to the overall context and scope of the computational model. The research context is of paramount importance, and can be thought of as the underlying thread of knowledge, scope and assumptions, that runs throughout the entire CoSMoS project. The scope, abstraction level, assumptions and constraints, provide the context behind how simulation results should be validated, interpreted and evaluated (Andrews et al., 2011).

The domain model is an abstract representation of the actual biological area of interest (the Domain), which documents our understanding of the domain into explicit statements, which may relate to assumptions, constraints and definitions of the underlying biological mechanisms and data, or indeed the structure of biologicals components and their relationships/interactions within real-world biology. Andrews et al. (2011) advise that as the domain model focuses on the real-world system, it should be free of any simulation language or hardware platform bias, as the decisions on appropriate tools and technologies should be left until the full scope of the functional requirements for the model are captured. One approach to documenting the domain model, which has been widely taken up by the computer science community, uses UML. Webb and White (2005) describe a process of diagrammatically modelling the content and functions associated with

biological systems (in their case a cell) before development of a corresponding computer simulation. The domain model should be developed in close collaboration with domain experts, to ensure a solid foundation of the science, which the subsequent platform model can build upon. The process starts with the identification of biological entities and their relationships with each other, progresses through the gradual addition of details, and ends with an executable programme that simulates biochemical pathways. Grizzi and Chiriva-Internati (2005) advise that complex systems can be viewed from many perspectives, and therefore can be described in many ways, each of which will be only partially true. Therefore, a progression of UML models are needed during the development of the domain model, as a starting point for more comprehensive exploration during the discovery and exploration phases of a CoSMoS project.

The platform model builds on the knowledge gained through development of the domain model, but focuses on the underlying systems engineering, and may be seen as the technical design of the computational model. As such, a standardised diagrammatic notation like UML and/or SBGN may again be used, however this time they are focused on the interactions, assumptions, and constraints relevant to the computational model. This is because the functionality (and mechanisms behind the functionality) within computational models do not always match biological reality due to technological constraints (e.g. programming languages or hardware architecture), thus requiring a computational workaround. Along with identification of component interactions (the network) and any associated functions and state transitions, it is also important to identify a set of parameters associated with the system as a whole, because all computational results have to be matched and tested against actual wet-lab experimental results (Kitano, 2001a). In the majority of cases, this parameter set may need to be estimated based on experimental data, due to the differences in scale between the real-world biological domain and its abstraction which has been represented through the domain model (Hamahashi and Kitano, 1999). Furthermore, in finding a parameter set, it must be recognised that this may be one of multiple parameter sets that generate simulation results equally fitted to experimental data, and that this is due to the abstractions taken in design and development of the model, i.e. there will be fewer parameters within the model than the real system in question. Additionally, the platform model also includes instrumentation and interfaces which may be required to visualise, record and analyse the simulation results.

The simulation platform encodes the platform model in software and hardware platforms, and can be thought of as the fully developed and calibrated computational model (simulator), whose execution culminates in a simulation. As discussed previously, simulations can be used for *in silico* experimentation, the results of which are captured within the results model. This results model has a relationship to the simulation platform that is analogous to the relationship betwen the real-world domain and the domain model. Following simulation and development of the results model, it is therefore compared to the domain model to establish whether the simulation platform provides an appropriate representation of the real-world domain being investigated. All four products of a CoSMoS project are dependent on the research context, and following the first iteration through the lifecycle, the research context may be adjusted according to the results model, or alternatively the domain model and/or platform model

and/or simulation platform may be updated as part of another iteration through the lifecycle to ensure the results model provides an appropriate representation of the underlying real-world domain. Once this is achieved, the computational model may be used for hypothesis-generation and design of new experiments on the real-world domain. See figure 2.6 for a diagrammatic representation of the CoSMoS phases, products and high-level activities.

Recent computational work, which used the CoSMoS process as the project lifecycle for developing biological models include: Experimental Autoimmune Encephalomyelitis (Read et al., 2013; Williams et al., 2013; Greaves et al., 2013), auxin transport in plants (Garnett et al., 2010), peyers patch formation in the immune system (Alden et al., 2012), and tumour formation (Bown et al., 2012).

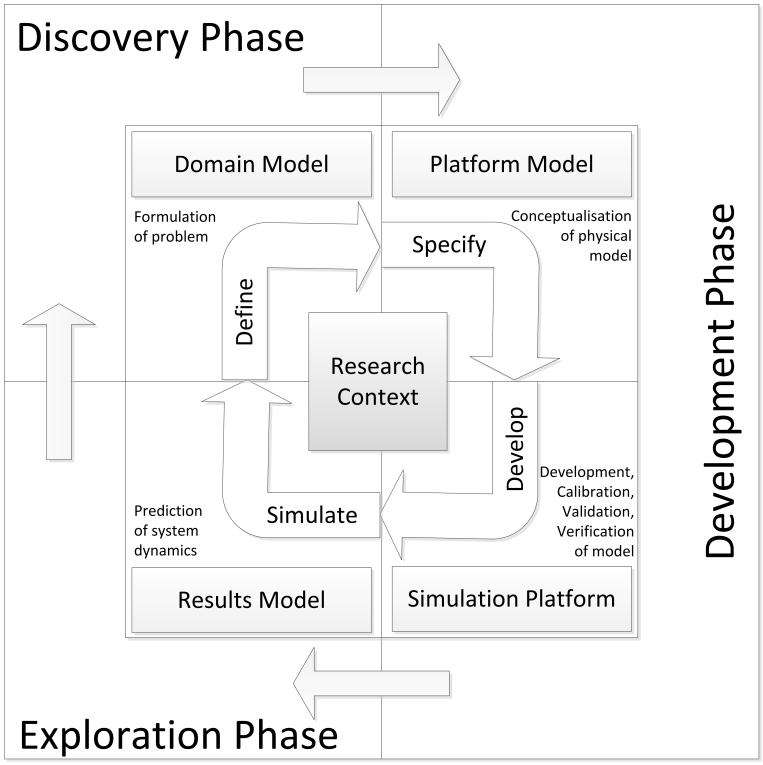

Figure 2.6: The CoSMoS process advocates an iterative lifecycle, consisting of three separate phases (discovery, development and exploration), and creation of four key project artefacts (domain model, platform model, simulation platform, and results model). The discovery phase focuses on formulation of the problems to be investigated through modelling, resulting in creation of a functional specification of the required biological behaviour (domain model). The development phase focuses on transforming the domain model into a technical specification (platform model) specific to the programming language(s) and computer architectures to be used, and actual development of the computational model (simulation platform), including calibration, validation and verification. The exploration phase focuses on the *in silico* experimentation to investigate the biological problems of interest, and generation of predictions (which are documented in the results model), which facilitate generation of novel hypotheses for subsequent testing in the biological arena. After Andrews et al. (2010)

2.8 Summary

Systems theory in relation to biology has progressed a long way from its generalised approach during the 1950s to its current targeted approach towards the various biological subdisciplines, for example systems immunology and systems phyisiology. Wolkenhauer et al. (2005) believe that we need to systematically perturb a biological system in order to ascertain its behaviour under varying conditions. Kitano (2007) however, takes this approach further by suggesting that the use of control theory and communication theory from engineering can be used to provide a theoretical foundation for a systems-oriented approach, with specific emphasis on understanding how we can control the robustness of living systems, in particular at the cellular level.

Over the past decade and a half, the use of computational approaches within systems oriented biological research has progressed from mere data analysis (bioinformatics) to a new wave of complementary research (with respect to wet-lab research) based around modelling and simulation. The techniques used within computational biology have evolved from mathematical-based deterministic approaches, to mathematical-based stochastic approaches, and have culminated in computational-based approaches such as agent-based modelling (ABM). An advantage of these computational approaches is that we have full control over the mechanisms of the model, thus once a model is established, equation solving or simulation allows quick studies of changes to the elements and associated parameter values. The results from *in silico* experimentation are therefore directly related to the abstraction level and design of the model, and not due to other factors outside of our control. This provides a more solid basis from which to test our hypotheses on the mechanisms behind biological functionality than wet-lab experimentation, where the large degree of variation, even within single cells of a clonal population, introduces a degree of additional complexity that reduces the certainty of our interpretations of experimental results (Elowitz et al., 2002).

Both mathematical and agent-based approaches can be used to model complex systems. However, whereas a mathematical model is solved, an ABM is simulated within a computer. Mathematical models require extensive use of system-level parameters for evaluating equations that produce evolutionary characteristics over time. As such, mathematical approaches are better suited to domains where we are interested in behaviour at a population level rather than an individual component level. ABMs by comparison provide an ability to use differing levels of parameters, and are much more efficient in modelling stochastic behaviour. This facilitates emergent system-level behaviour, which is due to the aggregated dynamics of individual agents, and therefore not defined in the model design via system-level parameters. Due to the object-oriented manner in which biology has been observed (through the reductionist approach), experimentalists are likely to feel more comfortable with modelling approaches that are visual and intuitively realistic rather than mathematically abstract. Furthermore, an ABM allows the modeller to take more precise (with respect to PDEs) 3-dimensional aspects into consideration for each agent, which therefore allows the model to more accurately reflect the biological system. Additionally, ABMs allow the collection of time-series data (of individual components) to facilitate examination of the dynamical nature of complex systems.

Simulation often requires integration of models that have been developed at multiple hierarchical levels, and therefore have orders of magnitude difference with respect to scale, for example, gene regulation and biochemical networks within cells, intercellular communication between cells, and the processes inherent within whole organs. Similarly, from a time perspective, some biochemical processes take place within a millisecond, with others taking hours or even days. Although some of these processes can be modelled by either stochastic computation or differential equations alone (e.g. biochemical networks); many cell phenomena require calculation of physical processes, such as structural dynamics, and therefore require a combination of both methods. Indeed, Duan et al. (2000) advise that the majority of attempts to model biological systems at various levels in the hierarchy have demonstrated that a combination of discrete and stochastic events need to be modelled, and classify these as *hybrid* systems.

The design and development of realistic computational models of biological networks is not a simple task, as the models will require multiple feedback signals, non-linear dynamics, the estimation of numerous parameters that have a degree of uncertainty (or even unknown) from wet-lab experimentation, and the introduction of stochastic noise to provide the range of dynamics seen across a population of molecules and cells within the system (Csete and Doyle, 2002). ABM provides a way to integrate multiple strands of data and knowledge (following the deep-curation approach to design), along with the ability to turn conceptual models into computational models that through simulation can test various hypotheses on underlying biological mechanisms. Furthermore, the inherent capacity to model non-deterministic and heterogeneous behaviours provides the opportunity to investigate the role of natural variations in biological populations and generate hypotheses for new wet-lab experiments (Walker and Southgate, 2009).

Advances in hardware architecture, capacity, and performance over the past two decades has facilitated the development of new software technologies that are able to cope with increasingly complex models that are simulating at scales much more closely aligned to actual biology. For example, the development of the FLAME simulation environment has the potential to run simulations for models that comprise millions of agents, which is orders of magnitude higher than previous approaches to simulating complex biological systems.

Following an increased adoption of computational approaches for investigating biological questions, a number of groups within the complex systems research community have advocated the use of a principled approach to design and development of the requisite computational tools. In particular, Kitano advocates a hypothesis-driven project lifecycle, which focuses on the biological questions of the research, and uses an iterative approach to developing the computational tools for advancement of biological knowledge. More recently, the CoSMoS project has built on the high-level cycle advocated by Kitano, by focusing on the development process of these computational tools. Like Kitano, CoSMoS advocates an iterative approach to development, but explicitly advocates close liaison between domain expert and modeller, to achieve a truly interdisciplinary approach, and not just that of a biologist dabbling with technology, or a computer scientist using their rudimentary knowledge of biology to design their models. The CoSMoS process also goes further by detailing a number of key activities and project deliverables associated with the various phases of its project lifecycle. Indeed, we believe this

new approach to software engineering within the academic research domain has begun to bridge the gap to comparable project lifecycles used in industry (such as the Oracle Unified Method).

This chapter has provided an overview of the theoretical underpinnings behind the technologies that will be used within this doctoral thesis. It began with an introduction to systems theory and network theory, before discussing systems biology, computational biology, and computational immunology. The chapter also provided the background to the various diagrammatic and computational approaches that are currently in use to develop computational models within biology, and in particular discussed the agent-based modelling paradigm and the FLAME simulation framework, which is based on communicating X-Machine architecture. Additionally, the chapter has introduced the concept of project lifecycles for managing the different phases within a computational biology project, and in particular has introduced the CoSMoS process, which provides a principled approach to design and development of computational models of complex systems. We believe this background discussion to be pivotal for setting the scene for this doctoral thesis, as we intend to follow the CoSMoS process for design and development of our agent-based model, using UML for the design and the FLAME simulation framework for development. The next chapter will provide the necessary background to the IL-1 stimulated NF-κB signalling pathway (our biological domain), and through discussion of the existing models will justify the need for our work.

3 The Domain: NF-κB Signalling Pathway

The immune system is a complex system of molecules and cells that are distributed throughout our bodies, whose main goal is to distinguish self and non-self (Voet and Voet, 1995), in order to provide us with a basic defence against pathogenic organisms and cells that have become malignantly transformed (Perelson and Weisbuch, 1997). There are two types of responses in the human immune system: the innate immune response, which is rapid, occurs on exposure to infectious organisms, but does not provide lasting immunity and is not specific for any individual pathogen; and the adaptive immune response, which takes days rather than hours to develop, but is capable of eliminating infections more efficiently, through for example the production of antibodies against a particular pathogen, and develops during the lifetime of an individual (Murphy et al., 2008). Many of the responses following identification of a foreign organism or toxic molecule created by them, are destructive in nature, and facilitate the removal of the infectious microorganism and/or parasite (Alberts et al., 1994). Following recognition of pathogen, the plethora of immune responses develop over time, and through stochastic interactions, signal amplification and feedback regulation, give rise to the gross properties of the immune system (Germain, 2001).

The human immune system is controlled by the action of about 10 main types of cells that participate in various aspects of the inflammatory response. Their activity is controlled by a wide range of regulatory and effector molecules, with many of these regulatory molecules being distributed across the various cell types, even though their function or regulator capacity may be slightly different. The elucidation of molecules important for the immune response is by no means complete, and new molecules continue to be discovered. To date, they essentially break down into three main groups: soluble molecules that transmit signals between cells; receptors; and intracellular signalling components that propagate the stimulation of the cell surface receptors through the cell to activate a cell response. Examples of such responses are expression of proteins to combat the infection, or indeed initiation of apoptosis (programmed cell death) for the cell in question (Perelson and Weisbuch, 1997). With responses including the immune system initiated death of cells, it is extremely important that responses are under tight control. With this in mind, Cohn and Langman (1990) suggest that the immune response must not be *too cold* (i.e. too insensitive, too slow, or too low a level of intensity) so that its response is overwhelmed by rapidly replicating pathogen. Conversely, Germain (2001) counters that the immune response must similarly not be *too hot*, because the responses used to combat pathogens are themselves capable of causing substantial destruction to the host. The response must therefore be *just right*, providing rapid, vigorous and properly modulated defences against pathogenic invasion and malignantly transformed host tissue.

3.1 Cell Signalling

Following the systems biology and network theory line of reasoning; for cells to combine into networks that achieve higher levels of organisation (e.g. tissue and organs), it is necessary for them to communicate. Communication in all organisms bar the most basic, is mediated through complex networks that integrate extracellular signals and intracellular processes for the generation of appropriate cellular responses (Pfeifer et al., 2008). The basis for this intercellular communication are the receptors in the cell membrane and the extracellular signals released by cells. Following the detection of extracellular signals, the transmission of information to the genes and gene-regulatory machinary occurs through a process known as signal transduction[1]. Here, the stimulus associated with an external signal received by a receptor at the cell membrane, is converted from one physical form into another using intracellular biochemical reactions, and thus promotes the relay of information, without physical flow of signal through the cell membrane (Lodish et al., 2000; Krauss, 2003). In most cases the receptors are transmembrane proteins on the target cell surface, which become activated following the binding of specific signalling molecules (ligands). Once activated, they propagate the signal through a cascade of intracellular reactions that culminate in changes to the rate of gene expression or enzyme activity (Downward, 2001). The consequence of signalling through ligand binding is in most cases a modification of the activity of intracellular enzymes or activation factors, e.g. transcription factors, which determine the reading and transcription of information encoded in the genome. There are a number of well-defined signalling pathways in humans that lead to activation of gene expression. Of these, the NF-κB pathway is believed to be unique in the speed of its activation (Karin and Ben-Neriah, 2000), and furthermore, it is central to regulation of inflammatory responses, as scientists have not yet discovered an inflammatory gene that is not controlled, at least in part, by NF-κB.

3.1.1 Example Signals within Cell Signalling

Every cell in the human body must have the ability to sense their surroundings for changes in conditions (for example through the detection of extracellular molecules), and must be able to generate a range of responses to these changes (Alberts et al., 1994). These detection and response systems need to be very tightly controlled, and in humans this is performed through a complex network of signalling pathways. There are a number of ways that cells may signal to each other, which mainly depend on the concentration interactions involved. If cells are adjacent, they may signal through membrane pores, or by membrane-bound ligands of one cell being identified by a membrane receptor on a neighbouring cell. If the cells are further apart, they may communicate through the release of

[1] All living cells sense and respond to their environment via a specific type of network termed a signal transduction network. These networks consist of cell receptors at the cell membrane for sensing the extracellular signals, various intermediate components within the cell for propagation of the signal, and culminate with components that upregulate the generation of specific proteins for the relevant cellular response. Tiger et al. (2012) advise that there are four main types of events within these networks: a) catalytic modification; b) bindings and interactions; c) degradation and synthesis; and d) changes in localisation.

signalling molecules, such as hormones, cytokines and growth factors (Hancock, 1997).

Hormones represent a broad category of signalling molecules that regulate metabolism, growth, and differentiation, and travel over relatively large distances within the body. They are produced in specialized cells of the endocrine glands, which release hormone molecules into the extracellular medium, and may be subsequently carried by the blood circulation to a new site of action, such as a different tissue (Creighton, 1999). Cytokines represent a group of protein molecules that produce a more localised response, having their effect on other cells which are close by, or even acting on the same cell that produced them (Krauss, 2003). Specific to the immune system are the sub-groups of cytokines, for example: interleukins (IL), tumour necrosis factors (TNF), and interferons (IFN). Examples of growth factors include platelet derived growth factor, epidermal growth factor, and fibroblast growth factor (Hancock, 1997). The term growth factor was originally used because the individual molecules in question were shown to be involved in the growth and differentiation of cells. More recently however, it has been shown that these molecules may also have other effects, including cell migration, and functions within the immune response (Creighton, 1999).

3.1.2 Cell Membrane Receptors

Specialised proteins, termed receptors, are utilised for the detection of signals. There are two principal ways for a signal to be detected: either the signal is detected externally through binding and activation of a cell membrane receptor, or it crosses the cell membrane and activates a receptor within the cell compartment or nucleus (two different internal compartments) of the cell. As mentioned above, most surface receptors are transmembrane in nature, and therefore span the membrane; thus having both external and internal facing domains (Downward, 2001). For the purposes of this research project, which is related to the NF-κB intracellular signalling pathway, we will restrict discussion to receptors of the cell membrane. Garrett and Grisham (1995) advise that there are three main types of cell membrane receptors: ion-channel-linked, which are involved in rapid signalling between synaptic cells (of the nervous system), and once excited through binding of the associated ligand (a neurotransmitter), permit the movement of ions across the membrane; G-protein-linked, which along with the extracellular binding site for the relevant ligand, also have an intracellular site that associates with a GTP (guanosine 5'-triphosphate) intermediary for propagation of the signal downstream; or enzyme-linked, which when activated, either function directly as enzymes, or are associated with enzymes.

3.1.3 Intermediate Components and Adaptors

Once a cell membrane receptor receives a signal, through binding to its external domain, it becomes activated, and in the case of G-protein-linked or enzyme-linked, the intracellular domain induces a specific reaction on the cytosolic side, which then transmits the signal onto other, intermediate components. G-protein-linked and enzyme-linked receptors both operate through phosphorylation and dephosphorylation cascades (Krebs, 1992). The former facilitate the binding of

GTP to an intermediate protein within the intracellular pathway, whilst the latter phosphorylates an intermediate protein through the action of a kinase enzyme (Downward, 2001). The phosphorylation of the first intermediate protein, results in a cascade of intracellular reactions to propagate the signal further downstream, and in a large number of cases towards the nucleus, where they alter the expression of genes and thereby alter the behaviour of the cell (Alberts et al., 1994).

Some proteins within a receptor complex, mediate the signal transduction between proteins of a signalling pathway by bringing them together, and are referred to as *adaptor* proteins[2]. They are generally signalling components that bind the receptor cytoplasmic domain and initiate the first steps of co-locating signalling molecules into multiprotein signalling complexes. Formation of the receptor complex and related events at the level of the receptor have a primary role in signal transduction (Downward, 2001).

3.1.4 Crosstalk

Signalling pathways are often depicted diagrammatically as linear sequences of events. This is a simplistic view, as signalling pathways often branch, and indeed multiple pathways may be integrated into a signalling network, much like the metabolic pathways network of Nicholson (2001). Branching and network formation may be mediated in two ways: a given signalling component can be activated and can receive a signal from different upstream components, thus integrating different signals into a pathway; or a given signalling component may generate different signals, thus branching into more than one downstream pathway.

Cells receive signals relating to multiple signalling pathways in parallel to each other. The integration of multiple upstream pathways and analogous divergence into multiple downstream pathways facilitate the cells integration of these multiple stimuli. Signalling in one pathway may therefore influence the activity of components of other signalling pathways. This interdependence of signalling is termed *crosstalk* (Downward, 2001), and enables the cell to coordinate multiple inputs at the level of signal transduction. See figure 3.1 for profiles of crosstalk from two different cell membrane receptors.

3.1.5 Cell Signalling in Innate Immunity

In general, infectious agents are met by cells that mount the innate immune response, such as macrophages[3] (for bacterial invaders). These immune cells try to contain the infection through a process known as *phagocytosis*, where the pathogen is engulfed and internally degraded by the immune cell (Murphy et al., 2008). Furthermore, immune response molecules, such as cytokines and chemokines[4] are secreted which initiate the process known as *inflammation*. In-

[2]Section 3.3 provides a detailed discussion of adaptor proteins involved within the NF-κB signalling pathway.

[3]Macrophages are a type of white blood cell which phagocytose (engulf and then digest) pathogens and cellular debris.

[4]Murphy et al. (2008) define cytokines as *"a general name for any protein that is secreted by cells and affects the behaviour of nearby cells bearing appropriate receptors"*. Similarly, they define chemokines as *"secreted proteins that attract cells bearing chemokine receptors, such as*

flammation is part of a complex physiological response to pathogenic invasion, and is beneficial to the combat of infection as it promotes eradication of the pathogen through recruitment of cells and molecules of the innate immune system out of the blood and into the infected tissue, where they act to destroy the pathogen directly (Ryan, 1977).

A similar mechanism to the degradation that occurs with phagocytosis exists to breakdown components within a cell. This mechanism is called autophagy (or autophagocytosis), which is a homeostatic process whereby cells dispose of intracellular organelles and protein aggregates (Levine et al., 2011). The autophagy pathway is linked to most cellular stress-response pathways, and crucially has a role in the control of inflammatory signalling (Levine and Kroemer, 2007). One way in which autophagy is induced is through NF-κB dependent cytokine production (Levine et al., 2011). In a simplified NF-κB model, spatial separation across the three major compartments is involved in propagation of the signal by extracellular signals binding to membrane receptors, a cascade of reactions occuring in the cytoplasm, and translocation of NF-κB to the nucleus for upregulation of gene transcription (Weng et al., 1999).

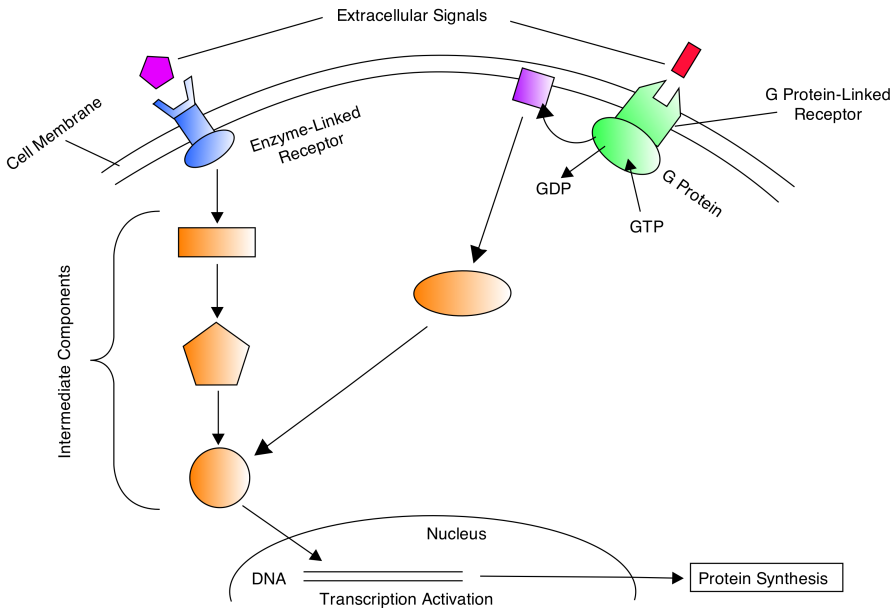

Figure 3.1: Generalised cell signalling pathway showing two distinct signals which activate different receptors. Once receptor is enzyme-linked and upon activation initiates a cascade of reactions, and the other is G-protein-linked, using a phosphate group from GTP to activate a co-receptor which then initiates a cascade. These two separate signalling pathways converge on a common intermediate component, which translocates across the nuclear membrane to activate transcription of specific genes, culminating through a number of other biological processes in the synthesis of proteins.

neutrophils and monocytes out of the bloodstream and into infected tissue".

3.2 Overview of Transcription Factor NF-κB

Nuclear factor-kappa B (NF-κB)[5] is a collective name for a family of inducible dimeric transcription factors, and as such is an essential intracellular messenger, which in conjunction with its signalling pathway connects various extracellular stimuli to the induction of gene expression. NF-κB plays a critical role in inflammation, immunity, cell proliferation, cell differentiation, and cell survival (Oeckinghaus and Ghosh, 2009). As a transcription factor, it activates the creation of an equivalent copy of RNA from a DNA sequence (transcription), which is subsequently translated into the corresponding amino acid sequence as part of protein synthesis. NF-κB was discovered just under 30 years ago by the laboratory of the Nobel Laureate David Baltimore. Sen and Baltimore (1986a) discovered multiple nuclear factor proteins that interacted with enhancers (regulators of DNA transcription) of the immunoglobulin heavy chain and the κ light chain in B-cells (a type of white blood cell, derived from the bone marrow, that produces antibodies); they called these NF-α, NF-μ, and NF-κ. NF-κB was initially considered to be B-cell specific, but was later shown to be ubiquitous - in all cells. It has been found to regulate a diverse range of biological processes, for example cell growth and survival, to immune response and inflammation (Sun and Liu, 2011). Two recent special issues of Cell Research (2011, vol 21) and Nature Immunology (2011, vol 12, no 8), were dedicated to NF-κB signalling and function.

3.2.1 Functions of the NF-κB Signalling Pathway

The choice between life and death of individual cells during infection of a host, is one of the key events in the immune response (Karin and Lin, 2002). The transcription factor NF-κB is a major player in the regulation of such life and death decisions, and is involved in the transcriptional regulation of a large number of genes, particularly those involved in response to infections (through the innate and adaptive immune systems), and other stressful situations (Tian and Brasier, 2003; Pahl, 1999). For example, it has been shown that NF-κB is activated by cytokines (Stylianou et al., 1992), reactive oxygen species (Bubici et al., 2006; Morgan and Liu, 2011), viral infection (Hiscott et al., 2006), bacterial cell wall products (Laflamme and Rivest, 2001), DNA damage (Bender et al., 1998), shear stress (Ganguli et al., 2005), shape changes (Nemeth et al., 2004), etc. In keeping with the large number of extracellular signals that activate the NF-κB signalling pathway, the transcription factor also upregulates a large list of target genes (Pahl, 1999). This association between NF-κB and expression of genes used within the immune response is thought to date as far back as insects in evolutionary terms (Siebenlist et al., 1994). Although NF-κB is ubiquitously expressed, its action is regulated in a well controlled, cell-type and stimulus-specific manner, which provides for a multitude of roles and effects (Brasier, 2006). In fact, one of the first genes that NF-κB activates is actually one of its own inhibitors, IκB (inhibitor of kappa B; Baeuerle and Baltimore 1988a,b; Sun et al. 1993). NF-κB activation is therefore an inducible, but transient event in normal cells.

[5]The nuclear factor-kappa B gains its name following its discovery by Sen and Baltimore (1986a) as an enhancer (regulator of DNA transcription) of the κ light chain (a sequence of amino acids) in B cells (a type of white blood cell, that makes antibodies).

The review by Pahl (1999) highlights that among other functions, NF-κB is important for the production and/or regulation of enzymes that generate prostaglandins (locally acting messenger molecules) and reactive oxygen species. Of particular note to the immune system and the inflammatory response, is that NF-κB regulates genes encoding many proinflammatory cytokines and chemokines, for example interleukin 1b, interleukin 2, tumour necrosis factor, and transforming growth factor (reviewed by Siebenlist et al. 1994), and interferon as part of the antiviral response (Boehm et al., 1997). Furthermore, an important early target of these cytokines and chemokines, is the vascular endothelium, whose constituent endothelial cells must recruit white blood cells in the circulatory system and then change their structural integrity to allow these blood cells to cross the vascular membrane into the site of infection. This is mediated through the expression of cell adhesion molecules, such as VCAM-I and ICAM-I (reviewed by Baldwin 1996).

As reviewed by Baichwal and Baeuerle (1997) and Sonenshein (1997), NF-κB activation protects most cells from apoptosis through induction of survival genes, such as the antiapoptotic factor Bcl-2, although under certain conditions and in certain cell types it may also induce apoptosis (Dutta et al., 2006). An example of this is NF-κB's involvement in development of mature B-cell and T-cell lymphocytes within the immune system. B- and T-cells that bind self-antigens with high affinity are likely to be self-reactive, and therefore need to be removed from the lymphocyte population through a process known as negative selection. NF-κB is believed to facilitate negative selection through the induction of proapoptotic genes. Surprisingly, NF-κB has also been linked to positive selection of T-cells as well, where it performs an antiapoptotic role (reviewed in Hayden et al. 2006).

Along with involvement in the developmental cycle of lymphocytes, NF-κB is also involved with activation of genes for immunoreceptors (e.g. T-cell Receptor and Major Histocompatibility Complexes I and II) at the cell surface of these lymphocytes (see review by Siebenlist et al. 1994), and in the upregulation of CD80/CD86 on antigen presenting cells (APCs; e.g. dendritic cells) which are co-stimulatory molecules, providing the second signal necessary for T-cell activation (Pahl, 1999). Furthermore, NF-κB also plays an important role in the development and function of primary and secondary lymphoid tissues. Primary (central) lymphoid organs include the bone marrow and thymus, whereas secondary (peripheral) lymphoid organs include the lymph nodes, Peyer's patches, and spleen (Hayden et al., 2006).

Due to the wide ranging gene products that are regulated through NF-κB and its associated signalling pathway, NF-κB has been deemed a *'Master'* regulator of inflammation (Brasier, 2006). See figure 3.2 for a high-level depiction of how the multitude of inducers for NF-κB activation, and resulting gene products, has led to the idea that NF-κB acts as the *hub* in a complex network of interactions that gives rise to a bow-tie motif.

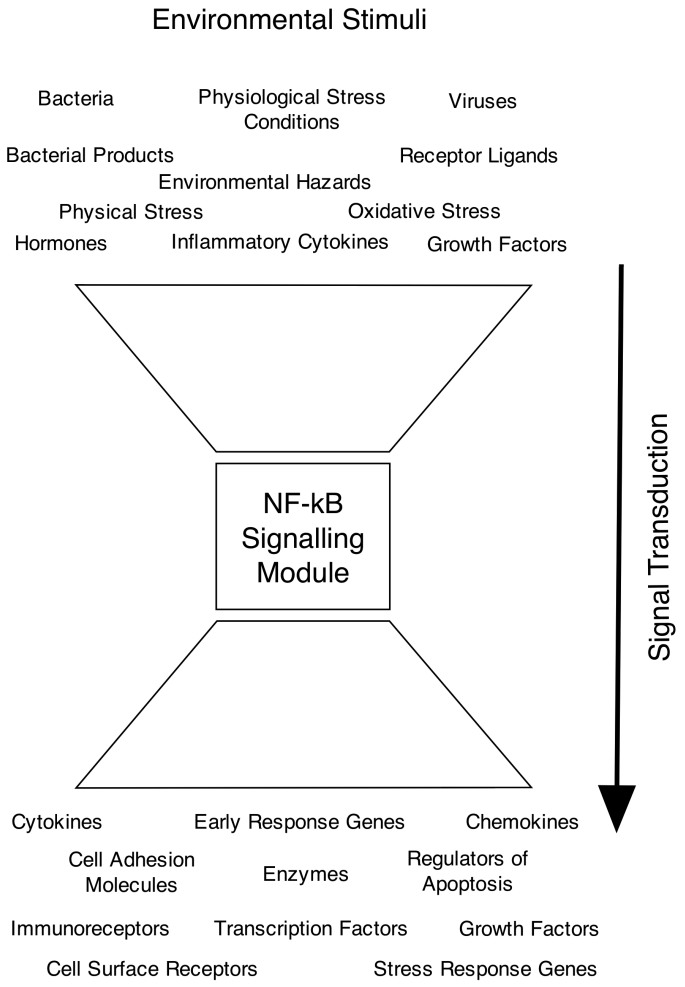

Figure 3.2: High-level representation of how the multitude of inducers of the NF-κB signalling pathway, which converge on the NF-κB signalling module (NF-κB, IκBα, and IKK), before diverging to regulate a multitude of gene products, has the topology of a Bow-Tie motif from systems biology. Pahl (1999) advises that there are over 150 known inducers of NF-κB activation, which can be categorised into: 1) bacteria, e.g. *Helicobacter, Salmonella, Staphyllococcus, and Listeria*; 2) bacterial products, e.g. lipopolysaccharides, exotoxins, and phospholipases; 3) viruses, e.g. Epstein-Barr virus, HIV, hepatitis, influenza, and measles; 4) inflammatory cytokines, e.g. various interleukins and tumour necrosis factors; 5) physiological stress conditions, e.g. hemorrhage, hyperglycemia, ischemia, liver regeneration, and shear stress; 6) physical stress, e.g. ultraviolet radiation and gamma radiation; 7) oxidative stress, e.g. hydrogen peroxide or ozone; 8) environmental hazards, e.g. lead and cobalt; 9) receptor ligands, e.g. CD28 ligand and CD40 ligand; 10) mitogens, growth factors and hormones. Similarly, Pahl (1999) also advises that there are more than 150 known genes that are regulated by NF-κB, which can be categorised into: 1) cytokines/chemokines, e.g. IFNγ, IL-1α, IL-1β, IL-2, TNFα, TNFβ; 2) immunoreceptors, e.g. CD48, CD80, IL-2 receptor α-chain, IgG heavy chain, MHC class 1, MHC class 2, TNF receptor; 3) cell adhesion molecules; 4) stress response genes; 5) cell surface receptors; 6) regulators of apoptosis; 7) growth factors; 8) early response genes; 9) transcription factors, e.g. A20, IκBα, nfkb1 (p105 precursor), nfkb2 (p100 precursor), p53 (tumour suppressor).

Class	Protein	Aliases	Gene
I	NF-κB1	p105 -> p50	NFKB1
	NF-κB2	p100 -> p52	NFKB2
II	RelA	p65	RELA
	RelB		RELB
	c-Rel		REL

Table 3.1: The two NF-κB subfamilies and their associated proteins and genes. The class I NF-κB subfamily contains the proteins p50 and p52, whereas the class II subfamily contains RelA (also known as p65), RelB, and c-Rel. Table was developed from the reviews by Baldwin (1996) and Oeckinghaus and Ghosh (2009).

3.2.2 Composition of NF-κB and the Role of its Inhibitor

NF-κB transcription factors are hetero- or homodimers containing members of the Rel family of proteins. This family is composed of five distinct gene products which are related by a conserved region known as the Rel Homology Domain (RHD) that is responsible for nuclear localisation, DNA-binding (to κB sites), and dimerisation (Ghosh et al., 1998). The NF-κB family is divided into two functionally distinct subfamilies: the first are encoded by large precursor proteins that are subsequently processed into smaller DNA-binding subunits with strong DNA-binding activity and weak transcriptional activation potential (p105/NF-κB1 and p100/NF-κB2); the second are translated as mature proteins that bind to DNA weakly and contain potent transcriptional activation domains (RelA, RelB, and c-Rel (Siebenlist et al., 1994; Baldwin, 1996)). See table 3.1 for mapping between the two subfamily classes and the various NF-κB subunits and their corresponding gene.

The subunit composition of the Rel/NF-κB dimers influences its subcellular localisation and mode of regulation (Tian and Brasier, 2003). For example, homodimers of NF-κB1 are primarily located within the nucleus, and have weak transcriptional activation potential, whereas heterodimers of the prototypical (first discovered, and most common) NF-κB complex consist of a p50 subunit (lacking an activation domain) and a RelA subunit (containing an activation domain; as noted in table 3.1, RelA is also known as p65), and are primarily cytoplasmic, due to regulation by their inhibitory subunits known as IκBs (Baeuerle and Henkel, 1994). Like NF-κB, IκB is a member of a larger family of molecules that include IκBα, IκBβ, IκBϵ, and IκBγ. The best characterised IκB is IκBα, mainly because it was the first member of the family to be cloned (Ghosh et al., 1998). The function and regulation of p50/RelA heterodimer of NF-κB is shown in figure 3.3. The inactive form is a trimeric complex of NF-κB (p50/RelA) bound to the inhibitor protein IκB, and remains within the cytosol due to masking of its nuclear localisation sequence (Shirakawa and Mizel, 1989). In response to cellular stimulation however, the IκB subunit dissociates, thus releasing the NF-κB complex, and due to unmasking of the nuclear localisation sequence, the released p50/RelA complex enters the nucleus, where it becomes activated and binds cognate binding sites to upregulate the transcription of target genes (Beg et al., 1992).

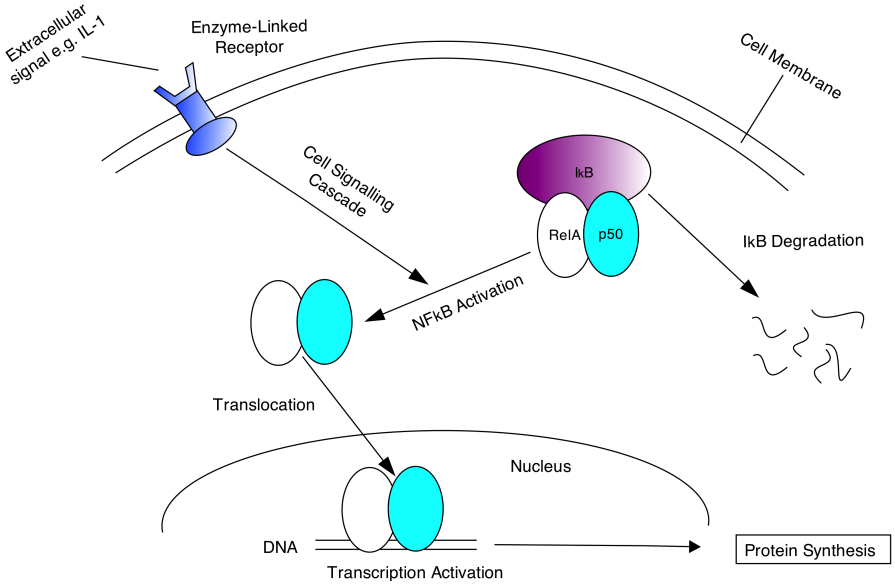

Figure 3.3: A high-level cartoon diagram of the p50/RelA NF-κB heterodimer and its associated signalling pathway. Following activation of an enzyme-linked receptor by an extracellular signal (such as IL-1), signal transduction is initiated through an intracellular signalling cascade and culminates with the dissociation (and degradation) of IκB from the inhibited NF-κB complex. The p50/RelA heterodimer is therefore released (due to the unmasking of its nuclear localisation sequence) and translocates to the nucleus where it activates transcription of target genes.

3.2.3 The NF-κB Signalling Pathway

Activation of the NF-κB transcription factor and signalling pathway is a tightly regulated event, involving phosphorylation of several members of the NF-κB and IκB protein families (Naumann and Scheidereit, 1994). NF-κB is normally sequestered in the cytosol of non-stimulated cells, and consequently must be translocated into the nucleus to function as a transcriptional activator of target genes. NF-κB activation is induced by a wide variety of different extracellular stimuli, including proinflammatory cytokines (such as TNFα and IL-1), bacteria, viruses, and physical and chemical stresses (Ghosh et al., 1998; Siebenlist et al., 1994). Currently, NF-κB activation is thought to be controlled by two distinct pathways, which have been termed the *canonical* and *non-canonical* pathways (Karin, 1999a,b; Senftleben et al., 2001). The two pathways are induced by different extracellular stimuli, controlled by different intracellular kinases, operate on different NF-κB complexes, and activate the transcription of different target genes (Brasier, 2006). The canonical pathway controls nuclear translocation of the p50/RelA (NF-κB1/RelA) heterodimer, whereas the non-canonical pathway controls translocation of the p52/RelB (NF-κB2/RelB) heterodimer.

Following activation of a cell membrane receptor and propagation of the signal via intracellular signalling, both the canonical and non-canonical pathways ultimately result in the signal reaching an IκB Kinase (IKK) complex. Israël (2009) advises that the composition of the IKK complex is key to the separation of canonical and non-canonical pathways. The IKK trimer within the canonical pathway consists of two kinase subunits IKKα and IKKβ, along with the regu-

latory subunit IKKγ (NF-κB Essential Modulator, or NEMO (Chen et al., 1996; Mercurio et al., 1997; Zandi et al., 1997)). This pathway involves activation of the IKK, with subsequent phosphorylation-induced degradation of IκB inhibitors for release of NF-κB. By contrast, the IKK trimer within the non-canonical pathway consists of two IKKα subunits and the NEMO subunit (Senftleben et al., 2001; Oeckinghaus and Ghosh, 2009), but also requires activation of an NF-κB-inducing kinase (NIK) for it to function (Xiao et al., 2004). NIK stimulates IKKα-induced phosphorylation of the NF-κB2 (p100) precursor to yield NF-κB2. See figure 3.4 for a high-level overview of the canonical and non-canonical pathways.

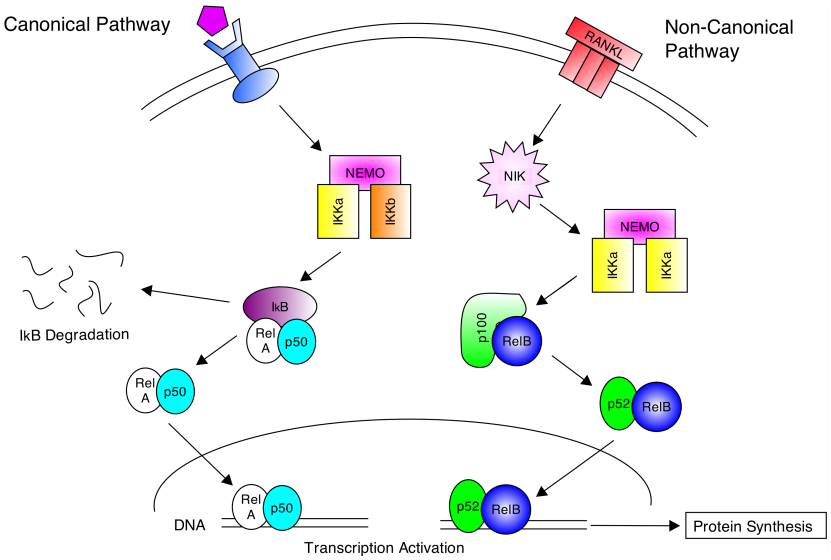

Figure 3.4: Simplified diagram depicting the two NF-κB signalling pathways. The canonical pathway on the left is initiated through extracellular stimuli, such as IL-1, and through activation of the IKK complex, leads to the degradation of IκB, and resultant release and translocation of the p50/RelA heterodimer into the nucleus for transcriptional activation. The non-canonical pathway is initiated through extracellular stimuli, such as Receptor Activation of NF-κB Ligand (RANKL; Novack 2011), and activates NIK, which then activates the IKK complex, before processing of the p100 NF-κB2 precursor into p52, and subsequent translocation of the p52/RelB heterodimer into the nucleus for transcriptional activation.

For the remainder of this thesis we will focus on the canonical pathway, and specifically signal transduction through the p50/RelA NF-κB heterodimer, as this is believed to be the most frequent Rel/NF-κB dimer in the majority of cells (Baeuerle and Henkel, 1994). Under normal conditions, IκBs bind to NF-κB dimers and sterically block their nuclear localisation sequences, thereby causing retention of the transcription factor within the cytosol. Following phosphorylation of the IκB (in this case IκBα) by the IKK complex (DiDonato et al., 1997), it undergoes a second modification called polyubiquitination[6], which then targets IκBα for rapid degradation (Brown et al., 1995). This degradation of NF-κB's inhibitor, exposes the nuclear localisation sequence, resulting in translocation of the transcription factor from the cytosol to the nucleus (Silver, 1991), where it

[6]It should be noted that neither phosphorylation nor ubiquitination alone is sufficient to dissociate the NF-κB-IκBα complex, hence, free NF-κB is only released after degradation of IκBα (Alkalay et al., 1995a; DiDonato et al., 1995)

is phosphorylated and becomes active (Sen and Baltimore, 1986b; Hoffmann and Baltimore, 2006). The active NF-κB may then bind to promoter regions[7] of target genes to activate their transcription. Interestingly, one of the first target genes to be transcribed is the inhibitor IκBα (Sun et al., 1993) itself. Newly synthesised IκBα accumulates within the cytosol before translocating to the nucleus, where it binds DNA-bound NF-κB and induces translocation back into the cytosol (Arenzana-Seisdedos et al., 1995). This negative feedback loop results in transient activation of the NF-κB signalling pathway.

3.2.4 Consequences of Malfunctions within the NF-κB Signalling Pathway

The most well defined function of NF-κB is its regulatory effects within the development and activation of the immune system. Indeed, NF-κB regulation is believed to be essential for the proper function of both the innate and adaptive immune systems. Under physiological conditions for normal (non-maligned) cells, NF-κB activation occurs transiently upon receiving a stimulus, due to the negative feedback regulation. Due to the very tight regulation involved with NF-κB activation and the subsequent downstream effects on gene expression, particularly in relation to immune responses to infection, any impairment of this regulation may result in NF-κBs direct involvement in a wide range of human disorders, including a variety of cancers (see special issues of Oncogene vol 18, 1999; and vol 25 2006), neurodegenerative diseases (Grilli and Memo, 1999), cardiovascular diseases (Brasier, 2006), arthritis (Foxwell et al., 1998), and numerous other inflammatory conditions (see special issue of Cell Research vol 21, 2010). In such cases, NF-κB is believed to have lost its transient nature of activation and to have become constitutively activated, leading to the uncontrolled expression of genes (Sethi et al., 2008). A good example is when the NF-κB pathway becomes dysregulated, and allows various cytokines to constitutively activate the pathway. This results in potential autoimmune responses, in particular within the central nervous system, and resultant onset of demyelinating diseases such as Multiple Sclerosis (MS) in humans, and the analogous Experimental Autoimmune Encephalomyelitis in animals (Karin and Lin, 2002). Bonetti et al. (1999) advise that the nuclear localisation of NF-κB is upregulated in various immune response cells (e.g. microglia and oligodendrocytes) that are located in active MS lesions. They propose that in such autoimmune diseases, the activation of NF-κB may exert an anti-apoptotic effect and contribute to the absence of apoptosis for maligned immune response cells.

Of particular note to various cancers is that several gene products which negatively regulate apoptosis in tumour cells (antiapoptotic, hence prosurvival) are controlled by NF-κB activation. Indeed, NF-κB has been linked to antiapoptotic function in tumours such as T-cell lymphoma and melanoma (Sethi et al., 2008), and recently Bivona et al. (2011) have identified NF-κB as a potential companion drug target for treatment of lung cancers. Their studies found that increased levels of IκB improved the response and survival of patients being treated for EGFR-mutant (epidermal growth factor receptor mutant) lung cancer. Simi-

[7]Every gene has an associated region of DNA called a promoter region, which is a sequence of DNA indicating where transcription of that gene should begin.

larly, Flavell et al. (2010) state that the M2 phenotype of tumour associated macrophages (macrophages stimulated by IL-3 or IL-4, and involved in tumour promotion) are mediated by the p50 NF-κB subunit, and that TGFβ down-regulates NF-κB in these macrophages. Furthermore, the metastasis of cancer requires the migration of cancerous cells both into, and out of, the vasculature that transport them around the body. As highlighted above (see section 3.2.1), one consequence of NF-κB activation is the penetration of cells across the vasculature through production of cell adhesion molecules (e.g. VCAM-I and ICAM-I) within the endothelial cells of blood vessels.

A number of viruses induce NF-κB activation, some of these intriguingly have also been found to contain NF-κB binding sites in their viral promoters (Pahl, 1999). As such, it is possible that viruses are able to *hijack* the transcriptional upregulation that results from NF-κB activation and rapidly replicate within an infected host. Thus the host's own defensive machinery (through NF-κB's ability to induce production of defensive compounds), becomes turned against itself. For example, it is thought that the presence of an NF-κB binding site in the HIV-1 promoter region may be one reason for the virulence of the virus in infected patients, and a low-level of NF-κB activation may help explain the mechanisms behind chronic infections (Kovacs et al., 1995).

NF-κB has wide-ranging effects controlled by a complex regulatory network of inhibitors and co-activators. Understanding the mechanisms that control NF-κB activation/cellular signalling is important for exploiting therapeutic approaches to treat human disorders due to its dysregulation. Specific targets for therapeutic agents could be the transcription factor itself, or any of the associated components within the pathway, including protein kinases and the IκB inhibitors themselves. Gilmore and Herscovitch (2006) reviewed the known inhibitors of NF-κB, which may provide a basis for research regarding pharmacological intervention.

3.3 Receptors, Co-Receptors, Adaptor Proteins and Kinases

In addition to the various NF-κB dimers, IκB inhibitor proteins and IKK complexes, there are a large number of upstream components that facilitate signal transduction through the NF-κB signalling pathway. Section 3.1.2 highlighted the three main types of cell membrane receptors within the immune response, and one particularly important family of receptors linked to the NF-κB signalling pathway within the innate immune system are the Toll-like receptors. Along with these *system signalling* cell membrane receptors, there are also co-receptors (additional cell membrane proteins), which are able to better recognise extracellular signals (Murphy et al., 2008). Upon extracellular signal recognition, the receptor becomes activated through a number of conformational changes being induced on its intracellular surface, which allows the binding of proteins within the cytoplasm. Once active, these receptor/co-receptor complexes are able to propagate signal transduction through a number of intermediate components, encompassing adaptor proteins and protein kinases, before reaching the IKK complex for subsequent activation of NF-κB.

3.3.1 Toll-like Receptors and the IL-1R Superfamily

The Toll-like receptor (TLR) family are important components of the innate immune system that recognise a wide range of microbial products, which possess pathogen-associated molecular patterns[8] (Doyle and O'Neill, 2006). There are 10 TLR proteins in humans[9], each of which recognises a distinct set of signal patterns, that are not represented within normal vertebrates, i.e. they recognise bacterial, viral and fungal proteins (Murphy et al., 2008). TLR1 & TLR2 and TLR2 & TLR6 form heterodimers and recognise bacterial lipoproteins; TLR3 recognises viral double-stranded RNA; TLR4 recognises Gram-negative bacterial lipopolysaccharides (LPS); TLR5 recognises bacterial flagellin (tail-like protrusion that functions in cell movement); TLR7 and TLR8 recognise viral single-stranded RNA; and TLR9 recognises specific DNA motifs that exist in viruses and bacteria (Doyle and O'Neill, 2006). Finally, TLR10 remains the only mammalian Toll-like receptor to which ligands have not yet been characterised, and forms heterodimers with TLR1 and TLR2 (Hasan et al., 2005).

TLRs were named following the discovery that they were homologous to a transmembrane protein named *toll*, which is necessary for embryo development (Anderson and Nusslein-Volhard, 1984) and defence against fungal infection (Lemaitre et al., 1996) in the fly *Drosophila melanogaster*. Furthermore, DNA sequencing of toll protein, highlighted high homology to the human interleukin 1 receptor (IL-1R; Gay and Keith 1991). The homology between toll and IL-1R was within their cytosolic (intracellular) regions, and the domain was termed the Toll/IL-1R resistance (TIR) domain (Doyle and O'Neill, 2006). Due to this homologous intracellular region, TLR and IL-1R have been grouped into a superfamily, with the TLR subgroup possessing an extracellular domain containing leucine-rich repeats, and the IL-1R subgroup possessing an extracellular immunoglobulin domain (O'Neill and Dinarello, 2000). Hultmark (1994) discovered that the macrophage differentiation marker MyD88 was a member of the TLR/IL-1R superfamily. It was later found that MyD88 connected the TIR domain of receptor proteins to downstream signalling components, and due to its lack of extracellular domain, it was deemed to be an adaptor protein (Wesche et al., 1997).

3.3.2 CD14, MD2 and TILRR Co-Receptors

The most well defined TLR protein is TLR4, and following ligation with LPS, it dimerises (Poltorak et al., 1998)[10] and associates with CD14 (a co-receptor for LPS) and an additional cellular protein MD2 (Murphy et al., 2008) in order to signal the presence of the bacterial product. There are two forms of the CD14 co-receptor: the membrane CD14 (mCD14), which is present at the surface of myeloid cells; and the soluble CD14 (sCD14), which facilitates the binding of LPS in cells that do not possess mCD14, such as endothelial cells, and is also secreted by the liver (Wright et al., 1990). Similarly, the MD2 protein associates with TLR4 on the cell membrane and also facilitates the response to LPS (Re and Strominger, 2002; Shimazu et al., 1999).

[8]Pathogen-associated molecular patterns (PAMPs) are a class of conserved microbial structure

[9]TLRs 3, 7 and 9 are intracellular, and therefore are not membrane-bound receptors.

[10]The laboratory lead for this paper, Bruce Beutler, was later awarded a Nobel prize in Medicine and Physiology for discoveries concerning the activation of innate immunity.

Recently, Zhang et al. (2010, 2012) discovered another co-receptor involved in the NF-κB signalling pathway. They have termed this co-receptor a Toll-like/IL-1 Receptor Regulator (TILRR), due to its association with (and activation of) IL-1R type I (IL-1RI). Furthermore, the TILRR/IL-1RI complex was also found to magnify the NF-κB inflammatory response by enhancing association with the MyD88 adaptor protein, and also to link TIR domain activation with mechanotransduction[11] through regulation of the Ras GTPase.

3.3.3 Adaptor Proteins and Kinases

A large degree of the specificity inherent within signal transduction pathways depends on the recruitment of multiple signalling components such as protein kinases and G-protein GTPases into short-lived active complexes in response to an activating signal such as a cytokine or bacterial product binding to its receptor. Adaptor proteins are proteins that lack any intrinsic enzymatic activity themselves, but instead mediate specific protein-protein interactions that drive the formation of protein complexes (Murphy et al., 2008). Following recognition of foreign material, the TLR and IL-1R superfamily utilise adaptor proteins to facilitate the formation of receptor complexes, many of which involve protein kinases, for propagation of signal via a cascade of phosphorylation steps within the NF-κB signalling pathway (Kawai and Akira, 2006)

One of the initial adaptor proteins encountered following ligand binding to TLR4 or IL-1RI, is MyD88. Once associated with the cell membrane receptor, MyD88 recruits members of the interleukin-1 receptor-associated kinase (IRAK) family (Wesche et al., 1997). To date, four IRAKs have been identified (IRAK1, IRAK2, IRAK4, and IRAK-M), with IRAK1 and IRAK4 possessing kinase ability. Upon recruitment to the receptor-MyD88 complex, IRAK4 and/or IRAK1 are sequentially phosphorylated (Cao et al., 1996a), before becoming dissociated from MyD88. The phosphorylated IRAK propagates the signal by activating tumour necrosis factor receptor-associated factor 6 (TRAF6; Cao et al. 1996b), which activates transforming growth factor-β-activated protein kinase 1 (TAK1; a member of the mitogen-activated protein kinase kinase kinase (MAP3K) family), that subsequently activates the IKK (IκB kinase) complex for degradation of IκBα and release of NF-κB (Doyle and O'Neill, 2006; Kawai and Akira, 2006). Along with MyD88, another adaptor protein termed Toll interacting protein (Tollip) mediates IRAK to IL-1R associations, however in this instance, the adaptor protein is believed to inhibit IRAK phosphorylation, and thus also inhibit the interaction of IRAK with downstream kinases (Burns et al., 2000).

Oda and Kitano (2006) present a comprehensive map of TLR and IL-1R signalling networks, which is based on published papers. They show that TLR signalling pathways can be approximately divided into four possible subsystems, of which the main subsystem contains MyD88-IRAK4-IRAK1-TRAF6 as a possible core process for the activation of NF-κB and the MAPK cascade, leading to the induction of the range of target genes (such as cytokines) that are essential for the innate immune response. Their map highlights extensive crosstalk and they propose MyD88 to be a non-redundant core element.

[11] Mechanotransduction is the mechanism by which cells transform mechanical stimulus into chemical activity.

3.4 Recent Experimental Approaches to Study NF-κB

The NF-κB signalling pathway has been extensively studied over the past 28 years since its first discovery. The majority of research has focused on large-scale biochemical dynamics, which require several thousands of cells at a time, that need to be killed in order for results to be obtained. Subsequent results gained are therefore at the level of the entire group, and essentially provide an average of the dynamics being studied (this is akin to the use of ordinary differential equation models discussed in section 2.6.2), and as such effectively hides the large differences in behaviour between cells. It is repeatedly observed within biology that genetically identical, individual cells in a standardised environment often display significant differences in their response to perturbations (Tijskens et al., 2003). To be able to look at these variations in detail, scientists have used *single-cell analysis*. Here, observations are made, and data collected, on individual living cells, allowing the variety of behaviours (between cells) throughout the time-course of experiments to become evident (this is akin to the use of agent-based modelling discussed in section 2.6.3).

The work of Carlotti et al. (1999) used an experimental technique for analysing signalling pathways, which involves the expression of exogenous fluorescently tagged proteins to see the effects on the cell. They used a set of enhanced green fluorescent protein (EGFP) and RelA constructs to analyse the IL-1 induced activation of NF-κB. Analysis of these effects was performed at the single-cell level using confocal laser scanning microscopy[12]. This was the first published example of single-cell analysis of the NF-κB signalling pathway. An example of the images generated by the Qwarnstrom group using fluorescent tagging and confocal microscopy is shown in figure 3.5. Through using a series of single-cell analysis experiments, they determined the kinetics of nuclear uptake of NF-κB in response to IL-1 stimulation[13]. Furthermore, they found that the signal transduction pathway propagating signals from IL-1RI to activation of NF-κB is saturable with respect to RelA concentration and displays modest amplification of signal throughout the signalling cascade. The authors built on this work a year later by using additional fluorescent protein constructs and a nuclear export inhibitor to study the two-way translocation between the cytosol and nucleus of NF-κB and IκBα (Carlotti et al., 2000). They developed a mathematical model of the association and dissociation of the NF-κB-IκBα complex, and import/export of components to demonstrate shuttling kinetics across the nuclear membrane. With this model they demonstrated a modest rate-limiting step for nuclear translocation, which was due to upstream effects (limitations) in the signal transduction pathway, and also predicted that $\sim$17% of cellular NF-κB is free of IκBα. This nuclear shuttling of NF-κB and IκBα, was confirmed by Schooley et al. (2003),

[12] Confocal microscopy requires molecules to be fluorescently tagged, and allows their behaviour to be observed as it is able to focus on a thin layer of a cell at a time, thus allowing the nucleus and cytoplasm to be distinguished.

[13] Interleukin (IL)-1 refers to two cytokines, IL-1α and IL-1β, which are key components in the regulation of inflammatory processes, as they promote recruitment of inflammatory cells at the site of inflammation through the expression of adhesion molecules and chemokines (Gabay et al., 2010).

using the technique of immunocytochemistry, who also showed that control of endogenous levels of NF-κB subunits exhibited the same cell-to-cell variation, and thus was a *real biological event* and not due to transfection.

Work within the group then focused on the degradation and translocation of IκBα and how this affected the dynamics of NF-κB. Yang et al. (2001) discovered that degradation and translocation of IκBα was dependent on the level of cytoplasmic expression, i.e. the level of degradation demonstrates a positive correlation with the level of the inhibitor protein within the cytoplasm (see figure 3.6). They later demonstrated that translocation, phosphorylation, and NF-κB-IκBα complex formation are all critical for regulation of IκBα steady-state levels by NF-κB (Yang et al., 2003). This latter work also highlighted that both the extent and affinity of the NF-κB/IκBα interaction plays a key role in system regulation, affecting both basal levels and those induced in response to incoming signals. This technique of using fluorescent constructs of components within the NF-κB signalling pathway as part of single-cell analysis, has since been used to assess the TNF induced activation of the pathway by several other groups who have confirmed the advantages it provides in obtaining detailed information of pathway regulation (Nelson et al., 2004; Ashall et al., 2009; Lee and Covert, 2010). Indeed, Lee and Covert (2010) suggest that single-cell analysis is advantageous over traditional biochemical techniques because they provide better time resolution of different responses between cells, higher sensitivity, automation, and richer datasets. Along with IL-1 mediated induction of NF-κB and IκBα dynamics, the group has also researched mechanotransduction (Ganguli et al., 2005).

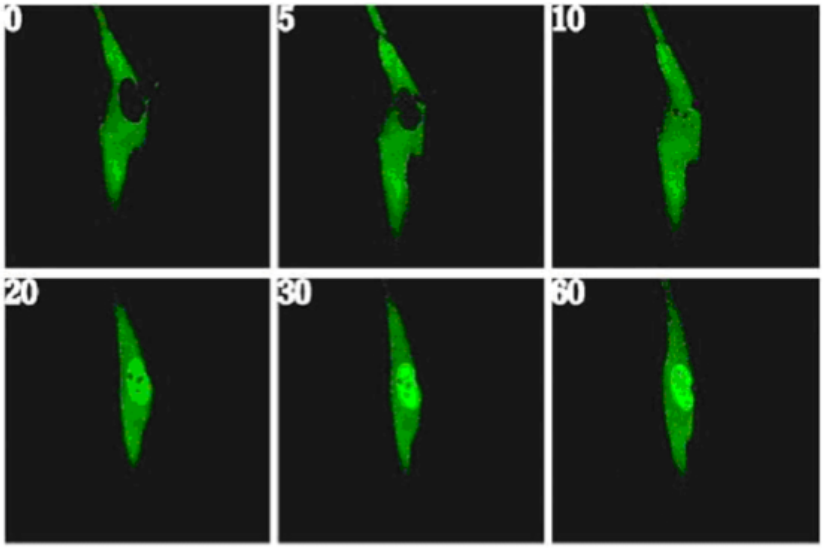

Figure 3.5: Image of an individual cell's NF-κB molecules using EGFP tagging and confocal microscopy following IL-1 induced stimulation. Images correspond to 0, 5, 10, 20, 30 and 60min following IL-1 induced stimulation, and the dark region at the centre of the cell at 0min is the nucleus, which can clearly be seen to be free of NF-κB. As time progresses, the intensity of fluorescence within the nucleus increases markedly, corresponding to an increase in NF-κB in this cell compartment. Figure from Carlotti et al. (1999).

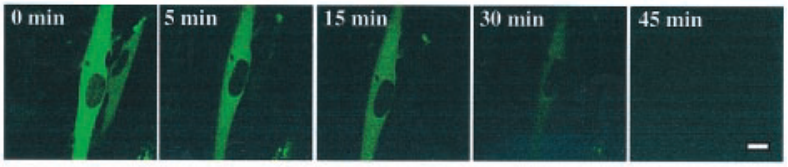

Figure 3.6: Five time-series images generated using confocal microscopy of fibroblasts following transfection with IκBαEGFP and RelA. The time-series clearly shows the degradation of IκBα following stimulation with IL-1β. Reproduced from figure 1a of Yang et al. (2003).

3.5 Current Computational Models of NF-κB

A formidable body of knowledge has been generated through wet-lab experimentation since the discovery of NF-κB in 1986, and more recently through interdisciplinary studies using computational models of the pathway. Segel (1995) has said that one of the hallmarks of a complex system is that it is a system that cannot be described by a single model. In this respect, we need to look no further than the recent work to model the NF-κB cell signalling pathway. Following the move by a large section of the scientific community away from the reductionist approaches and towards a systems-level understanding, the use of computational modelling and simulation has been adopted by a growing number of well known labs within the NF-κB arena. Various individuals, and indeed different research groups, have made use of fundamentally different computational methods to explore NF-κB transcriptional activation. As the field of computational modelling is maturing, the computational approaches that have been used are beginning to pay dividends. Recently, equation-based and agent-based models have been used within a predictive capacity to generate hypotheses for testing through additional wet-lab experimentation. We will review these existing computational models throughout the rest of this section.

3.5.1 Deterministic Ordinary Differential Equation Models

It is generally accepted that the first mathematical model of NF-κB dynamics was developed by Carlotti et al. (2000). This model focused on IκBα association and dissociation rates, along with IκBα and NF-κB translocation between the cytoplasm and the nucleus (see figure 3.7). Their reaction kinetics model consisted of six variables (NF-κB, IκB, and NF-κB-IκB concentrations in both the cytoplasm and nucleus) and ten reaction processes. They examined each pair of processes in turn to determine the net effect on system variables, and extrapolated the system conditions required for dynamic equilibrium.

We believe the first model of the wider NF-κB signalling pathway however, was that of Hoffmann et al. (2002). They developed an ODE-based computational model using the Gepasi software (version 3.1; Mendes (1993, 1997)). They were particularly interested in the temporal control of NF-κB activation by the coordinated degradation and synthesis of IκB proteins. The model therefore focuses on reactions that govern IκB metabolism, including synthesis, degradation, cellular localisation, and association/dissociation with NF-κB.

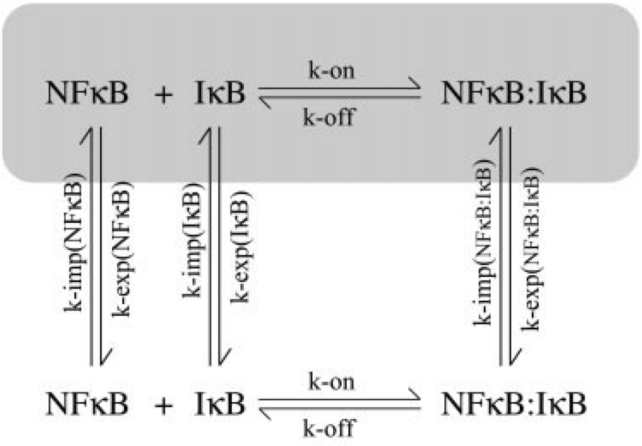

Figure 3.7: Reaction kinetic model of nuclear-cytoplasmic shuttling by Carlotti et al. (2000). This model consists of 6 variables and 10 processes. The variables correspond to concentrations of NF-κB, IκB, and NF-κB-IκB in both the cytoplasm and nucleus. The processes correspond to the kinetics of NF-κB and IκB association and dissociation, along with the translocation kinetics of each component across the nuclear membrane.

This mathematical model comprised a system of 24 ODEs describing reaction kinetic equations of the change in concentration (with respect to time) of cytoplasmic and nuclear NF-κB, cytoplasmic and nuclear IκB (-α, -β, and -ϵ), cytoplasmic IKK, and resultant complexes that they may form. A number of biochemical reactions were grouped to reduce complexity of the model, two examples are: (i) IκB proteins are degraded via phosphorylation by IKK, then polyubiquitination and subsequent degradation by the proteasome, however the model combines these into a single IKK-dependent degradation reaction; (ii) mRNA synthesis, translocation to cytoplasm, and translation into protein, was treated as a single-step process of IκB protein synthesis. The model also contained 30 parameters, including: (i) synthesis of each IκB isoform; (ii) stability of free NF-κB-bound IκB proteins; (iii) formation of IKK-IκB-NF-κB complexes; (iv) enzymatic rate constants of IKK-containing complexes; and (v) transport rates affecting localisation of IκBα, β, ϵ and NF-κB and their complexes. The parameter values used were gained using a mixture of biochemical experimentation within their lab, review of published literature, and estimation through model fitting techniques. An example of this latter approach are the five parameters that were used to capture steady-state nuclear and cytoplasmic localisation of NF-κB and IκB proteins.

Alongside the computational modelling, the authors performed wet-lab experimentation to study the dynamics of NF-κB-IκB signalling through measuring NF-κB activity (over time), not by single-cell analysis, but instead by using electrophoretic mobility shift assay (EMSA) and western blots for IκB. They were interested in the differential functions of IκBα, β, and ϵ isoforms on regulation of NF-κB activation, and used knockout mice for the three isoforms, and cross-bred these to yield double-knockout cells where NF-κB was inhibited by a single IκB isoform. These cells were stimulated with TNFα in order to induce the NF-κB signalling response. Their conclusions from wet-lab experimentation were that coordinated degradation, synthesis and localisation of all three IκB isoforms is required to generate the characteristic NF-κB activation profile.

Three computational models were developed to represent these double knockouts: $\beta^{-/-}\epsilon^{-/-}$ (IκBα), $\alpha^{-/-}\epsilon^{-/-}$ (IκBβ), and $\alpha^{-/-}\beta^{-/-}$ (IκBϵ), which were *calibrated* so that simulation results were consistent with wet-lab dynamics. The three computational models were then combined to form a single model that was believed to represent *wild-type* cells, and used as a basis for predicting wild-type dynamics. Through varying the relative contributions of IκB isoforms, the model was used to predict the discrete functional roles of IκB proteins in NF-κB regulation. The combined model has IκBβ and -ϵ synthesised at a steady rate throughout simulations, and increasing IκBα synthesis following NF-κB activation. The IκBβ and -ϵ genes within the model are therefore not under the control of NF-κB. Because of this subtlety, the authors were able to use model fitting to determine IκBβ and -ϵ mRNA synthesis parameters to calibrate simulation responses to those of experimental wild-type responses. Simulation yielded two very different results, predicting that (i) IκBα mediates rapid NF-κB activation and strong negative feedback regulation, resulting in an oscillatory NF-κB activation profile, and (ii) that IκBβ and IκBϵ respond more slowly to IKK activation and act to dampen the long-term oscillations of the NF-κB response.

The authors believe that the outputs of the wild-type model are in good agreement with their wet-lab data, and therefore that the model may be used for predictive purposes. We agree with this statement in so far as the predictions are to be made in relation to the very small scope of the original model. Hoffmann et al. (2002) modelled a very small subset of the NF-κB signalling pathway, which they termed the IκB-NF-κB signalling module (IKK, NF-κB, and IκB), and therefore quite heavily abstracted the model away from biological reality. Their reasoning is that the majority of upstream activities converge at the IKK complex, and therefore activation of the complex would be an appropriate input for the model. Similarly, as NF-κB may act as a transcriptional activator for a number of target genes, they deemed its activation and translocation to the nucleus to be an appropriate output for the model. Although we understand, and indeed agree with this reasoning, the model has become very generalised regarding NF-κB responsive genes (there is no detail regarding actual gene transcription other than IκBα), and furthermore, although the authors state that the IKK complex becomes activated following TNFα stimulation, the amount of abstraction inherent to the model means that this could actually be any stimulatory signal. We therefore believe that the initial model may have been heavily *tuned* in order to align with their wet-lab experimental dynamics.

Recently, Cheong et al. (2006, 2008) have reviewed this initial work and argue that the functions of the three IκB isoforms combine to allow the signalling module to distinguish between short and long lasting stimuli. Furthermore, Kearns et al. (2006) through augmentation and simulation of the model, discovered that IκBϵ provides negative feedback to control NF-κB dynamics, and Kearns and Hoffmann (2009) have also updated the model to contain 73 parameters. Like the original model, they have used various sources for the parameter values: a third coming from *in vitro* measurements and quantitative cell biology (in particular reactions of IKK-mediated IκB phosphorylation and degradation), a third from published literature (e.g. half-lives, transport rates, and NF-κB-IκB affinities), and the remaining third from parameter fitting. Following additional wet-lab experimentation, they also hypothesise that the relative strength of IκBα and

IκBϵ feedback mechanisms and their temporal relationships to each other may account for cell type-specific regulation of NF-κB dynamics. Basak et al. (2007) have also updated the model to introduce the p100 protein, which is part of the non-canonical signalling pathway, and LPS induction of IKK-mediated IκB degradation. They showed that p100 acts as an inhibitor of NF-κB activation, and termed this IκBδ as it was found to be a bona fide IκB protein. They used the model to explore feedback regulation and dynamics of p100, and suggest there may be crosstalk between the canonical and non-canonical signalling pathways. More recently, Shih et al. (2009) have further augmented the model to simulate the effects of the newly discovered inhibitor IκBδ, and found that it provides negative feedback in the presence of persistent, pathogen-triggered signals, by dampening NF-κB responses during sequential stimulation events. The authors subsequently updated this model to make specific to dendritic cells (DC), and discovered that RelB regulation during DC activation integrates the canonical and non-canonical pathways (Shih et al., 2012). We believe that this computational model is the first to model the kinetics of both RelA and RelB containing dimers.

Despite the successes of the two models (initial and revised) from the Hoffmann lab, they appear to have a number of weaknesses. The models were calibrated against wet-lab experimentation that used the average of multiple cells rather than single-cell analysis; as such, they did not reflect the large variation in behaviours between individual cells that is inherent to the NF-κB signalling pathway. With this in mind, the work by Nelson et al. (2004) built on the early population-based model from the Hoffmann group to incorporate functionality for single-cell dynamics, as seen in their wet-lab results. They used fluorescence imaging of RelA and IκBα to study oscillations in RelA nucleus-cytoplasm localisation in HeLa[14] cells and SK-N-AS cells (human s-type neuroblastoma cells that possess deregulated NF-κB signalling). Their work showed that single-cell time-lapse imaging and computational modelling of RelA localisation showed asynchronous oscillations following cell stimulation that decreased in frequency with increased IκBα transcription. The authors used the ODE-based model developed by Hoffmann *et al* and validated against their own single-cell experimental data. Through simulation, they were the first to discover that nuclear-cytoplasmic oscillation, damping, and peaks of RelA (through TNFα stimulation), are highly evident in single-cell analysis, but that this was not apparent at the *averaged* population-level due to phase differences in the dynamics of the individual cells. They also discovered that continuous TNFα stimulation elicited oscillations in IκBα expression, which were out of phase with RelA nuclear-cytoplasmic oscillations. Furthermore, they highlighted that changes to parameter values for free IKK and IκBα were intimately coupled to the oscillation dynamics of nuclear NF-κB, and suggested that these NF-κB oscillations could repeatedly deliver newly activated NF-κB into the nucleus, maintaining a high nuclear ratio of active-inactive NF-κB. They tested this in the wet-lab using a nuclear export inhibitor (leptomycin B) to trap RelA in the nucleus, with results supporting the earlier work of Carlotti et al. (2000), who showed that rapid dephosphorylation of NF-κB in the nucleus may be a key factor in the switch-off of NF-κB-dependent

[14]HeLa cells are human cervical carcinoma cells, named after Henrietta Lacks, from whom the cell line was derived (Lucey et al., 2009)

gene expression. Using these simulation and wet-lab experimental results, they proposed that oscillations in NF-κB localisation, coupled to cycles of RelA and IκBα phosphorylation, maintain NF-κB-dependent gene expression.

Lipniacki et al. (2004) used the Hoffmann model as a baseline, and amended in three main ways. Firstly, the kinetics of the nuclear and cytoplasmic compartments were refined to take into account differences between their volumes when calculating concentrations of components - Hoffmann et al. had assumed the sum of cytoplasmic and nuclear concentrations of diffusing substrates to be constant; this holds true when cytoplasmic and nuclear volumes are equal, but in most cases the cytoplasm is substantially larger than the nucleus (Alberts et al., 1994). Secondly, they made use of additional information (i.e. assumptions and considerations from publications by external groups) regarding IκBα interactions within the system; for example Carlotti et al. (1999) and Rice and Ernst (1993) suggest that only 10-15% of total NF-κB is not complexed to IκBα in a resting cell[15]. Thirdly, they re-estimated the mRNA transcription and translation coefficients through the use of published molecular level data. This latter point reflected the need to introduce granularity around formation and dissociation of complexes, mRNA synthesis (transcription), translocation to cytoplasm, and translation into proteins.

Lipniacki et al. (2004) utilised a high degree of abstraction, and again focused on the NF-κB-IκB signalling module by using 15 ODEs to model kinetics of IKK, NF-κB, IκBα, and the IKK inhibitor A20 within the nuclear and cytoplasmic compartments of cells. They removed IκBβ and IκBϵ and approximated the collective action of all IκB isoforms by IκBα, which is the most active and abundant one. They also added three new assumptions around the dynamics of the IKK complex: (i) resting cells have a *neutral* form IKK_n; (ii) stimulated cells have an *active* form IKK_a, which phosphorylates IκBα; and (iii) Krikos et al. (1992) showed that IKK can be inactivated by a protein named A20, which like IκBα, experiences strong transcriptional activation by NF-κB. This inhibition of IKK was incorporated into the model, with A20 converting IKK_a into the *inactive* form IKK_i, which no longer phosphorylates IκBα.

The model by Hoffmann et al. had assumed that ubiquitination and degradation of IκBα immediately followed its phosphorylation by IKK. Lipniacki et al., through reference to Yang et al. (2001), believe this to take several minutes in reality, and therefore modified the NF-κB nuclear transport coefficient to incorporate a longer time delay between the IKK-mediated phosphorylation of IκBα within the NF-κB-IκBα complex, and the translocation of free NF-κB from the cytoplasm into the nucleus. This modification enabled the model to also take into account the time needed for ubiquitination and proteolysis, within a single coefficient. Additional assumptions were built into this revised model around the IKK complex: (i) each form of IKK (IKK_n, IKK_a, and IKK_i) were to degrade with the same rate; (ii) IKK_a would form transient complexes with IκBα, but following IκBα phosphorylation, the IκBα degrades, whereas the IKK molecule remains; (iii) there would be constant synthesis of new IKK to replenish those that have degraded; (iv) A20 synthesis is also induced by NF-κB nuclear activity,

[15]The work by Pogson et al. (2008) suggesting that 66% of IκBα binds to the cytoskeleton had not taken place yet.

but when synthesised (located within the cytoplasm) it inhibits IKK, and has the effect of decreasing IKK-mediated degradation of IκBα, therefore resulting in more IκBα and thus less activated NF-κB. Their model therefore incorporates two negative feedback loops for NF-κB activation, one involving inhibition of free NF-κB via the binding of IκBα, and the other involving the inhibition of IKK by A20. The amendments were made to the existing code (from Hoffmann et al. 2002) which was written in Gepasi, and additional scripts to analyse the resulting simulation data were also developed, but this time in Matlab[16]. The model was validated against published data of Lee et al. (2000) regarding A20-mediated dynamics, and the model dynamics of Hoffmann et al. (2002). After parameter fitting, the authors believed that the model successfully reproduced time behaviour of wild-type and A20 deficient cells.

Following the publication of this augmented computational model, the Hoffmann group (Cheong et al. (2006)) reimplemented their model using Cellerator (Shapiro et al., 2003), updated their parameter values to take account of the revisions from Lipniacki *et al*, and extended the model scope to incorporate IKK activation. They complemented this computational work with additional wet-lab experimentation (at the population-level of cells) that focused on constant stimulation with varying TNFα doses. They found that both duration and dose of TNFα have little effect on the duration of the initial NF-κB response, and that NF-κB responds sensitively to a wide range of TNFα concentrations. Their *in silico* experimentation predicted that these signal transduction properties were crucially dependent on the transient nature of IKK activity. Furthermore, they found the dynamics of IKK activity to be non-linear in nature, which they conjecture could be the basis for ensuring robust TNFα-induced NF-κB responses to offset limitations imposed through diffusion of the ligand within the extracellular environment. More recent work by Tay et al. (2010) and Turner et al. (2010) who incorporated a number of stochastic processes within their computational models (see semi-stochastic models below), have shown that in contrast to these population-level studies, that the NF-κB response is heterogeneous at the single-cell level, with fewer cells responding to lower doses of TNFα, due to the individual cell responses being controlled by a stochastic threshold, and thus yielding an *all-or-nothing* response (Turner et al., 2010).

Finally, the most recent deterministic model that we are aware of was developed by Choudhary et al. (2013), who predicted through *in silico* experimentation (and validated using siRNA knockdown wet-lab experiments), that the canonical and non-canonical pathways are coupled through the action of TNF associated factor 1 (TRAF1) and NF-κB inducing kinase (NIK). They demonstrated that TNF stimulation (of the canonical pathway) induces TRAF1 expression, and that the newly expressed TRAF1 binds to NIK (of the non-canonical pathway) with high avidity. They conjecture that the TRAF1-NIK complex is a central component for cross-talk between the two NF-κB pathways, as the TNF-induced delayed activation of the non-canonical pathway is dependent upon a feed-forward mechanism activated by TRAF1 expression from the canonical pathway.

[16]http://www.mathworks.com/products/matlab/

3.5.2 Semi-Stochastic (Hybrid) Differential Equation Models

Following on from these initial ODE models, Lipniacki et al. have published two subsequent increments of their model, by adding stochastic gene activation (Lipniacki et al., 2006), and then stochastic receptor activation by TNFα (Lipniacki et al., 2007); with the wider group also publishing two additional increments, by adding granularity at the single-cell level (Tay et al., 2010), and incorporating positive feedback through TNFα expression (Pekalski et al., 2013). The first increment continued to use ODEs (14 this time around) with the Gepasi software and also introduced a *stochastic switch* to account for the activity of the genes relating to A20 and IκBα. They did this by using 4 equations which account for binding and dissociation probabilities of NF-κB molecules to regulatory sites (in DNA) for A20 and IκBα promoters, the gene transcription into mRNA and the resultant translation into A20 and/or IκBα proteins. This stochastic gene activation facilitates the variability of protein concentrations within simulated cells that is akin to that seen in biology, e.g. the stochastic switch allows large variances in simulations due to amplifying effects of mRNA synthesis and protein translation. We believe Lipniacki et al. (2006) were the first to have successfully modelled the amplification cascade (of the NF-κB signalling pathway) following gene activation.

The second increment by Lipniacki et al., was again built on their previous work, but was entirely re-written using Matlab. They once again used differential equations (15 this time around) to model the kinetics of NF-κB signalling, but this time further increased the scope to incorporate the TNFα cell membrane receptor and the enzyme IKK kinase (IKKK), which activates the IKK complex. The TNFα receptor was incorporated into the amplification cascade described above (for the second generation Lipniacki model) through its activation being stochastic. The authors performed manual parameter fitting for this receptor activation so that 90% of simulated cells are activated in the first 10min of TNFα stimulation, and claim that predictions from their single-cell level simulation results agree qualitatively with published IKK and NF-κB activity data of the population-level results from Cheong et al. (2006). However, unlike the population-level results of Cheong et al. that suggest TNFα dose has little effect on the duration of the initial NF-κB response, their *in silico* experiments show that at low TNFα dose only a fraction of cells are activated, but that in these activated cells the amplification mechanisms assure that the amplitude of the NF-κB nuclear translocation remains above a threshold. Additional *in silico* experiments showed that low nuclear NF-κB concentration only reduces the probability of gene activation, but does not reduce the gene expression of those responding. They hypothesise that the two effects provide stochastic robustness in responding cells, allowing cells to respond differently to the same stimuli, but causing their individual responses to be unequivocal. This suggests that amplification-saturation dynamics are present within the model, due to the final cell response at the level of NF-κB target genes being approximately equal, regardless of whether a single receptor or 100 receptors were activated. It is currently unknown whether a single activated receptor is sufficient to initiate a response of this magnitude in biological systems, so further wet-lab experimentation is required in parallel to the computational work of this group.

The third increment, was developed by Tay et al. (2010), who augmented the

mathematical model with additional parameters so that it may be used with high-throughput single-cell resolution wet-lab experimental data. The model contains 16 differential equations (some of which are stochastic in nature) and 34 rate constants, 20 of which used fixed constants from published data, and the remaining 14 were manually fitted in order to calibrate simulations against single-cell traces. The most intriguing finding from their wet-lab and *in silico* experiments, which were also independently discovered by Turner et al. (2010) during a similar time period, was that not all cells responded to TNFα and that the fraction of activated cells decreased with decreasing TNFα dose, thus representing the discrete nature of single-cell activation. The also found that early gene expression was not dependent on the intensity of the inducing signal, but instead relied on the high amplitude of the NF-κB, which supported their earlier hypothesis of robust NF-κB responses, even with relatively few cells responding. This model has been used by Fallahi-Sichani et al. (2012) as the intracellular basis of a hybrid model, by linking to an existing high-level agent-based model of macrophage responses to *Mycobacterium tuberculosis* (TB) infection. They surmise that through manipulating NF-κB mediated responses (particularly macrophage activation and TNFα expression), you can improve the function of a TB granuloma to contain the infection.

The fourth and final increment from the Lipniacki group, was developed by Pekalski et al. (2013), who built on the model of Tay et al. (2010) by adding the positive feedback associated with the expression of TNFα. Their premise was that the first phase of the innate immune system detects pathogens through membrane and cytoplasmic receptors, and that this leads to activation of transcription factors (such as NF-κB) and the production of proinflammatory cytokines, such as TNFα. The secretion of these cytokines, then leads to the second phase of the innate immune response in cells that have not yet encountered the pathogen; with these cytokine-activated cells producing and secreting the same cytokine, and thus propagating the immune response through this positive feedback. They discovered that the introduction of positive feedback changes system dynamics, and may lead to long-lasting NF-κB oscillations in wild-type cells and persistent NF-κB activity in A20-deficient cells.

Additional work using hybrid models at the single-cell level has been performed by Ashall et al. (2009) and Paszek et al. (2010). Ashall et al. built on the findings of Nelson et al. (2004) who had used real-time fluorescence imaging and computational modelling to show that the NF-κB system can be oscillatory to TNFα stimulation. Instead of using the computational modelling in a predictive capacity, Ashall et al. first commenced with wet-lab experimentation, and then developed both deterministic and semi-stochastic models to simulate the cellular behaviours. During the wet-lab experimental phase, they exposed individual cells to pulses of TNFα at various time intervals between the pulses, to mimic the pulsatile nature of inflammatory signals. They discovered that lower frequecy stimulations using 200min intervals generated synchronous translocations of NF-κB across the nuclear membrane of equal magnitude with successive pulses; whereas higher frequency stimulations using 60min or 100min intervals generated synchronous translocations with reduced magnitude for successive pulses. They suggest that this indicates a failure of the system to reset at these higher frequency intervals, and conjecture that this is due to the length of time associated

with the negative feedback by IκBα within the system. Following wet-lab experimentation, they then developed a new deterministic model, which modified the core network of IKK-NF-κB-IκBα along with the A20 inhibitor, that was the basis for Lipniacki et al. (2004). This new model was able to replicate the TNFα pulsatile stimulation data at the population-level, using a single parameter set. As the deterministic model was unable to elucidate the heterogeneity of single-cell responses to TNFα pulsatile stimulation, they augmented the model with stochastic processes for three negative feedback mechanisms: transcription of IκBα, transcription of A20, and delayed transcription of IκBϵ. Like the deterministic model, this semi-stochastic model was able to simulate the wet-lab experimental results, but also predicted: persistent oscillations of similar amplitude in both wild-type and IκBϵ-deficient cells after TNFα stimulation; and that stochastic variation due to the delayed transcription of IκBϵ may generate increased cell-to-cell heterogeneity in wild-type cells as compared to IκBϵ-deficient cells.

Paszek et al. (2010) subsequently built on this work of Ashall et al. by augmenting the transduction pathway of the computational model with the kinase IKKK, and calibrating the semi-stochastic behaviour to the results of their new single-cell (wet-lab) experimentation into the kinetics of IκBα and IκBϵ activation. Through the use of TNFα and IL-1β cytokines, and the synthetic stimulus PMA (phorbol ester differentiation factor phorbol 12-myristate 13-acetate), they discovered a transcriptional delay between IκBα and IκBϵ (under TNFα and IL-1β stimulation); with IκBα transcription increasing immediately after stimulation and reaching a maximum at 30min (under IL-1β stimulation), but showing no transcription for IκBϵ before 35min and a maximum at approximately 120min (under either TNFα or IL-1β stimulation). Furthermore, they discovered that PMA stimulation results in a 45min delay to transcription of both IκBα and IκBϵ genes, and that this delay did not substantially change the average timing or amplitude of NF-κB oscillations at the population-level, but instead affected the single-cell timing of the oscillations to maximise the heterogeneity of oscillation phasing between individual cells. They therefore hypothesise that the network topology of the NF-κB signalling system is stimulus-dependent, and that the generation of cellular heterogeneity maintains functional responsiveness of individual cells, to promote tissue robustness.

Through these separate iterations, the models of Lipniacki et al., Tay et al., Ashall et al. and Paszek et al. have now refined the initial Hoffmann model, and indeed investigated new areas of the signalling pathway from a computational perspective. We believe that the major achievements are: consideration of the different nuclear and cytoplasmic volumes; introduction of stochasticity regarding cell membrane receptor activation and gene activation, to facilitate modelling at the single-cell level; the use of more accurate data as parameter values; and perhaps most importantly, the rewriting of the model using Matlab, which is a more contemporary software development tool for computer scientists.

3.5.3 Stochastic Agent-Based Models

Although quantitative mathematical models are well established tools for modelling complex biological phenomena, they require an exhaustive set of precise parameters to be specified for each variable. This is fine for small models that align to dynamics of a few components, however when the scale is increased to capture a more realistic scope at the system-level (Acerbi et al., 2012), they begin to suffer from limitations in accuracy as data for use in analysing the effectiveness of differential equation models is often unavailable from wet-lab immunological experiments (Perelson and Weisbuch, 1997). As such, equation-based methods are unable to fully model system dynamics at the individual component-level, and therefore suffer considerably from their inability to capture the natural variation inherent to all biological processes. An alternative to these equation-based approaches is to model molecules or cells individually and assign probabilities to each possible interaction or state change through rule-based techniques. The individual component behaviours may then be aggregated up to system-level dynamics, which are then extrapolated in order to make predictions of the system-level behaviours in real biology (Cohn and Mata, 2007; Stark et al., 2007).

Following on from the early successes with equation-based computational models, Pogson et al. (2006) took the modelling of the pathway's stochastic behaviour one step further, by developing an agent-based model of the NF-κB signalling module. This computational model was developed in Matlab, and utilised the concept of communicating X-Machines (Barnard et al., 1996) to represent the individual agents and their associated interactions. They chose to break away from the pattern of using differential equations which relied on the assumption that a cell is homogenous with respect to its chemical constituents. We agree with this stance, and believe that such an assumption will not be valid due to the cells internal compartmentation, and non-uniform distribution of key molecules (Alberts et al., 1994). An important aspect of communicating X-Machines is that each agent has *memory*, which in this instance holds current physical location and current state, and may also contain a set of randomised parameter values (from a given distribution of suitable parameter values), to further instil stochasticity into the model. Pogson et al. used a set of autonomous agents within their model that correspond to: NF-κB dimer, IκBα, cell membrane receptor, nuclear importing receptor, and nuclear exporting receptor. Furthermore, they used a greater level of detail regarding translocation of molecules across the nuclear membrane, and added a 3-dimensional spatial dimension using continuous space to the model, which the differential equations used by Hoffmann et al. and Lipniacki et al. are unable to represent, and to a higher degree of granularity than the partial differential equations (PDEs) used by Terry and Chaplain (2011) and Ohshima et al. (2012) could represent.[17]

[17]Terry and Chaplain (2011) developed a 2D (X,Y cartesian co-ordinate) spatial model of a cell, where the movement of molecules by diffusion within the cytoplasmic and nuclear compartments, and translocation of molecules across the nuclear membrane, was explicitly considered. Through simulations using a variety of cell geometries, they discovered that sustained oscillations in NF-κB are robust to changes in the shape of the cell, or the shape, size and location of the nucleus within the cell. Ohshima et al. (2012) subsequently built on this work through development of a 3D spatial model. They discovered that the default kinetic parameters for previous ODE and PDE models were unable to replicate NF-κB oscillations,

The mathematical underpinning of the agent-based model by Pogson et al. is based on generalised biochemical reaction kinetics of two substrate molecules interacting and forming a product molecule. Graphical visualisation of the reactions allows us to view the increase of product (e.g. NF-κB-IκBα complex) and associated decrease in substrate molecules (e.g. free NF-κB and free IκBα) during an inhibitory reaction. As the relevant components are located within 3-dimensional space (X, Y, Z co-ordinates), the authors built in an interaction radius to each molecule to ensure a chemical reaction only occurs when the requisite substrate molecules come within a certain distance of each other. They abstracted the interaction radius of molecules to be a sphere; clearly, a number of biochemical components have orientations, through for example polarity and non-symmetrical quarternary structure, however we believe this to be an appropriate approximation to biology for the purposes of the model.

Simulations (using the model) run as a discrete-event system, with each iteration of the simulation run representing a defined period within time, i.e. constant time step. The authors validated the model against single-cell analysis wet-lab data of Carlotti et al. (1999, 2000) and Yang et al. (2003), which used fluorescent tagging and confocal microscopy to observe IκBα degradation and NF-κB shuttling dynamics. Successive time-step iterations were therefore able to relate directly to real-time single-cell analysis; and the model therefore reflects the discrete stochastic nature of subcellular events.

Careful review of the model has exposed six main limitations. It may be argued that three of these (points 3-5) may be due to the abstraction level taken (i.e. like the previous mathematical models, a very limited part of the pathway), however they are detailed here for completeness:

1. Only possibility of interaction between NF-κB and IκBα was considered, and occurs as *fait accompli*. The process of actual interaction (inhibition of NF-κB by IκBα) e.g. stochasticity of binding was not considered.

2. If two IκBα molecules are within the interaction radius of a NF-κB molecule, the model currently chooses which one to interact with at random. A more physically realistic decision making process would be beneficial, through for example interacting with the closest IκBα molecule.

3. Considers TIR-mediated activation of NF-κB signalling pathway, although due to the level of abstraction taken, this could be any receptor associated within the pathway. Furthermore, although the generic cell membrane receptors are modelled as agents, the corresponding extracellular signals are not, therefore activation of the receptors is assumed to have occurred at the start of simulations, and to have been uniform in signal strength, i.e. variation in ligand concentration and length of stimulation is not considered.

and that a new set of parameter values needed to be ascertained in order to calibrate this 3D model to system dynamics. They also discovered the system to be sensitive to the nuclear-to-cytoplasmic ratio of NF-κB and IκBα, the translocation coefficients, and the diffusion coefficients. Unfortunately, the 3D environment was not a true sphere (did not use continuous space), because in order to make the computation of local diffusion gradients tractable, they used a number of cubes that were oriented to approximate to a sphere.

4. No explicit transcription, translation and translocation of IκBα back to nucleus following translocation of free NF-κB to nucleus. Uses creation of a *temporary agent* with a time delay, which then changes state to IκBα in the cytoplasm.

5. At the beginning of simulations, following TIR activation, temporary agents are created with internal time delays to account for the cell signalling cascade upstream of the NF-κB signalling module (from receptor to IKK that triggers IκBα phosphorylation).

6. The model was scaled down (regarding the number of agents) to improve computation time. The model had approximately 1,000 molecules in total, whereas biological cells have approximately 60,000 NF-κB molecules and 180,000 IκBα molecules. This is justified however as the model was calibrated against wet-lab data, but does position the model at a highly abstracted level.

Pogson et al. later updated the original 2006 model and performed *in silico* experimentation to predict internal cell structural components in regulating the NF-κB signalling pathway (Pogson et al., 2008). This revised model made the IKK complex explicit (as an agent), and introduced the transcription and translation of the three IκB isoforms (IκBα, IκBβ, and IκBϵ) as per the Hoffmann et al. model. A detailed review of the scope and design of this model, has identified an additional limitation (over and above those described previously) regarding these isoforms however. Although Pogson et al. have introduced IκBβ and IκBϵ into this augmented model, these do not appear to have any function other than to introduce stochasticity in inhibition of NF-κB, i.e. an IκBα molecule within the NF-κB interaction radius will form a bond and inhibit NF-κB; IκBβ or IκBϵ within the interaction radius do not bind, but have created an effective *collision* in 3-dimensional space for the relevant time-step(s), so reduce the overall probability of NF-κB-IκBα complex formation. As per the 2006 model, this augmented model was also validated against the single-cell analysis data generated from the Qwarnstrom lab, and sensitivity analysis demonstrated a narrow range of acceptance for IKK levels, robustness to NF-κB and IκBα levels, and importantly greater sensitivity to IκBα than for NF-κB. Furthermore, simulation results yielded the negative feedback loop consistent with the Hoffmann et al. and Lipniacki et al. models.

The authors hypothesised that the actin cytoskeleton might be a depot for the excess IκBα within cells. They estimated the number of binding sites on the cytoskeleton that could be available for the interaction and then used the agent-based model to compare simulation results under different concentrations of free IκBα. Simulations of actin-IκB interactions yielded a maximal NF-κB-IκB complex formation at a 1:3 ratio NF-κB:IκBα. Without actin, the maximal was reached at a 1:1 concentration ratio, which confirmed previous work. The authors hypothesise that this key role for actin-IκBα interactions, sustains optimal pathway regulation by adjusting the NF-κB-IκB complex formation at the steady-state, and to control negative feedback following activation of the signalling pathway. Additional wet-lab experiments from the Qwarnstrom lab supported these findings and demonstrated a ratio of actin-bound to free IκBα

of 2:1 in unstimulated cells, indicating two-thirds of IκBα may be bound to the cytoskeleton at steady-state.

3.5.4 Peer-Validation of these Computational Models

Adverse effects of over-transfection was initially observed by Carlotti et al. (1999) and Yang et al. (2003), who compared control of exogenous, tagged intermediates, with results from immunocytochemical experiments to identify transfection limits in regards to cell-to-cell variations and regulation of NF-κB. This risk of over-transfection was leveraged by Barken et al. (2005) as a limitation of the work by Nelson et al. (2004). However, this debate has recently been resolved, and the initial observations by Nelson et al. confirmed, through the work of Tay et al. (2010) and Sung et al. (2014) on stable cell lines, and the work of Sung et al. (2009) and Zambrano et al. (2014a) on p65 EGFP knock-in transgenic animals.

We also believe that continued refinement and augmentation of these existing models will in all likelihood yield a computational model that will be able to accurately predict observed biological phenomena across both single-cell and averaged population levels, and suggest that the recent work of Tay et al. (2010) and Turner et al. (2010) has made an important contribution towards this end. They have both shown through single-cell analysis with TNFα, and associated stochastic modelling, that NF-κB activation is heterogeneous and is a digital process, with fewer cells responding at low doses. Additionally, Turner et al. also found that cells display analogue processing to modulate system dynamics, by controlling NF-κB peak intensity, response time, and number of oscillations.

Further work, building on the model by Hoffmann et al. has been performed by a number of groups around the world. Sung and Simon (2004) removed the IκBβ and IκBϵ components, before augmenting the model with a number of novel inhibitors to examine components within the signalling network that could be targeted for therapeutic interventions. They found the system to behave differently depending on which molecule is targeted within the signalling network, in particular by upstream inhibition (IKK activity and IκB degradation) producing similar system dynamics, whereas direct inhibition of NF-κB translocation resulted in distinct dynamics. Werner et al. (2005) focused on the temporal control of IKK activity, by looking at TNFα and LPS induction and the effects of A20 deficiency. Through k-means cluster analysis of the 627 distinct IKK activity profiles, they suggest that the NF-κB signalling module is particularly sensitive to sustained low IKK activity, but robust to transient perturbations. Furthermore, they predict that long-lasting NF-κB activity may be mediated by unexpectedly low amplitudes of IKK activity, whereas transient NF-κB activity requires a much higher increase in IKK activity. Hayot and Jayaprakash (2006) incorporated the Gillespie algorithm (Gillespie, 1977) into the model for the study of cell-to-cell variability, along with the replacement of the quadratic (second order) nuclear NF-κB-induced synthesis of IκBα with a linear (first order) term, following the reasoning of Nelson et al. (2004) and Lipniacki et al. (2004). This augmented model found the fluctuations of NF-κB to agree with Hoffmann et al. (2002), and predicted that fluctuations are small when transcription is strong (as used within the Hoffmann model), whether transcription is linear or quadratic, but that intrinsic fluctuations can be strong, and indeed as significant as any extrinsic noise

if the rate of promoter binding is small. They therefore advise that it is essential for transcription rate constants to be determined experimentally if models are to be developed using differential equations.

Additionally, the model has been used by other groups to validate their *in house* wet-lab experimental data, including: Ihekwaba et al. (2004, 2005), who performed a parameter refit for IκBβ and IκBϵ, and through using sensitivity analysis discovered the model is particularly sensitive to parameters relating to IKK and IκBα, and that overexpression of IκBα in transfected cells would yield a minimal perturbation to the system; Mathes et al. (2008) who added IκBα degradation functionality for both IKK-mediated degradation of NF-κB-bound IκBα, and IKK-independent degradation of free IκBα; and Wang et al. (2011) who adapted the model to simulate single-cell data using short and strong pulses of TNFα, and found that weak pulses provide a rich variety of non-linear temporal signalling dynamics, which may account for the diversity of expression patterns for the multitude of target genes in the signalling pathway.

Similarly, the model of Lipniacki et al. (2004), has also been used with data from other labs, including: Yue et al. (2006) who performed a sensitivity analysis, focusing in particular on local sensitivities through bifurcation/pairwise studies; Joo et al. (2007) who propose that the IKK-NF-κB-IκBα-A20 interaction network is the key portion of the wider pathway, as these components have critical kinetic rate variables for IKK activation; and Ashall et al. (2009) who treated cells with short pulses of TNFα to mimic pulsatile inflammatory signals, and predicted that negative feedback loops regulate both the resetting of the system and cellular heterogeneity. See figure 3.8 for a diagrammatic representation of the evolution and chronology of these computational models.

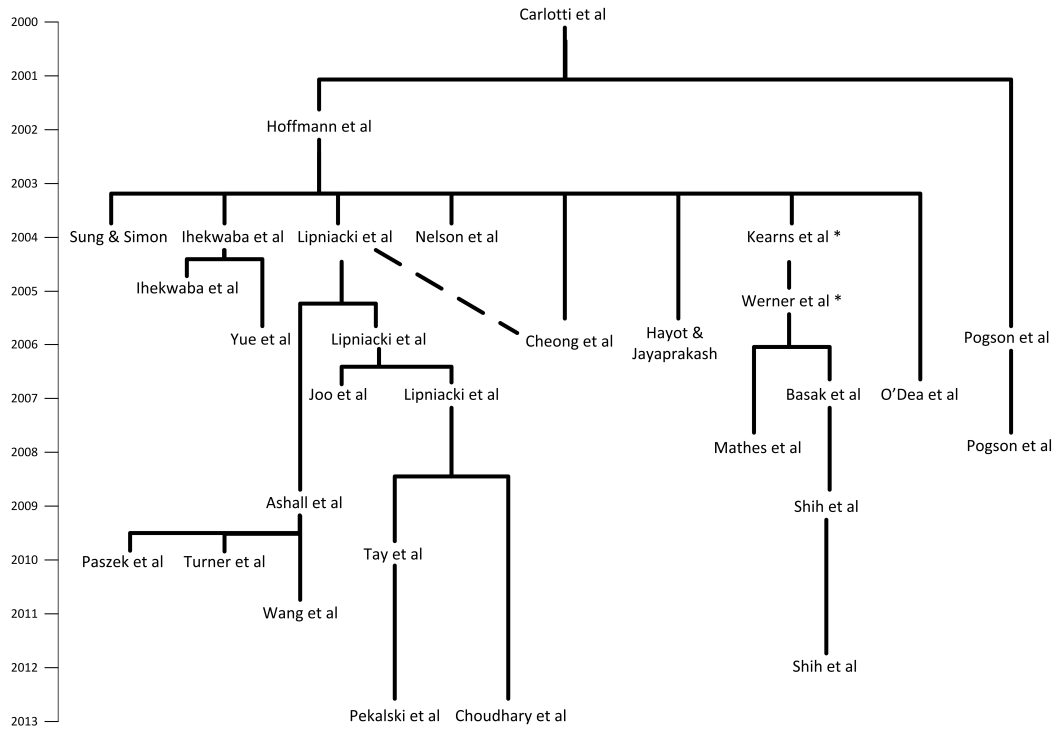

Figure 3.8: Dendrogram of the existing computational models of the NF-κB signalling pathway. It can clearly be seen that all roads lead from the nuclear-cytoplasmic shuttling model of Carlotti et al. (2000), however the ODE-based model of Hoffmann et al. (2002) is generally considered to be the first computational model involving the wider signalling pathway. It can also be seen that the Hoffmann model has been the basis for considerable reimplementations and developments over the decade since it was first published, with a number of groups using the equation-based approach as the basis for their modelling. More recently however, Pogson et al. (2006) have used an agent-based approach to model the spatio-temporal aspects of the system, and to allow the agents (instantiations of individual NF-κB and IκB molecules) to possess stochastic behaviour. * Werner et al. (2005) actually builds upon the work of Kearns et al. (2006) - we therefore assume that the Kearns paper was submitted first, but took longer to become published, hence the non-intuitive order (with respect to publication year) in the dendrogram. The dotted line from Cheong et al. (2006) represents the fact that although this model was an extension of Hoffmann et al. (2002), it also took the updated parameter values from the work of Lipniacki et al. (2004).

3.5.5 Minimal Models

The models discussed above, have represented mathematical and computational abstractions of the major components of the NF-κB signalling pathway. It has recently been argued by a number of different groups however, that the complex nature of these models makes them computationally expensive to run, and also requires complicated analysis techniques in order to make inferences from simulation results. As such, these groups have developed what have been termed *minimal models* that are able to replicate the majority of the phenomenological behaviours of the more complex computational models, but use the minimum number of equations possible. We therefore discuss within this section the four minimal models from Krishna et al., Yde et al., Longo et al., and Zambrano et al., along with the model reduction algorithm developed by West et al.

To the best of our knowledge, the first minimal model after that of Carlotti et al. (2000) was developed by Krishna et al. (2006), and focused on a small core network of the pathway that drove oscillatory behaviour. Through reducing the model of Hoffmann et al. down to a core feedback loop of three coupled ODEs, they were able to ascertain the minimal model required to generate oscillations. This simplified model was validated against the work of Hoffmann et al. and Nelson et al., and was able to simulate: the sustained oscillations obtained with only the IκBα isoform of IκB; the damped oscillations in wild-type cells that included the other forms of IκB (e.g. IκBβ and IκBϵ); the spikiness of nuclear NF-κB and asymmetry of cytoplasmic IκB oscillations; and the phase difference between NF-κB and IκB. Their key finding was that saturation degradation of cytoplasmic IκB in the presence of IKK was crucial for oscillatory behaviour, because it sets an upper limit within the system for the degradation rate, and thus allows IκB to accumulate and remain in the system for longer than with linear degradation rates. They conjecture that this effectively introduces a time delay into the negative feedback loop, which is known to generate oscillations.

The authors were subsequently involved in the development of another minimal model (Yde et al., 2011). Here they modelled amplification of the immune response, following cytokine mediated activation of the NF-κB pathway, and the positive feedback that occurs through upregulation of cytokine expression - as previously modelled by Werner et al. (2005). The ODE model was developed at the tissue-level (cell population level), with the detailed network dynamics of the NF-κB signalling pathway within individual cells abstracted away. In fact, they discovered that tissue-level immune responses were able to emerge within the minimal model through using only three (high-level) variables, relating to: NF-κB, a generic regulator, and a generic inhibitor. The model predicts that cytokines produced by the stimulated cell(s) at the site of infection, diffuse away from this primary infection site and trigger the transient response of the NF-κB pathway for the production of cytokines in neighbouring cells. They conjecture that this generates a *propagating wave* of NF-κB induction and cytokine production throughout the infected tissue.

Like the minimal model of Krishna et al. (2006) above, Longo et al. (2013) also used the Hoffmann model as the basis for developing a minimal model of negative feedback within the NF-κB signalling pathway. This new model used a single delayed compound-reaction, to replace the cascading reactions that had been utilised previously (Hoffmann et al., 2002). They used this to investigate

the underlying mechanisms involved with oscillations and the dampening of oscillations that emerged through the two negative feedback loops of the inhibitors IκBα and IκBϵ. Through investigating the oscillatory responses using the single negative feedback associated with IκBα, they discovered that both the frequency and decay rate of the oscillations are highly dependent on the internal parameters of the core network of the signalling module, but are not sensitive to extracellular stimuli levels. They suggest that the oscillatory frequency within the system can therefore not be encoding information about the stimulus, and conjecture that stimulus-specific gene expression is therefore unlikely to be determined by stimulus-specific frequencies of NF-κB oscillations, and therefore may involve amplitude modulation over time. Furthermore, upon introduction of IκBϵ into the model, the authors were able to reproduce the findings of their previous work (Kearns et al., 2006), where the oscillations caused by persistent stimulation were now significantly dampened. They suggest that the second negative feedback mechanism (IκBϵ), may have evolved to produce dampening of the oscillatory behaviour of the first feedback mechanism (IκBα).

The three minimal models above all utilised the technique of reducing the total number of equations by developing a smaller set of compound equations that were grouped according to sequences of reactions within the signalling network. An alternative approach has been taken by Zambrano et al. (2014b) however, whereby the ODE-based model is developed according to the different *layers* of processes within the cell environment. They have used a three layer approach, with the first layer representing the transcription and translation of new IκBα molecules; the second layer representing the NF-κB signalling module and its negative feedback by IκBα; and the third layer representing the commencement and propagation of signal following the recognition of extracellular stimuli at the cell membrane, down to activation of IKK. Their model was able to replicate both the heterogeneity of the system seen in the models of Tay et al. (2010), Paszek et al. (2010), and Sung et al. (2009), along with the spiky oscillations shown by Krishna et al. (2006). As such, the authors believe that the community should make more use of minimal models as a basis for investigating the underlying mechanisms of the NF-κB signalling pathway, and that the reliance on complicated computational models should be minimised because they incur the risk of being overfitted to the specific wet-lab experiments from individual groups, and thus may not be used by the wider community who use different *in vitro* models.

Finally, along with the explicit development of minimal models by individual groups, recent work by West et al. (2014) has focused on the development of algorithmic methods to enable a principled approach for reducing existing differential equation based models into new minimal models. Their algorithmic approach is based on quasi-steady-state approximations, and produces a set of ranked variables according to how quickly they approach their momentary steady-state. Once ranked, the developer may then eliminate variables at each step within the related equation-based model, whilst preserving the system-wide dynamics. The authors tested their reduction algorithm on the two feedback (IκBα and A20) model of Ashall et al. (2009) and the original ODE model by Krishna et al. (2006) (upon which the minimal model was extracted), and report that the system dynamics that emerge from the reduced models compare favourably with the original, more complex models.

3.5.6 The Need for a New Computational Model of NF-κB

The complexity of the innate immune system, and in particular the inflammatory process has been difficult to fully investigate using reductionist and linear approaches alone, since it is characterised by non-linear kinetics as well as numerous feedback loops. Several signalling pathways relating to the combating of various pathogenic infections converge on NF-κB activation, resulting in a highly complex regulatory system for the innate immune response. We therefore believe the NF-κB signalling pathway to be a good candidate for research that follows a systems biology approach to further our understanding. Computational modelling to date, has successfully complemented traditional wet-lab techniques to generate hypotheses on the mechanistic behaviour of the various components of the signalling pathway (see table 3.2 for the major computational modelling advancements). As such, these *in silico* derived predictions are helping to further our understanding of *in vitro* system dynamics, in particular, how signals received by the IKK complex are propagated down the signal transduction pathway through activation of the NF-κB signalling module. This computational experimentation has given rise to the belief that IKK regulatory mechanisms may represent sensitive clinical targets in diseases with aberrant innate immune system activity. There appears to be opportunity to expand the scope of current models however, to also encompass upstream events prior to IKK activation, thus incorporating a greater degree of granularity for cell membrane receptor complexes in order to capture the induction of the pathway through various extracellular signals, and to include downstream events to capture actual activation of transcription for target genes (other than IκB isoforms) that are relevant to the inflammatory response, for example various cytokines.

NF-κB has wide-ranging effects controlled by a complex regulatory network of inhibitors and co-activators. Understanding the mechanisms that control NF-κB activation/cellular signalling is important for exploiting therapeutic approaches to treat human disorders due to its dysregulation. Specific targets for therapeutic agents could be the transcription factor itself, or any of the associated components within the pathway, including protein kinases and the IκB inhibitors themselves. Gilmore and Herscovitch (2006) reviewed the known inhibitors of NF-κB, which may provide a basis for future research regarding computational modelling, and subsequent pharmacological intervention.

We therefore believe that computational models tightly coupled to wet-lab experimental analysis will be indispensable to furthering our understanding of the NF-κB signalling pathway. Furthermore, with the advances in software modelling technologies and computational power over the past decade, we believe that the timing is right for development of agent-based models of the NF-κB signalling pathway on a scale that has not been seen before. New technologies, such as Java Mason (Luke et al., 2005), which allows large agent-based simulations to be developed that can be manipulated in real-time during simulations, or FLAME (Coakley et al., 2006b, 2012), which has recently been used to model the European economy, and the GPU version of FLAME (Richmond et al., 2009), allow massively parallel agent-based simulations to be run. Along with these advances in technology, recent developments in cell biology have also advanced our corresponding understanding behind the pathway. We believe a number of key questions may be answered through development of large-scale computational models

of NF-κB. From a cell biology perspective, these relate to research objectives 2, 4, 5 and 6 of this thesis, and are: (i) what can *in silico* experimentation tell us about the relative roles of the intermediate components within the signal transduction events; (ii) what can *in silico* experimentation tell us about the various receptors, co-receptors, and adaptor proteins, and their role(s) in signal transduction events; (iii) what can *in silico* experimentation tell us about the dysregulation that can occur (regarding NF-κB activation and IκBα degradation and resynthesis) when cells are in *diseased* states, and how can we perturb the system back to a *healthy* state; and (iv) what can *in silico* experimentation tell us about the cross-talk that occurs when various extracellular signals (e.g. LPS, TNFα, IL-1) and other environmental stimuli (e.g. bacteria, UV radiation, phsyical stress) converge on the NF-κB signalling module. There are also a number of key questions from a modelling perspective, which relate to research objectives 2, 3 and 5 of this thesis, and are: (i) does abstraction level affect the accuracy of simulation-driven predictions; (ii) does the resolution level (e.g. number of agents) affect the accuracy of simulation-driven predictions; (iii) does the simulation platform (different technology, e.g. equation-based versus agent-based) affect the accuracy of predictions; (iv) what are the relative merits of averaged population data versus single-cell data for calibration and validation of computational models; and (v) what are the advantages and limitations of using massively parallel computing architectures, for reproducing the large-scale variation (as seen in biological systems) into computational simulations.

3.6 Summary

This chapter has provided the background to the NF-κB signalling pathway in general; has highlighted the central role that NF-κB plays in a large number of human diseases and illnesses; and has defined our *Domain* of interest (as per the CoSMoS process), which is the IL-1 stimulated NF-κB signalling pathway. Additionally, the chapter has reviewed the existing computational models of the NF-κB signalling pathway, and has identified the need for a new computational model specific to IL-1 stimulation, which uses the agent-based modelling paradigm.

The next chapter will define our domain model of the IL-1 stimulated NF-κB signalling pathway, which will be used as functional specification for our computational model. The chapter will also reflect on the use of UML and statistical techniques for domain modelling.

Year	Authors	Modelling Paradigm	Stimuli	Cell Type	Level	Pathway Components	Key Advances
2000	Carlotti et al. (2000)	Deterministic	IL-1	Monkey Smooth Muscle	Single-Cell	NF-κB, IκBα	First known model of NF-κB and IκBα dissociation and translocation dynamics.
2002	Hoffmann et al. (2002)	Deterministic	TNFα	Mouse Fibroblast	Population	IKK, NF-κB, IκBα, IκBβ, IκBϵ	First model of TNFα induced activation of the signalling pathway.
2004	Lipniacki et al. (2004)	Deterministic	-	-	Population	IKK, NF-κB, IκBα, A20	Incorporated 2-feedback loop (IκBα and A20) to Hoffmann model and parameter refit.
2004	Nelson et al. (2004)	Deterministic	TNFα	Human HeLa and SK-N-AS	Single-Cell	IKK, NF-κB, IκBα	First model to show NF-κB oscillations at single-cell level. Augmented Hoffmann model using single-cell data for calibration and validation.
2006	Pogson et al. (2006)	Agent-Based	IL-1	-	Single-Cell	TIR, IKK, NF-κB, Nuclear Membrane Transporters	First known agent-based model of signalling pathway.
2006	Lipniacki et al. (2006)	Semi-Stochastic	TNFα	-	Single-Cell	IKK, NF-κB, IκBα, A20, mRNA, genes	Stochastic transcription to generate cellular heterogeneity.
2006	Cheong et al. (2006)	Deterministic	TNFα	Mouse Embryonic Fibroblast	Population	IKK, NF-κB, IκBα	Incorporated temporal profiles of IKK activation.
2006	Kearns et al. (2006)	Deterministic	TNFα	Mouse Embryonic Fibroblast	Population	IKK, NF-κB, IκBα, IκBβ, IκBϵ	Reimplementation of Hoffmann model in Matlab. Showed that IκBϵ provides negative feedback control of NF-κB oscillations.
2007	Basak et al. (2007)	Deterministic	LTβR	Mouse Embryonic Fibroblast	Population	IKK1, IKK2, NF-κB, IκBδ	Incorporated fourth IκB inhibitor (nfkb2 p100, or IκBδ). Models cross-talk of canonical NF-κB/RelA activity in response to non-canonical IKK1-induction.
2007	O'Dea et al. (2007)	Deterministic	TNFα	Mouse Embryonic Fibroblast	Population	IKK, NF-κB, IκBα, IκBβ, IκBϵ	Distinguished between NF-κB bound and *free* IκB pools. Investigated steady-state regulation of NF-κB signalling module.
2007	Lipniacki et al. (2007)	Semi-Stochastic	TNFα	-	Single-Cell	TNFR1, IKKK, IKK, NF-κB, IκBα, A20, mRNA, genes	Incorporated stochastic switches for cell membrane receptor activation by TNFα ligand, and transcription of IκBα and A20 genes.
2008	Pogson et al. (2008)	Agent-Based	IL-1	Human HeLa	Single-Cell	TIR, IKK, NF-κB, IκBα, genes, cytoskeleton, Nuclear Membrane Transporters	Updated earlier ABM with transcription and translation of IκBα to provide negative feedback. Also incorporated sequestration of excess IκBα to cytoskeleton.
2009	Shih et al. (2009)	Deterministic	TNFα, IL-1, LPS	Mouse Embryonic Fibroblast	Population	IKK, NF-κB, IκBα, IκBβ, IκBδ, IκBϵ	Modelled the 4 distinct IκB variants and utilised TNFα, IL-1 and LPS stimulation to determine signal specificity for the negative feedback loops.

Continued on next page

Table 3.2 – continued from previous page							
Year	**Authors**	**Modelling Paradigm**	**Stimuli**	**Cell Type**	**Level**	**Pathway Components**	**Key Advances**
2009	Ashall et al. (2009)	Semi-Stochastic	TNFα	SK-N-AS and Mouse Embryonic Fibroblast	Single-Cell	IKK, NF-κB, IκBα, IκBϵ, A20, genes	Incorporated delayed stochastic transcription of IκBϵ, and stochastic transcription of IκBα and A20.
2010	Tay et al. (2010)	Semi-Stochastic	TNFα	Mouse Fibroblast	Single-Cell	TNFR1, IKKK, IKK, NF-κB, IκBα, A20, mRNA, genes	Updated to reflect the heterogeneous, digital response of single cells, and the analogue dynamics of peak NF-κB intensity, response time and oscillation number, to modulate the overall population response.
2010	Paszek et al. (2010)	Semi-Stochastic	TNFα	Mouse Embryonic Fibroblast	Population	IKKK, IKK, NF-κB, IκBα, IκBϵ, A20, mRNA, genes	Aggregated large-scale single-cell dynamics to show that cellular heterogeneity (for timings of NF-κB oscillations) is important for population-level robustness.
2010	Turner et al. (2010)	Semi-Stochastic	TNFα	SK-N-AS	Single-Cell	TNFR1, IKK, NF-κB, IκBα, A20, mRNA, genes	Updated Ashall model to incorporate stochastic processes for IKK activation.
2012	Fallahi-Sichani et al. (2012)	Semi-Stochastic	TNFα	Macrophage	Population	TNFR1, IKKK, IKK, NF-κB, IκBα, A20, mRNA, genes	Merged their previous ABM of granuloma formation (did not model NF-κB) with the ODE model of Tay, to develop a multi-scale hybrid model.
2013	Choudhary et al. (2013)	Deterministic	TNFα	Human Epithelial	Population	TRAF1, NIK, TRAF2, NF-κB, IκBα, IκBδ, A20, mRNA, genes	Integrated canonical and non-canonical pathways, using TRAF1-NIK as a feed-forward complex.
2013	Pekalski et al. (2013)	Semi-Stochastic	TNFα	3T3	Single-Cell	TNFR1, IKKK, IKK, NF-κB, IκBα, A20, mRNA, genes, TNFα	Built on Tay model to integrate negative and positive feedback loops due to IκBα and A20, and TNFα respectively.

Table 3.2: The top 20 computational models (in chronological order) that have facilitated our increased understanding of the NF-κB signalling pathway. Along with the year of journal article publication, we have also documented: the modelling paradigm used (deterministic, semi-stochastic, or agent-based), the specific extracellular stimulus (if any), the specific cell type (if any), the hierarchical level of the model (single-cell or population), the pathway components that were explicitly modelled, and the key advances of the model.

Part II

Design and Development of the Agent-Based Model

4 Domain Model of IL-1 Stimulated NF-κB Signalling Pathway

As discussed previously, biological systems are complex, with behaviours and characteristics that result from a highly connected set of interaction networks that function through time and space. The first product developed during a project following the CoSMoS process is the *domain model*. The domain model documents the complex behaviours and characteristics of the system, and as stated by Andrews et al. (2010): *"encapsulates understanding of appropriate aspects of the domain into explicit domain understanding"*. Whereas the domain is the general area of study, which in this case is the IL-1 stimulated NF-κB signalling pathway, the domain model is the description of the domain that reflects the scientific basis for development of a simulation. By its very nature, the domain model is therefore an abstraction of the area of interest, and should capture at an agreed level, the scientific details of relevance to the future computational model. This chapter will define our domain model of the IL-1 stimulated NF-κB signalling pathway, and also explore the role of diagrammatic and statistical techniques for domain modelling, to fulfill research objective 1.

4.1 Overview of Domain Modelling

The iterative process of domain modelling helps the modeller to explore the biological domain in conjunction with one or more *domain experts* before development of the computational model. Once complete, and validated by the domain expert(s), the domain model acts as the functional specification for the computational model, and provides a comprehensive and transparent understanding of the domain that underpins the *in silico* experimentation performed as part of the exploration phase of a CoSMoS project. As such, the domain model is an essential project deliverable that provides an audit trail on how the real-world biology is linked, through abstractions, assumptions, and constraints, to the functionality of the computational model.

Although domain models of complex biological systems may require many diverse sources of information, the actual domain modelling process may be categorised into either the data-driven or deep-curation approaches. The data-driven approach utilises techniques from bioinformatics and relies on the network topology, reaction kinetics and other model parameters being inferred from large high-throughput datasets, whereas the deep-curation approach develops a domain model through manual integration of knowledge from published literature, databases and empirical data gained from more traditional, non-omic scale experiments (Ghosh et al., 2012).

The data-driven approach uses computational algorithms to infer causal relationships among system components from high-throughput experimental data.

These high-volume datasets are usually collected under varying experimental conditions where functionality of components (both individual and groups of) are systematically perturbed (Schadt et al., 2009). Once the network topology is inferred from the datasets, it is then developed further into a domain model that focuses on the mechanisms of interactions between the components (Ghosh et al., 2012). This approach is expected to be useful for developing computational models over multiple scales, as the algorithms are able to find causal relationships both within a given hierarchical level (e.g. gene-gene or protein-protein) and between hierarchical levels (e.g. gene-protein).

The deep-curation approach also requires generation of a detailed topology of system components and their interactions, but instead of generating these automatically from high-throughput datasets, relies on manual creation from readily available data in the form of journal articles, databases, and publicly available datasets (Ghosh et al., 2012). Although potentially more time-consuming than the data-driven approach, deep-curation does have the advantage that modellers can add their own mechanistic details to the network structure, to act as hypotheses on how the various components may interact for the generation of system-wide behaviours. This, along with the need for precise and detailed knowledge of component interactions for the design and development of computational models, makes the deep-curation approach more relevant to the needs of modellers.

The domain model may be a collection of informal notes relating to relevant aspects of the domain, but may also include informal sketches (such as cartoons), more formal diagrams (such as those produced with UML), mathematical equations, scientific constants (e.g. biochemical rate constants), and physical descriptors (such as size, quantity, location, and speed). The key constraint of the domain model is that it should remain free from an implementation specific focus and should therefore not contain any reference to the programming languages or workarounds which may be required during development of the simulator. As such, the domain model should be focused on the scientific domain, and not design considerations for the resulting computational model and simulator. One final point is that the domain model must be developed at the correct abstraction level to answer the questions of scientific interest; for example, the single-cell data relating to NF-κB pathway dynamics, means that the initial model should be at the level of subcellular interactions and biochemical reactions within a single cell.

4.2 UML and Domain Modelling

UML has been used by a number of groups to semi-formally define domain models relating to biological systems of interest. For example, Webb and White (2005) used UML to model the biochemical reactions within a cell that relate to sugar metabolism, through the glycolytic metabolic pathway. They developed the model using class inheritance and containment diagrams, sequence diagrams, containment cartoon diagrams, state machine diagrams, and cartoon activity diagrams. More recently, Read et al. (2009a, 2014) used a number of diagrammatic techniques to develop a domain model of experimental autoimmune encephalomyelitis. They used a cartoon diagram to capture the cell interaction network; an expected behaviours cartoon diagram to capture the phenomena that we as scientists may observe from the biological domain, with cell interactions

that are hypothesised to occur in order to produce these observable phenomena; tables to document the timings of key events within the domain; and when using UML focused their attention to class, activity and state machine diagrams.

We have chosen to follow the approach of Webb and White (2005) and Read et al. (2009a, 2014) in using UML as the basis to semi-formally define the domain model of of our biological system. Along with a number of UML diagrammatic notations, we have also chosen to deviate from the somewhat rigid framework and have used less formal cartoon diagrams where necessary to ensure the biological meaning can be conveyed efficiently. As advised above, the domain model may be developed using conventions outside of diagrammatic ones, and where necessary we have used a number of mathematical and statistical approaches to complement the diagrammatic models of the NF-κB signalling pathway.

4.3 The Domain Model

This section presents a domain model of the IL-1 stimulated NF-κB signalling pathway. The first subsection focuses on the use of diagrammatic domain modelling techniques, largely expressed using UML, whereas the second subsection focuses on the use of statistical techniques for modelling the dynamic system-wide observable behaviours.

The current state of scientific understanding of the domain dictates that certain aspects of the signalling pathway are unknown, or contested, within the literature. The process of developing a rigorous and comprehensive domain model highlights these areas of inconsistency within the literature, and gaps in the scientific knowledge, which must be overcome through the use of documented assumptions of the domain.

There exists substantial quantities of literature on the NF-κB signalling pathway, and various aspects of the pathway are independently studied by a wide variety of labs. It is generally understood that representing every aspect of a real-world system in models and simulations is computationally intractable. As such, a subset of the properties and behaviours from the real-world system need to be defined for subsequent investigation. One of the primary purposes of the domain model is to capture this subset of real-world system properties, and therefore provide a definition of the abstraction level taken for the modelling project.

The various iterations of the domain model are validated by the domain expert, facilitating a common understanding that underpins the subsequent *in silico* experimentation, to ensure simulation results are interpreted within the correct biological context. As such, the domain model is a key CoSMoS project deliverable, that provides an auditable link between the simulation work and the real-world domain. As discussed in chapter 2, the platform model, which will be presented in the next chapter, provides additional details (from a technical perspective) that define how the domain model will be implemented using suitable programming frameworks to yield the actual computational model (i.e. the simulation platform).

The present domain model represents the most recent iteration, and has focused on the subset of signalling components investigated by the Qwarnstrom lab. Focus has been applied to the observations of NF-κB and IκBα dynamics from Carlotti et al. (1999, 2000) and Yang et al. (2001, 2003).

4.3.1 Diagrammatic Domain Modelling

The diagrammatic model has been developed using the deep-curation approach, with particular attention paid to the literature surveyed in chapter 3 and personal communication with the domain expert. Initial focus will be the emergent system-wide behaviours of the pathway, before increasing the level of detail to the interactions between system components, and then the dynamics of individual components. The diagrammatic domain model is presented in a top-down manner, comprising three levels of abstraction, as defined below:

1. A system-level overview of the domain model. This highly abstract level provides an outline of the biology of the NF-κB signalling pathway. Particular focus is made to the behaviours of the system following induction by extracellular signal, and how these are believed to correspond to phenomena observed in the real-world domain. This abstraction level of the domain model does not make use of UML, but instead utilises less formal cartoon diagrams to convey system-wide properties.

2. Modelling component-level interactions of the domain model. This medium level abstraction, decomposes the IL-1 stimulated NF-κB signalling pathway into its constituent molecular components. This level models an abstracted view of the key biochemical and molecular interactions between the components, that together give rise to the emergent behaviours of the system. A cartoon diagram, along with UML activity, class, and sequence diagrams have been used in modelling these component-level interactions.

3. Modelling individual component dynamics. This level of abstraction provides the greatest detail within the domain model, through modelling the dynamics of individual components within the system. UML state machine diagrams have been used in expressing these models.

Modelling System-Level Properties

Since its discovery by the laboratory of the nobel laureate David Baltimore in 1986, there has been a prolific amount of published research into the NF-κB signalling pathway, with focus on: the NF-κB dimer; its inhibitors; the extracellular stimuli inducing the pathway; and indeed the associated genes that are upregulated. For example, a PubMed Central literature search in March 2014 indicated that there were over 49,000 articles and 4,500 reviews published on NF-κB. It is computationally intractable to represent every aspect of the real world domain in computational models, and therefore a subset of the system needs to be defined in order to represent an abstracted view of the system.

Following the approach of Andrews et al. (2010) and the example of Read et al. (2009a), we have chosen to commence the domain modelling process with a cartoon-like diagram (i.e. not UML), termed an *expected behaviours* diagram (see figure 4.1). This diagram depicts the observable phenomena of the IL-1 stimulated NF-κB signalling pathway, along with the known interactions between system components that generate system-wide behaviours. The diagram also provides us with an opportunity to define a number of hypotheses on how these

known component interactions may yield the observable phenomena. The expected behaviours diagram therefore provides a diagrammatic view of the relationship between the real-world domain and the domain model (Read, 2011).

The top section of figure 4.1 defines the observable phenomena of the IL-1 stimulated NF-κB signalling pathway, in that the pathway results in an inflammatory response against extracellular stimuli, and that after a period of time, this inflammatory response ceases. The dotted horizontal line demarcates these observable phenomena from hypotheses that are believed to be responsible for their emergence. These hypotheses consist of expected behaviours (boxes annotated with '<<expected>>' tags) that emerge through the interactions of the underlying system components. The known interactions between system components are represented through a set of solid, directed lines, whilst the expected behaviours are linked to these system components through a set of dashed lines.

As discussed in chapter 3, wet-lab experimental research into NF-κB since its discovery, has identified that a large number of inflammatory signals (extracellular stimuli) activate cell membrane receptors to initiate its signalling pathway. It is believed that signal transduction through the intracellular network, via activation of various intermediate signalling components, amplifies the signalling cascade so that a short, transitory burst of stimuli, may result in a more prolonged immune response, through transcription of inflammatory genes and translation of the corresponding mRNA into inflammatory proteins. Furthermore, it is also believed that one of these inflammatory response proteins, is the inhibitor IκBα, which results in negative feedback to dampen the inflammatory response.

The intracellular interactions outlined above, and within figure 4.1, are purposefully defined at a very abstract level of detail in order to provide a simplified system-level overview of the IL-1 stimulated NF-κB signalling pathway. Increased levels of detail, along with an examination of how the expected behaviours may emerge from the underlying intracellular component interactions, will follow within the subsections on 'modelling component-level interactions' and 'modelling individual component dynamics'.

The spatial relationships of the components detailed within the expected behaviours diagram can be seen in the cartoon diagram of figure 4.2, which provides an abstract representation of a Eukaryotic cell. For the purposes of modelling the IL-1 stimulated NF-κB signalling pathway, the cellular structure can be abstracted away to contain just three containment structures: the cell membrane, which for our purposes contains the cell membrane receptor and co-receptor proteins; the cytoplasm, which contains the cytosol (intracellular fluid) that further contains the adaptor proteins, intermediate signalling components, NF-κB, and IκBα, and the mRNA generated from gene transcription; and the nucleus, which contains DNA, its nuclear membrane, which houses the nuclear membrane transporter proteins involved in translocation, and the NF-κB and IκBα that have been translocated from the cytoplasm.

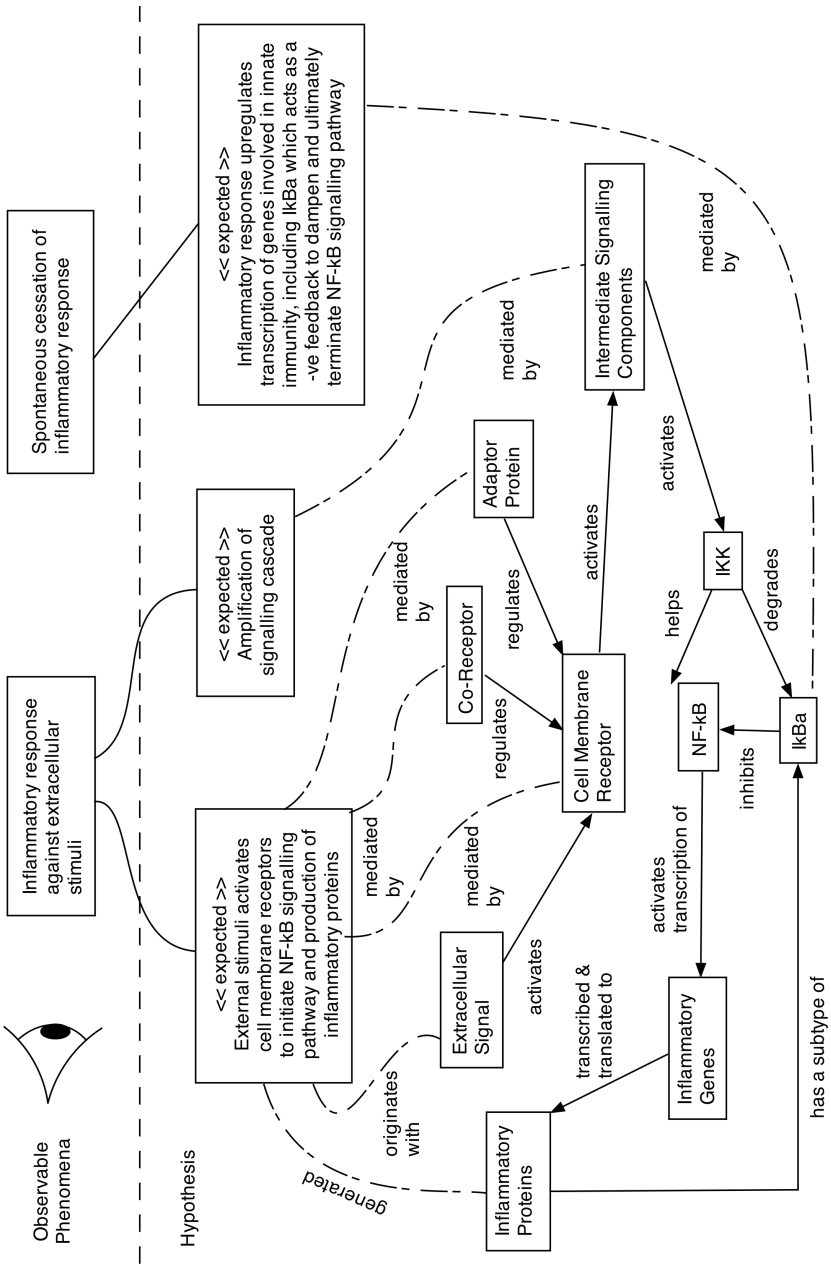

Figure 4.1: Expected behaviours diagram depicting the observable phenomena of the IL-1 stimulated NF-κB signalling pathway; the behaviours that are hypothesised to be responsible for these phenomena; and at an abstracted level the components of the complex system that are believed to be responsible for the development of these emergent behaviours. At the highest level of the system, activation of the NF-κB pathway initiates a transitory inflammatory response (the system dynamics automatically cease the response). It is hypothesised (expected) that these phenomena occur through interaction of three functional modules which relate to activation of cell membrane receptors, amplification of the signalling cascade, and upregulation of transcription. Developed from the reviews of Baldwin (1996), Baeuerle and Henkel (1994) and Brasier (2006).

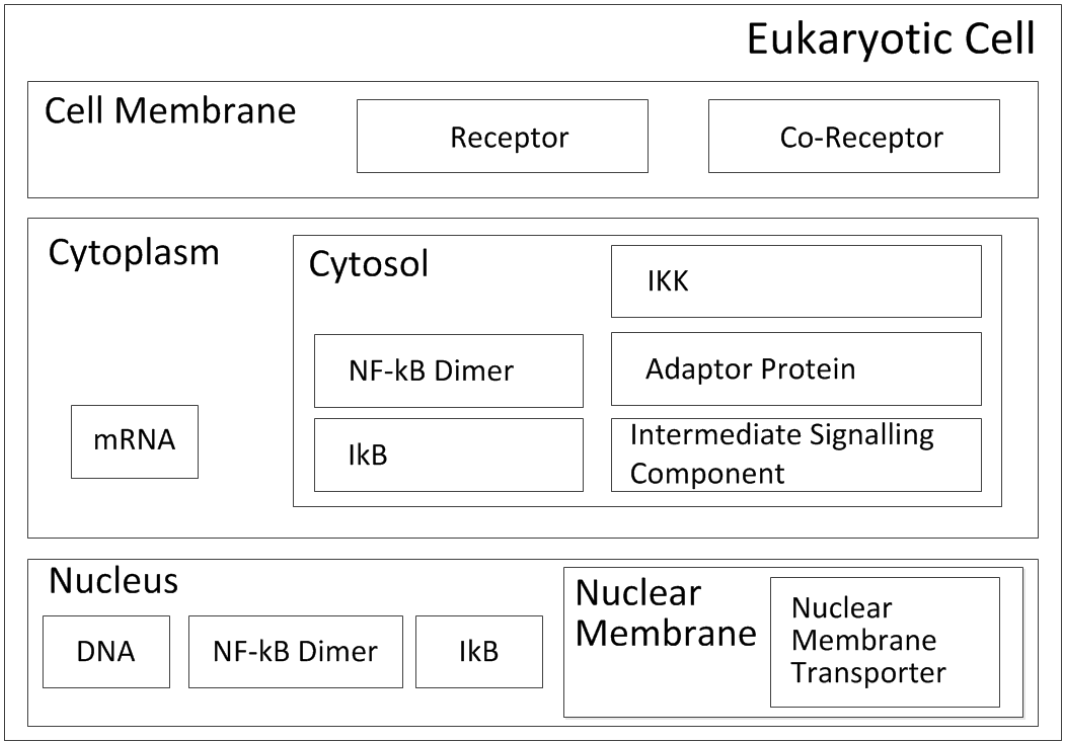

Figure 4.2: Cartoon-like containment diagram showing the components involved in the NF-κB signalling pathway and the physical environment in which they are situated within a Eukaryotic cell. Developed from the review of Hoffmann and Baltimore (2006).

Modelling Component-Level Interactions

We believe that the quintessential diagrammatic notation within science broadly, and biology in particular, is the cartoon diagram. The NF-κB signalling pathway can be described from a high-level perspective using such a convention to communicate the interactions between system components, and in this instance the diagram can also act as a network map and illustrate the sequence of interactions between components (see figure 4.3). As discussed in chapter 3, following an extracellular ligand (signalling molecule) binding to a member of the TLR or IL-1R superfamily, the receptor dimerises, and co-receptors such as CD14 and MD2 (in the case of TLR4) help facilitate and amplify the receptor response. In situations where the Tollip adaptor protein binds, it mediates association of IRAK to IL-1RI, but inhibits IRAK. Conversely, in situations where the MyD88 adaptor protein binds, it mediates association of the receptor complex with IRAK protein kinase, which in turn activates TRAF6 through phosphorylation. Once activated, TRAF6 continues signal transduction through activation of TAK1, which subsequently activates the IKK complex. The activated IKK complex may then dissociate the IκBα molecule from the inhibited NF-κB complex. The IκBα molecule is then degraded by the proteasome (not shown), whereas the released NF-κB is able to translocate from the cytosol to the nucleus, where it is subsequently activated and upregulates the transcription of target genes. Furthermore, if TILRR binds to (the extracellular portion of) IL-1RI, it potentiates (enhances) the recruitment of MyD88, which leads to amplification of the signalling cascade.

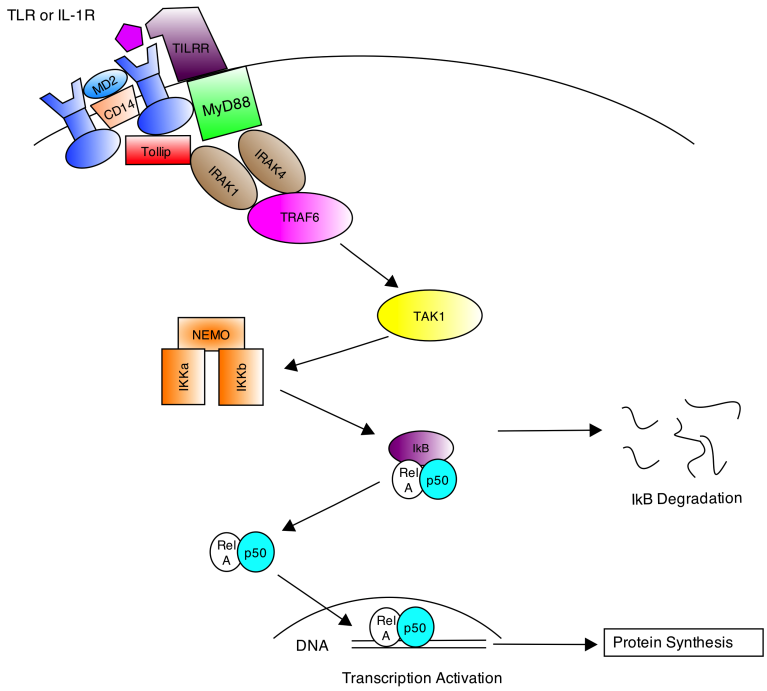

Figure 4.3: Simplified diagram depicting the high-level interactions between the TLR or IL-1R superfamily of receptors, the co-receptors and adaptor proteins, and the protein kinases within the NF-κB canonical signalling pathway. Diagram developed from findings of Akira and Takeda (2004); Burns et al. (2000); Doyle and O'Neill (2006); Kawai and Akira (2006); Zhang et al. (2010).

These relationships between components can be described in more detail using the UML class and sequence diagram notations (see figures 4.4 and 4.5 respectively). The class diagram defines the relationships between components within the system, and can be used to convey aggregation and inheritance. Indeed, development of the class diagram facilitated discussion on the relationships between signalling pathway components responsible for the initiation and propagation of the inflammatory response. An example of aggregation within the system is that the inhibited NF-κB-IκBα complex contains an NF-κB dimer along with an IκBα molecule. Similarly, an example of inheritance within the system is that the exporting and importing nuclear receptors are both sub-types of a more generic nuclear receptor for translocation of molecules across the nuclear membrane. The sequence diagram, represents another view of the system by explicitly defining the order of events within the signalling pathway.

Both of the diagrams commence with activation of the cell membrane receptor complex by extracellular stimuli. The receptor complex is then able to bind to adaptor proteins, which mediate the association of the complex to IRAK[1]. Furthermore, if a TILRR co-receptor is part of this cell membrane receptor complex, an amplification of the receptor response occurs through the increased recruit-

[1]Note, that as the domain model is an abstract representation of the real-world domain, we have purposefully chosen to omit the detailed biochemical reactions inherent to the binding of molecules, such as phosphorylation of specific amino acid residues, or conformational changes of the protein's structure, and have instead chosen to consider the high-level relationships between molecules.

ment of MyD88 adaptor protein. IRAK may then propagate the signal through a cascade of intermediate molecules along the signalling pathway (abstracted to contain TRAF6 and TAK1 in our domain model), before activation of the kinase IKK. Active IKK may then bind to the inhibited NF-κB-IκBα to form the trimeric complex IKK-NF-κB-IκBα, where it acts to facilitate degradation of the IκBα molecule[2] and release *free* NF-κB. The released NF-κB dimer may then move freely around the cytoplasm, and if it moves into the interaction boundary of an importing nuclear receptor, may bind to the receptor, for subsequent translocation into the nucleus. Once in the nucleus, the NF-κB dimer may become activated, bind to the promoter region of a target inflammatory response gene, for transcription of the gene into its corresponding mRNA sequence, and subsequent translation into the inflammatory response protein[3]. As discussed previously, one of the first target genes is that of the inhibitor IκBα, and following the generation of new IκBα molecules in the cytoplasm, they are able to translocate to the nucleus (via the importing nuclear receptors), bind to the activated NF-κB complex to reinhibit the molecule and release it from the promoter region of DNA, to stop transcription of any more mRNA (resulting in negative feedback). The released NF-κB-IκBα molecule is then translocated out of the nucleus and back into the cytoplasm via an exporting nuclear receptor, for completion of the signalling pathway.

Although the UML class diagram (figure 4.4) and sequence diagram (figure 4.5) provide useful notations for defining the overall relationships between components and associated events within the NF-κB signalling pathway, they lack the ability to show the concurrency of activities inherent to the system. Specifically, the UML sequence diagram incorrectly provides the perception that all molecules of a give type (e.g. IRAK molecules) are in the same biological state at any given moment in time. As such, the implication is that once the receptor complex has been activated, that all IRAK molecules are activated at the same time, and that the cascade of reactions is performed in a synchronised manner for all molecules involved. This could not be further from the truth, as complex biological systems invariably have molecules that are out of phase (i.e. differing biological states) to other molecules of the same molecular type. For example, any given NF-κB dimer within this signalling pathway may be in either inhibited, free, or active states, and if free, may also be located in either the cytoplasm or the nucleus. Fortunately, the UML standard contains another notation which may help convey the sequence of activities for any given parts of the system.

[2]Note that chapter 3 advised IKK functions to dissociate the IκBα molecule from the NF-κB dimer, and that the proteasome subsequently degrades it following polyubiquitination. For the purposes of our domain model however, we have chosen to asbtract away some of these details in the UML class diagram, and have merged the dissociation by IKK, polyubiquitination and degradation by the proteasome into a single degradation step mediated by IKK.

[3]The processes of transcription and translation are complicated, requiring a large number of other components than those described above. As the focus of our computational model will be the dynamics relating to NF-κB activation and IκBα degradation, we have chosen to abstract away the details into simple one-step transcription and translation processes.

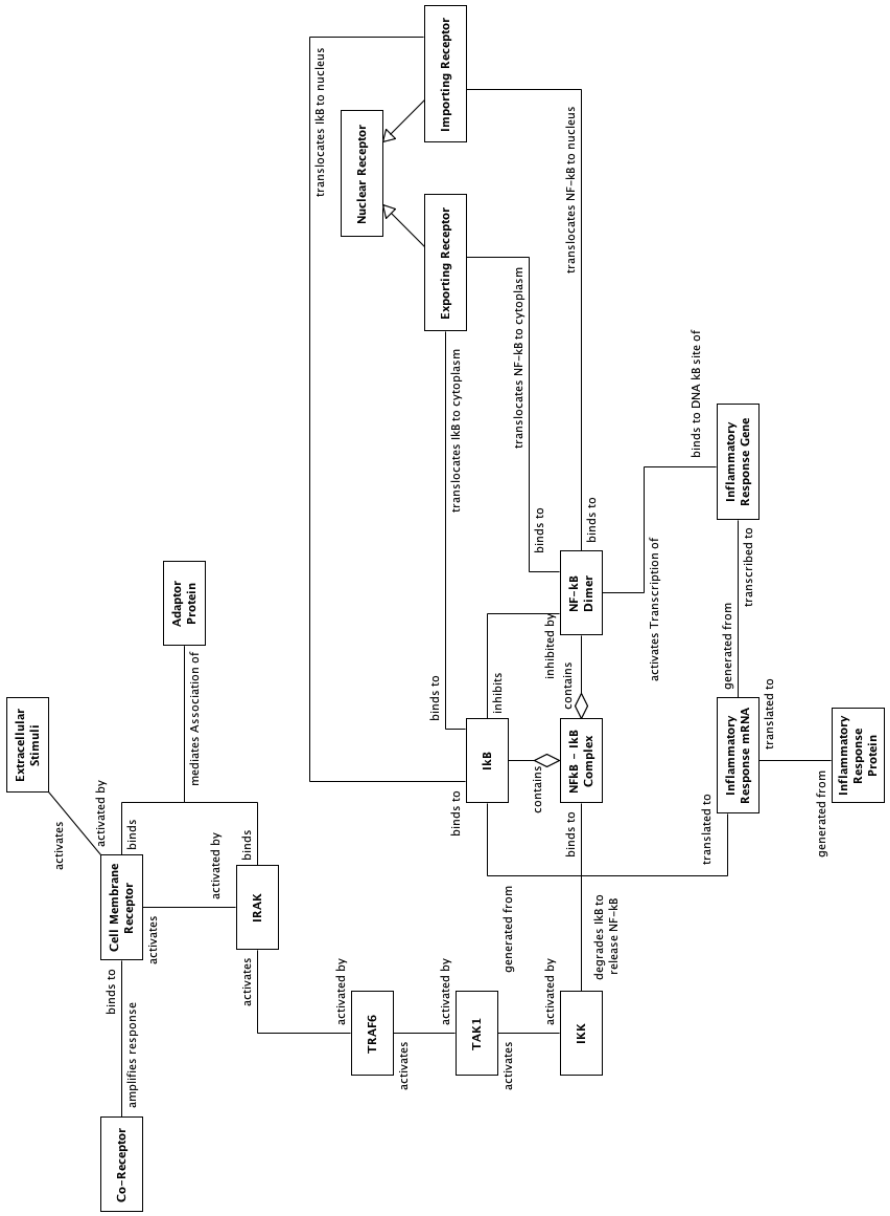

Figure 4.4: UML class association diagram for the IL-1 stimulated NF-κB signalling pathway. The layout follows the spatial aspects of the cascade of component relationships involved in the biological system from cell membrane activation down to gene transcription within the nucleus. The diagram conveys directed relationships by the labels at each end of the connectors. The arrowheads pointing towards *Nuclear Receptor* convey inheritance, such that *Exporting Receptor* and *Importing Receptor* are sub-types, and the unfilled diamonds on the NF-κB-IκB complex denote aggregation. Developed from the reviews of Hayden et al. (2006) and O'Neill and Dinarello (2000).

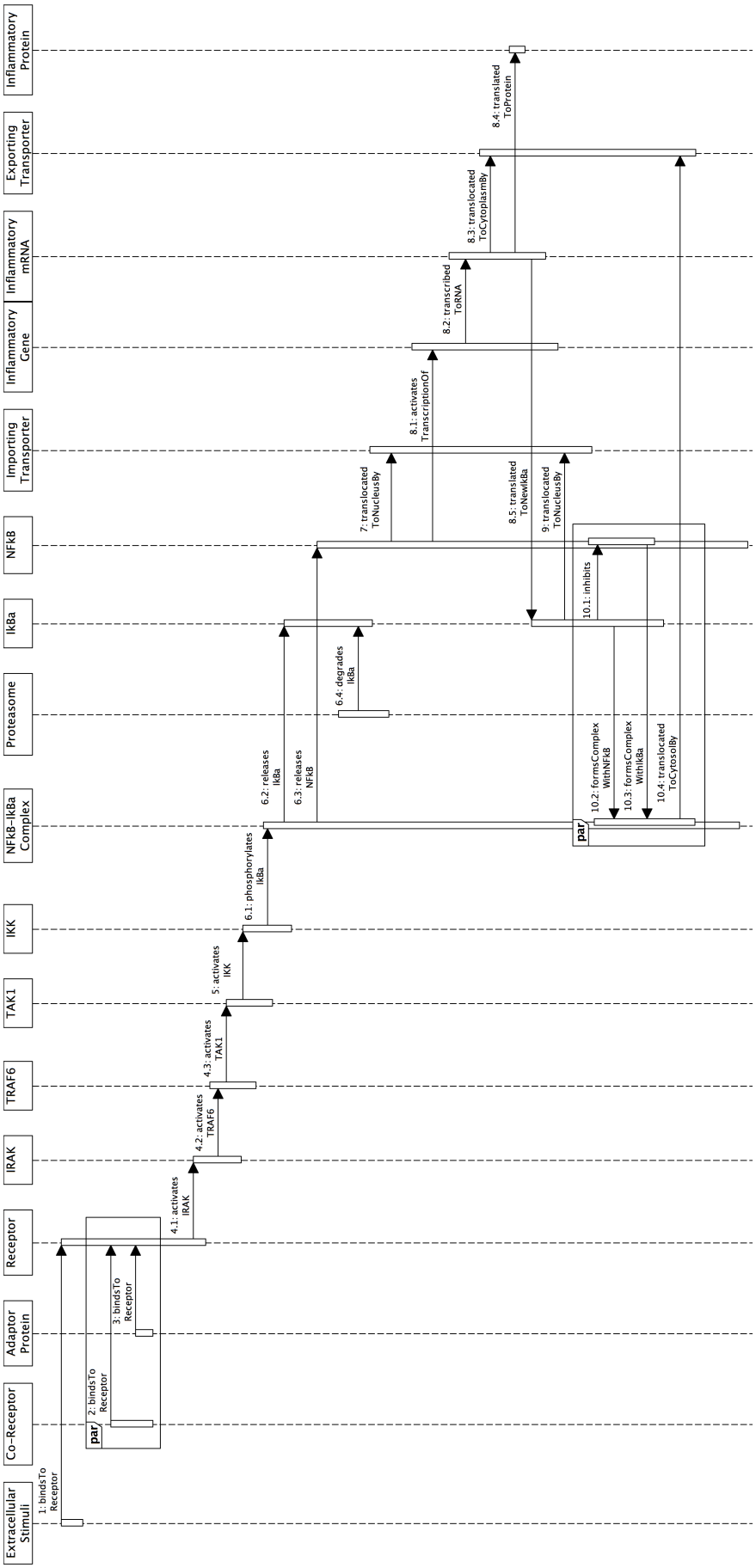

Figure 4.5: UML sequence diagram for the IL-1 stimulated NF-κB signalling pathway. The diagram commences with extracellular stimuli binding to the cell membrane receptor and sequentially moves through each of the steps before transcription of inflammatory genes and resultant translation into inflammatory response proteins, culminating in inhibition of NF-κB again by IκBα. Developed from the reviews of Hayden et al. (2006) and O'Neill and Dinarello (2000).

As UML activity diagrams focus on activities and not components, they are able to convey the individual interactions (expressed as activities) that give rise to the emergent behaviour of the system. Furthermore, the focus on activities allows us to aggregate sets of individual interactions into functional modules. As discussed in chapter 3, the NF-κB signalling pathway can be logically separated into three key functional modules relating to cell membrane receptor activation, activation of the NF-κB signalling module, and generation of new IκBα to dampen the response through negative feedback regulation. The activities inherent to the functional modules are modelled using UML activity diagrams in figures 4.6, 4.7 and 4.8 respectively.

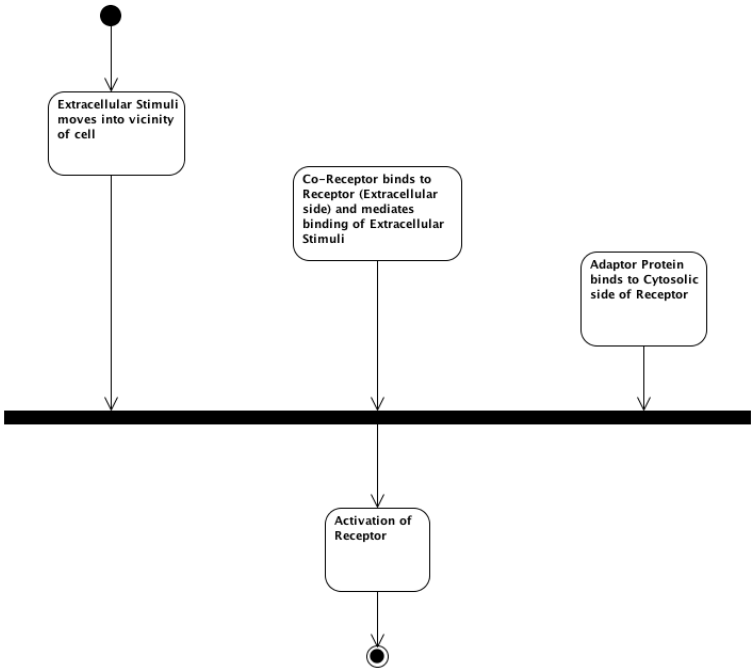

Figure 4.6: UML activity diagram of cell membrane receptor activation within the IL-1 stimulated NF-κB signalling pathway. Developed from the review of O'Neill and Dinarello (2000).

As per the UML sequence diagram in figure 4.5, activities within the system begin with extracellular stimuli and the formation of the active receptor complex (figure 4.6). The associated signal transduction then follows, with the first activity being the activation of IRAK, which then propagates the stimuli-related signal through the pathway via phosphorylation of intermediates. Upon phosphorylation of the IκBα inhibitor which is bound to NF-κB, the activity splits into two branches: a) phosphorylated IκBα is released from the NF-κB complex and becomes degraded via the proteasome, and b) the NF-κB dimer is released, binds to an importing nuclear receptor, is translocated from the cytosol to the nucleus, and is then activated (figure 4.7). Once active, the NF-κB dimer may bind to the promoter region of an inflammatory response gene and initiate transcription, which ultimately generates new inflammatory response proteins. As we are only concerned with generation of new IκBα within our research, we ignore generic inflammatory response proteins, and only model (in the UML) the activities that culminate in the negative feedback regulation of the pathway (figure 4.8). As shown in figure 4.9, UML also allows multiple activity diagrams to be linked together, and the swim-lane convention provides us with an opportunity to convey the biological location of these activities.

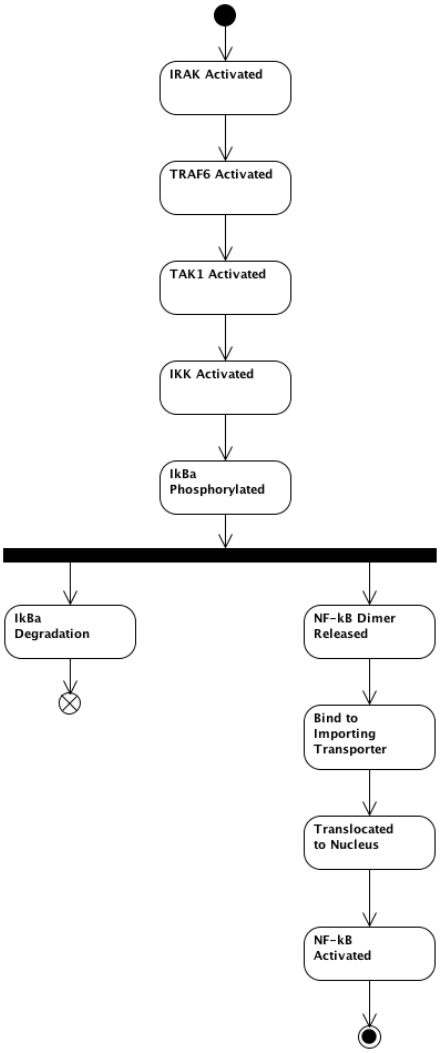

Figure 4.7: UML activity diagram for activation of the NF-κB signalling module. Developed from the review of Hayden et al. (2006).

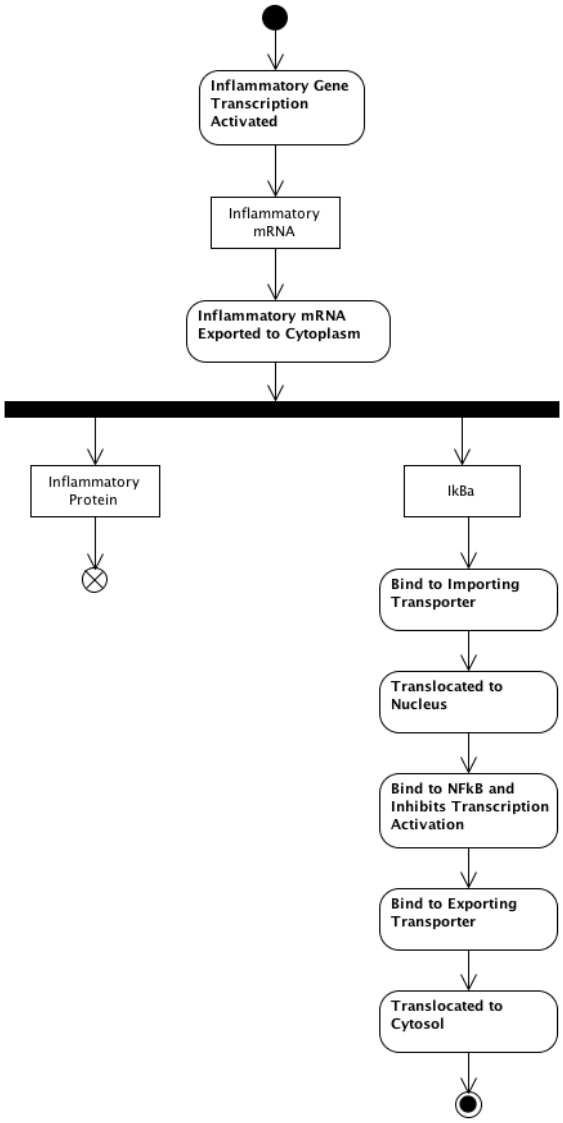

Figure 4.8: UML activity diagram for inflammatory gene transcription and translation, which culminates in the generation of new IκBα inhibitor proteins, for reinhibition of the NF-κB complex. Developed from the review of Brasier (2006).

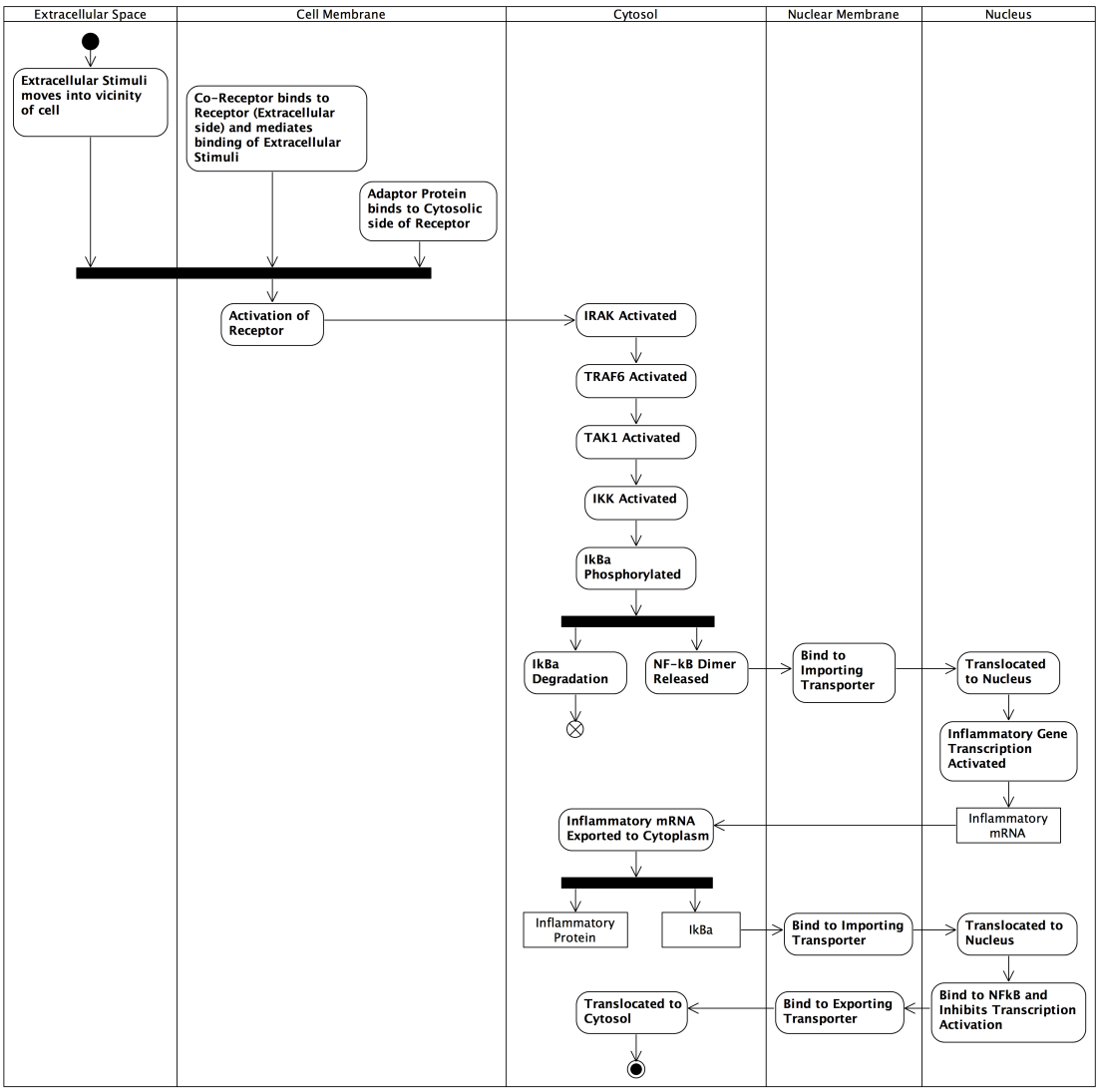

Figure 4.9: Full UML activity diagram for the NF-κB signalling pathway using the concept of swim-lanes to convey sub-cellular location of components. Developed from the reviews of O'Neill and Dinarello (2000), Hayden et al. (2006), and Brasier (2006).

Modelling Individual Component Dynamics

The final set of UML diagrams that represent the domain model, refer to low-level dynamics of individual components and use the UML state machine diagram notation. Figures 4.10 to 4.16 represent the state machine diagrams for the flow of components within the NF-κB signalling pathway, from the cell membrane receptor, through to the intermediate components (including IKK), IκBα, NF-κB, nuclear membrane transporters, and the inflammatory gene and mRNA that give rise to the inflammatory response.

Figure 4.10 defines our abstracted view of the cell membrane receptor within the signalling pathway. It can be seen that the receptor initially starts off in a *dormant* state, but may become *active* upon binding of co-receptor and MyD88 adaptor protein (defined using an adapted version of UML guard notation), along with subsequent binding of the external stimuli. Conversely, and as discussed previously, the cell membrane receptor may also become *inhibited* upon binding of Tollip. As defined in the previous cartoon and UML diagrams, following activation of cell membrane receptor, the extracellular signal is propagated through the signalling pathway through activation of intermediate components, culminating with activation of IKK. For the purposes of the domain model, we have abstracted away the IRAK, TRAF6, TAK1 and IKK components to that of a *generic* intermediate component, which by default is *dormant*, but becomes *active* following phosphorylation as the signal is propagated through the transduction cascade (figure 4.11).

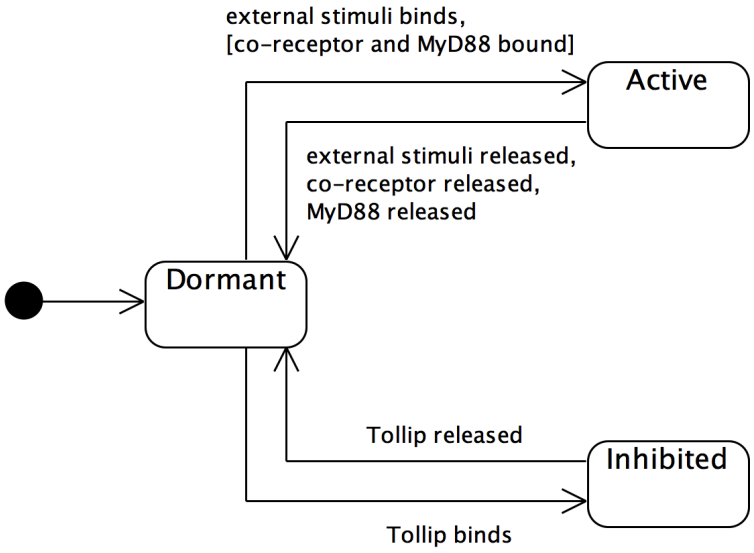

Figure 4.10: UML state machine diagram for a cell membrane receptor involved in the NF-κB signalling pathway Burns et al. (2000) and O'Neill and Dinarello (2000).

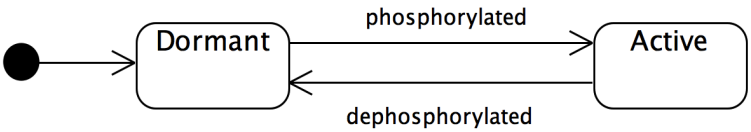

Figure 4.11: UML state machine diagram for a *generic* intermediate component involved in the propagation of signal from the cell membrane receptor to NF-κB.

Within the NF-κB signalling module, the IκBα inhibitor molecule by default (i.e. following creation through transcription and translation) is unbound (*free*), but may probabilistically bind to NF-κB when it enters an interaction boundary, to enter an *inhibiting* state. Following activation of the IKK enzyme, and its binding to the inhibited NF-κB-IκBα complex, the IκBα becomes phosphorylated and releases the NF-κB dimer, to again enter the *free* state. This time however, the free IκBα undergoes polyubiquitination and becomes degraded by the proteasome (figure 4.12). Similarly, the NF-κB dimer, is by default in an *inhibited* state within the system due to IκBα inhibition (figure 4.13). Following phosphorylation of IκBα, it becomes *free*, and able to translocate to the nucleus where it may become *active* (note the guard condition), to facilitate the upregulation of inflammatory gene transcription. Should the NF-κB dimer spontaneously unbind from the promoter region of the inflammatory gene, it will once again enter the *free* state, or alternatively new IκBα molecules (this time within the nucleus) may also bind, to return the NF-κB dimer to an *inhibited* state.

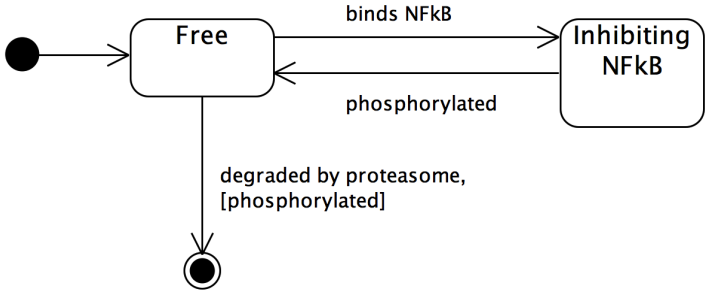

Figure 4.12: UML state machine diagram for the inhibitor IκBα within the NF-κB signalling pathway. By default, IκBα is *free* (indicated by the blacked-out circle), however it can probabilistically bind to NF-κB, thus entering an *inhibiting NF-κB* state. Upon phosphorylation by IKK, it returns to a *free* state for degradation by the proteasome, and removal from the system (indicated by the concentric circle symbol). Developed from Baeuerle and Baltimore (1988a) and Karin and Ben-Neriah (2000).

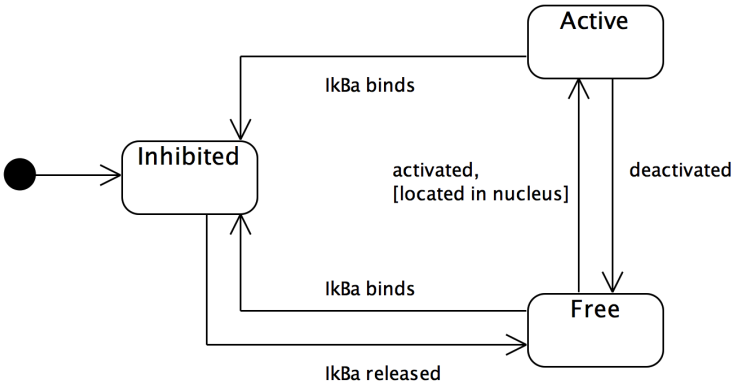

Figure 4.13: UML state machine diagram for the NF-κB dimer. By default, within our abstracted view of the system, NF-κB is in an *inhibited* state. Following IKK-mediated degradation of IκBα, it enters a *free* state, and if located within the nucleus may become *activated*. If it becomes bound again to IκBα, it again enters an *inhibited* state, or alternatively it may naturally dissociate from the DNA sequence to become *free*. Developed from Baeuerle and Baltimore (1988a,b) and Siebenlist et al. (1994).

As per the UML class association diagram (figure 4.4), sequence diagram (figure 4.5) and activity diagram (figure 4.9), binding of an IκBα molecule or NF-κB dimer to a nuclear membrane transporter, transitions the transporter protein from a *dormant* to an *active* state for the translocation of the ligand from either the cytoplasm to the nucleus, or vice versa (figure 4.14). Following translocation of an NF-κB dimer to the nucleus and its binding to the the promoter region of an inflammatory gene, the gene transitions from a *dormant* to an actively *being transcribed* state (figure 4.15) for generation of mRNA. Upon creation, the new mRNA is translocated to the cytoplasm (see figure 4.5), where it is *translated* into new inflammatory protein by the ribosome (figure 4.16).

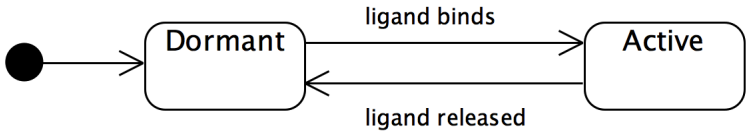

Figure 4.14: UML state machine diagram for the nuclear membrane transporter within the NF-κB signalling pathway.

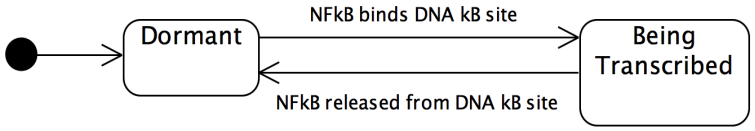

Figure 4.15: UML state machine diagram for a *generic* gene within the NF-κB signalling pathway.

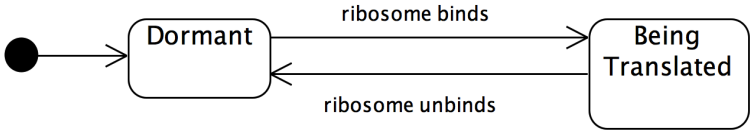

Figure 4.16: UML state machine diagram for a *generic* mRNA molecule within the NF-κB signalling pathway.

Modelling Numerical Aspects of the System

The three different views outlined above provide a top-down perspective of the NF-κB signalling pathway, which reflects the hierarchical nature of complex systems per Oltvai's pyramid of life (Oltvai and Barabasi, 2002), discussed in chapter 3. The UML and cartoon-like diagrams used so far, have been useful for semi-formally defining the relationships and dynamics at the system, component, and intra-component levels, however they have not been able to appropriately capture the numerical aspects of the signallng pathway.

The diagrammatic technique of *mindmapping* was therefore used to convey a number of detailed aspects of the domain that could not be modelled above (see figure 4.17). This mindmap documents (within a single diagram) a number of key rates, ratios, and physical attributes associated with the NF-κB signalling pathway. For example, we have found UML to be deficient in its ability to model details regarding the ratios of NF-κB molecules (in free and inhibited states) against free IκBα molecules across the cytoplasmic and nuclear compartments. Similarly, we have been unable to convey nuclear translocation dynamics or details of IκBα degradation within UML in a form that would be intuitive to biologists. The following points from Carlotti et al. (1999, 2000) and Pogson et al. (2008) are key to the activation and propagation of the signalling pathway:

- The cytoplasmic to nuclear ratio of NF-κB is 10:1 under non-stimulated conditions
- There is approximately 17% free NF-κB (not bound to IκBα) in resting cell
- The 'bindable' NF-κB to IκBα is approximately 1:1
- The total NF-κB to IκBα is 1:3 with the excess sequestered to the actin cytoskeleton
- Under IL-1 stimulation, there is a 20% decrease in cytoplasmic NF-κB; a 40-fold increase in nuclear NF-κB (moves from 1:10 to 4:1 location ratio); and an 8-fold increase in transfected versus endogenous NF-κB
- There are approximately 60,000 RelA (NF-κB) molecules in an endogenous cell; approximately 66,000 IκBα molecules (10% excess with respect to NF-κB); and approximately an additional 135,000 IκBα molecules in an endogenous cell, which are sequestered to the actin cytoskeleton
- The approximate cell volume of a fibroblast cell is 2000 μm^3. The approximate volume of the nucleus is 100 μm^3
- Phosphorylation of IκBα molecules peaks at 10min, and ubiquitination peaks at 30min post IL-1 stimulation
- IκBα degradation rates are approximately as follows: 40% degraded after 10min, 60% after 30min, and 80% after 60min

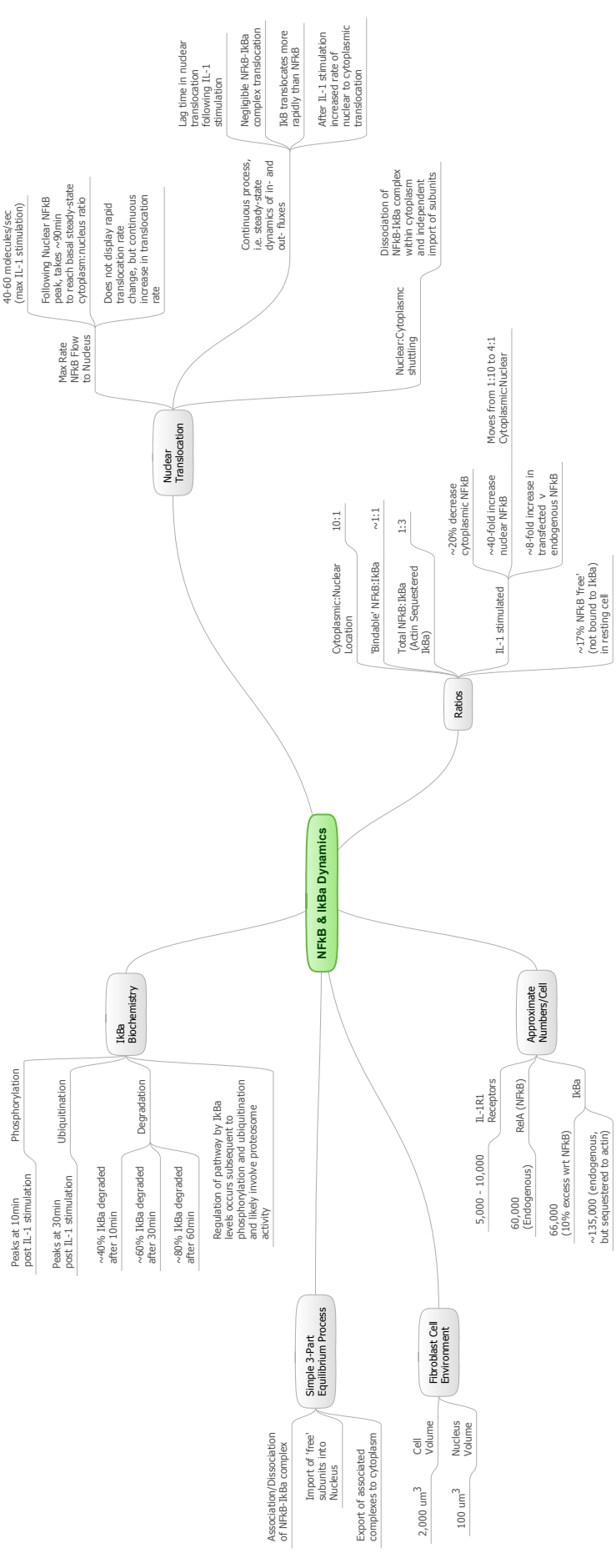

Figure 4.17: Mindmap of the key rates, ratios and constants of the IL-1 stimulated NF-κB signalling pathway. The mindmap provides key details for the cell environment, approximate number of key molecules, IκBα biochemistry, nuclear translocation, and ratios of NF-κB and IκBα molecular states within the cytoplasm and nucleus. Developed using Carlotti et al. (1999, 2000) and Pogson et al. (2008).

4.3.2 Statistical Analysis to Complement the UML Domain Model

The use of cartoon and UML diagrams are an essential first step towards development of a domain model, however we believe that in isolation they are not enough to provide a comprehensive model. In particular, they have been unable to convey the dynamics of IκBα degradation (along with the associated NF-κB release and subsequent activation), or indeed model the quantitative aspects of the signalling pathway. We have therefore used a number of univariate and multivariate statistical techniques to complement the UML diagrams, in order to develop a more comprehensive domain model of the NF-κB signalling pathway.

Initial Univariate Statistical Analysis

As discussed within chapter 3, a degree of variance is inherent to all aspects of biology due to the underlying stochastic physiological events of individual cells. The single-cell fluorescence data of Yang et al. (2001, 2003) encapsulates this stochasticity, and we feel that this should be explored further within the domain model. For predictive purposes, and indeed to indicate the types of statistical tests that should be used in the future (and which graphical representations are best suited), it is often desirable to understand the shape of the underlying distribution of the data. To determine the underlying distribution, it is common to fit the observed data to a theoretical distribution by comparing the frequencies observed in the data to the expected frequencies of the theoretical distribution. We have used Chi-squared (χ^2) goodness of fit tests (see Appendix A) to ascertain that the single-cell analysis (IκBα degradation) fluorescence data (at time 0min) of Yang et al. (2001) approximates to a Negative Binomial distribution, which we believe follows the usual patterns in biology of variation due to stochasticity (Bliss and Fisher, 1953; White and Bennetts, 1996; Tijskens et al., 2003).

The data contained measurements from single-cell analysis performed on 88 cells: 52 were transfected with IκBα Enhanced Green Fluorescent Protein (EGFP) and stimulated with IL-1; and 36 were transfected with IκBα-EGFP, but not stimulated with extracellular ligand, thus representing a *control* group. Single-cell analysis occurs on live cells, with the same set of cells being followed over time. All measurements within the data related to cytoplasmic fluorescence and were taken over a period of one hour, at intervals corresponding to 0, 10, 30 and 60 min. Figure 4.18 represents the control data with integer binning and a superimposed curve that follows the negative binomial distribution with median average of 1.947153 (calculated from the control data). Similarly, figure 4.19 represents the IL-1 stimulated data with integer binning and a superimposed curve that follows the negative binomial distribution with median average of 1.729876 (calculated from the IL-1 stimulated data).

The purpose of this statistical domain modelling is to document the statistical summaries of the single-cell data, to describe the differences between control and IL-1 stimulated cells, and to define whether there are any natural groupings within the observations. Both univariate and multivariate statistical analysis was performed, with the multivariate analysis comprising both supervised and unsupervised techniques, on account that the dataset size is borderline acceptable for creating the training and test sets used within supervised techniques.

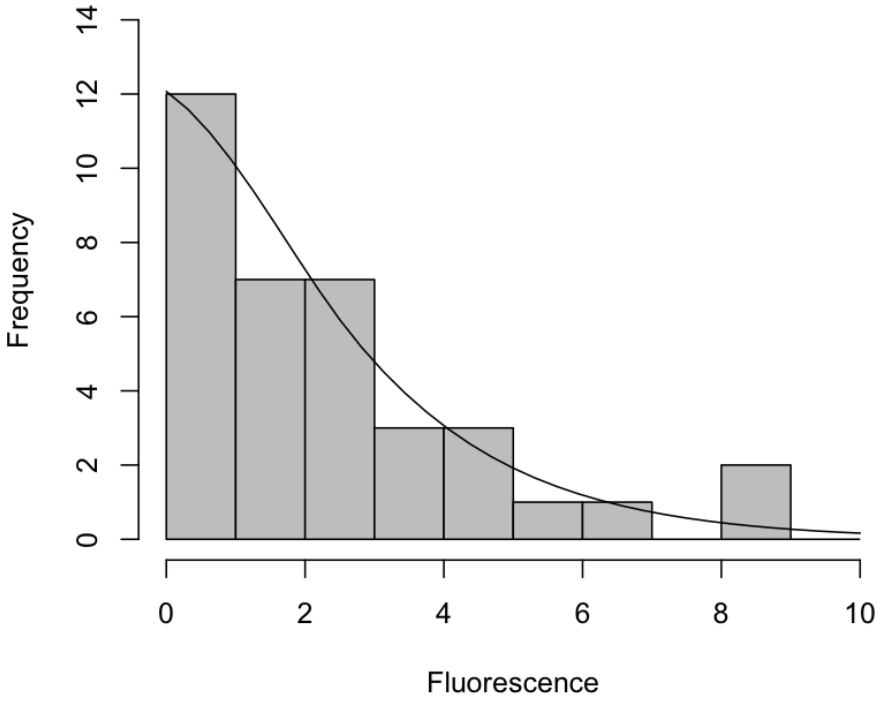

Figure 4.18: Histogram of control observations that have been binned (grouped) using an integer interval of initial (time 0min) fluorescence. The superimposed line represents a negative binomial distribution, using the median calculated from the raw data (Yang et al., 2003). The median average has been calculated as 1.947153.

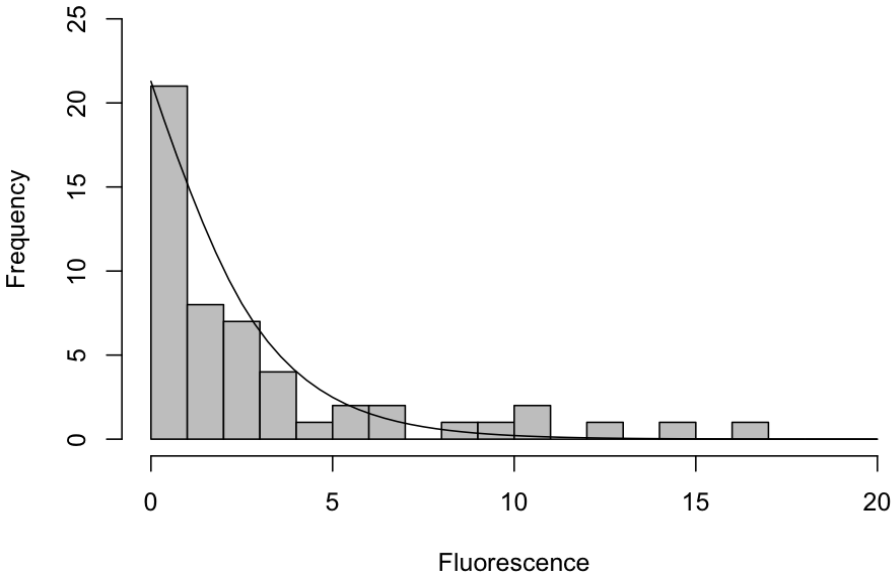

Figure 4.19: Histogram of IL-1 stimulated observations that have been binned (grouped) using an integer interval of initial (time 0min) fluorescence. The superimposed line represents a negative binomial distribution, using the median calculated from the raw data (Yang et al., 2003). The median average has been calculated as 1.729876.

The negative binomial distribution is an alternative to the Poisson distribution, and is especially useful for data over an unbounded positive range whose sample variance exceeds the sample mean. Due to the large variance inherent in a negative binomial distribution; plotting a time course graph of mean average cytoplasmic fluorescence with standard deviation or variance as error bars, may be misleading due to the potential for the error bars to represent a negative fluorescence. We have therefore used box-whisker plots to represent a simple picture of cytoplasmic fluorescence based on the range of the data and values of the quartiles (Upton and Cook, 1996). Box-whisker plots provide a convenient way to compare multiple distributions, as evident in figure 4.20, which represents the cytoplasmic fluorescence change over a 1hr period for the transfected cells from Yang et al. (2003). These plots clearly show the large variability in fluorescence of cells at particular time points, the increased variability following IL-1 stimulation (depicted by the increased number of outliers), and also that those cells that were transfected with RelA or RelA-p50 along with IκBα show a marked decrease in fluorescence after a 1hr period - corresponding to a marked increase in the degradation of IκBα over the time course.

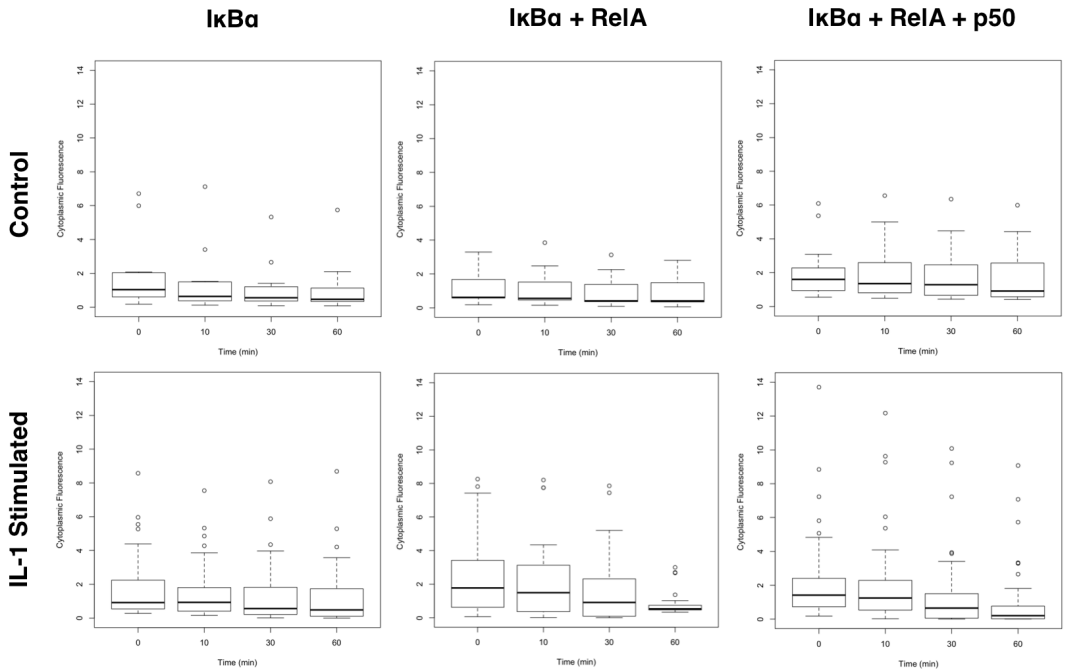

Figure 4.20: Box-whisker plots of cytoplasmic fluorescence for the control and IL-1 stimulated observations over a 1hr period from Yang et al. (2003). There were three levels of transfected cells representing: 1) IκBα, on the left hand side; 2) IκBα + RelA, down the centre; and 3) IκBα + RelA + p50, on the right hand side. The increased number of outliers highlight that the data is more dispersed under IL-1 stimulation. Furthermore, it can be seen that the addition of RelA or RelA-p50 generates a marked decrease in fluorescence, corresponding to an increase in the degradation of IκBα.

The data suggests that there is a residual/basal level of IκBα degradation within the system, and that following IL-1 stimulation this degradation rate increases. However, as a first step to allow IκBα degradation to be compared within unstimulated and IL-1 stimulated environments, we believe that the fluorescence data should be transformed so that each observation (an individual cell) become its own control, by dividing the cells' fluorescence reading at various

time-steps by its initial fluorescence reading (at time 0min). This converts the fluorescence for time 0min for each cell to 1.0 (arbitrary units), and standardises the ensuing degradation over a 1hr period[4]. Figure 4.21 is a graph of the control (unstimulated) and IL-1 stimulated data using a subset of data that had initial fluorescence upto and including 1.5 arbitrary fluorescence units using median average[5] and variance bars for interquartile ranges (25th to 75th percentiles). It can be seen that good separation is gained at 30min onwards, with a little overlap still apparent at 10min. The rates of degradation are 0.366 fluorescence units per hour for control and 0.864 fluorescence units per hour for IL-1 stimulated.

An alternative view in analysing the *full* time-series data is to visualise the four time measurements for each of the 88 observations using a scatterplot matrix (as suggested by Maindonald and Braun (2010)). The most striking feature is that the data does not separate very well into simple groupings by stimulation status, i.e. control versus IL-1 stimulated (figure 4.22), although there may be small regions of clustering at the lower levels of cytoplasmic fluorescence. This suggestes that the dataset is either not independent, and as such may have dependencies between the underlying components of the system, or that the inherent stochasticity within the signalling pathway has yielded variation within the fluorescence measurements that prevents conclusions to be drawn using simple univariate statistical analysis.

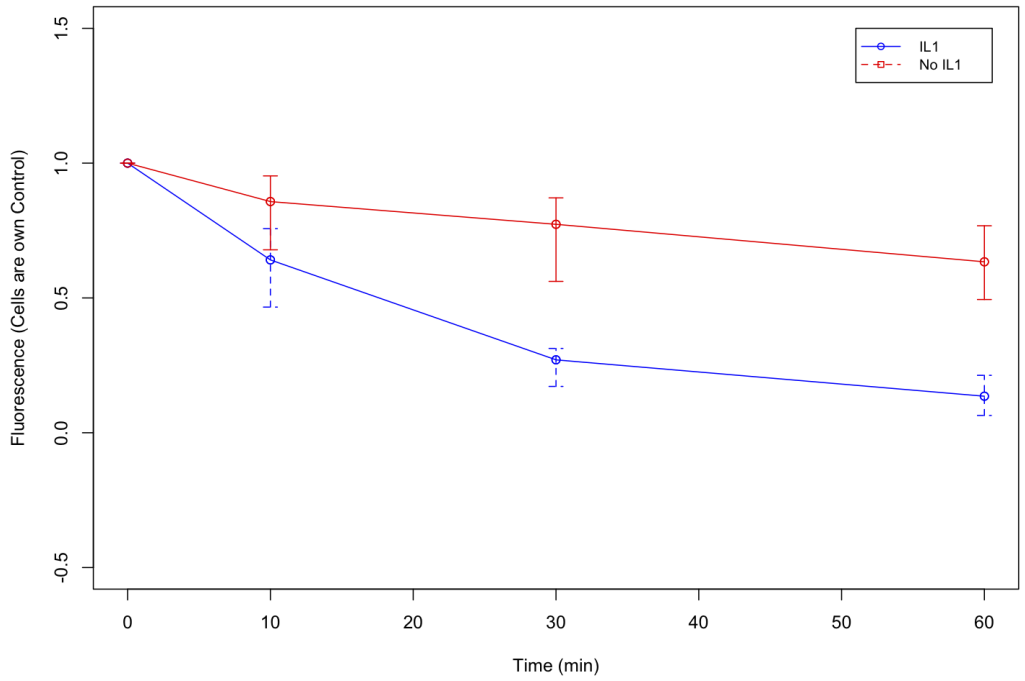

Figure 4.21: Graph of median average fluorescence for control and IL-1 stimulated observations from Yang et al. (2003). The data has been transformed so that each cell has become its own control. The error bars illustrate the spread of observations between the 25th and 75th percentiles.

[4]Our transformation of the data so that each observation becomes its own control at time 0min, is consistent with Bliss and Fisher (1953). They advise that an adequate fit of data to the negative binomial distribution provides a justification for transformation of the data to stabilize the variance, as a preparatory step for further statistical analysis by other techniques.

[5]The representation of this data as an average is not novel to this thesis, as Yang et al. (2003) displayed as mean averages, however it does represent another dimension to the domain model.

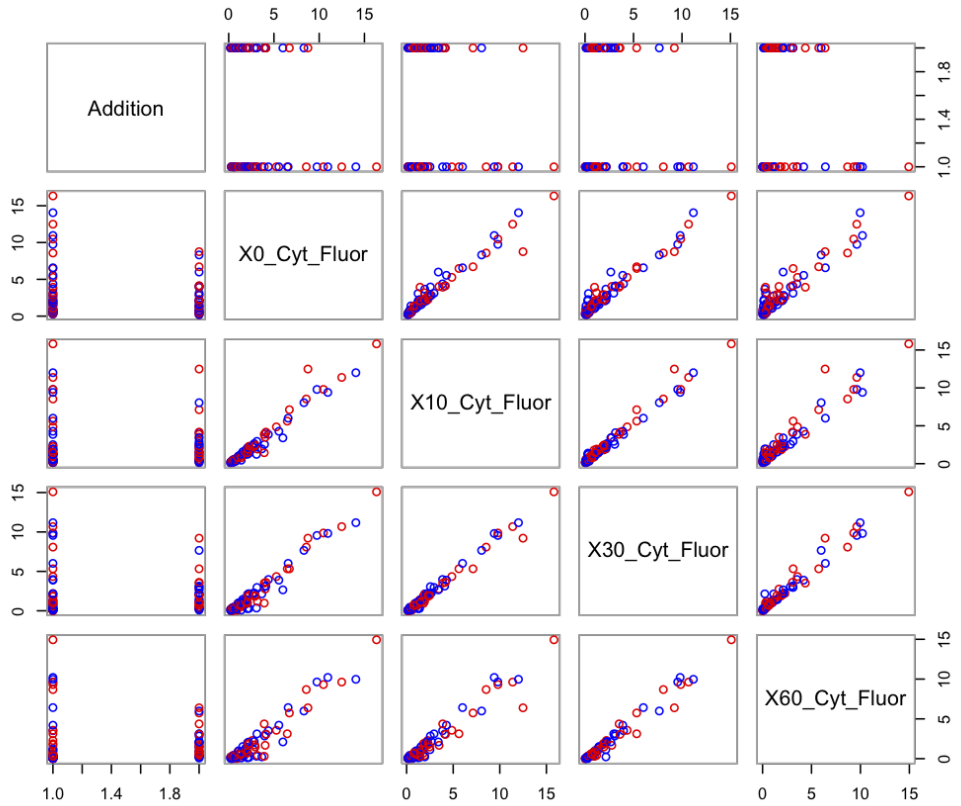

Figure 4.22: Scatterplot matrix of the single-cell fluorescence data from Yang et al. (2003) coloured by IL-1 stimulation (red) and control (blue), with both axes therefore representing fluorescence. Although no clear separation between the measurements is visible, it can be seen that partial clustering occurs within the lower levels of cytoplasmic fluorescence.

Hierarchical Cluster Analysis

Due to the poor separation gained using the scatterplot matrix, hierarchical cluster analysis[6] was used so that data may be quickly visualised in an alternative way using a multivariate technique (Bridges, 1966; Fraley and Raftery, 1998). Hierarchical clustering was performed using seven different clustering algorithms (Ward, single, complete, average, McQuitty, median and centroid). The resulting dendrograms for each method were consistent, in that no clear clustering was evident to group control cells and to group those stimulated with IL-1. There were however a few areas where clustering may be evident, and therefore appropriate for further investigation to ascertain any natural groupings within the data. Of note are the three main clusters obtained by scaling the data and using the *complete* clustering method (figure 4.23). The first cluster (left) groups IL-1 stimulated observations with an initial cytoplasmic fluorescence (at 0min) less than 3 fluorescence units (IL1_1 to IL1_34) and control observations with an initial cytoplasmic fluorescence less than 2.079 fluorescence units. There are a few anomalous observations (IL1_37 to IL1_40), but overall the cluster looks very

[6]The raw data was transformed so that the meaurement category (IL1 v No IL-1) and repetition number were encoded, to produce a single label for each observation, e.g. observation 1 was coded as IL1_1 and related to the first single-cell observation with IL-1 stimulation.

clean. The second cluster, groups control observations 21 to 24 (No IL1_21 to No IL1_24) with initial cytoplasmic fluorescence of 2.079 to 8.0 and IL-1 stimulated observations IL1_41 to IL1_45, along with IL1_32, IL1_35 and IL1_36, which may be anomalies. The final cluster (actually clusters 3 and 4, but cluster 3 only has a single observation - IL1_52) groups both IL-1 stimulated and control observations with initial cytoplasmic fluorescence greater than 8.0 fluorescence units.

Scaling the data, slightly decreases the correlation coefficient of the test data (between the distance matrix and the cophonetic distance) from 0.9159 to 0.8982, but enhances the clustering of IL-1 stimulated observations with each other, and control observations with each other (i.e. IL-1 stimulated observations group together and control observations group together), within the three main clusters. We believe that this slight reduction in correlation may be due to an amplification of standard error within the observations, however the correlation score is still high, and therefore we believe that the resulting dendrogram may be deemed an appropriate summary of the data.

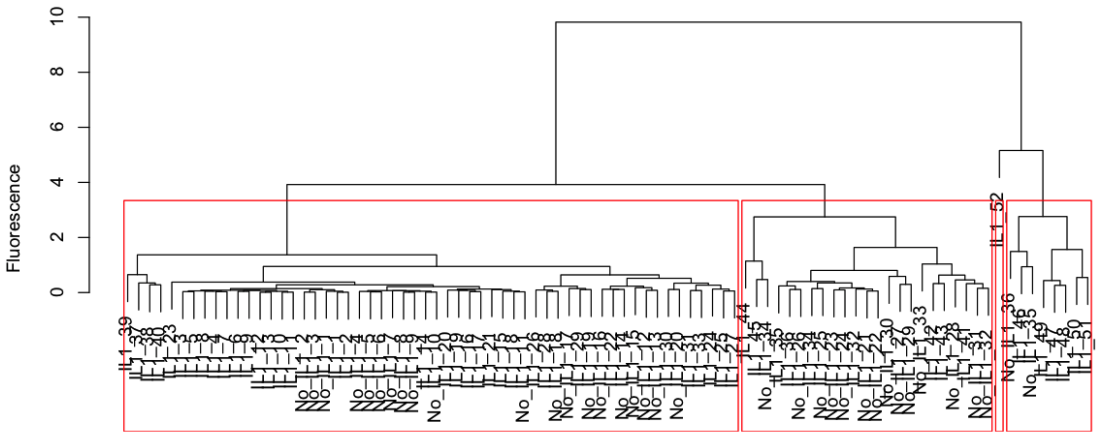

Figure 4.23: Dendrogram representing the clustering of observations from Yang et al. (2003) by hierarchical cluster analysis. The red boxes indicate that hierarchical cluster analysis identifies the three forced clusters as observations having an initial cytoplasmic fluorescence less than 3.0, between 3.0 and 8.0, and above 8.0 fluorescence units.

Principal Component Analysis

The scatterplot matrix and hierarchical cluster analysis provide preliminary indications of groupings/relationships between the data, however as no clear clustering is evident, a more powerful technique is required. Principal Component Analysis (PCA) is an unsupervised multivariate technique, which determines new variables (inherent to the dataset) based on the direction of maximum variance (Pearson, 1901). As such, it can be used to reduce the dimensionality of data for detecting underlying structures (Wold et al., 1987). The scree plot (figure 4.24) shows the variance corresponding to each *principal component* within the data of Yang et al. (2003). It flattens out after principal component 1 (PC1), indicating that only a single principal component is required for the variation seen in the data. The cumulative proportions of variance (table 4.1) reinforce this hypothesis.

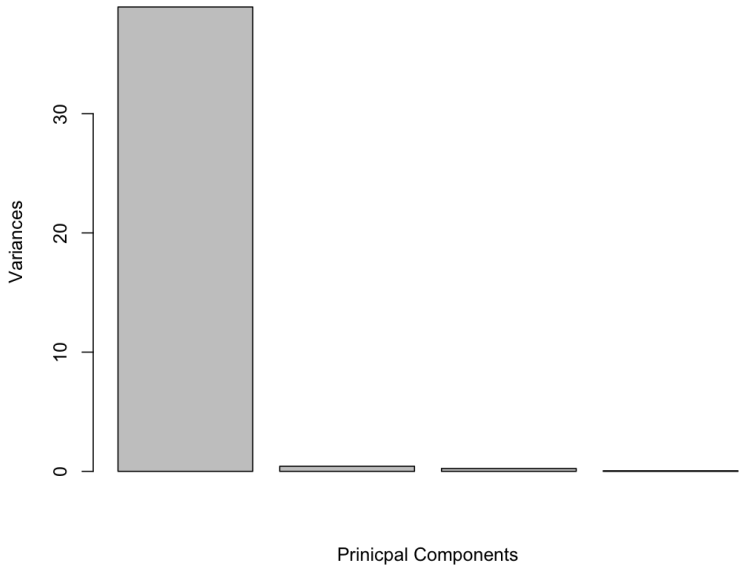

Figure 4.24: Scree plot of the principal components from principal component analysis of observations from Yang et al. (2003). Each bar corresponds to its respective principal component; bar heights are the variances of the principal components.

	PC1	**PC2**	**PC3**	**PC4**
Standard Deviation	6.241	0.65322	0.48826	0.21399
Proportion of Variance	0.982	0.01076	0.00601	0.00115
Cumulative Proportion	0.982	0.99283	0.99885	1.00000

Table 4.1: Summary of principal component analysis of the single-cell fluorescence data, showing the standard deviation, proportion of variance and cumulative proportion of variance for each principal component.

A bi-plot of PC1 and PC2 provides another visual representation of these principal component loadings. Figure 4.25 shows that all four time measurements contribute to the separation of the data (through PC1), with times 0, 10 and 30min having the greatest effect. Furthermore, it was found that: PC1 was dominated by measurements at times 0, 10 and 30min; PC2 was dominated by measurements at times 0 and 30min; PC3 was dominated by measurements at times 0 and 60min; and PC4 was dominated by measurements at times 0 and 10min.

Further analysis was performed on all pair-combinations (six in total) of principal components, and colour-coded depending on the relevant category to which they belong. Initially only two categories were used, representing *control* and *IL-1 stimulated* observations, however this did not yield a clean separation of observations. Analysis on pair-combinations was therefore extended to comprise additional categories, which correspond to control and IL-1 stimulated observations further divided by initial (time 0min) cytoplasmic fluorescence ranges. The best separation by principal components occurred using initial fluorescence ranges of 0-1.5, 1.5-3.0 and greater than 3.0 fluorescence units.

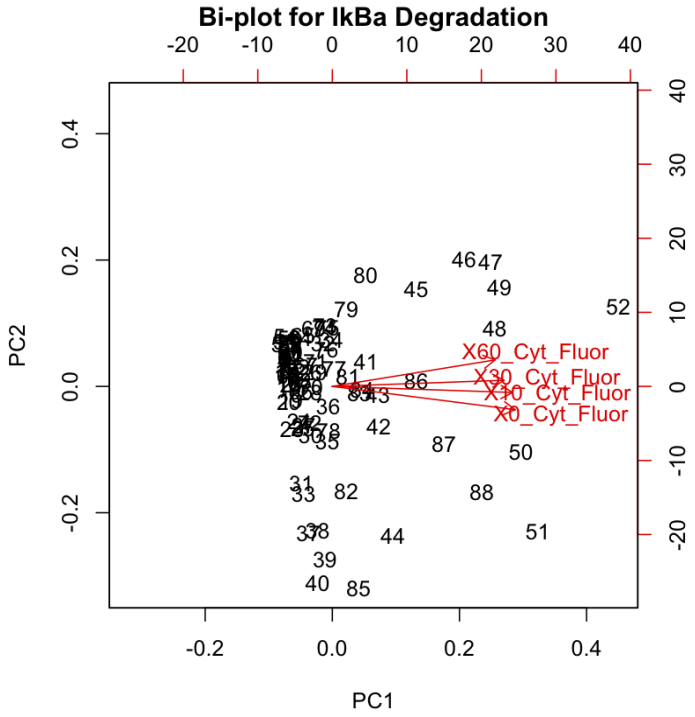

Figure 4.25: Bi-plot of PC1 and PC2 from principal component analysis of observations from Yang et al. (2003). This plot shows that measurements for times 0, 10 and 30min contribute equally to the separation of PC1 due to their virtually equivalent arrow lengths. They are not fully parallel to the PC1 axis however, and therefore also contribute slightly to PC2.

Complete separation does not occur for any of the six pair-combinations, however separation emerges for initial cytoplasmic fluorescence upto 1.5 fluorescence units against the rest of the data (i.e. observations with initial fluorescence of 0-1.5 units, separate away from observations with initial fluorescence greater than 1.5 units). Figure 4.26 represents the plot of PC1 versus PC2, which has been colour-coded to categorise control and IL-1 stimulated observations that have been grouped by their initial cytoplasmic fluorescence. Along with the separartion of control and IL-1 stimulated observations with initial fluorescence less than 1.5 fluorescence units, there is also partial separation within the range 1.5 to 3.0 fluorescence units. Observations with initial cytoplasmic fluorescence greater than 3.0 fluorescence units show no identifiable separation between control and IL-1 stimulated conditions. Evidence for the separation of control and IL-1 stimulated observations with initial cytoplasmic fluorescence upto 3.0 fluorescence units is further provided through the plot of PC1 loadings against observation number (figure 4.27). This plot of loadings also confirms the good separation of observations with initial cytoplasmic fluorescence upto 1.5 fluorescence units.

Multidimensional scaling and Kruskal's non-metric scaling were also performed on the data, and these tests yielded very similar results to the PCA analysis above (data not shown).

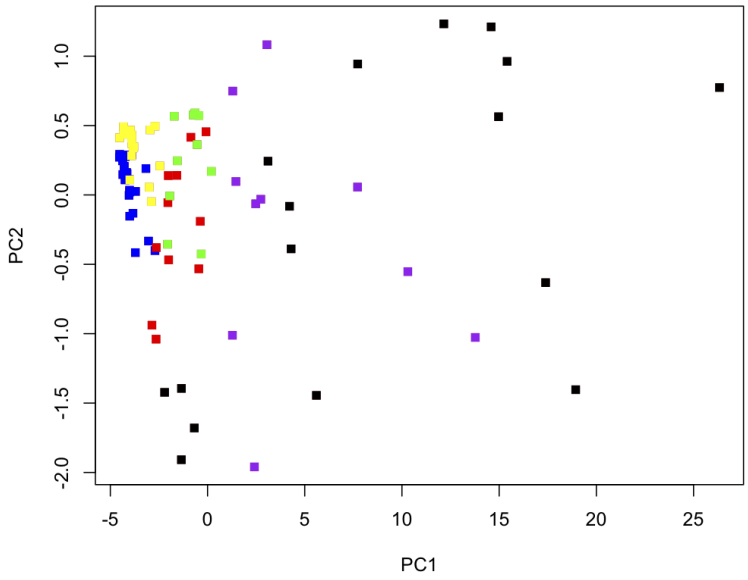

Figure 4.26: PCA plot of principal components 1 and 2, colour-coded by observation category, i.e. control versus IL-1 stimulated and range of initial cytoplasmic fluorescence. The six categories are: IL-1 stimulated/0-1.5 = Blue, IL-1 stimulated/1.5-3.0 = Red, IL-1 stimulated/>3.0 = Black, control/0-1.5 = Yellow, control/1.5-3.0 = Green, and control/>3.0 = Purple. The plot shows good separation of observations with initial fluorescence < 1.5 units from the rest of the data, with only a small amount of overlap between the control and IL-1 stimulated observations. There is also a degree of separation between observations with initial fluorescence values of 1.5-3.0 units from the rest of the data, however the amount of overlap between control and IL-1 stimulated is more significant here.

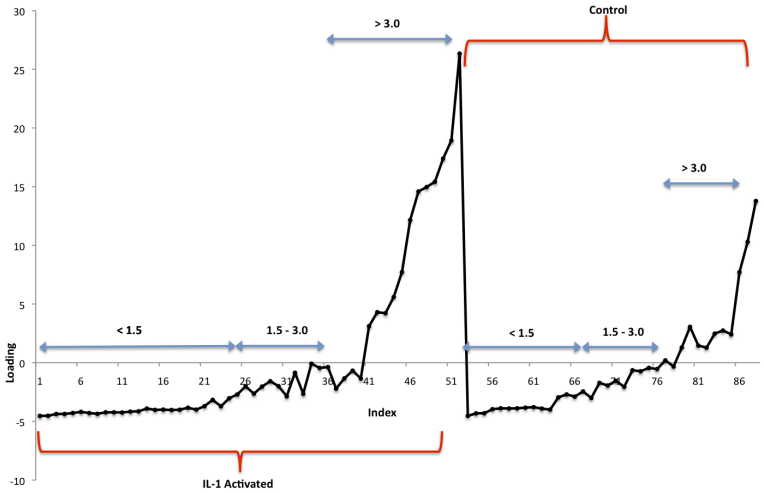

Figure 4.27: Plot of loadings for principal component 1 following PCA. PC1 was chosen because this is the component which contributes most to separation of the data. It can be seen that observations with initial fluorescence between 0-3.0 and >3.0 can be separated easily as the observation between 0-3.0 units have negative loadings and >3.0 have positive loadings. Furthermore, observations for cells with initial fluorescence between 0-1.5 tend to have relatively stable loadings (around -4.5), whereas those between 1.5-3.0 begin to have more variable loadings.

4.4 Reflections on the use of UML and Statistical Techniques for Domain Modelling

This reflective section draws upon the experience of developing the domain model of the IL-1 stimulated NF-κB signalling pathway. Emphasis is focused on the strengths and weaknesses of UML for domain modelling complex biological systems, and statistical techniques have been found to complement UML, thus facilitating the development of a more complete domain model.

The reflective analysis commences with section 4.4.1, through an examination of the benefits of a top-down approach to developing the domain model, which increased the granularity of detail from system-wide behaviours, down to component interactions, and finally individual component dynamics. Section 4.4.2 reflects on the suitability of UML for developing domain models of complex biological systems, with particular emphasis on the use of UML class, activity, sequence and state machine diagrammatic notations. Finally, section 4.4.3 reflects on the use of univariate and multivariate statistical techniques to complement UML in developing domain models of complex biological systems.

4.4.1 Reflections on the Process used for Domain Modelling

As discussed in chapter 3, the NF-κB signalling pathway is a complex intracellular network that manifests in stochastic and dynamic responses to inflammatory stimuli. The system-wide behaviours, generated as an inflammatory response to pathogenic invasion and other physiological perturbations, emerge through the cumulative effect of low-level intracellular interactions within an individual cell, being amplified to a large-scale across a population of immune response cells. As such, the inherent complexity of the signalling pathway and its associated stochasticity and dynamics, renders the process of domain modelling both time consuming and non-trivial in nature. We have therefore chosen to use a top-down approach for domain modelling, which allows a succession of models to be developed in an iterative manner, at increasing levels of detail.

Domain modelling, as specified by the CoSMoS process, follows an iterative process, where the various views of the system (in this case biological views) are refined until agreement is gained between the modellor and domain expert. As discussed previously, we have chosen to use the *deep-curation* approach for domain modelling. This relies upon the manual curation of relevant facts and information from the published literature, experimental datasets, and interactions with the domain expert, to develop an initial model in a quasi-top-down manner. As described in section 2.7, the domain model is akin to a functional requirements document used in software engineering, and therefore provides a transparent definition of the real world domain behaviour (i.e. functionality) that will be required in future iterations of the resulting computational model. This iterative approach, with close liaison between modellor and domain expert, ensures that an extended period of time is set aside for comprehensive exploration of the real-world domain before simulation development, and furthermore ensures the modellor has sufficient understanding of the scientific domain before design and implementation of the resulting *in silico* experimentation that completes the CoSMoS lifecycle (as per figure 2.6).

The top-down approach for developing the domain model is represented through the three levels of UML models in section 4.3.1. We have commenced by modelling the system-level properties using a cartoon-like *expected behaviours diagram* (see figure 4.1) as advocated by Andrews et al. (2010) and Read et al. (2009a). We believe that this provides a concise way to diagrammatically convey not only the component interactions within the complex biological system, but also an ability to define the observable phenomena from the system-wide dynamics. Importantly, this diagram also links the two concepts together through the ability to define hypotheses on how the interactions generate the emergent behaviours of the system. Although the diagram is not part of a UML notation, and therefore has no formal requirements, it is useful in identifying the behaviours, interactions, and components in the biological domain that are to be included in the domain model. Furthermore, as shown in section 4.3.1, a number of additional diagrammatic views were developed to complete the system-level domain model.

Following the domain modelling of system-level properties, the top-down approach led us to model the component-level interactions of the NF-κB signalling pathway. UML was used as the basis for modelling, and the process of semi-formally defining the molecular interactions within the signalling pathway gave rise to a large number of questions regarding: the abstraction level to be used for the computational model; the dynamics of interactions between the different types of molecular species; the quantities for each component within the system; and whether there were any rate constants for more generalised biological processes that are involved in the signalling pathway (for example, diffusion kinetics or Brownian motion). The process of modelling these molecular interactions also served to inform and guide the functional specification of individual component-level dynamics.

Our reflections above are therefore in agreement with Read (2011), who argues that the process of domain modelling using a top-down approach is more intuitive than a bottom-up approach. This is because the concept of expanding functional aspects of complex systems through increasing the detail of lower level interactions, e.g. from their system-level properties, through to component interactions and individual component dynamics, is more intuitive than the ability to conceptualise system-wide behaviours from individual component-level dynamics by reporting them first, and then building up levels of the functional hierarchy.

4.4.2 Reflections on the Suitability of UML for Domain Modelling

As mentioned previously, the domain model is analogous to a functional specification from software engineering and development projects from industry. As such, the primary purpose is to unambiguously capture our abstracted view of the functionality of the real world domain, which will be incorporated within future iterations of the resulting computational model. This ensures that the developer and domain expert are in full agreement on the scope of the CoSMoS project. In addition to capturing the functional requirements, we have also discovered that the actual process of developing the domain model in conjunction with the domain expert has facilitated a much more in-depth understanding of the domain, than would have been gained through reading published literature alone. Fur-

thermore, due to the complex, stochastic nature of the NF-κB signalling pathway, we have been unable to develop a single diagrammatic view which could capture the various components, interactions, and dynamics of the system. It has therefore been necessary to utilise a number of different cartoon and UML notations in order to complete the domain modelling exercise, thus confirming the experiences of Read et al. (2014). These different diagrammatic views allow us to capture the initiation and propagation of the signalling pathway, across the inherent hierarchies of the system (i.e. system-wide, component interactions, and individual component dynamics). We believe this to be a natural progression when domain modelling, and suggest that the different views reflect the modular nature of biology, as emphasised through systems biology.

Perhaps one of the simplest diagrammatic views of a complex system such as the NF-κB signalling pathway is that of its physical environment, which in this case is a Eukaryotic cell. Unfortunately, UML being a very prescriptive notation is not very effective in succinctly capturing the essence of component containment within a cell membrane that we require (figure 4.28). As the language is more relevant for designing software systems, it may be of more use within this goal for the subsequent *platform model*, we therefore believe that the cartoon-type diagram used for figure 4.2 better serves our needs for presenting a view of physical containment within the domain model.

Following on from containment and physical structure, another UML diagram focused on the types of components (termed classes) within a system is the inheritance class diagram (see figure 4.29). We believe that this UML notation is useful for illustrating the functional similarities between the components, but again, seems very technical for the intended audience (i.e. biological domain experts) at the stage where we prepare the domain model. We believe that a table may suffice to express the same information, and would also be more intuitive to a biologist. As per the UML class containment diagram, this notation had its merits when it comes to defining the technical requirements of the resulting computational model as part of the platform modelling process. Once the diagrammatic views representing containment and inheritance have been developed, the next step in the process of eliciting component-level interactions is to develop a full UML class association diagram (as per figure 4.4). This class association diagram depicts the full spectrum of relationships between components within the domain.

To make the diagram more intuitive to an audience of biologists, we have ordered the associations sequentially, to follow the spatial cascade of the NF-κB signalling pathway, from the cell membrane to the nucleus. We have also purposefully omitted the concept of cardinality from the relationships to ensure the diagram is not too cluttered. This does however expose a problem with interpretation. As it stands, the class diagram incorrectly suggests that every cell will become stimulated by extracellular signal, however this is not correct, as there may be some cells that are never exposed to inflammatory stimuli throughout their lifetime. Conversely, for a cell to become stimulated, it must simultaneously perceive a number of IL-1 molecules exceeding a given threshold. We therefore believe the use of class diagrams for modelling stochastic and dynamic systems may lead to a degree of confusion and misleading interpretations of system behaviours. As such, we believe that they should be constrained to providing a view

of static relationships between system components.

The UML notations used thus far, have focused on the static relationships between system components. Although this has been useful to gain a feel for the physical structure of the signalling pathway and functional relationships between the components, the class diagrams (full association, inheritance, and containment) have not allowed us to express any of the dynamic relationships relating to the components following activation of the cell membrane by a suitable stimulating ligand. UML sequence diagrams provide a mechanism to show the order of events that occur within the complex system and how these relate to interactions between the individual components (see figure 4.5).

Although easy to follow from a chronology of interactions perspective, we believe that sequence diagrams may not be the most suitable of UML notations for expressing biological information succinctly. A good example relates to the phenomena of feedback within complex systems. As the sequence diagram conveys time linearly along one dimension (in our case vertically, from top to bottom through the use of *lifelines*), this prevents the diagrammatic expression of feedback loops. Specifically to our domain, the IL-1 stimulated NF-κB signalling pathway is unlike most complex systems, in that it does not possess a positive feedback loop which propagates the signal, but instead contains negative feedback, through the production of its inhibitor IκBα as part of the upregulation of inflammatory genes during the immune response (see steps 8.1 to 8.5 in figure 4.5). Fortunately, UML employs another notation that may be used in the form of communication diagrams (see figure 4.30). Here the developer is free to position the individual components where they deem appropriate within the diagram (unlike the rigid temporal sequence from top-to-bottom used in sequence diagrams) and the order of interactions is denoted through numbering the events on the links between components. As per the association class diagram (figure 4.4), we have positioned the components within the communication diagram to match the spatial dimension with a Eukaryotic cell to make the diagram more intuitive to biologists.

An alternative view to the UML sequence and communication diagrams, which convey the interactions between components, is the activity diagram, which conveys the network of activities being performed within the system. Figures 4.6, 4.7 and 4.8 illustrate respectively the activities being performed within cell membrane receptor activation, the NF-κB signalling module, and transcription (and translation) of inflammatory response genes (and proteins)[7]. The UML activity diagram notation has been updated and refined since its initial draft standard, and now benefits from a number of advanced concepts for conveying complex aspects of activities within a system. Figure 4.31 utilises the concepts of discrete parameter sets for receptor complex formation, and decision points for the ligand-induced activation or inhibition of receptor.

The negative feedback within the system is hard to convey using UML notation when the activities of the whole system are divided into a number of discrete functional modules. The individual UML activity diagrams (figures 4.6 - 4.8)

[7]The activities on the activity diagrams are expressed at the level of a single cell. It must be noted however, that the inflammatory response acts upon a population of cells, and therefore a large number of cells will be undergoing similar processes concurrently, and may be out of phase.

can therefore be linked to form a large end-to-end UML activity diagram that commences with the activation of a cell membrane receptor and culminates with the generation of new IκBα and inhibition of the NF-κB complex. Furthermore, another concept within the UML notation standards that may be beneficial for us to introduce the sub-cellular location within our diagrams is that of swim-lanes[8]. Figure 4.9 links the three individual activity diagrams into a single end-to-end activity diagram, and also incorporates the concept of swim-lanes to convey the location of these activities within the cell.

Although portraying temporal interactions as per UML sequence diagrams, we believe that UML activity diagrams are more intuitive for non-Computer Science audiences as they are more flexible in relation to the position of components within the system, thus allowing the positioning of components to approximate to the spatial locations within a Eukaryotic cell. We believe that end-to-end activity diagrams that utilise swim-lanes provide a very useful diagrammatic view of the NF-κB signalling pathway, but also due to the ability to print onto a single page using A3-sized paper, provide the ability for the developer and domain expert to discuss the scope of the computational model and the finer details of the underlying wet-lab biology.

It must be noted however that care is required when interpreting activity diagrams. We believe they are a very useful notation to convey the order in which activities are performed by system components, but may mistakenly provide the impression that a component performing an activity may become inactive afterwards. This is not always the case, as components may go on to perform other functions that are not depicted on an individual activity diagram, that is being used to convey a particular view of the complex system.

The final UML diagrammatic notation that we have found useful for development of the domain model are state machine diagrams. These provide a useful mechanism to convey low-level dynamics within individual components of the system. Individual state machine diagrams have been generated for all components within the pathway that actively change state within the system (see figures 4.10 to 4.16). As per the individual activity diagrams, they can also be joined sequentially and through embedding within a high-level state machine diagram for the entire Eukaryotic cell, can also convey the location of the state machines (figure 4.32).

[8]Activity diagram swim-lanes are utilised within the UML standards to convey responsibility for an activity. We believe however, that a degree of flexibility should be utilised in the precise semantics of the UML standards when modelling biological systems. As such, we believe that swim-lanes are an ideal mechanism to convey location of activities within a UML activity diagram.

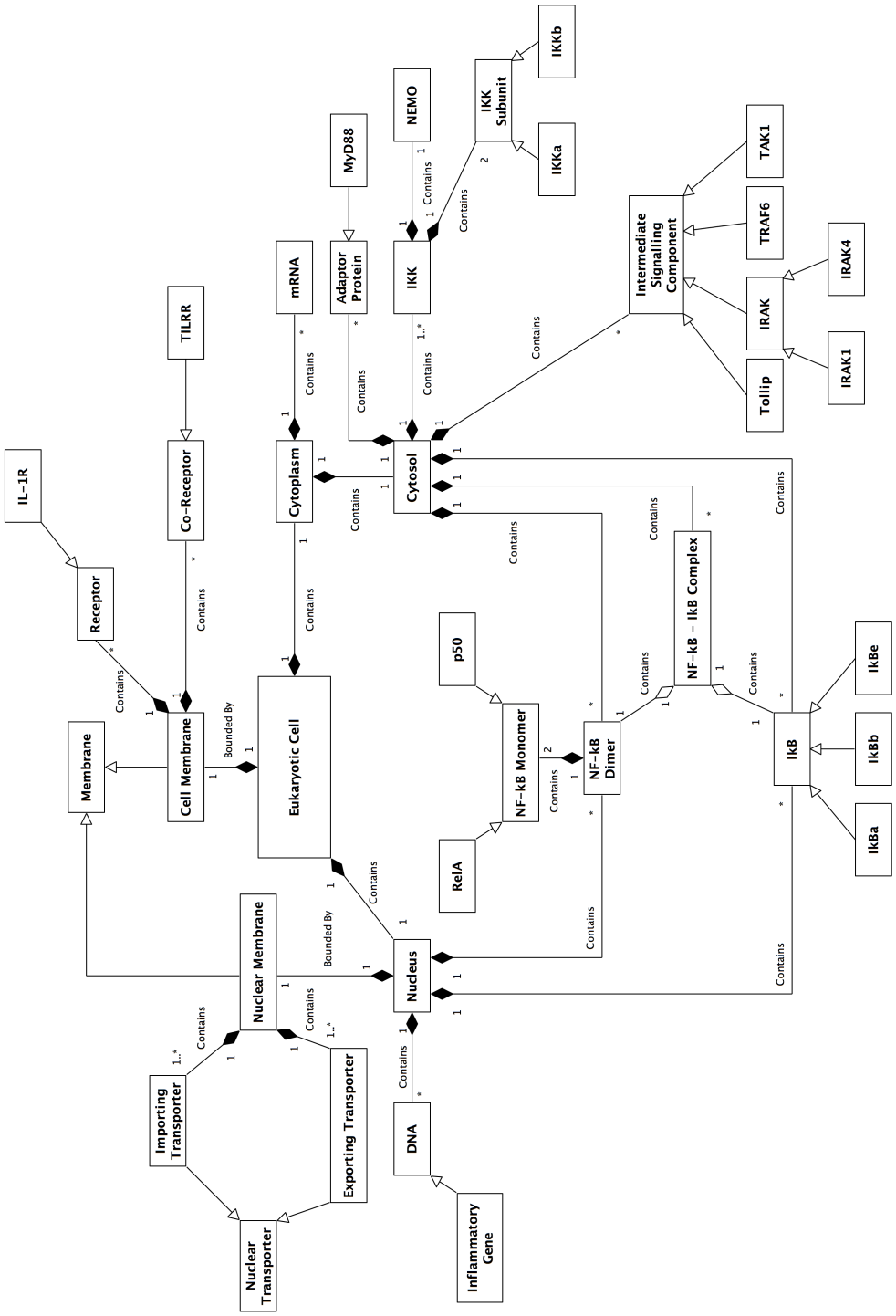

Figure 4.28: UML containment diagram representing the components involved in the pathway and their physical environment in which they are situated within a cell. Developed using Burns et al. (2000), O'Neill and Dinarello (2000), Hayden et al. (2006) and Karin and Ben-Neriah (2000).

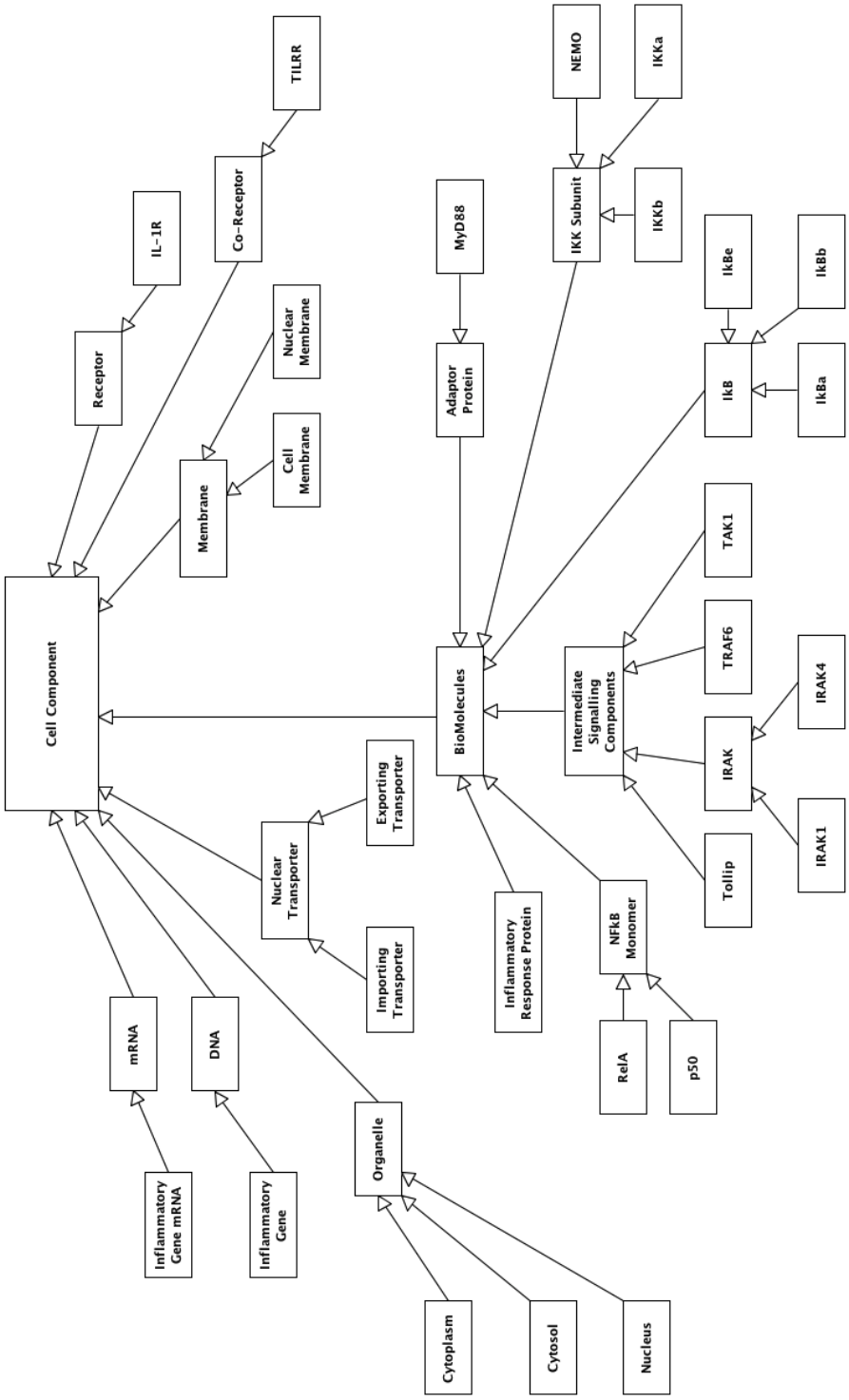

Figure 4.29: UML class inheritance diagram highlighting the functional similarities of components involved in the pathway. Arrows indicate the order of inheritance, for example the components *cell membrane* and *nuclear membrane* are more specific instances of the *membrane* component, and therefore inherit the attributes that are associated with *membrane*. Developed using Burns et al. (2000), O'Neill and Dinarello (2000), Hayden et al. (2006) and Karin and Ben-Neriah (2000).

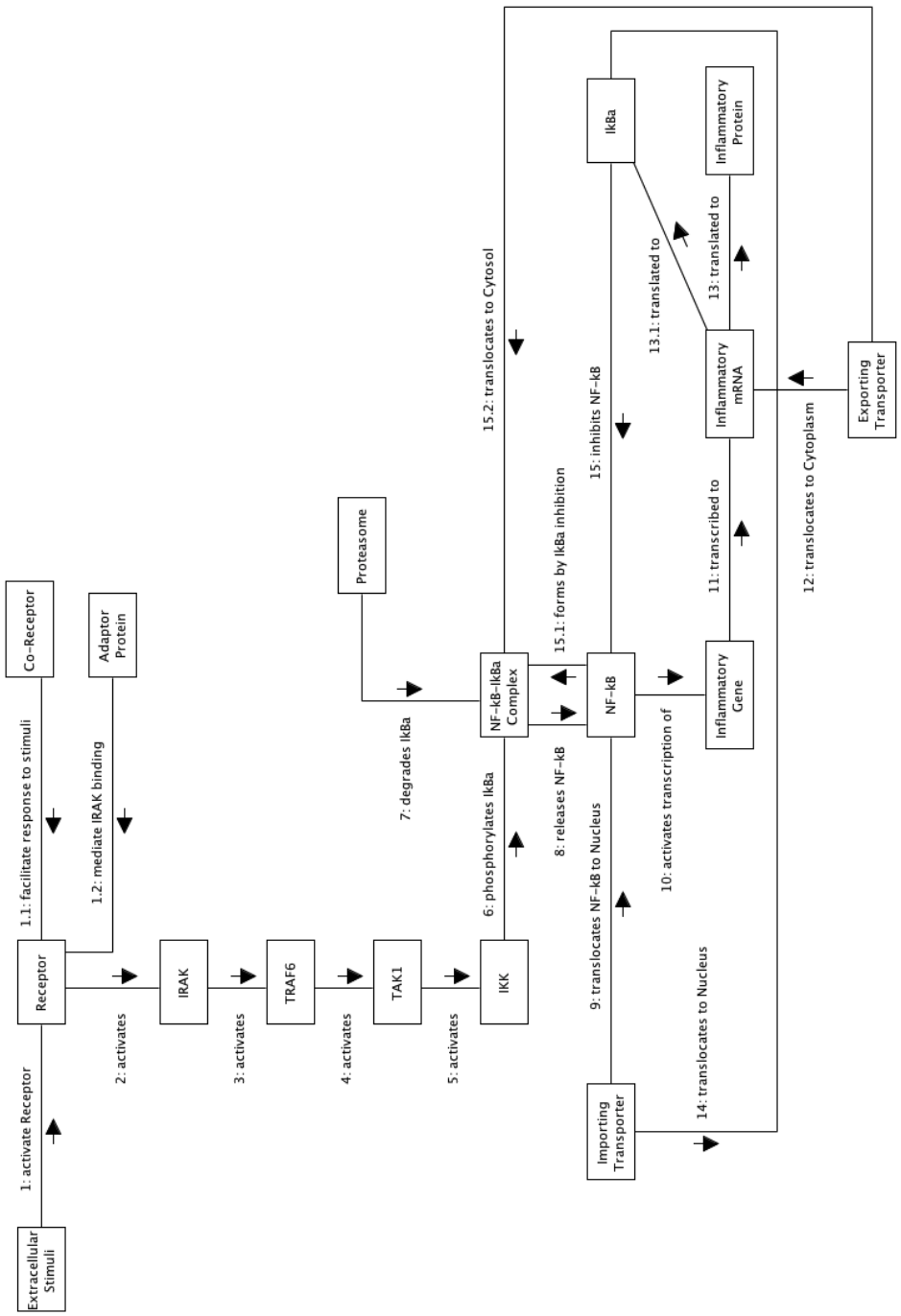

Figure 4.30: UML communication diagram for the pathway. Although portraying temporal interactions as per sequence diagrams, we believe these are more intuitive for non-Computer Science audiences as they are more flexible in relation to the position of system components, thus allowing the positioning of components to approximate to the spatial locations within a cell. Developed from reviews of Hayden et al. (2006) and O'Neill and Dinarello (2000).

Figure 4.31: Alternative UML activity diagram for activation of the cell membrane receptor associated with the NF-κB canonical pathway. The activity diagram starts with extracellular stimuli (LPS or IL-1) having been expressed, and the first activity is the movement of this signal into the vicinity of the cell. As per figure 4.6, there is also the requirement for co-receptor (e.g. CD14 & MD2 for TLR4 or TILRR for IL-1R) and adaptor protein (e.g. MyD88) to also bind with the receptor for amplification of the stimuli-induced activation of the NF-κB canonical pathway, however this time the activity diagram is split using a decision point for the respective extracellular stimuli and uses discrete parameter sets for the actual receptor complex formation and resultant receptor activation. For bacterial lipopolysaccharide stimuli, this entails the association of TLR4 receptor, CD14 and MD2 co-receptor and My88 adaptor protein. Similarly, for IL-1 stimuli, this entails the association of the IL-1R receptor, TILRR co-receptor and MyD88 adaptor protein. Following formation of the relevant receptor complex, the receptors are then activated, and the signal is propagated through the signal transduction chain of intermediate signalling components, as per figure 4.7. Developed from O'Neill and Dinarello (2000), Burns et al. (2000) and Zhang et al. (2010).

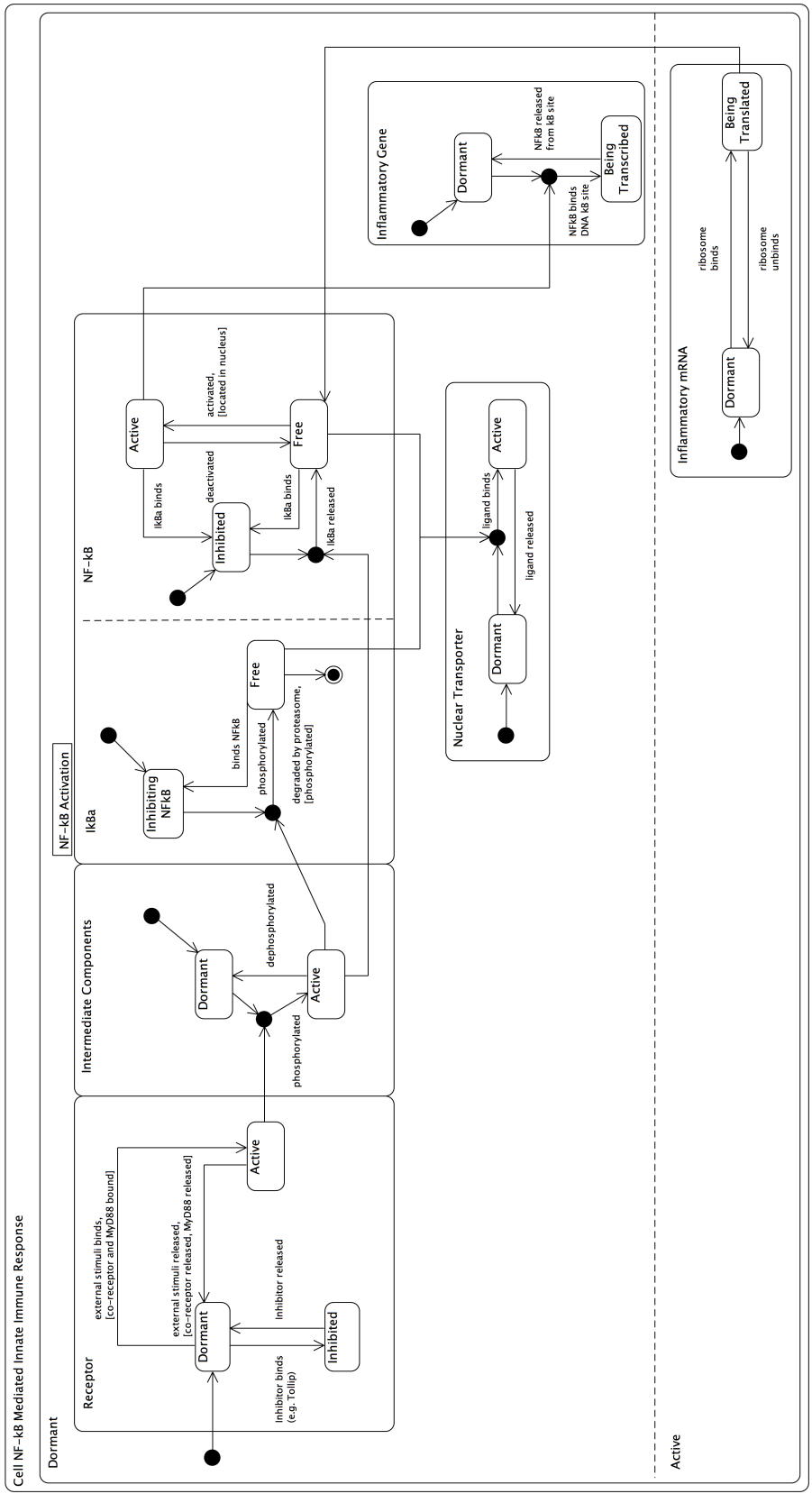

Figure 4.32: Embedded state machine diagrams for the NF-κB signalling pathway. This provides an alternative view to the individual state machine diagrams in figures 4.10-4.16. These individual state machines are explicitly linked using UML join notations and are embedded within a single large state machine that represents the cell. Here the cell has two states relating to dormant or active. Developed using Burns et al. (2000), O'Neill and Dinarello (2000), Hayden et al. (2006), Baeuerle and Baltimore (1988a,b), Karin and Ben-Neriah (2000) and Siebenlist et al. (1994).

4.4.3 Reflections on the use of Statistical Techniques to Complement UML in Domain Modelling

We believe that univariate and multivariate statistical analysis of the data has provided a complementary view to the domain model that UML alone would not be able to capture. Through visualising the data as histograms (figures 4.18 & 4.19), it quickly became evident that the data was highly skewed, and that any assumptions to a normal distribution, which requires a symmetrical distribution around the modal value, would be incorrect. Due to the uncertainty about direction of difference, two-tailed χ^2 goodness of fit tests were therefore used to identify the type of distribution that the data approximated to.

It appears that although the full dataset is clearly not normally distributed, there is a subset of observations within the 0-3.0 fluorescence units range, which does tend to Normality[9]. The χ^2 scores therefore need to be interpreted in context to the dispersal across the x-axis. Although the χ^2 scores show no significant difference from the expected values that relate to a normal distribution (for the 0-3.0 fluorescence range), the actual normal distribution with mean and standard deviation calculated from the data subset would contain negative fluorescence values. As this will not occur in reality, we are left assuming the data approximates to a normal distribution, with the caveat that this only occurs for a portion of the distribution. Therefore, although the χ^2 tests allow an approximation to normality, we believe that it is more accurate to approximate to a negative binomial distribution, which can be approximated across the full range of observable fluorescence. This is in keeping with the findings of White and Bennetts (1996) and Bliss and Fisher (1953), who advise that biological populations, be that cell or organism level, often approximate very closely to negative binomial distributions. As such, future statistical tests on the data, and indeed any simulation level data produced from *in silico* experimentation, should be non-parametric in nature as these are applicable to any distribution, and do not assume normality. Furthermore, due to its non-parametric nature, the central measure used should be the median average, as this is not affected to the same extent from skewed data as the mean average (Siegel, 1957).

Results of the hierarchical cluster analysis highlighted evidence of separation of observations into three groups relating to initial cytoplasmic fluorescence: 0-1.5, 1.5-3.0 and >3.0 fluorescence units (see figure 4.23). The correlation between the distance matrix and cophonetic distance for the dendrogram with unscaled data (complete method) was 0.9159 and with scaled data (again using the complete method) was 0.8982 and therefore provides evidence that the dendrograms are a reasonable graphical summary of the dataset. Although the scaled data provided the dendrogram with the best clustering, it was only slightly better than that gained using unscaled data, and the correlation was also slightly worse than that gained with unscaled data. This may have been due to the scaling process amplifying stochastic variation, and therefore subsequent analysis was performed using unscaled data.

[9]Transfection is an unnatural state for a living cell. The system appears to cope with a certain degree of excess protein, but at some stage (i.e. with fluorescence units more than 2.0-3.0), the system shuts down. Therefore, in the group of functional cells, with fluorescence between 0-3.0, the distribution approximates to Normality.

Principal component analysis, being an unsupervised technique may be relied upon to be a good summary of the data. Analysis suggested that a single principal component was needed for separation of the data; therefore separation was visually displayed by using PC1 and PC2 (see figure 4.26). As per the hierarchical cluster analysis, PCA showed good separation with respect to the observations that could be grouped into 0-1.5, 1.5-3.0 and >3.0 ranges of initial cytoplasmic fluorescence. Furthermore, PCA showed reasonable separation between control and IL-1 stimulated observations in the 0-1.5 initial cytoplasmic fluorescence range, with a small amount of overlap between the two groups. Similarly, there was partial grouping of observations within the 1.5-3.0 and >3.0 fluorescence ranges, however within these groups there was no discernible separation of control versus IL-1 stimulated observations. The bi-plot (see figure 4.25) shows that PC1 is dominated by measurements at times 0, 10 and 30min, with those for 10 and 30min being virtually parallel with the PC1 axis, indicating that these measurements are the most important for separation of observations. The plot of PC1 loadings (see figure 4.27) provides very strong evidence that observations relating to cells with initial cytoplasmic fluorescence between 0-1.5 units can be separated from the rest of the observations.

Overall, it is possible to conclude that multivariate statistical techniques may be used to classify observations into groups dependent on their initial cytoplasmic fluorescence, and to separate control from IL-1 stimulated observations within the range 0-1.5 fluorescence units. Through analysing the dendrogram, bi-plot and PC1 versus PC2 plot, it can be deduced that there is evidence of partial separation of control versus IL-1 stimulated upto 3.0 fluorescence units, however the inherent variance associated with the data becomes too great beyond that. We believe that the large variability of observations is due to normal stochasticity within cell and molecular biology, and also more specifically in this case, a certain degree of over-transfection of the fluorescent protein constructs (e.g. IκBαEGFP), as pointed out by Carlotti et al. (2000) and Yang et al. (2001, 2003) because the cell effectively *stops* working. We therefore believe that in the short term, for the goal of developing an initial agent-based model of NF-κB dynamics, it would be appropriate to focus on a subset of the experimental data. As the advantage of single-cell analysis is lost if you pool the data and calculate an average, and with the results of the above multivariate statistical tests in mind, we propose that data relating to initial fluorescence <1.5 be used for development and calibration of the model. We also believe that in order to get rational results, each cell needs to form its own control (which was also the approach taken for the model of Pogson et al. (2006)), in order to eliminate the wide variations observed when averaging dynamics over multiple cells, and by implication simulations. Furthermore, it is believed that such an approach would yield more consistent results as cell time-course dynamics would be expressed as a percentage of initial fluorescence for each cell.

4.5 Summary

We believe that domain modelling is an essential part of the scoping and designing phase of simulation development within computational immunology. Although the use of UML is a tried and tested approach to semi-formally define software

implementations in industry, and indeed is a useful tool for domain modelling, it can only take us so far. Like Read et al. (2014), we have found UML to be expressive for static and relational aspects of complex biological systems, but deficient when trying to convey the stochastic and dynamic aspects of the system. Furthermore, following unsuccessful attempts to define the signalling pathway in a single UML diagram, we have concluded that the domain model should be developed using multiple *views* of the system, which build upon, and complement each other to provide the comprehensive view that we require.

Through the process of domain modelling described above, we have discovered that the diagrammatic domain model using UML, benefits from the additional perspectives that statistical analysis provides, and furthermore also benefits from the flexibility offered through other diagrammatic techniques such as the cartoon diagram, or mindmapping. We therefore suggest that on its own, UML is not enough for developing comprehensive domain models of complex biological systems, and should be complemented by other approaches.

This chapter has addressed research objective 1: Explore the role of diagrammatic and statistical techniques for developing a domain model of the NF-κB case study. Furthermore, it has contributed to the wider field of computational biology, through the critical reflections on the ability of UML to model the complex behaviours and dynamics of intracellular signalling pathways, and the way in which various statistical techniques may be used to complement UML.

5 The Platform Model

The second product of a CoSMoS project, as described by Andrews et al. (2010), is the *platform model*, which "comprises design and implementation models for the simulation platform, based on the *domain model* and research context". As discussed previously, whereas the *domain* is the general area of study, which in this case is the IL-1 stimulated NF-κB signalling pathway, and the domain model is the description of the domain that reflects the scientific basis for development of the simulation; by its very nature, the platform model is an implementation specific abstraction of the domain model, and is focused on the technical aspects of the future simulator.

Due to the incremental approach for development of the computational model in this thesis, we have attempted to capture within our domain model *all* of the components of the signalling pathway that will be relevant to us during the lifetime of this project. Conversely however, the iterative approach to development, will result in a number of iterations of the platform model, which define the functionality required of the various releases of the computational model. The initial version, represented in this chapter, will in effect reproduce the functionality of the model from Pogson et al. (2006), following reverse engineering of the published functionality of the model[1], and redevelopment using the FLAME agent-based simulation framework. The second version of the computational model will incorporate increased granularity of components at the cell membrane, to extend the scope of *in silico* experimentation to components upstream of the NF-κB signalling module.

5.1 Overview of Platform Modelling

As per the domain model, the platform model may be a collection of informal notes relating to relevant aspects of the domain, but may also include informal sketches (such as cartoons), more formal diagrams (such as those produced with UML), mathematical equations, scientific constants (e.g. biochemical rate constants), and physical descriptors (such as size, quantity, location, and speed) of simulation components and functions. Unlike the domain model, the key differentiator of the platform model is that it is required to include an implementation specific focus and should therefore document programming language specific requirements and/or workarounds, which may be required due to the programming language or computer architecture constraints during development of the simulator. For example, FLAME was developed using the concept of X-Machines, and in particular communicating stream X-Machines (see figure 5.1), therefore the platform model will need to cater for this technical framework. The platform

[1] We have been unable to gain the code for this Matlab model as the publication preceeded the now generally agreed principle of publishing your code with the empirical research paper.

model also has a wider scope, outside of the specific simulation platform, in that it also includes the concept of instrumentation, which allows observation of system dynamics (through for example a visual front-end); the extraction of simulation data from the *simulation platform*; and analysis of this data as the basis for the *results model*.

The aims and objectives of this first version of the platform model are to semi-formally document the technical requirements for the simulator, which will be developed as part of the first iteration. Briefly, this first iteration will reproduce the functionality inherent in the agent-based model previously published by Pogson et al. (2006). This model was developed at a fairly high-level of abstraction, and contained five main agent types: cell membrane receptor, NF-κB, IκBα, nuclear transporter protein, and the enzyme IKK. Unfortunately, this work was published before the widespread use of supplementary material and open access to computer programme code, and therefore access to the underlying detail of the model has been limited. Initial development of the platform model has therefore relied on the process of reverse engineering their computational model via a detailed manual walkthrough of the functionality described in the two publications (Pogson et al., 2006, 2008). As such, reproduction of functionality in the resultant *simulation platform* will be based at a qualitative level through the process of calibrating system dynamics to published results (i.e. curve fitting).

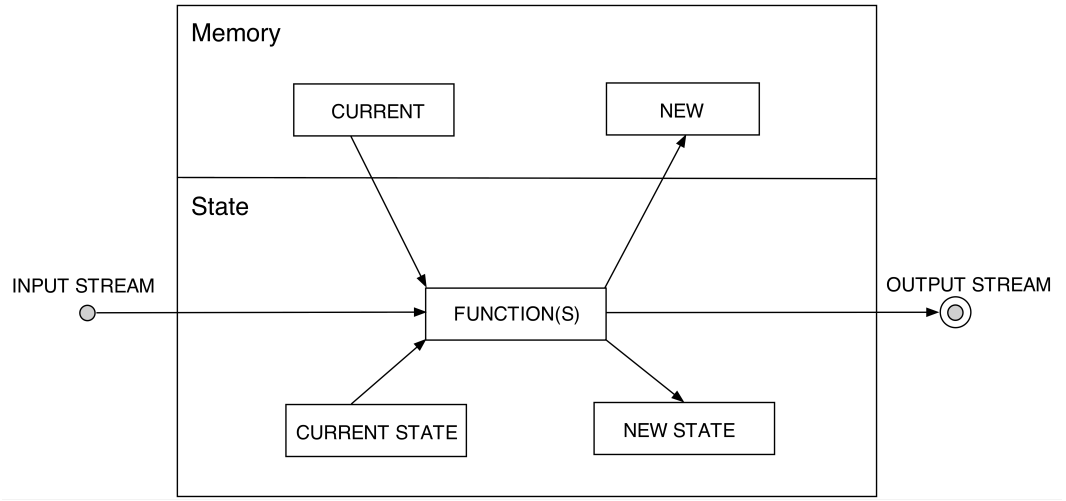

Figure 5.1: Generic communicating stream X-Machine diagram. Here it can be seen that a set of input messages are read by an agent, and state transitions may be produced using information from these messages, along with the agents current state and current memory. Following the functions being performed, the agents memory will be updated, along with its state if applicable.

5.2 The Diagrammatic Platform Model

Discrete event simulation is an approach in which the state of the computational model is updated at discrete times within the lifetime of an individual simulation. It is particularly suited to simulations with a large number of stochastic events, such as those defined in agent-based modelling. Indeed, the underlying basis of an agent-based model is the actual agents themselves, and it is through the interactions of the individual agents that the emergent behaviour of the system arises. There are five main agents (cell membrane receptor, IKK, IκBα, NF-κB, and nuclear transporter) within the simulator for iteration 1, along with a number of other components that essentially represent simulation compartments and space (cell, cell membrane, nuclear membrane, cytoplasm and nucleus). These are represented through a class inheritance diagram (figure 5.2) and a class containment diagram (figure 5.3).

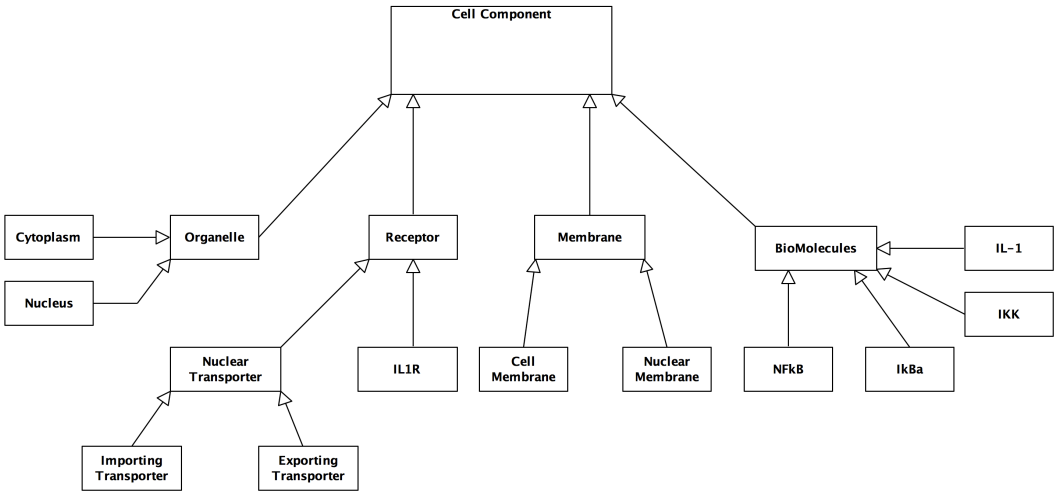

Figure 5.2: UML inheritance class diagram for the platform model. The functionality of iteration 1 can be achieved using a small number of agent types, with essentially two main categories of agents representing receptors (cell membrane and nuclear transporter) and biomolecules (IL-1, NF-κB, IκBα, and IKK). There are also environmental factors relating to membranes (cell membrane and nuclear membrane) and cellular organelles, which in this case are restricted to the cytoplasm and the nucleus.

From a technical perspective, all system components are generalisable as *cell components*, and will contain a standard set of attributes (e.g. 3D coordinates, along with movement parameters and functions). Below this system-level generalisation, there are four key categories of components, which relate to: an *organelle*, which for our purposes has been abstracted away to consist of only the cytoplasm and the nucleus; a *receptor*, which can be either the IL-1R cell membrane receptor or a nuclear transporter (importing or exporting); a *membrane*, which has been asbtracted away to only incorporate the cell membrane or the nuclear membrane; and *biomolecules*, which reflect the key signalling proteins of IL-1 extracellular stimuli, NF-κB, IκBα, and IKK.

Regarding the containment of agents, the cell membrane, nuclear membrane, cytoplasm and nucleus effectively operate as the simulation environment as they provide the necessary cell structure within which the NF-κB, IκBα and IKK

agents interact. The cell membrane receptors (IL-1R) are confined to the cell membrane, and the nuclear transporters to the nuclear membrane. The IKK and NF-κB-IκBα complex agents are confined to the cytoplasm, however free NF-κB and IκBα agents are able to move between the cytoplasm and nucleus, mediated by the nuclear transporters.

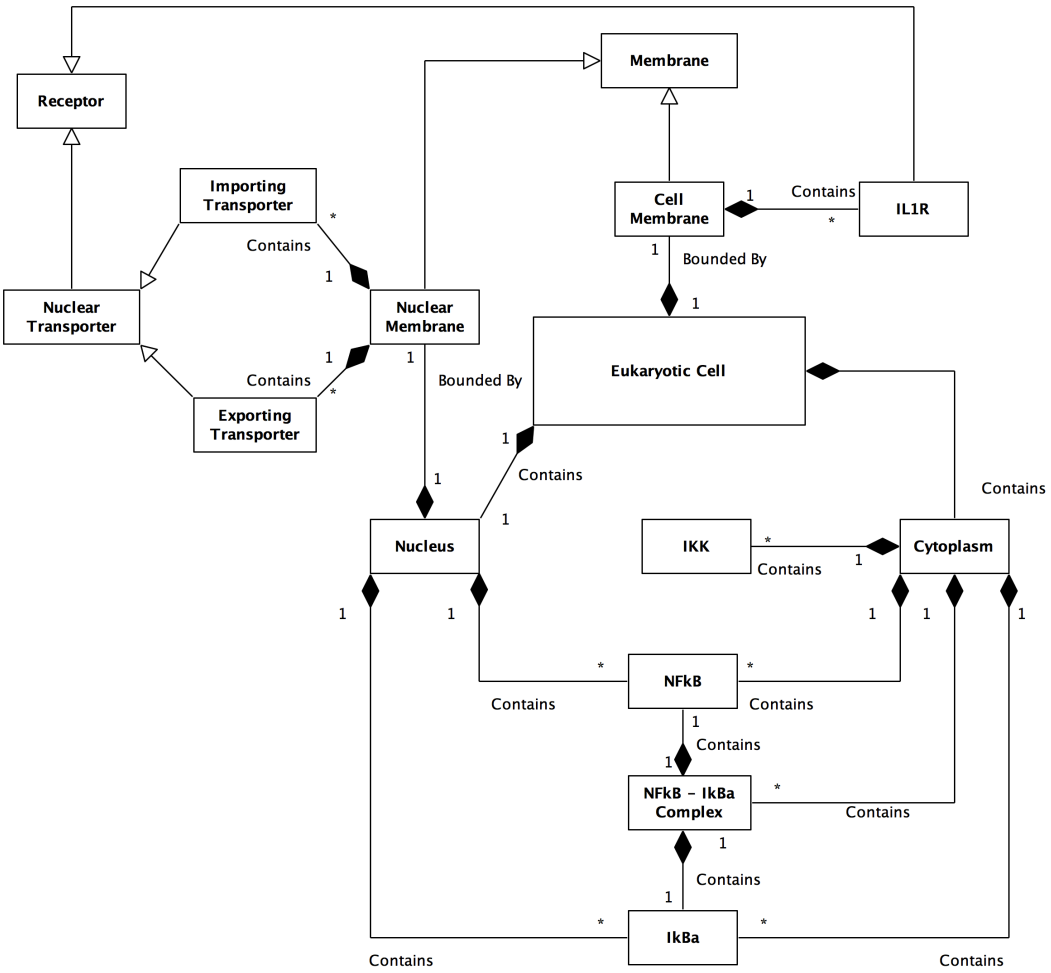

Figure 5.3: UML containment class diagram for the platform model. The cell membrane, nuclear membrane, cytoplasm and nucleus can be thought of as cellular structures within the computational model, and are static components, forming the physical environment of the computational model, within which the other agents are spatially located. The cell membrane receptor is localised to the cell membrane, and the nuclear transporters are localised to the nuclear membrane. The IKK and NF-κB-IκBα complex agents are restricted to the cytoplasm, whereas the unbound NF-κB and IκBα agents are able to translocate across the nuclear membrane and may be located in either the cytoplasm or nucleus.

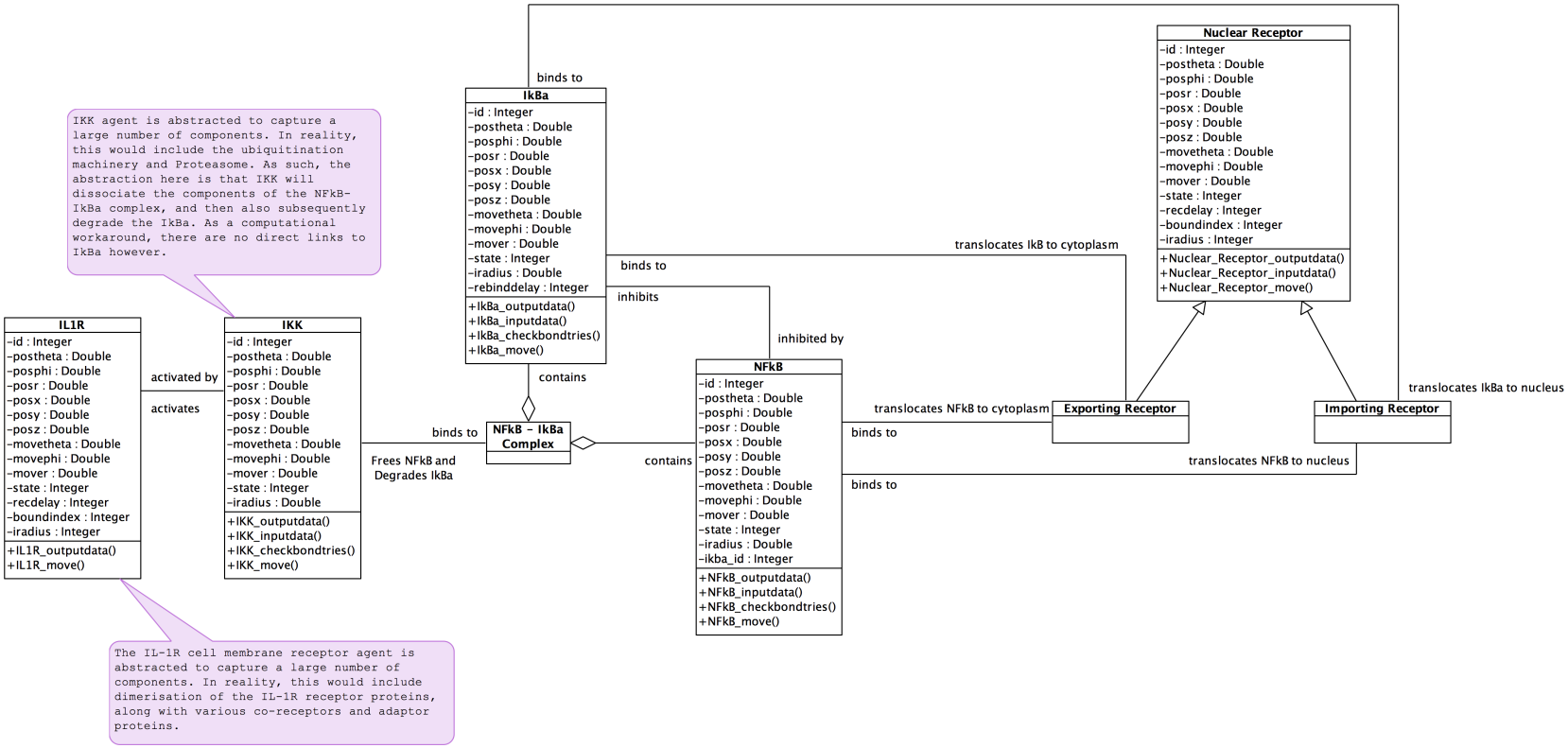

Figure 5.4: UML class association diagram for the platform model. The main static associations between different agent types is conveyed following activation of the cell membrane receptor (IL-1R) by external stimuli. The IL-1R agent type is an abstraction of biology, and is used to incorporate the IL-1R receptor, but also the co-receptor and adaptor proteins. Similarly, the IKK agent type is another high-level abstraction of the biological system, and is used to represent all intermediate components between the cell membrane receptor and the NF-κB signalling module (e.g. IRAK, TRAF6, TAK1). Following activation of IKK, it can function to dissociate the components from the NF-κB-IκBα complex, and free NF-κB agents are then able to translocate from the cytoplasm to the nucleus. Furthermore, any free NF-κB and IκBα agents are also able to translocate between the cytoplasmic and nuclear compartments.

The class association diagram (figure 5.4) specifies the high-level interactions between the different agent types within the system. This stage of the technical specification is intentionally focused on the static relationships between agent types. It can be seen that a number of components and events defined in the domain model (from chapter 4) have been abstracted away to simplify the initial computational model, and to allow focus on the dynamics of NF-κB activation and IκBα degradation. As such, we have abstracted away the detail of: the co-receptor and adaptor proteins at the cell membrane, which are captured within the IL-1R cell receptor component; the intermediate components between the cell membrane receptor and the NF-κB signalling module, which are represented by IKK (acting as a generic intermediate) component; and the transcription and translation processes, which are not modelled, as we focus on the signalling pathway up to and including the activation of NF-κB. Pogson et al. (2006) state that their model used a temporary agent for IKK that used an 'internal time delay', however there was no detail regarding the actual time of the delay, how long the agent was active, or indeed how many agents became instantiated following activation of the cell membrane receptor. We believe that a more suitable approach for agent-based modelling is to incorporate an internal time delay for activation into the individual IL-1R agents, and once active, propagate the signal transduction cascade through their probabilistic binding with IKK, which facilitates the emergent and stochastic behaviour of the system through the various downstream interactions.

As per the domain model, the order of interactions within the system is highlighted through use of sequence and communication diagrams. The sequence diagram (figure 5.5) specifies the desired system functionality under control conditions, where there is no IL-1 stimulation. As such, the dynamics provide a continuous degree of IκBα inhibition of NF-κB, a basal level of dissociation back to free NF-κB and IκBα, and a small level of shuttling of free NF-κB and IκBα between the cytoplasm and nucleus. Following the system achieving *control* starting dynamics, where the required ratio of agents and their associated states has been achieved between the cytoplasm and nucleus, a copy of the resulting XML file was taken to act as the *starting* parameters file for IL-1 stimulation simulations. Here, the association of NF-κB and IκBα, along with their basal dissociation continues, however the presence of external stimulation via IL-1 will activate the IL-1R cell membrane receptors, which facilitate the activation of IKK. Once IKK is active, it will probabilistically bind to inhibited NF-κB-IκBα complex, to release NF-κB and degrade the IκBα (see figure 5.6). The free NF-κB may then continue movement within the cytoplasm, and once it binds to an importing transporter will be translocated to the nucleus, for subsequent activation. Although UML sequence diagrams were found to have limited benefits with respect to communication diagrams during the domain modelling phase, they are more powerful when platform modelling, as they are able to convey the relevant messages sent by agents (e.g. location and finalbond) and received by agents (e.g. newbond) during each time-step of a discrete event simulation.

As previously discussed in the domain model, the communication diagram (see figure 5.7) provides a UML notation, which allows the designer more freedom in where the various components (classes) are located on the page. As such, we have positioned the various components in their approximate locations for the relevant

interactions within a cell. The functionality follows that of the IL-1 stimulation sequence diagram. An alternative view of the required system dynamics can be seen in the activity diagram of figure 5.8, which along with a focus on activities, also utilises the concept of swim-lanes to specify the spatial aspects of behaviours within the computational model.

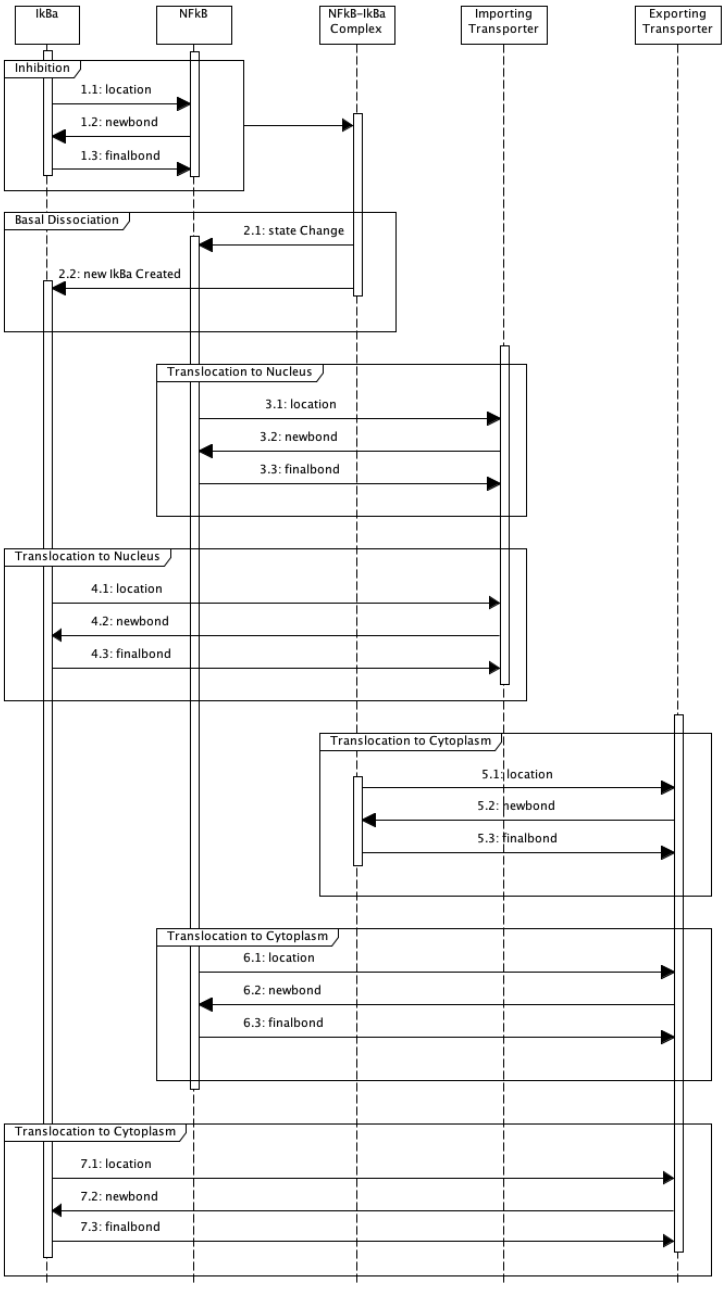

Figure 5.5: UML sequence diagram for control conditions, representing the sequential order of agent interactions for iteration 1. The system dynamics begin with all NF-κB and IκBα agents being *free* within the cytoplasm, and moving randomly in 3D space. Once a free NF-κB agent moves within the interaction radius of a free IκBα agent, a probabilistic binding can occur, which creates an NF-κB-IκBα inhibited complex. There is also a basal probability of a spontaneous dissociation back to free NF-κB and free IκBα agents. Furthermore, free NF-κB and IκBα agents may also translocate across the nuclear membrane via the nuclear transporters.

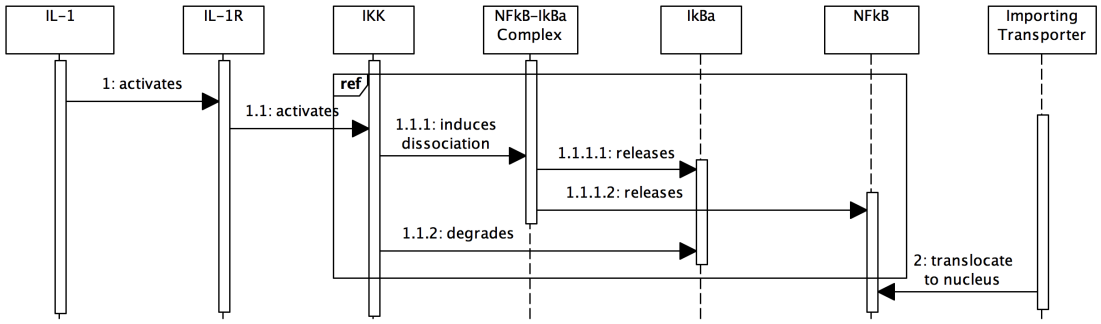

Figure 5.6: UML sequence diagram for IL-1 stimulated conditions, representing the sequential order of agent interactions for iteration 1. Activation of the system begins when the extracellular stimuli (IL-1) rises above a given threshold, which activates the IL-1R cell membrane receptor agents. We have approximated this biological behaviour through the use of a countdown timer from commencement of the simulation, which after a pre-defined period of time (number of simulation time-steps) makes all IL-1R agents active. Once activated, these IL-1R agents initiate the signalling cascade through activation of IKK, which then dissociates the NF-κB-IκBα complex, upon which the free NF-κB is able to continue random movement and ultimately translocate to the nucleus. Conversely, following dissociation, the IκBα agent becomes degraded, and is therefore removed from the simulation. In a similar way to IL-1R activation, we have also used a countdown timer to approximate the lag time before IKK activation.

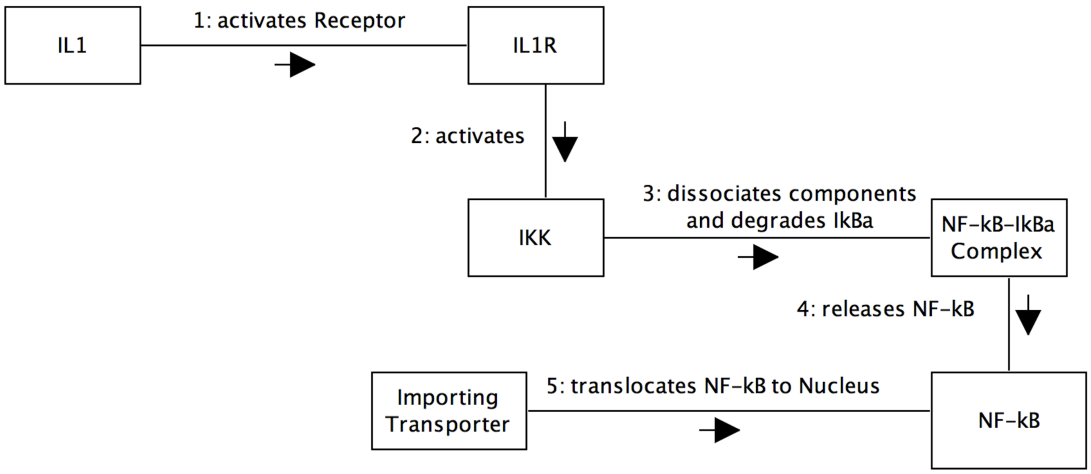

Figure 5.7: As per the domain model, a UML communication diagram can be used as an alternative representation of agent interactions within the computational model. As per the sequence diagram for IL-1 stimulated dynamics, activation of the IL-1R agents propagates a cascade of interactions, culminating in the activation of NF-κB in the nucleus, which in biology would lead to the transcription of inflammatory genes, but in our model is out of scope and therefore not represented in the platform model.

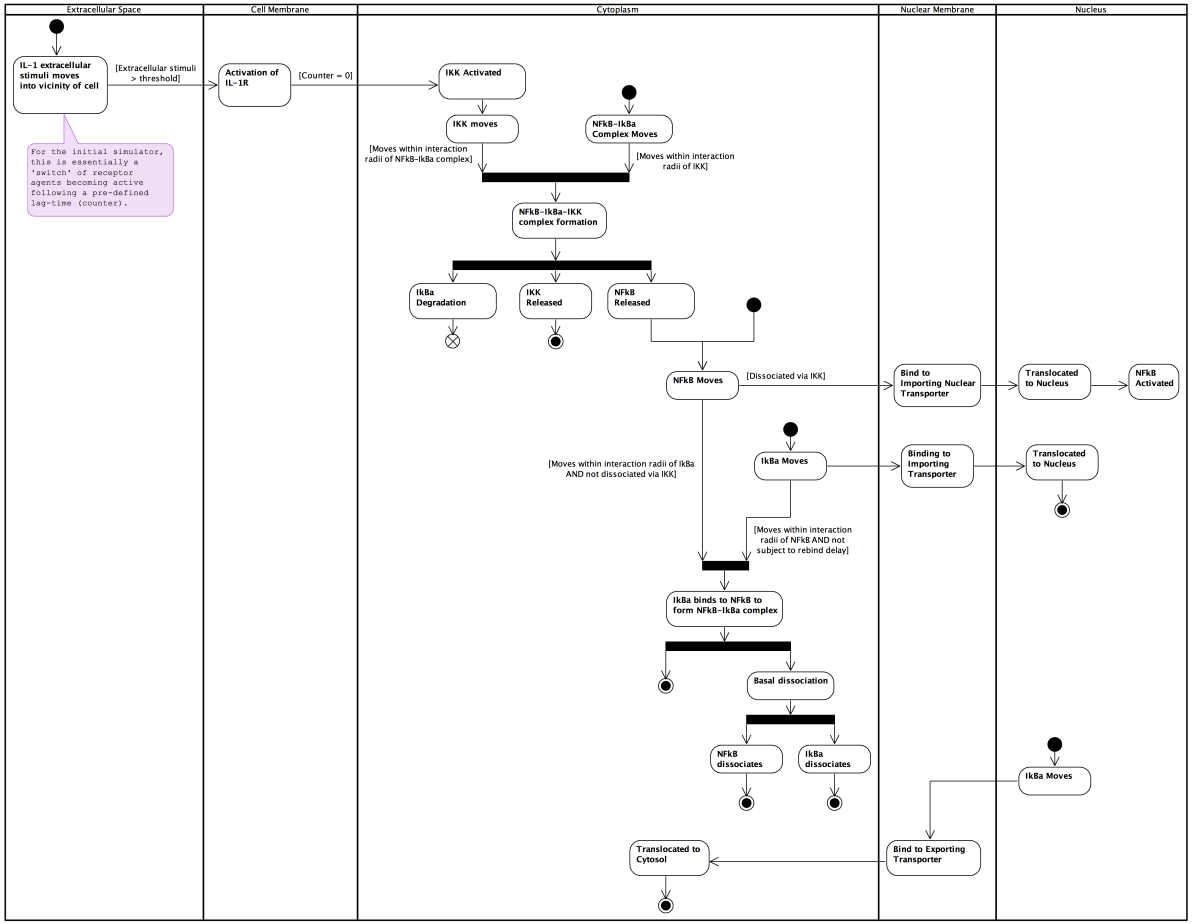

Figure 5.8: UML activity diagram utilising swim-lanes to convey the spatial location of activities within the simulator following IL-1 stimulation. This represents the full set of activities within the computational model from cell membrane receptor activation, through activation of the NF-κB signalling module, translocation of NF-κB to the nucleus, and its subsequent activation. Downstream activities following NF-κB activation have been abstracted away to allow us to focus on NF-κB activation and IκBα degradation dynamics.

Finally, the low-level detail of system components can be specified by either UML state machine diagrams or X-Machine diagrams. As we have decided to use the FLAME agent-based simulation framework, which is based on the concept of communicating stream X-Machines, we have utilised X-Machine diagrams to specify the internal states and associated messages of the agents (see figures 5.9), along with a set of linked UML state machine diagrams to represent the system as a whole (figure 5.10). This approach of modelling the low-level system detail of the signalling pathway as discrete signalling modules follows Neves and Iyengar (2002), who advocate the use of modules to focus on the receiving of inputs, processing of the relevant information, and generation of outputs.

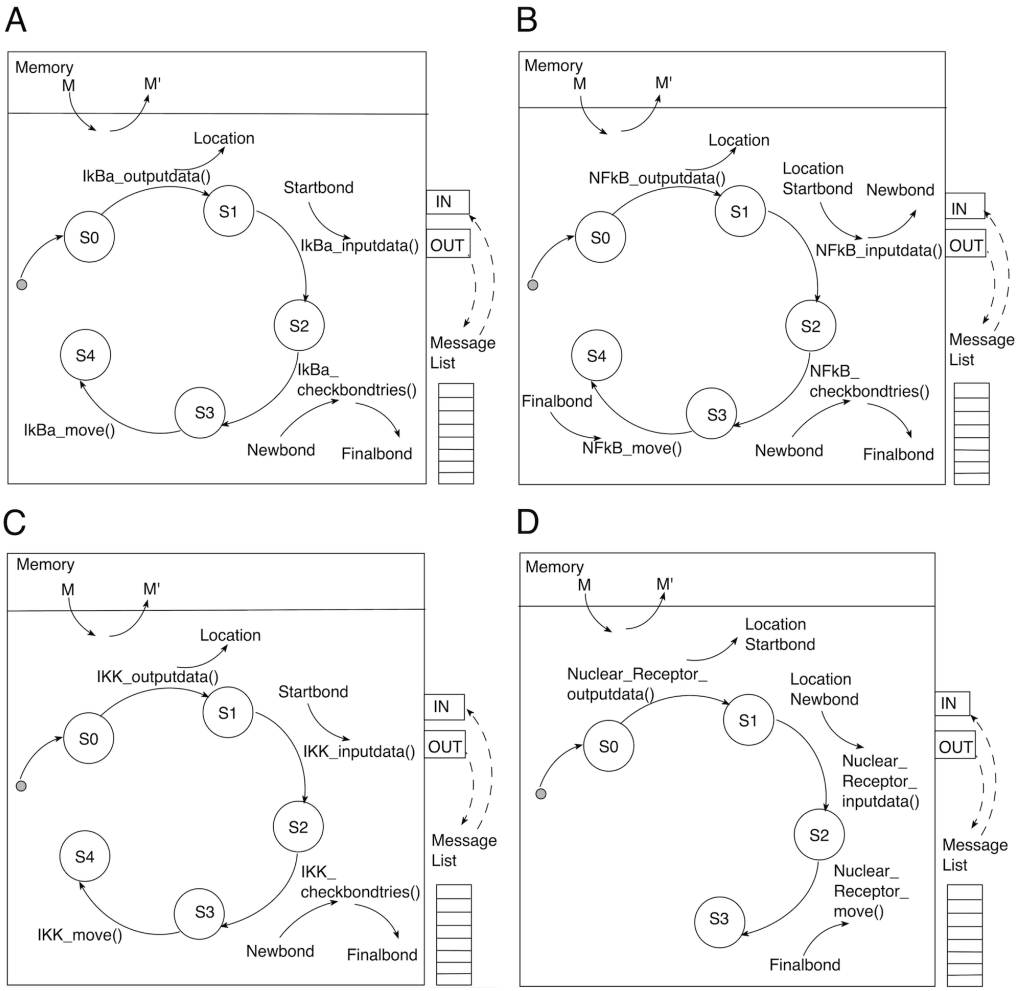

Figure 5.9: X-Machine diagrams for the platform model, showing the internal state changes of the individual agents. Each iteration within a FLAME simulation commences with internal state 0, generates a set of messages, and may potentially update its internal memory at each internal state transition. The generated messages are sent to an external message list, and may be read by other agents during the input data stage (S1 to S2) of their internal state transitions. (A) represents an IκBα X-Machine, (B) represents a NF-κB X-Machine, (C) represents an IKK X-Machine, and (D) represents a nuclear receptor X-Machine.

In general, the various agents within an agent-based model form the network of components of a communicating and processing system. These agents are able to hold information pertinent to their own status, receive external information, process this information, act upon any results that may change their status, and then send this updated information externally so that other agents may respond accordingly. As advocated by Barnard et al. (1996), a set of communicating X-Machines were constructed in a modular fashion to ensure that the various functions within the computational model can be constructed independently. This approach is further supported by the fact that communication between the X-Machines is indirect, as the modelling paradigm makes use of a centralised message board. Thus, each X-Machine can be constructed and modified without it affecting other processes/functions as long as the interfaces between components remains the same.

The diagrammatic model of an X-Machine relates to the system state transitions and not the biological (domain specific) states. These internal state transitions are via transition functions, which contain rules that change the agents memory, and send and receive messages via the input and output ports of the X-Machine. These in turn are connected to message lists that hold the messages used for communication between the agents. There is a global simulation timestep after all the X-Machines in a model have completed one transition function, with the end and start of transition functions, effectively acting as a way to synchronise the processing of X-Machines and the communication of messages.

When communicating X-Machines are used to represent agents in an agent-based model, communication is usually restricted to interactions with neighbouring agents that are located close to one another (Coakley et al., 2006b). The types of messages used within our signalling pathway model would fall into two broad categories relating to location messages and bonding messages. Location messages would be used to determine whether an agent is within the interaction zone, and once the most suitable agent within the interaction zone is identified, bonding messages would then be used to communicate this bond to the relevant agents.

An alternative representation, which incorporates the various agents, their internal states, and the messages used for communication is the X-Machine state-graph (figure 5.11). This latter diagram provides a more detailed representation of the internal dynamics of each X-Machine, and how they communicate with each other through various message types. The separation of agent behaviours into individual functions such as outputting location and state data, checking to see if a bond is available, or movement, allows the modularisation of agent functionality so that different versions can be easily swapped in and out with minimal changes needed to the rest of the model description.

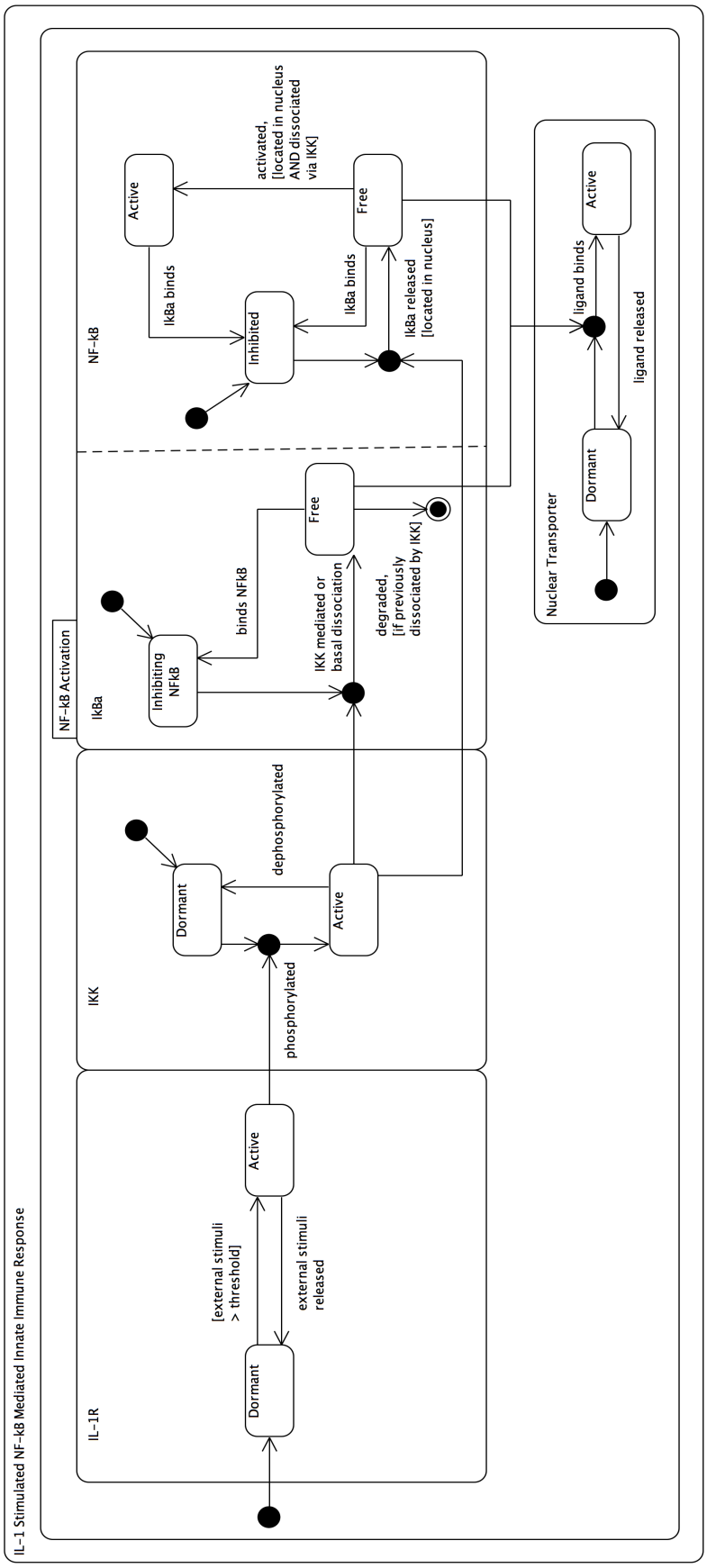

Figure 5.10: UML state machine diagram depicting the full set of detailed state changes for each agent type within iteration 1. As per the domain model, we have chosen to link each individual agent's state machine diagram, in order to form an end-to-end state machine diagram for the computational model.

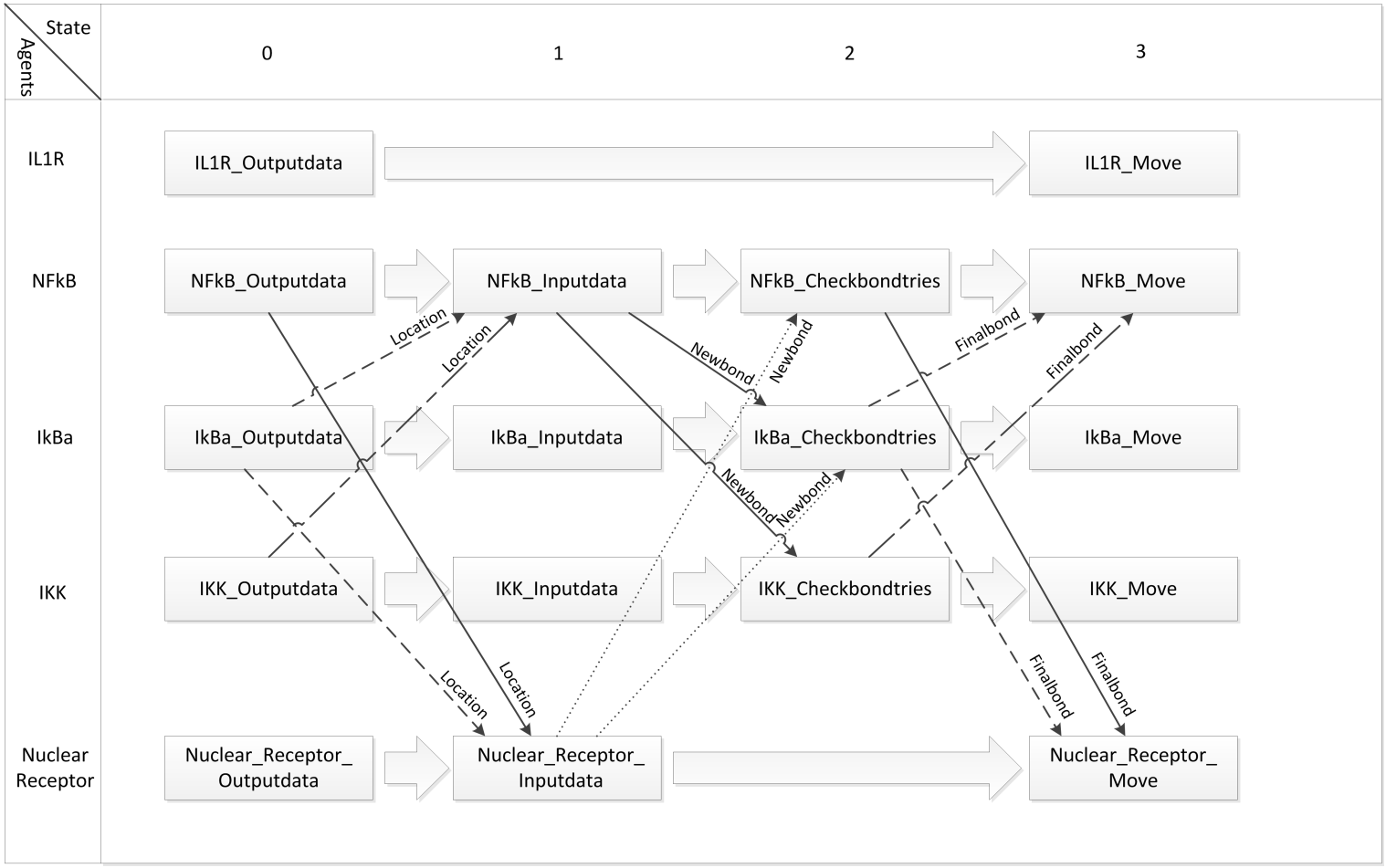

Figure 5.11: X-Machine stategraph diagram depicting a matrix of the five agent types (X-Machines) versus their internal states. Communicating stream X-Machines require that each agent (X-Machine) has a different function being performed at each internal state, which may produce messages or read messages produced by other agents, and also update their internal memory for attributes such as 3D coordinates and biological state.

5.3 Biological State Changes

Along with the internal transition states, which relate to the X-Machine underpinnings of FLAME (see figure 5.9), there are also biological states, which the agents need to represent. These are held within the *memory* of the individual X-Machines, in order for the simulation to generate emergent behaviour analogous to biology. For instance, the various agents may transition through a number of biological states, which are related to their binding status and location. A good example is that of NF-κB (figure 5.12), which can be *free* in either the cytoplasm or nucleus, inhibited by IκBα in the cytoplasm, active in the nucleus, or bound to one of the nuclear transporters. Similarly, IκBα can be free in the cytoplasm or nucleus, inhibiting NF-κB in the cytoplasm, or bound to a nuclear transporter (Baeuerle and Baltimore, 1988a; Karin and Ben-Neriah, 2000).

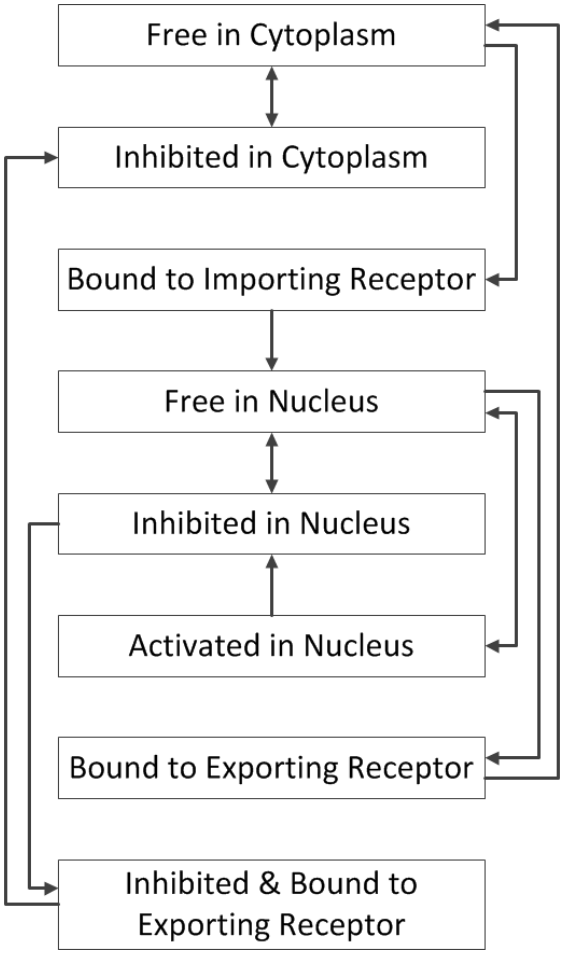

Figure 5.12: NF-κB biological state transition flowchart diagram for the platform model. It can be seen that NF-κB in the cytoplasm can be either free or inhibited, and that free cytoplasmic NF-κB can bind to the importing receptor for translocation to the nucleus. Similarly, NF-κB in the nucleus can be free, inhibited, or activated (transcribing). This time however, both free and inhibited NF-κB can bind to the exporting receptor for translocation into the cytoplasm.

5.4 Assumptions and Constraints

Along with the intracellular relationships and individual component dynamics as defined in the previous UML diagrams, there are also a number of assumptions and constraints, which underpin the platform model. These assumptions predominantly relate to the abstractions required when moving from the domain model to the platform model, but also a number of technical workarounds to make our model computationally tractable (Neves and Iyengar, 2002). The assumptions and constraints specific to our platform model are listed below:

1. We are modelling the NF-κB intracellular signalling pathway and not the actual innate immune response. We do not therefore need to cater for the actual cell being *dormant* or *active* as a high-level entity itself, because we are developing a stochastic model at the level of the individual cell, and not a multi-scale model where the status of individual cells is aggregated to form the basis of an organism-level response. *Refers back to Domain Model figures 4.4, 4.5, 4.9, and 4.17.*

2. NF-κB is modelled as a dimer (RelA and p50), as the functionality related to the individual monomers or alternative dimers between other rel family members, is not the focus of our research. *Refers back to Domain Model figure 4.4.*

3. IκBα is the only isoform of IκB that is considered in the model. Although there is natural turnover of molecules within resting cells (i.e. degradation and regeneration), this is assumed to be at equilibrium, and therefore no degradation of IκBα will occur in *control* (unstimulated) conditions within our computational model. This is in line with von Bertalanffy (1969) who advises that although chemical reactions are continually going on within the system, they are regulated by the law of mass action, so that the formation and disappearance is in balance. *Refers back to Domain Model figures 4.5 and 4.17.*

4. Only IκBα that has undergone IKK-mediated dissociation from NF-κB can be degraded. Although there is a lag-time between dissociation and degradation of IκBα due to phosphorylation and polyubiquitination, for the purposes of our computational model we will remove the IκBα agent from the system immediately, following the IKK-mediated dissociation. Furthermore, due to our focus on network dynamics up to and including NF-κB only, we do not model the synthesis (transcription and translation) of new IκBα. *Refers back to Domain Model figures 4.5 and 4.7.*

5. Only a single type of cell surface receptor is modelled (IL-1R). Dimerisation of the individual IL-1R agents, along with the binding of co-receptor and adaptor proteins is abstracted away, as we merge these processes into the single activity of cell membrane receptor activation. *Refers back to Domain Model figures 4.4 and 4.10.*

6. There are no agents for extracellular signals. We propose to use a concentration parameter that increases each time-step from the beginning of a simulation. The IL-1R agents become active once a specific threshold

concentration is reached. We therefore treat the extracellular stimuli as a whole chemical entity, where the temporal rise in its concentration effectively works as a *switch*, with a time lag before switching. *Refers back to Domain Model figure 4.10.*

7. *In vivo*, molecules will have orientation due to shape and polarity, however as per Andrews and Bray (2004), we approximate to a sphere in the computational model by using a point in 3D space that has a spherical interaction zone. This point uses cartesian (X,Y,Z) coordinates, which represent the centre of mass of the agent.

8. Constant volume is assumed for the cell environment. Most systems that we identify in systems biology correspond to some biological entity. Such entities may be an organelle, like the nucleus or the mitochondria, or it may be the whole cell. Cells and cellular compartments typically have fluctuations in their volume, however this is mathematically difficult to model (Palsson, 2011). Furthermore, as the focus of our research is the activation dynamics of NF-κB and the degradation dynamics of IκBα within the signalling pathway, there is no need to incur this level of computational overhead in our simulations.

9. The finer architecture of the cell, over and above the cell membrane, nuclear membrane, cytoplasm and nucleus is not considered. Cells are highly structured, however rapidly diffusing compounds such as metabolites and ions will distribute quickly throughout the compartment and one can justifiably consider the concentration to be relatively uniform once a steady-state has been found. *Refers back to Domain Model figure 4.2.*

10. The direction of movement of molecules within a cell may be affected by a number of factors, including chemical concentration gradients and subcellular structure. However, as our focus is on the dynamics of NF-κB and IκBα, we have decided to abstract away the detailed mechanisms, and instead approximate movement within the cell to 3D Brownian motion. Furthermore, the size of molecules is assumed to be sufficiently small that collisions between them can be neglected for the purposes of movement.

11. Cell surface receptors and nuclear membrane transporters also move, however these movements are restricted to the corresponding membrane. We approximate their movement within our computational model through random movements along an orbital plane (i.e. along the surface of a sphere).

12. The intermediate signalling components between the cell membrane receptor and the IKK molecule will be abstracted away. As such, IKK agents may be considered *intermediate* components, which link active IL-1 receptors to the NF-κB signalling module. Within biology, the signal becomes amplified as it is propagated, with each component in the pathway having the ability to interact with more than one downstream component. As such, the IKK enzyme will become activated once a given concentration threshold of upstream components has been reached. As per IL-1R activation, we also utilise a countdown timer to approximate the lag-time associated with

reaching the concentration threshold of upstream components. *Refers back to Domain Model figure 4.11.*

13. Binding occurs through agents entering each others interaction boundary (calculated as a sphere around the cartesian co-ordinates of the agent), generation of a probability (0.0 to 1.0) using a pseudo-random number generator, and comparison against a binding probability parameter that has been assigned to the agent. Binding will occur if the generated probability falls within the probability range of binding for the agents in question.

14. Both the binding of NF-κB and IκBα to form the inhibited complex, and the basal dissociation (without IKK mediation) back to free components can occur in the presence or absence of external stimuli. The dynamics of basal dissociation will be gained through calibration of the computational model against *control* observations from Carlotti et al. (2000). *Refers back to Domain Model figures 4.12, 4.13 and 4.17.*

15. All NF-κB and IκBα agents are located in the cytoplasm at the start of a control simulation. Following calibration of parameter values, movement (via 3D Brownian motion) and stochastic binding will result in the simulation reaching the required ratios of free NF-κB, free IκBα, and NF-κB-IκBα complex across the cytoplasmic and nuclear compartments. The XML output file for this iteration will be used as the starting parameters file for subsequent IL-1 stimulated simulations.

16. NF-κB and IκBα agents are able to translocate across the nuclear membrane. The dynamics of translocation during *control* conditions will be calibrated against wet-lab data. Conversely, the NF-κB-IκBα complex does not translocate into the nucleus, but instead only into the cytoplasm following inhibition of NF-κB within the nucleus. *Refers back to Domain Model figures 4.4 and 4.9.*

17. Constant temperature is assumed, as larger organisms have the capability to control their temperatures. As rate constants are normally a strong function of temperature, we have chosen to treat cells as isothermal systems, to allow a simplification whereby kinetic properties (system dynamics) emerge through agent interactions.

18. Osmotic balance is ignored. Molecules possess various physico-chemical attributes such as osmotic pressure and electrical charge, which impact the dynamic states of networks. For instance, in cells that do not have rigid cell walls (i.e. non-plant cells), the osmotic pressure has to be balanced inside and outside the cell, so that it does not split/burst. As our focus is specific to NF-κB activation and IκBα degradation dynamics within the signalling pathway, we believe it is suitable to abstract away details relating to osmotic balance.

19. Electroneutrality is assumed. Molecules tend to be charged positively or negatively. As our focus is restricted to the translocation of NF-κB and IκBα agents between the cytoplasmic and nuclear compartments, we do not

need to model the detailed physico-chemical properties of agents, and therefore assume the system remains in electroneutrality when agents translocate across the nuclear membrane.

20. System-wide simulation parameters that give rise to the emergent behaviour of the signalling pathway will be restricted to: speed of agents, basal dissociation probability of NF-κB-IκBα complex, binding probability of agents, rebind delay of nuclear transporters following translocation of NF-κB or IκBα, rebind delay of IκBα following basal dissociation, rebind delay of IKK following dissociation of NF-κB and IκBα from the inhibited complex. These will need to be calibrated as part of the development of the computational model in order to approximate qualitatively to biological system dynamics from Carlotti et al. (1999, 2000) and Yang et al. (2003).

5.5 The Cell Environment

The cell environment, within which the simulation will be situated has been abstracted to be two concentric spheres, with the spherical nuclear compartment being situated within a larger spherical cell. Both compartments are bounded by the notion of membranes, being the nuclear membrane and cell membrane respectively. The membranes not only represent physical barriers to the respective compartments, but are also where the relevant receptors, e.g. nuclear receptors (importing and exporting) and IL-1 cell surface receptors are located.

As a physical instantiation of the actual membrane in the form of an *agent*, is not possible within FLAME, we have instead developed within our computational model the notion of the relevant membrane barriers through the use of cartesian coordinates to define the surface of the spherical membranes in 3D space. A number of rules have then been developed for the various signalling components to ensure that their movements are realistic with biology, for example agents within the cytoplasm are not able to move outside of the cell (see section 5.6), or membrane-bound agents are not able to move away from the membrane surface (see section 5.7).

The volume of a sphere has been used to extrapolate the relevant radii for these two compartments so that the relative ratio of Cell:Nuclear volume is consistent with the domain model (i.e. Cell:Nucleus $\sim$20:1 for a fibroblast cell). As the volume of a sphere is defined by the equation $\frac{4}{3}\pi r^3$, and we have arbitrarily defined our cell radius as 10,000 points in simulation space, the approximate radii of the nucleus to maintain the 20:1 volume ratio, will be 3,750.

Along with the assumptions and constraints listed above (see section 5.4), there are also a number of numerical requirements for the platform model. In particular, it would be too computationally expensive to model biologically realistic numbers of agents (e.g. 60,000 RelA molecules as per the domain model), therefore we are forced to proportionately scale down the number of agents in our simulations. Furthermore, the stochasticity inherent to biological systems dictates that we do not calibrate the model to an absolute value, such as a median average, but instead calibrate within a defined range of values around the median. The mindmap (see figure 5.13) documents the relevant numerical requirements of the platform model regarding the size of the cellular and nuclear compartments

and the number of NF-κB and IκBα agents in their associated states in both *control* and IL-1 stimulated conditions. An arbitrary provision has been made for the number of IL-1R and IKK agents, as the emergent dynamical behaviour of the system will be calibrated using these numbers.

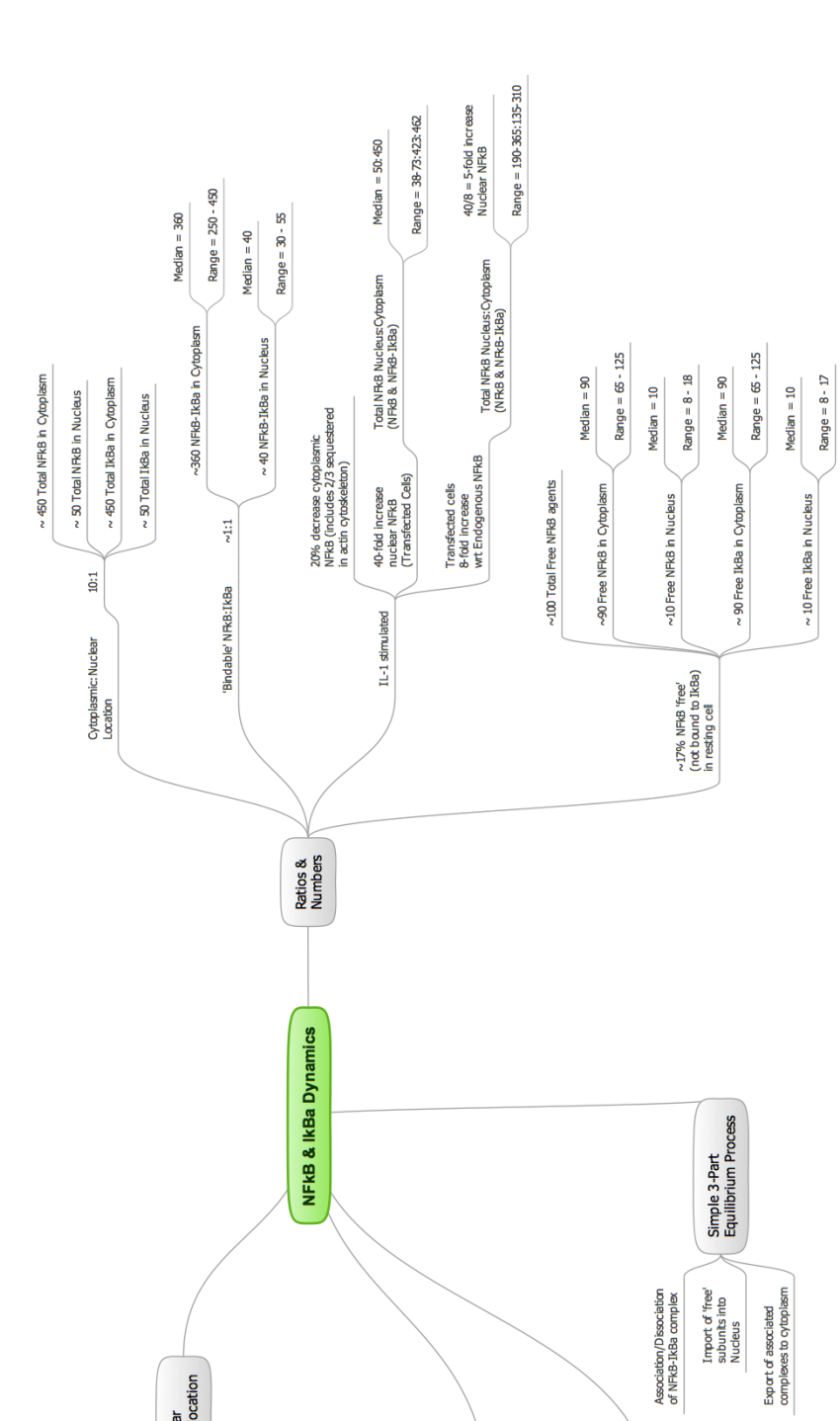

Figure 5.13: Mindmap of numerical requirements for the platform model for iteration 1. The cell radius has been arbitrarily fixed to provide a physical environment for the agent-based interactions to occur, however the ratio of cell volume to nucleus volume has been maintained from biology. Similarly, the number of agents has been scaled down to make the simulation computationally efficient, but has maintained the ratios of various agent types, their states, and their locations

5.6 Movement of Agents within the Cell

As defined above, the cell environment is abstracted to consist of just the nucleus within the cell (two concentric spheres), and therefore does not contain any of the additional internal organelles or cytoskeleton inherent to a biological cell. We are therefore also able to abstract away the movement of agents being affected by diffusion gradients or any affects by organelle structures, so that the movement of agents may approximate to Brownian motion (random movement) in 3D space (Einstein, 1905). As such, agents that are located within either the cytoplasm or nucleus can move randomly in their respective 3D compartment, however should they inadvertently move outside of the given compartment within an iteration, functionality is required to mirror their location back into the relevant compartment. For example, if an agent is within the cytoplasm, but moves outside the cell membrane boundary, we will need to mirror the coordinates back into the cytoplasm, and similarly, we will mirror an agent back into the cytoplasm if it moves into the nucleus, or back into the nucleus if it moves into the cytoplasm (figure 5.14).

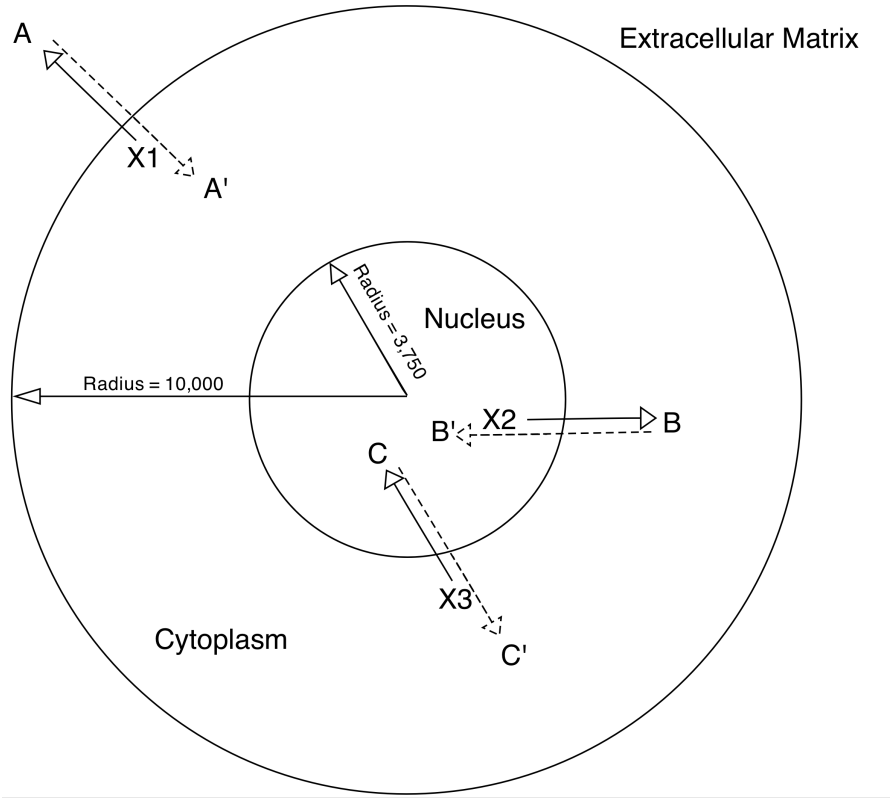

Figure 5.14: Mirroring functionality following the random 3D movement of an agent outside of its associated compartment (cytoplasm or nucleus). For example, agent X1 is an NF-κB molecule located in the cytoplasm that through random 3D movement at a given time-step has new coordinates outside of the cell (position A). We therefore need to mirror these new coordinates back inside the cytoplasm at the same angle and distance that it is outside of the cell (position A'). Similar mirroring will be required for agents at positions B and C, where they have moved from the nucleus to the cytoplasm, and cytoplasm to nucleus respectively.

5.7 3D Orbital Movement of Membrane Receptors

In a similar way to the NF-κB, IκBα and IKK agents, the cell membrane receptor and nuclear transporter agents move randomly within simulations, however this time they move randomly across a 3D orbital plane (sphere). Spherical coordinates determine the position in 3D space based on the distance ρ (rho) from the origin (i.e. radius of cell for cell membrane receptor positions) and two angles θ (theta) and ϕ (phi). Figure 5.15 depicts these relationships in 3D cartesian coordinate (X,Y,Z) space. The distance ρ (represented by the red line) is the distance from point P to the origin. If the point Q is the projection of point P to the X-Y plane (represented by the blue line), then θ is the angle between the positive X-axis and the line segment from the origin to point Q (represented by the green line). Lastly, ϕ is the angle between the positive Z-axis and the line segment (red line) from the origin to point P (Nykamp, 2011).

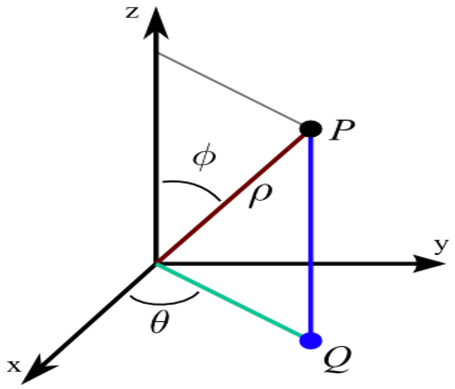

Figure 5.15: Definition of 3D spherical coordinates ρ, θ, and ϕ for a point P, which provides the underlying mathematical basis for 3D orbital movement of membrane receptors within our computational model. Used with permission of D Q Nykamp under a Creative Commons License (Nykamp, 2011).

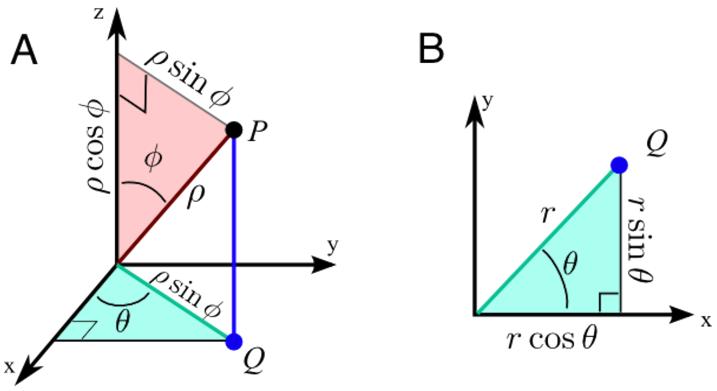

Figure 5.16: Diagrammatic definition of the process to transform spherical coordinates ρ, θ, and ϕ for a point P, to their respective cartesian coordinates X,Y,Z. Used with permission of D Q Nykamp under a Creative Commons License (Nykamp, 2011).

The relationship between the cartesian coordinates (X,Y,Z) of point P and its spherical coordinates (ρ,θ,ϕ) on a 3D orbital plane can be calculated using trigonomtery. Figure 5.16a highlights the relationship between the spherical coordinates and the cartesian coordinates. The red right-angled triangle at the top of part A is defined by the vertices at the origin, the point P, and its projection onto the Z-axis. Due to the hypotenuse having length ρ, and ϕ is the angle between the hypotenuse and the length of the side along the Z-axis, the Z-coordinate of P, representing the height of the right-angled triangle is Z = ρcosϕ. Similarly, the length of the third side of the right-angled triangle is the distance from P to the Z-axis, which is ρsinϕ.

The cyan right-angled triangle in the 3D coordinate system of figure 5.16a shows that the distance of point Q from the origin is the same as the distance for point P to the Z-axis. This is further emphasised in figure 5.16b, which defines the right-angled triangle in a 2D XY-plane, whose vertices are the origin the point Q, and its projection onto the X-axis. Here, the distance of point Q to the origin, which is the length of the hypotenuse for the right-angled triangle is labelled r. As θ is the angle that the hypotenuse makes with the X-axis, the other two sides are represented by the distances along the X-axis and Y-axis, which are the same as the X- and Y- components of the point P (from figure 5.16a). These distances are defined by X = r$\cos\theta$ and Y = r$\sin\theta$. As the distance from point P to the Z-axis and point Q to the origin (distance r) are the same distances, we can also state that r = $\rho\sin\phi$. As such, we can substitute the definitions for r, and rewrite the distances along the X-axis and Y-axis as X = $\rho\sin\phi\cos\theta$ and Y = $\rho\sin\phi\sin\theta$. Therefore, in summary, the formulae for cartesian coordinates in terms of spherical coordinates are:

$$\begin{aligned} X &= \rho\sin\phi\cos\theta \\ Y &= \rho\sin\phi\sin\theta \\ Z &= \rho\cos\phi \end{aligned}$$

5.8 Reversible Inhibition of NF-κB

In biology, when NF-κB and IκBα interact to form an inhibited complex, the individual molecules are conserved, i.e. the two NF-κB molecules (dimer) and the IκBα molecule remain within the complex. It would be ideal to use this approach within the computational model, however this would incur significant overheads due to the X-Machine nature of how FLAME functions. As such, if we were to try and mirror the biology, we would be required to link the 3D coordinates of the NF-κB dimer to the 3D coordinates of the IκBα agent. Following movement of the inhibited complex within the simulation, we would then be required to have a step that takes the new 3D coordinates of one agent and attach these to the remaining agent, to ensure that both agents within the inhibited complex (NF-κB dimer and IκBα) occupy the same point in cartesian space. As this approach would incur considerable computational overhead, we decided to simplify the technical aspects of the computational model by using the NF-κB dimer as the driving agent, so that upon binding with IκBα, its biological state is updated to *inhibited* and the IκBα agent is immediately removed from the simulation. Thus,

movement of the inhibited complex now solely relies on a single NF-κB agent, and removes the need for computationally expensive mapping of 3D coordinates between NF-κB and IκBα agents.

Due to the nature of biological reactions being reversible, we have had to build in functionality for the inhibited NF-κB-IκBα complex to spontaneously dissociate back to its constituent components, i.e. NF-κB + IκBα $\rightleftharpoons$ NF-κB-IκBα. This is straightforward for the NF-κB agent as it just requires an update to its biological state attribute; the IκBα agent however is required to be newly generated. We make the assumption here that the newly generated IκBα will initially take the same 3D coordinates as the respective NF-κB agent that it dissociated from. This does pose a small issue however, in that the individual NF-κB and IκBα agents are immediately within each others interaction zone following dissociation, so that they instantaneously bind back to the inhibited complex at the next simulation time-step. This is clearly unacceptable, so a rebind delay counter is linked to the newly generated IκBα to enable it to move out of the interaction zone of the corresponding NF-κB agent (Andrews and Bray, 2004).

Movement of molecules within the simulation is approximated to 3D Brownian motion. At each time-step within the simulation, random movement along the 3 axes within cartesian space is calculated through taking the distance to be moved across the X, Y, and Z axes, and then calculating the new 3D coordinate by adding these to the current location. Molecules are assumed to collide with the compartment boundary (i.e. cell membrane or nuclear membrane). We assume elastic collision, so that if the molecules are destined to move outside of their current compartment, they will be *bounced* back into the cytoplasm or nucleus.

As discussed previously, the individual agents are located in the simulation as points in space. In order to simulate interactions between agents, the concept of an interaction zone is used. Interactions between agents are possible if they move within each others interaction zone and are of complementary type (e.g. NF-κB and IκBα agents, but not two NF-κB agents). If more than one complementary agent is within anothers interaction zone, then the model will need a method to *choose* the most suitable agent to bind to. One approach would be to randomly choose an agent out of a set of complementary agents, and with a large number of agents this would be of little consequence, however as the initial models are expected to use in the region of 2,000 - 3,000 total agents, we need a more physically realistic *decision* process.

A spherical interaction volume is an approximation to the actual region inside which another agent must be to initiate binding. In biology, the orientation of molecules due to shape and polarity is vital for interaction, however we have deemed knowledge of the orientation of molecules to be out of scope for the purposes of our model, as it does not form the basis for any of our research questions, and can therefore be abstracted away. It could be argued that as attraction of molecules in biology will not be the same in all directions, we could utilise a randomness in the decision of interaction to account for the mutual orientation of agents, as well as how close they may have been during the simulation time-step. We have instead chosen to capture this stochastic nature through enforcing the choice of binding agent to be the *nearest neighbour*, through simple geometry calculations on all complementary agents within the interaction boundary, and the use of a probabilistic binding parameter to convey the likelihood of actual

binding. Additionally, the size of molecules is assumed to be sufficiently small that collisions between them can be neglected for the purposes of movement. This allows us to calibrate simulation dynamics to the wet-lab data of Carlotti et al. (2000, 1999) and Yang et al. (2003) based on interactions at the level of the agent, and negate the need to enforce complex physical rules for binding at the level of individual amino acids within the agent.

5.9 Instrumentation

Development of an agent-based computational model is a major necessary step, however on its own, it will not allow us to perform *in silico* experimentation. In order to run simulations and analyse the resultant data, we also require development of a number of peripheral computational tools (termed *instruments*). For this project, we require the following instrumentation tools to be developed: Python[2] scripts to allow the generation of XML starting parameters files for running simulations in FLAME; Ruby[3], MS DOS batch, and Unix shell scripts to allow the submission of simulation runs to computing resources; Python and Matlab[4] scripts to transform and process output XML files to CSV files; Matlab and R[5] scripts to analyse the data and automatically generate various graphs; a visualisation front-end to provide an animated view of system dynamics over time; and Python and Matlab scripts to analyse the output data for statistical significance using the Kolmogorov-Smirnov test (Massey, 1951), and effect magnitude using the A-Test (Mann and Whitney, 1947; Vargha and Delaney, 2000) (see Appendix B for supporting material on these two statistical techniques). These instrumentation tools along with the actual agent-based model are represented by the UML Package diagram (see figure 5.17).

5.10 Summary

As per the domain model, class diagrams were used in our platform model to represent the containment, inheritance and association characteristics of agents. These UML diagrams are now much more useful (than when used in the domain model) as they are able to convey technical specifications for the computational model. The developed class diagrams are intentionally not fully complete (with respect to the UML standard) as they omit a number of key details, such as multiplicity (i.e. how many of each agent will be used within simulations), constraints, and dependencies, but this is to ensure the readability of the diagrams, with the additional information being conveyed elsewhere within the platform model. Regarding the containment of agents, the cell membrane, nuclear membrane, cytoplasm and nucleus, effectively operate as the simulation environment as they provide the necessary structure, within which the NF-κB, IκBα and IKK agents interact and yield the emergent behaviour of the signalling pathway.

The order of interactions within the system has again been documented through

[2]www.python.org
[3]www.ruby-lang.org/en/
[4]www.mathworks.co.uk/products/matlab
[5]www.r-project.org/

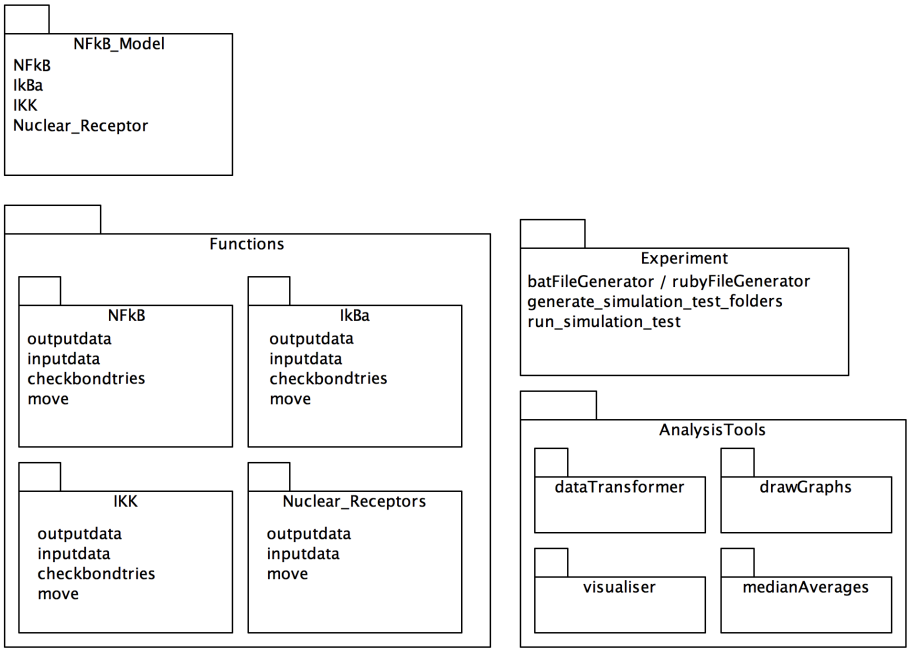

Figure 5.17: UML package diagram for the platform model, representing the four agent definitions within the agent-based model, along with the associated instrumentation tools for generating parameter files, running simulations, transforming and analysing the output data, and visualising the simulation.

UML sequence, communication and activity diagrams. State machine diagrams were also used to express the detailed biological state changes of individual system components, however as we will be using the FLAME simulation framework to develop the computational model, these were complemented with X-Machine diagrams to express the detailed internal state changes of individual system components. Development of these diagrams has raised a number of key questions around the temporal dynamics of the system, in particular the lag times between cell membrane receptor activation and dissociation of the NF-κB-IκBα complex, the degradation of IκBα, and the subsequent translocation of NF-κB to the nucleus. These will be resolved through the calibration exercise, once the compuational model has been developed.

As found when developing the domain model, we believe that the activity diagram with swim-lanes has been the most useful notation for conveying the technical specifications of the system regarding the consequences of interactions between components, and that state machine diagrams are the most useful notation for defining the technical specification of individual system components. Unlike the domain model, the platform model also includes implementation specific details and as per Andrews et al. (2008) we have found it useful to document the various assumptions and constraints (see section 5.4) regarding the technical scope of the computational model. These bullet points, along with the UML diagrams (in section 5.2) and mindmap (in section 5.5) define the most recent platform model for iteration 1. However, as the basis of the CoSMoS process is an iterative approach to design and development, future iterations that augment the functionality within the computational model will require the platform model to be updated.

One of the key strengths of the CoSMoS process is the advocation of separating the abstracted view of biology (documented within the domain model) from the technical specification of the computational model (documented within the platform model). This separation ensures the abstracted view of biology and the technical specifications of the system remain discrete models, and thus aims to minimise confusion during the development of the computational model around what aspects of the programming code relate to biology requirements, and what aspects are necessary as technical workarounds due to constraints of the specific programming frameworks being used (e.g. communicating X-Machines and FLAME). As such, we believe the process of platform modelling to be an integral part of the development lifecycle for computational models of biological systems, and believe that the platform model presented in this chapter will provide an unambiguous specification for the *simulation platform* which is developed and calibrated in the next chapter.

6 Development and Calibration of the Simulation Platform

The third product of a CoSMoS project, is the *simulation platform*, which "encodes the platform model into a software and hardware platform upon which simulations can be performed". Whereas the platform model is the implementation specific abstraction of the domain model, which in this case was developed using UML and other diagrammatic and statistical approaches, the simulation platform is the instantiation of this in actual code. The completed simulation platform will encode the default settings for the architecture and parameters, to realise the *control* (unstimulated) and IL-1 stimulated dynamics of the IL-1 stimulated NF-κB signalling pathway, as abstracted within the platform model for iteration 1. This chapter will develop a computational model (research objective 2) that is calibrated against biological data (research objective 3) and provides functionality to facilitate *in silico* experimentation around the interactions of various signalling pathway components, and how these give rise to the emergent behaviour of the system (research objective 4).

6.1 Development of the Simulation Platform

The agent-based computational model was developed using the FLAME simulation framework, according to the technical specification defined in the platform model (see chapter 5). The FLAME simulation framework has been developed in line with the agent-based modelling paradigm, by utilising the concept of X-Machines to represent logical entities (the agents) within the programme. Furthermore, through the use of communication streams between these X-Machines, they are able to maintain their own individual states. As discussed in section 2.6.3, FLAME agents are defined within an XML template, which specifies the agent attributes and internal (X-Machine) states, and requires an additional set of C files, where the rule-based functionality associated with agent interactions is defined. Following model specification within these templates, the FLAME modelling framework parses the XML and C code to generate executable simulation code (see figure 2.4). The FLAME simulation framework also provides a simulation engine that manages the execution of simulations, and the interactions between the X-Machine agents (through a centralised message board). Time is discretized into *time-steps*, and within every time-step the individual X-Machines iterate through their internal states (as per figure 5.9), which may culminate in interactions with other X-Machines, and updates to their biological states.

Two main technical issues were encountered during the development of the computational model, which were due to the background design decisions taken during the development of the FLAME simulation framework (Coakley, 2007) and the associated mechanisms for processing the internal state transitions of the X-

Machines. The first issue related to the efficient handling of basal dissociation of the NF-κB-IκBα inhibited complex back to free NF-κB agents and IκBα agents. The second issue related to the use of a pseudo-random number generator, and the need to set an associated *seed value* at the simulation level. These two issues will be discussed in turn, below.

Basal Dissociation Issue

An agile approach to development (Janzen and Saiedian, 2005; Shore and Warden, 2008) was used to increase the functionality of the computational model in small incremental steps. Although minor issues relating to the normal coding lifecycle were encountered (and quickly resolved) along the way, a major issue was encountered when functionality was added for basal dissociation of the NF-κB-IκBα inhibited complex. For computational efficiency reasons, upon binding of IκBα with NF-κB to form the inhibited complex, instead of the IκBα agent having its 3D cartesian (X, Y, Z) coordinates set at each time-step to mirror that of the the relevant NF-κB agent, we instead chose to update the status of the NF-κB agent to be that of the inhibited (bound) complex, and removed the IκBα agent from the system. This results in a reduction in the total number of individual agents within the system, but as the new status of the NF-κB agent models that of the inhibited complex, it can be argued that the total number of NF-κB and IκBα *molecules* within the system from a biological perspective remains consistent, and therefore respects the principle of the conservation of mass (Beard et al., 2004). Functionality to provide a small degree of basal dissociation of the complex following its formation, required the updating of the NF-κB agent's state from *inhibited* to *free*, via a straightforward update to the X-Machines internal memory, but it was necessary to introduce *new* IκBα agents back to the simulation to model the release of inhibitor back in to the system.

The simplest approach to provide this functionality would be to use a global counter, initially set to the highest agent ID of the set of IκBα agents at the start of a simulation. Upon inhibition of NF-κB by a free IκBα agent, the NF-κB agent state is updated to reflect that it is now *inhibited* and the corresponding IκBα agent is removed from the system. Upon basal dissociation, the NF-κB agent would return to a *free* state, and a new IκBα agent would be added back in to the simulation, with an agent ID equivalent to the global counter value incremented by one (for this new agent). Unfortunately, as FLAME was developed to run across multiple clusters using parallel processing functionality, the concept of a *global mutable parameter* does not exist, as this would make the running of simulations extremely inefficient due to the issue of concurrency locking[1].

We were therefore unable to use a global counter for generating the ID of the new agents, and therefore tried to generate the ID using a pseudo-random number generator, which generated numbers above an arbitrary value of 5,000,000 to ensure we did not overwrite existing agents[2]. Unfortunately, due to the number of

[1]Individual nodes may be dependent on the current value of a global parameter, which may be being processed by another node. This would require nodes to *wait* until they ascertained the global parameter was not being processed, thus incurring a significant overhead regarding the wallclock time of simulation time-steps, and offsetting the benefits of parallel processing.

[2]FLAME requires that each agent within a specific agent type (e.g. IκBα) possesses a unique agent ID. If a new agent is created with the ID of an existing agent, the original agent becomes

IκBα agents within the simulation, and the number of iterations required to reach steady-state-like dynamics, the calibration exercise for basal dissociation yielded a large number of basal dissociation reactions. These basal dissociation reactions, in turn generated a large set of new IκBα agents. As discussed above, the IDs of these new agents were generated using a pseudo-random number generator, however due to the large number of basal dissociation reactions per simulation time-step, there was a correspondingly large set of random numbers generated, which resulted in the generation of duplicate IDs. This had the effect of overwriting a number of the newly generated IκBα agents, therefore reducing the total number of IκBα agents over time, as they were essentially removed from the simulation. We therefore encountered an issue regarding the conservation of mass, as the dynamics of the computational model resulted in a reduction in the total amount of IκBα under control (non-stimulated) conditions (see figure 6.1). This issue was finally resolved by adding a local attribute to the NF-κB definition, which would store the relevant ID of the IκBα agent that it became inhibited by. If basal dissociation was to occur under this scenario, the stored ID could be retrieved and assigned to the newly generated IκBα agent.

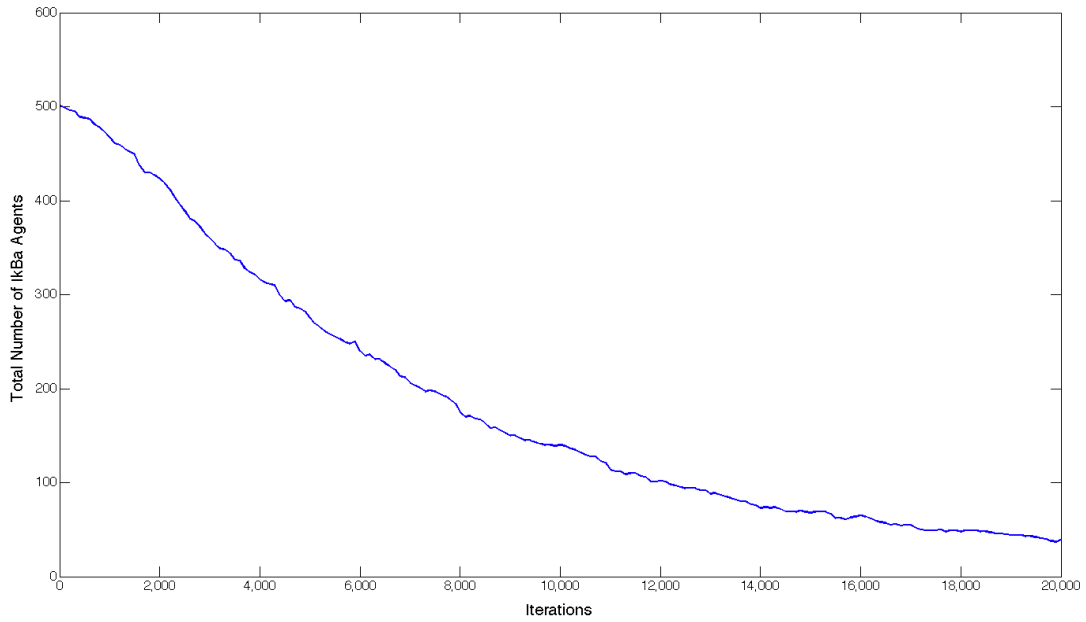

Figure 6.1: IκBα Conservation of Mass Issue. Due to the FLAME simulation framework being developed for a parallel processing architecture where simulations are run over multiple nodes, we were unable to use a global mutable parameter and therefore had to find a workaround for generating new agent ID values. The initial workaround, which used a pseudo-random number generator, resulted in duplicate IDs being generated, and the overwriting of newly created agents (their removal from the simulation).

overwritten, and thus removed from the system.

Pseudo-Random Number Generation Issue

The use of a Pseudo-Random Number Generator (PRNG) is required to build in probabilistic behaviour, such as agent interactions, to our computational model. Furthermore, in order to ensure our computational model is not overtuned to the data during the calibration exercise, we will run multiple replicate simulations that utilise different PRNG seed values, in order to generate the stochasticity over multiple cells (as per figures 4.18 and 4.19). The theory behind setting a seed for a pseudo-random number generator (PRNG) is that you set once, and use many times (Barry, 2011; van Niel and Laffan, 2011). Unfortunately, FLAME does not allow you to do this very easily, as the only way to set the PRNG seed is within the definition of an agent (the C functions file). As discussed above, due to the parallel nature of FLAME, there is no opportunity to set a global constant from within a simulation run, as it would be inefficient for this to propagate across the multiple nodes within the computer cluster/grid. As such, not only can you not update a global variable, such as total counts, but you are unable to set the PRNG seed at the global level. You therefore encounter the issue of either not being able to set your own PRNG seed, and thus relying on the seeds generated using the system clock, or to develop a workaround where the PRNG seed is set within an individual agent's definition (see figure 6.2).

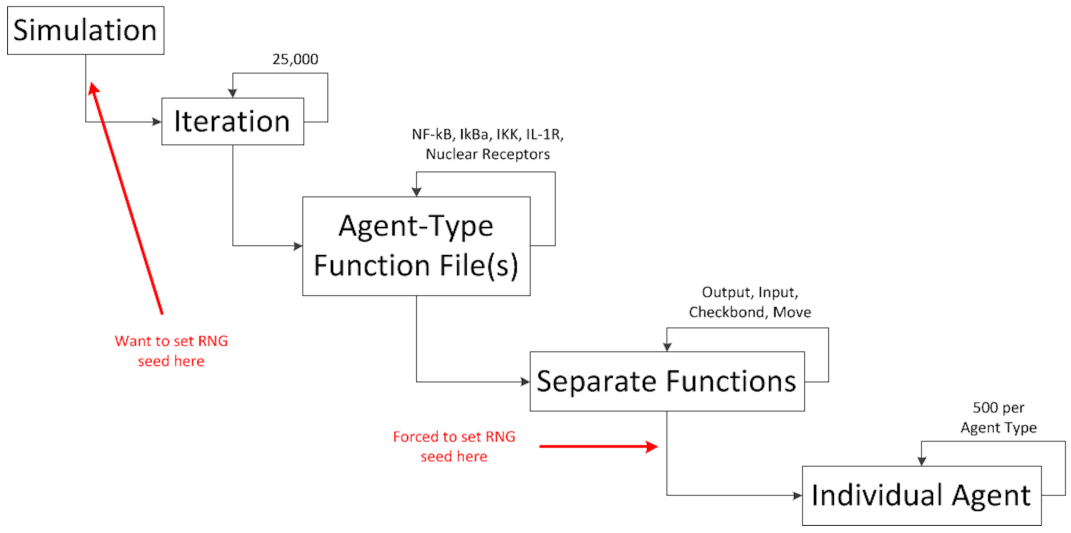

Figure 6.2: The levels at which the pseudo-random number generator seed value can be set within FLAME. Although we require to set the seed value once per iteration, but use many times for the generation of pseudo-random numbers, FLAME unfortunately does not allow the concept of a global mutable parameter. As such, FLAME requires all functionality to be written at the level of agents, be that agent-types, i.e. function called once per agent-type per iteration, or indivdual agents, i.e. function called once for each agent and for each iteration.

Unfortunately, by setting the seed at the level of an agent function, you encounter the issue of resetting the seed each iteration, and thus do not gain any real stochasticity. This is because the process of setting the seed value generates a deterministic (hence the use of the prefix *pseudo*) list of *random numbers* to

be used within probabilistic functions (van Niel and Laffan, 2003). The resetting of the seed value within an agent function (with the same seed value), therefore results in the same pseudo-random number list being generated as per the previous iteration, which is further compounded by the simulation returning to the beginning of the list, and not commencing from where you got up to previously. For example, when setting the seed value at the level of an individual agent for a simulation that contains 1,000 NF-κB and IκBα agents over 25,000 iterations, you effectively reset the seed value 25,000,000 times. Similarly, when setting the seed value at the level of an agent type, you effectively reset the seed value 50,000, which although orders of magnitude smaller, still affects the ability of the PRNG to incorporate stochasticity into the simulation.

One approach to gain stochasticity between simulation runs is to run simulations in *Production mode* within FLAME, which uses the system clock to generate the seed value. This is not ideal however, as we will not be able to exactly reproduce individual simulation runs (for example, when debugging during model development), as the system clock is perpetually changing with the progression of time. The only workaround that was found to resolve this issue was to create a *dummy* agent as part of our system design, which sets the PRNG seed value in the first iteration of an individual simulation. This functionality was achieved by utilising an agent-level counter (with respect to the iteration number), and logic to set the PRNG seed value when the counter equals zero, but does nothing when the counter is equal to 1 or more. Furthermore, to keep the simulator tidy of computational artefacts, the agent is also removed from the simulation once the counter reaches a pre-specified number (see figure 6.3).

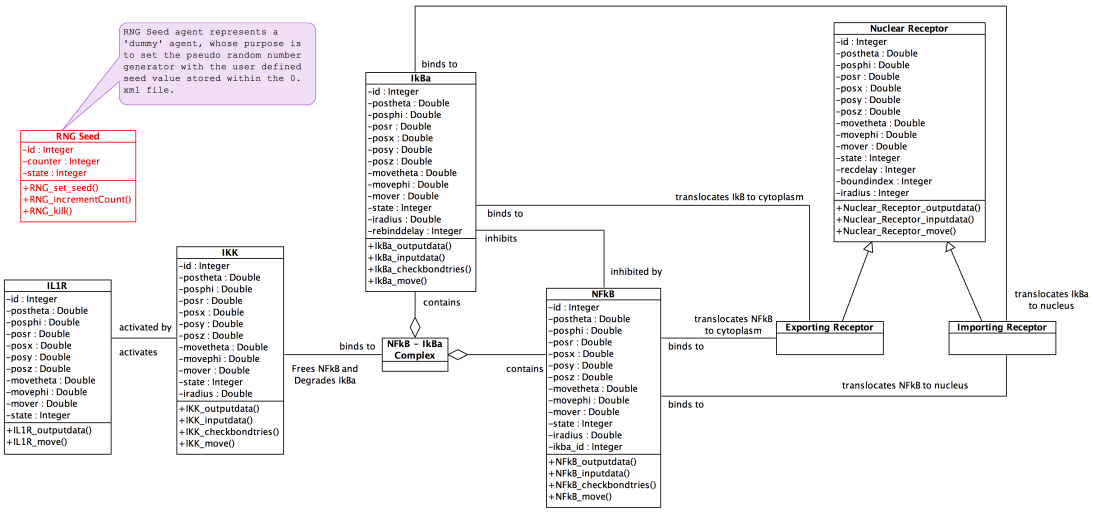

Figure 6.3: Updated UML class association diagram, which represents the addition of the dummy agent as a separate class (in red), which is not associated to any other classes. This dummy agent has the sole purpose of setting the PRNG seed value in the first iteration of a simulation, and then disappears from the simulation.

6.2 Calibration of Simulation Platform for Control Dynamics

Before *in silico* experimentation can begin, the various parameter values, such as rate constants (e.g. speed of agents), concentrations (i.e. number of agents in cytoplasm and nucleus), and probabilities of particular events (e.g. binding, dissociation and translocation) need to be estimated. This step, termed calibration, is an important activity within simulation platform development, as it aims to align the simulation dynamics with the emergent behaviour of the underlying biological system.

6.2.1 Calibration Process

Andrews et al. (2010) advise that in the context of a CoSMoS project, calibration is achieved through a series of *in silico* experiments that adjust the parameters and mechanistic behaviours of the simulation platform, in order to discover the relationships between the simulator and the behaviours observed in the real-world domain. The calibration of our agent-based model of the NF-κB signalling pathway was a collaborative effort between the modeller (Richard Williams) and the domain expert (Eva Qwarnstrom). The computational model was calibrated against published literature and the domain expert's understanding of the IL-1 stimulated NF-κB signalling pathway. The contribution of the domain expert in this calibration activity has been essential to ensure the simulation platform is not only grounded in the domain, but also adheres to the scope of the domain and platform models. As discussed by Kirschner et al. (2007), there are several approaches for estimating the parameter values of computational models during the calibration process: 1) direct experimental determination of a parameter; 2) simultaneous estimation of several parameters at once by fitting experimental data to a model; and 3) estimation of a parameter based on known values for a similar system.

The calibration process that we utilised to align the behaviour of our computational model with that of the underlying biological system used a mixture of the three approaches discussed by Kirschner for parameter value estimation. Environmental parameter values such as cell and nuclear diameter have been approximated from the literature. A number of parameter values regarding agent interactions were arbitrarily set, such as the interaction radius within which agents need to enter before they are eligible for probabilistic binding, and the delay applied to nuclear receptors following translocation of an agent before they may translocate another agent. Additionally, the parameters that we believe are fundamental to the emergent behaviour of our computational model were estimated through a process of varying parameter values during multiple simulation runs, until simulation dynamics approximated (through qualitative curve-fitting) to those of the wet-lab data of Carlotti et al. (2000) and Yang et al. (2003). Calibration through varying parameter values, was therefore iterative, with each parameter being the focus of investigation, and required consultation between the modeller and domain expert until a parameter space was found that yielded qualitative alignment to the underlying biological data on which the computational model was designed.

The six key parameter values that are required for generating an approximation of simulation results to wet-lab biology are:

1. **Differential Time** - The multiplication factor used within the system to calculate the distance moved by agents per time-step.
2. **Rebind Delay** - The time delay introduced to an IκBα agent following its basal dissociation from an NF-κB agent. This ensures that the agents have sufficient time to move outside of each others interaction boundary, so that they do not rebind at the next time-step.
3. **Association Rate** - The probability of binding following the movement of an agent into the interaction boundary of a complementary agent.
4. **Basal Dissociation Rate** - The probability of spontaneous dissociation of an NF-κB-IκBα complex to form free NF-κB and free IκBα agents.
5. **Nuclear Import Rate** - The probabilistic translocation of free NF-κB and free IκBα agents from the cytoplasm in to the nucleus.
6. **Nuclear Export Rate** - The probabilistic translocation of free NF-κB and free IκBα agents from the nucleus in to the cytoplasm.

Stepney (2012) advocates the need for a translation step in ascertaining values required for calibration of computational models. For example, the domain model may comprise both parameter values and experimental data (d_i) derived from wet-lab experimentation, which following scientific analysis yields domain results data (d_r) (see figure 4.17 domain model mindmap). To move to the simulated world, the domain model data needs to be translated to appropriate simulation platform values s_i through the generation of a platform model (see figure 5.13 platform model mindmap). A simulation experiment, given input data s_i, and using the simulation parameters, functions and methods (T_{ds}) called within the computational model during simulation runs, will produce raw simulation results data s_r. Through the use of multiple simulation replicates, this raw simulation results data is then processed to generate simulation output data (s_o), which could comprise of the median averages of the individual s_r data. The calibration exercise is to adjust the simulation parameters and translation functions T_{ds} to achieve an approximation of domain results data to simulation results data ($d_r \sim s_r$). The relationship between domain and simulation results does not need to be exact, because a certain degree of variation is required to produce the stochasticity inherent to biology; it does need to be statistically similar, or show qualitative agreement however. The approach taken here is qualitative in nature, as we are essentially curve-fitting the simulation results to that of the wet-lab data. For example, following calibration to control dynamics, we would need the NF-κB and IκBα agents apportioned between the cytoplasm and nucleus in their associated biological states, as shown in the translation table below (table 6.1). Additionally, as expected for control conditions, all of the IKK agents will all be located within the cytoplasm in an inactive state; all of the IL-1R receptor agents will be orbiting the cell membrane in an inactive state; and all of the nuclear receptors (both import and export) will be orbiting the nuclear membrane.

Agent	Location	d_r	s_i	$\widetilde{s_r}$	s_r Range
Free NF-κB	Cytoplasm	9,000	498	90	65 - 125
	Nucleus	1,000	1	10	5 - 20
Free IκBα	Cytoplasm	9,900	496	90	65 - 125
	Nucleus	1,100	2	10	5 - 20
NF-κB-IκBα	Cytoplasm	45,000	3	360	250 - 450
	Nucleus	5,000	0	40	30 - 55

Table 6.1: Calibration translation table as advocated by Stepney (2012). This translates/maps: 1) the approximate numbers and locations of components from wet-lab experimentation, which forms the domain results data d_r; 2) the initial simulation platform values (s_i), which are used to commence simulation runs during the calibration process; and 3) the desired simulation results data s_r (median average and range), which needs to be attained during the calibration process. Values for d_r are extrapolated from Carlotti et al. (2000) and values for s_r were calculated as part of platform modelling.

6.2.2 Differential Time

There is a global constant within the model which relates to the speed of movement of non-membrane bound agents, i.e. IKK, NF-κB and IκBα agents. Figure 6.4 illustrates how the speed at which agents move within a single time-step can affect the behaviour of the system as a whole. Agents moving at high speeds may pass through the interaction boundary of reciprocal agents within a given time-step, and therefore reduce the likelihood of inhibitory and translocation interactions, thus severely affecting the emergent behaviour of the system.

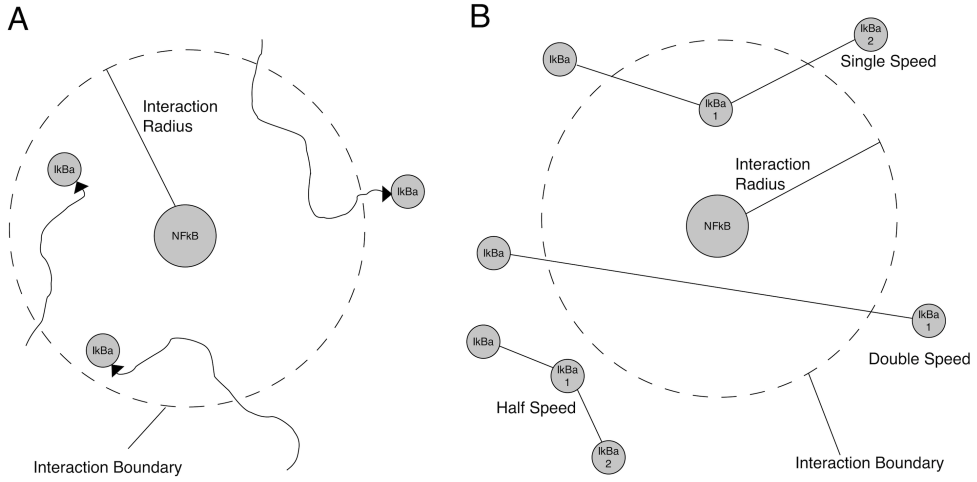

Figure 6.4: Cartoon illustration of the affects that the speed of agent movement within a given time-step (differential time) may have on the dynamics of the system as a whole. **A** shows an abstracted view of *normal* behaviour in the biological system. **B** shows an abstracted view of movement within our computational model, using IκBα agents at three different speeds. An IκBα agent with single DT may move into the interaction boundary of a reciprocal NF-κB agent within a given time-step, and if it does not probabilistically bind, may move outside of the boundary at the next time-step. Conversely, an IκBα agent with a smaller or higher DT, may not enter the interaction boundary, or may pass straight through, within a single time-step.

One-at-a-time factorial analysis (Cuthbert, 1973) was performed to understand how the distance moved per simulation time-step affected the dynamics of the computational model. The Differential Time (DT) parameter value, which essentially acts as a multiplier for the distance moved at each time-step, was adjusted to 0.5, 1, 2, 2.5, 4, 5, and 10. It was discovered that DT=2 provides an optimal balance between the speed of the simulation versus the quality of the simulation output. With DT=2, we were able to reduce the lag period to reach control-like dynamics, where the ratio of free and complexed NF-κB and IκBα agents (located across the cytoplasmic and nuclear compartments) were beginning to approximate to the s_r range (with respect to table 6.1). This can be explained through reference to the previous cartoon diagram (figure 6.4). A DT <2 means that agents may not enter the interaction boundary within a given time-step, and furthermore, may alter direction at the next time-step, thus not having a chance to probabilistically bind, and therefore requiring longer simulation runs (increased numbers of time-steps) to generate the required control dynamics. Conversely, a DT >2 means that agents may pass straight through the interaction boundary of a reciprocal NF-κB agent in a single time-step, which also reduces the likelihood of probabilistic binding within the simulation.

6.2.3 Rebind Delay

A rebind delay is required to ensure that the free NF-κB and IκBα agents resulting from basal dissociation at a given time-step do not immediately rebind at the next time-step. This is achieved by setting a countdown timer, which acts as a delay to rebinding, and thus allows the agents to move outside of each others interaction boundaries.

Two-at-a-time factorial analysis (Saltelli et al., 2009) was performed to find an appropriate rebind delay through varying the basal dissociation probability and rebind delay parameter values (see figure 6.5). As expected, it was found that a rebind delay of zero yielded no difference in simulation dynamics when varying basal dissociation as every NF-κB-IκBα complex that dissociated, would probabilistically bind in the next time-step as there was no ability for the agents to move outside of each others interaction boundary. Interestingly, it was also discovered that simulation dynamics when using rebind delays of 1 or 2 time-steps suffered from an issue with the conservation of IκBα agents, with simulation runs appearing to *lose* IκBα agents temporarily. Investigations highlighted that the issue was an artefact of the computational model, which was due to the underlying mechanism of FLAME when adding new agents to a simulation. It appears a lag period is introduced by the FLAME simulation framework between functions being called to generate a new IκBα agent (following basal dissociation), and the actual incorporation of the new agent in to the simulation - the new IκBα agent is added to the simulation at the end of the next time-step and not at the end of the current one in which it was requested. Further analysis shows that this is an artefact of the system at low rebind delay values however, as when values >3 are used, the computational model conserves the IκBα agents as expected. A rebind delay of 3 was therefore chosen as part of the calibration process for control dynamics.

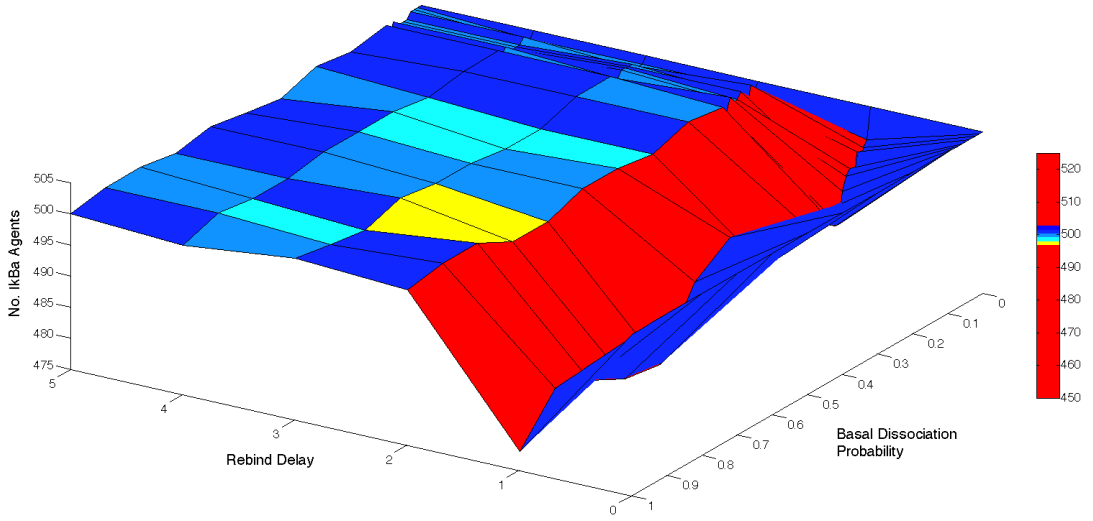

Figure 6.5: 3D surface heat map showing the two-at-a-time factorial analysis for basal dissociation and rebind delay, with specific reference to total IκBα within the system. A rebind delay of zero does not alter the dynamics of the system, however a rebind delay of 1 or 2 time-steps appears to be affected by a system artefact introduced by the underlying mechanisms of the FLAME simulation framework. This results in an issue with the conservation of mass within the system, through the temporary loss of IκBα agents. Rebind delay parameter values >3 do not appear to suffer from this system artefact.

6.2.4 Basal Dissociation versus Association

The basal dissociation and association rates are required to provide functionality that approximates to the reversible reactions between molecules, i.e. NF-κB + IκBα $\rightleftharpoons$ NF-κB-IκBα. For example, when an IκBα agent moves within the interaction boundary of a free NF-κB agent, there is a probability that it will bind to form the inhibited NF-κB-IκBα complex - the probability that binding occurs is captured by the association parameter. Similarly, following binding, there is also a probability, albeit a lot smaller, that the inhibited complex will dissociate to yield free NF-κB and free IκBα - this probability is captured by the basal dissociation parameter.

Two-at-a-time factorial analysis was performed, through varying the association and basal dissociation parameter values. An initial set of 121 experiments using 0.1 increments (between 0.0 and 1.0) for both parameters, found that system dynamics were incredibly sensitive to changes in basal dissociation. In fact, these initial experiments showed that although a basal dissociation of 0.0 provided agents that spanned the desired range (s_r range with respect to table 6.1), from 0.1 onwards there appeared to be no NF-κB-IκBα complexes, as they had all dissociated to the free NF-κB + IκBα agents.

As per the parameter estimation for the rebind delay parameter, we also encountered a conservation of mass issue regarding IκBα agents within the system when basal dissociation rates were greater than 0.1, as the total number of IκBα agents were falling below the expected 501 (see figure 6.6). However, this system

artefact disappeared when basal dissociation parameter values between 0.000001 - 0.0001 range were used, indicating that the FLAME computational model is more stable at these very low probability ranges of basal dissociation. Further investigation identified that basal dissociation probabilities between the 0.000001 - 0.00001 range provide variable system dynamics within the required range for calibration. Basal dissociation parameter values >0.1 generate no NF-κB-IκBα complexes in the cytoplasm, suggesting that the system is fragile at these higher levels of basal dissociation.

System dynamics in relation to the association parameter indicate an upward trend in NF-κB-IκBα complexes within the cytoplasm as the parameter value increases, and a downward trend in NF-κB-IκBα complexes within the nucleus. Association parameter values between 0.8 - 1.0 have the highest number of NF-κB-IκBα complexes in the cytoplasm, and the lowest number in the nucleus, however these are outside of our desired calibration range (s_r range with respect to table 6.1), which suggest a degree of co-dependency with the import and export parameters associated with nuclear translocation.

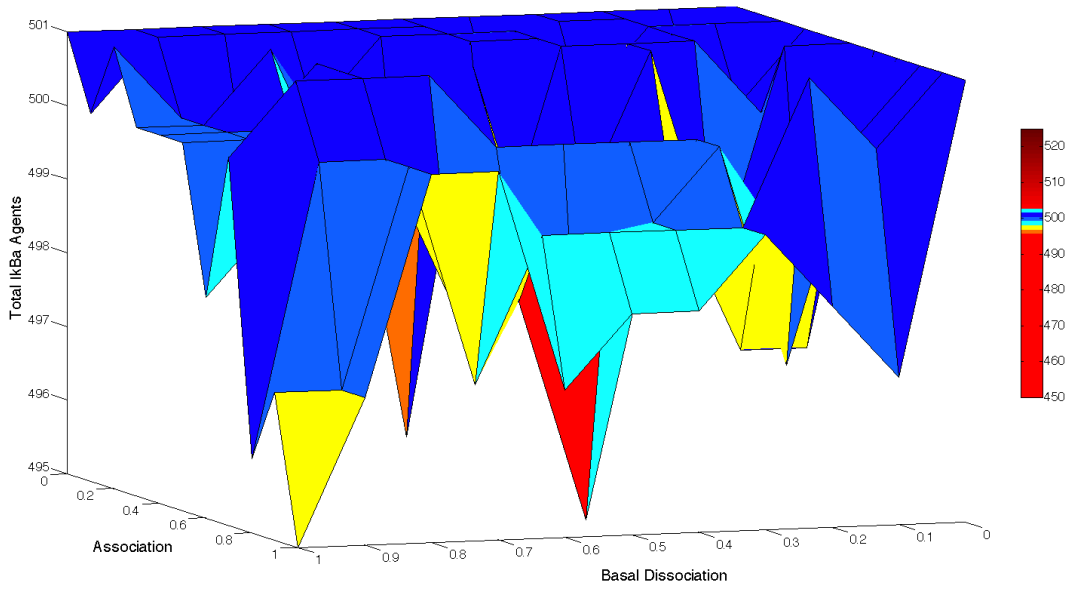

Figure 6.6: 3D surface heat map showing the two-at-a-time factorial analysis for association and basal dissociation, with specific reference to total IκBα within the system. Basal dissociation rates >0.1 encounter a conservation of mass issue regarding IκBα agents, which like the similar issue encountered during rebind delay parameter estimation, is again a system artefact caused by the underlying mechanisms of FLAME.

6.2.5 Nuclear Import versus Export Rates

Investigations to estimate parameter values for import and export probabilities (across the nuclear membrane) highlighted that the export parameter is co-dependent on the import parameter, and is in effect redundant regarding its effects on system dynamics under control conditions. The import probability appears to affect dynamics across the full probability range however (0.0 - 1.0). With low import probabilities, we have correspondingly higher levels of free NF-κB and free IκBα in the cytoplasm, but NF-κB and IκBα within the nucleus are within the required calibration range. It was also discovered that at low import probabilities, the system is very sensitive to dissociation rate, but comparatively stable to association rate, suggesting that the import parameter is co-dependent to varying degrees on both the basal dissociation parameter value and the association parameter value.

The export parameter value was therefore arbitrarily set to 0.5 to allow the maximum scope for movement during any subsequent *in silico* experimentation following the calibration exercise for IL-1 stimulated dynamics. Additional analysis was performed on the co-dependencies between the import, basal dissociation and association parameters. The desired range of agent states and locations (s_r range with respect to table 6.1) were generated by the computational model with basal dissociation values between 0.0000025 and 0.000005, association values between 0.6 and 0.75, and an import value between 0.03 and 0.05.

6.2.6 Calibrated Control Dynamics

Following parameter estimation experiments, the following parameter values were used to generate control dynamics of the NF-κB signalling pathway: DT = 2; Rebind Delay = 3; Basal Dissociation = 0.0000025; Association = 0.65; Import = 0.03; and Export = 0.5 (see figure 6.7).

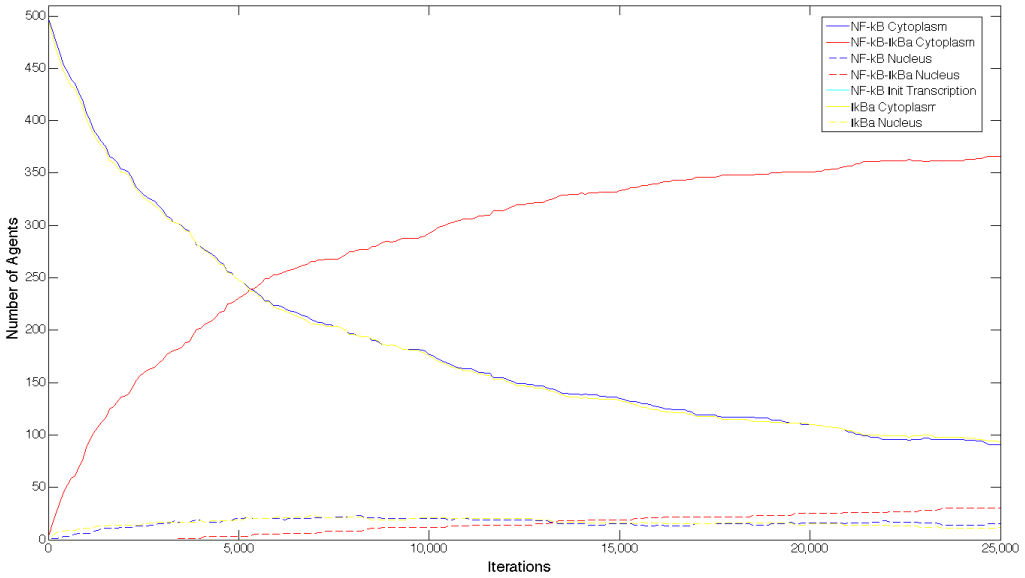

Figure 6.7: Calibrated control dynamics for the NF-κB signalling pathway. The computational model, using the parameters estimated in sections 6.2.2 to 6.2.5 generates emergent behaviour over a 25,000 iteration simulation that corresponds to the desired calibration range of agent states and locations (s_r range with respect to table 6.1).

6.3 Uncertainty Analysis

As discussed by Read et al. (2012) and Alden et al. (2013), there are two sources of uncertainty within computational models of real-world domains that require analysis before the results of *in silico* experimentation may be used within a predictive capacity. The first source, *epistemic* uncertainty, arises because our knowledge of the real-world system that we are modelling is incomplete, and therefore we do not have complete assurance that the parameter values derived through the literature, the domain experts professional opinion, and the calibration process, are wholly accurate (Helton, 2008). The second source, *aleatory* uncertainty, arises from the inherent stochasticity in systems, be that the computational model itself, or the underlying biological system (Helton, 2008).

6.3.1 Epistemic Uncertainty Analysis

As it is intractable to model every aspect of the real-world domain, our computational models are simplified versions of reality, with no direct one-to-one translation between the computational model and the real-world domain. The calibration process therefore compensates for this loss of granularity, to ensure that simulation dynamics approximate to the dynamics of the real-world domain. The relationship between the abstracted computational model and the biological system (in our case the IL-1 stimulated NF-κB signalling pathway) is critical for using the results of *in silico* experimentation in a predictive capacity to generate novel hypotheses for testing within the wet-lab. As such, epistemic uncertainty analysis focuses on the lack of knowledge regarding certain parameter values (Helton, 2008), the abstracted nature of computational models, and the effects of varying the absolute parameter values that have been defined through the calibration process (Alden et al., 2013).

Our epistemic uncertainty analysis followed the recent work of Read et al. (2012), but instead of using a global sensitivity analysis, such as the latin hypercube (McKay et al., 1979), we were forced to use[3] a mixture of robustness analysis, through the perturbation of parameter values for a single parameter (one-at-a-time approach), and local sensitivity analysis for co-dependent parameters through a two-at-a-time approach (Saltelli et al., 2009).

Our focus will remain on five of the parameters from the calibration process: Differential Time, Rebind Delay, Basal Dissociation, Association, and Nuclear Import. By considering the results of this robustness and sensitivity analysis in the context of the domain specific knowledge of NF-κB, we are able to qualify the implications of epistemic uncertainty on *in silico* experimentation and the subsequent simulation derived predictions of the underlying biological system.

[3]Due to the resource intensive nature of FLAME, in particular the extremely large amount of fast disk space required to store the XML output files generated with each simulation, we were not able to perform a global sensitivity analysis, which would adjust all of the parameter values in parallel, so instead utilised local sensitivity analysis approaches. Although we do not believe that this has detracted from our epistemic uncertainty analysis, the reduction in storage requirements, had a concomitant increase in experimental design for the sensitivity analysis simulations, as we could not rely on any preceded tools, but had to manually design the one-at-a-time and two-at-a-time perturbations to simulation parameter values.

Differential Time

Differential Time (DT) was perturbed around the calibrated value of 2 time-steps by performing sensitivity analysis experiments between 0.25 to 10 time-steps at 0.25 increments[4]. Figure 6.8 shows the results of this sensitivity analysis, with respect to the total number of NF-κB and IκBα agents in their corresponding states and locations. The profiles for cytoplasmic NF-κB, NF-κB-IκBα and IκBα are relatively smooth and show a progressive change with increasing DT. The profiles of agents within the nucleus are very noisy however, further highlighting how small changes in absolute numbers of agents in the nucleus are amplifying noise. This sensitivity analysis also shows that the computational model is very sensitive to DT, as the cytoplasmic NF-κB and IκBα agent numbers are only within our calibration range when DT parameter values are between 1.25 and 3.0, implying a fragility for agent movement (as per figure 6.4).

Rebind Delay

The rebind delay parameter was initially perturbed around the calibrated value of 3 using one-at-a-time factorial analysis, by running simulations with rebind delay values between 0 and 15 time-steps (see figure 6.9). The results of this sensitivity analysis show the computational model is robust to perturbations of the rebind delay value with respect to the cytoplasmic agents, but sensitive with respect to the nuclear agents. We initially suspected that this was another example of the low absolute numbers of nuclear agents in the calibrated model, however further investigation suggested a co-dependence on another parameter may result in this complex non-linear behaviour with respect to agents located within the nucleus.

Through two-at-a-time sensitivity analysis, it was discovered that this co-dependence was between the rebind delay parameter value and the association parameter value, as shown in the 3D heatmaps of figure 6.10. All cytoplasmic agents show significant robustness to rebind delays between 2-5 time-steps and association >0.3. However, the agents located within the nucleus display robustness when the rebind delay and association parameter values are constrained between specific ranges: NF-κB-IκBα agents within the nucleus are robust with rebind delay between 3-5 and association between 0.25-0.65; NF-κB agents within the nucleus are robust with rebind delay between 2-4 and association between 0.35-0.75; and, IκBα agents within the nucleus are robust with rebind delay between 3-5 and association > 0.55.

[4]As discussed in section 6.2.2, the DT parameter acts as a *multiplier* to calculate the distance moved per time-step, as such it does not need to be an integer value.

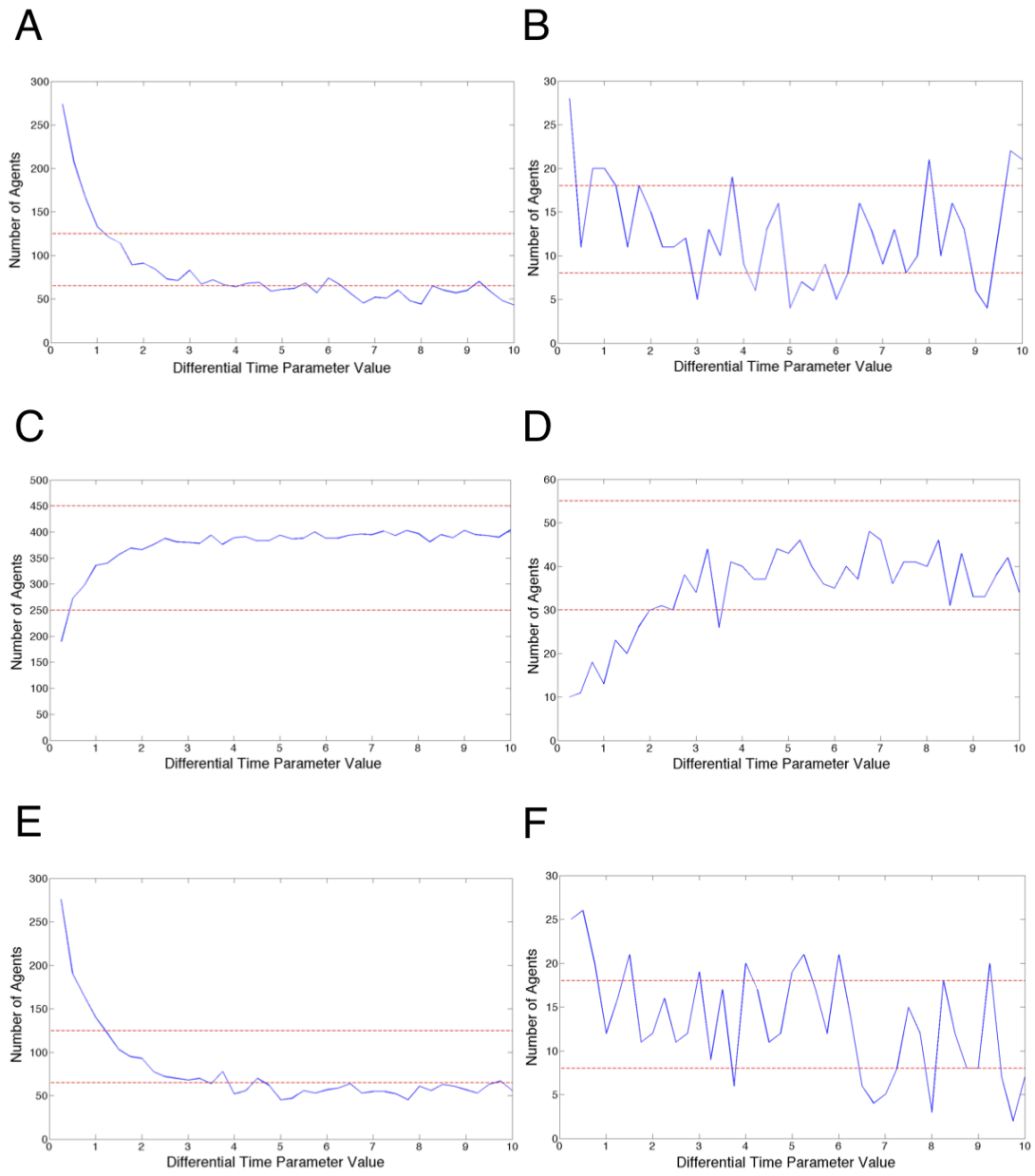

Figure 6.8: Sensitivity analysis of the differential time parameter, and its effects on the total number of NF-κB and IκBα agents and their respective states and locations after 25,000 iteration simulation runs. **A** shows the variation in the number of NF-κB agents within the cytoplasm. **B** shows the variation in the number of NF-κB agents within the nucleus. **C** shows the variation in the number of NF-κB-IκBα agents within the cytoplasm. **D** shows the variation in the number of NF-κB-IκBα agents within the nucleus. **E** shows the variation in the number of IκBα agents within the cytoplasm. **F** shows the variation in the number of IκBα agents within the nucleus. The dotted red lines show the calibration range for the agents. It can be seen that the computational model is relatively robust to DT parameter value changes as a whole, but is very sensitive for cytoplasmic NF-κB and IκBα, where DT needs to be between 1.25 and 3.

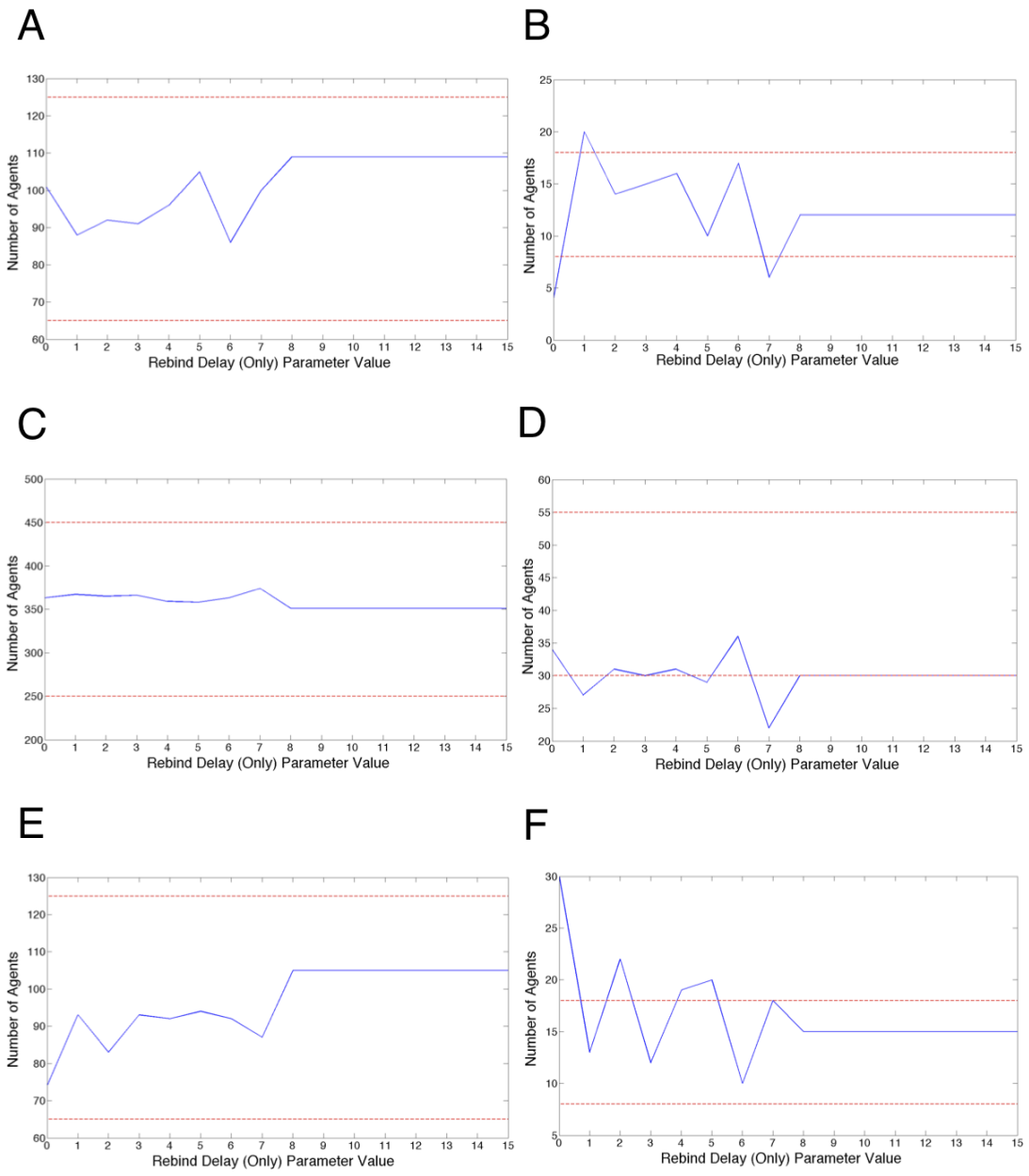

Figure 6.9: Sensitivity analysis of the rebind delay parameter, and its effects on the total number of NF-κB and IκBα agents and their respective states and locations after 25,000 iteration simulation runs. **A** shows the variation in the number of NF-κB agents within the cytoplasm. **B** shows the variation in the number of NF-κB agents within the nucleus. **C** shows the variation in the number of NF-κB-IκBα agents within the cytoplasm. **D** shows the variation in the number of NF-κB-IκBα agents within the nucleus. **E** shows the variation in the number of IκBα agents within the cytoplasm. **F** shows the variation in the number of IκBα agents within the nucleus. The dotted red lines show the calibration range for the agents. It can be seen that the computational model is relatively robust to rebind delay parameter value changes for agents located within the cytoplasm, but sensitive for agents located within the nucleus.

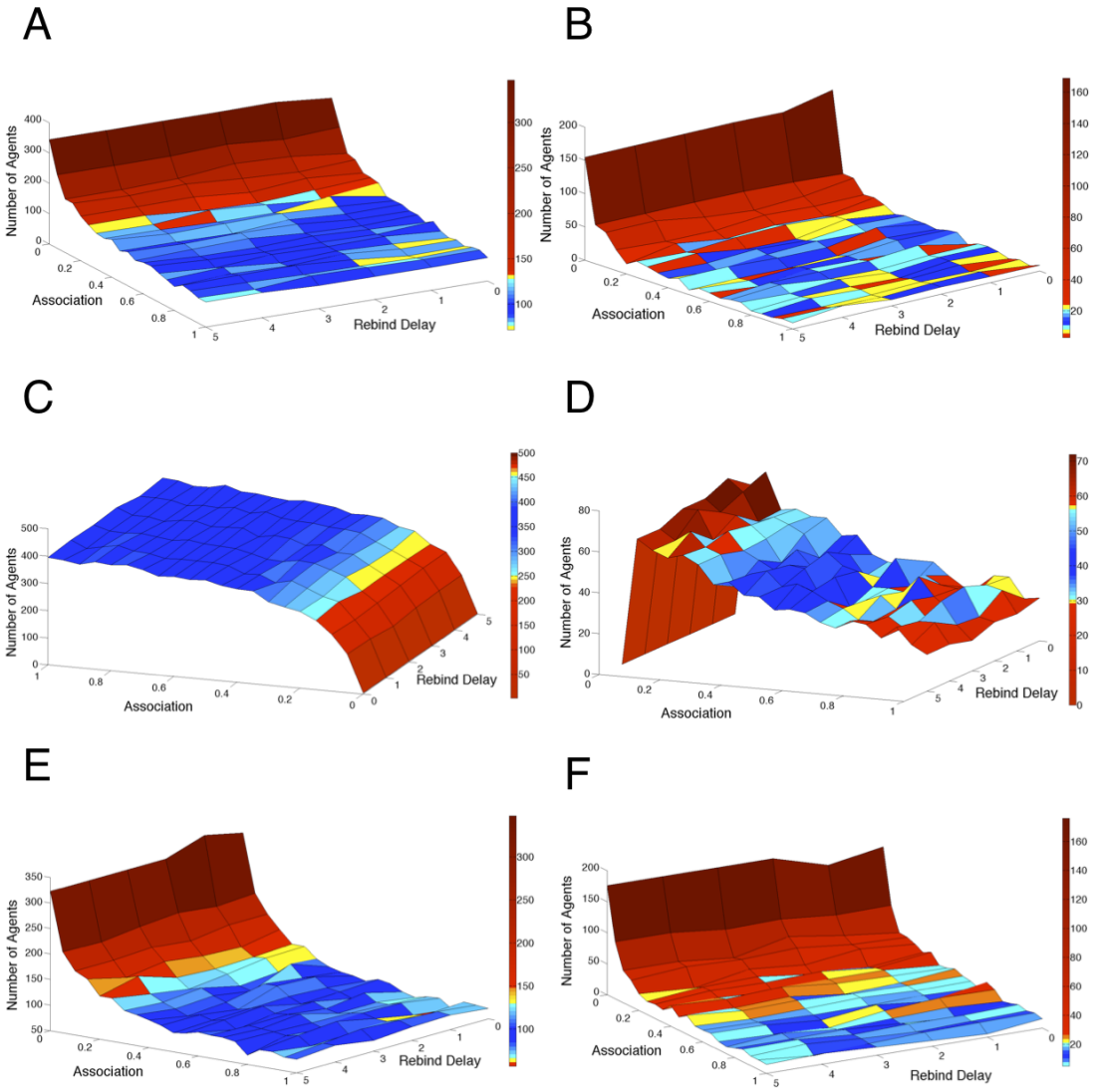

Figure 6.10: Sensitivity analysis of the rebind delay and association parameters, and the effects of their co-dependence on the total number of NF-κB and IκBα agents and their respective states and locations after 25,000 iteration simulation runs. **A** shows the variation in the number of NF-κB agents within the cytoplasm. **B** shows the variation in the number of NF-κB agents within the nucleus. **C** shows the variation in the number of NF-κB-IκBα agents within the cytoplasm. **D** shows the variation in the number of NF-κB-IκBα agents within the nucleus. **E** shows the variation in the number of IκBα agents within the cytoplasm. **F** shows the variation in the number of IκBα agents within the nucleus. As per figure 6.9, it can be seen that the computational model is relatively robust to rebind delay parameter value changes for agents located within the cytoplasm. In addition however, it can be seen that the computational model is robust for: NF-κB-IκBα agents located within the nucleus with rebind delay between 3-5 and association between 0.25-0.65; NF-κB agents in the nucleus with rebind delay between 2-4 and association between 0.35-0.75; and IκBα agents in the nucleus with rebidn delay between 3-5 and association greater than 0.55.

Basal Dissociation versus Association

Due to the co-dependence between the basal dissociation and association parameters, two-at-a-time perturbations were performed around the calibrated values. As the calibrated association parameter value was 0.65, we ran sensitivity analysis using parameter values rising at 0.05 increments across the full probabilistic range. As the basal dissociation parameter was found to be extremely fragile during the calibration process, we performed sensitivity analysis within close proximity of the calibrated value (0.000001-0.000009), and then at orders of magnitude higher ranges (e.g. 0.00001-0.00009, 0.0001-0.0009, 0.001-0.009, 0.01-0.09, 0.1-1.0). 3D heatmaps of this sensitivity analysis (figure 6.11) clearly show the fragile nature of the system with respect to the basal dissociation parameter values.

The 2D heatmaps within figure 6.12 display a subset of data from the basal dissociation sensitivity analysis. As per previous sensitivity analysis, the computational model appears to be relatively robust for cytoplasmic agents, with the NF-κB agents being robust when association >0.50 and dissociation <0.00003; the NF-κB-IκBα agents being robust when association >0.50 and dissociation <0.00007; and the IκBα agents being robust when association >0.35 and dissociation <0.00003. The NF-κB-IκBα agents within the nucleus were found to be robust when association is between 0.3-0.65 and dissociation <0.0001. The NF-κB and IκBα agents within the nucleus showed very complex dynamics, which we again believe are due to the small fluctuations in absolute numbers giving rise to large variances away from the desired calibration range. It can be seen however that association between 0.4-0.65 and dissociation <0.00001 provides an extremely small range of parameter space where the computational model remains within the calibration range.

Nuclear Import

The overwhelming feature of translocation across the nuclear membrane following analysis through the calibration process, was that the nuclear export parameter appeared to be redundant under control (non-stimulated) conditions, with the export dynamics being dependent on the nuclear import parameter value. We initially suspected an issue with the underlying code, however code walkthroughs by the developer and various members of the York Computational Immunology Lab, and debugging tests, indicated that this was not the case. We therefore believe that this may be an artefact of the computational model under control conditions, which may be due to the small absolute numbers of agents within the nucleus with respect to the cytoplasm.

Due to this redundancy, focus was applied to the nuclear import parameter, through perturbations around the calibrated value using one-at-a-time sensitivity analysis (figure 6.13). It is clear that the system is very sensitive to changes in the import parameter value, with the numbers of cytoplasmic agents falling outside of the calibration range with import parameter values >0.1, and the numbers of nuclear agents falling outside of the calibration range with import parameter values >0.05. As such, the computational model is fragile with respect to the nuclear import parameter, with calibration dynamics only being gained through a very narrow range of values (0.03-0.05).

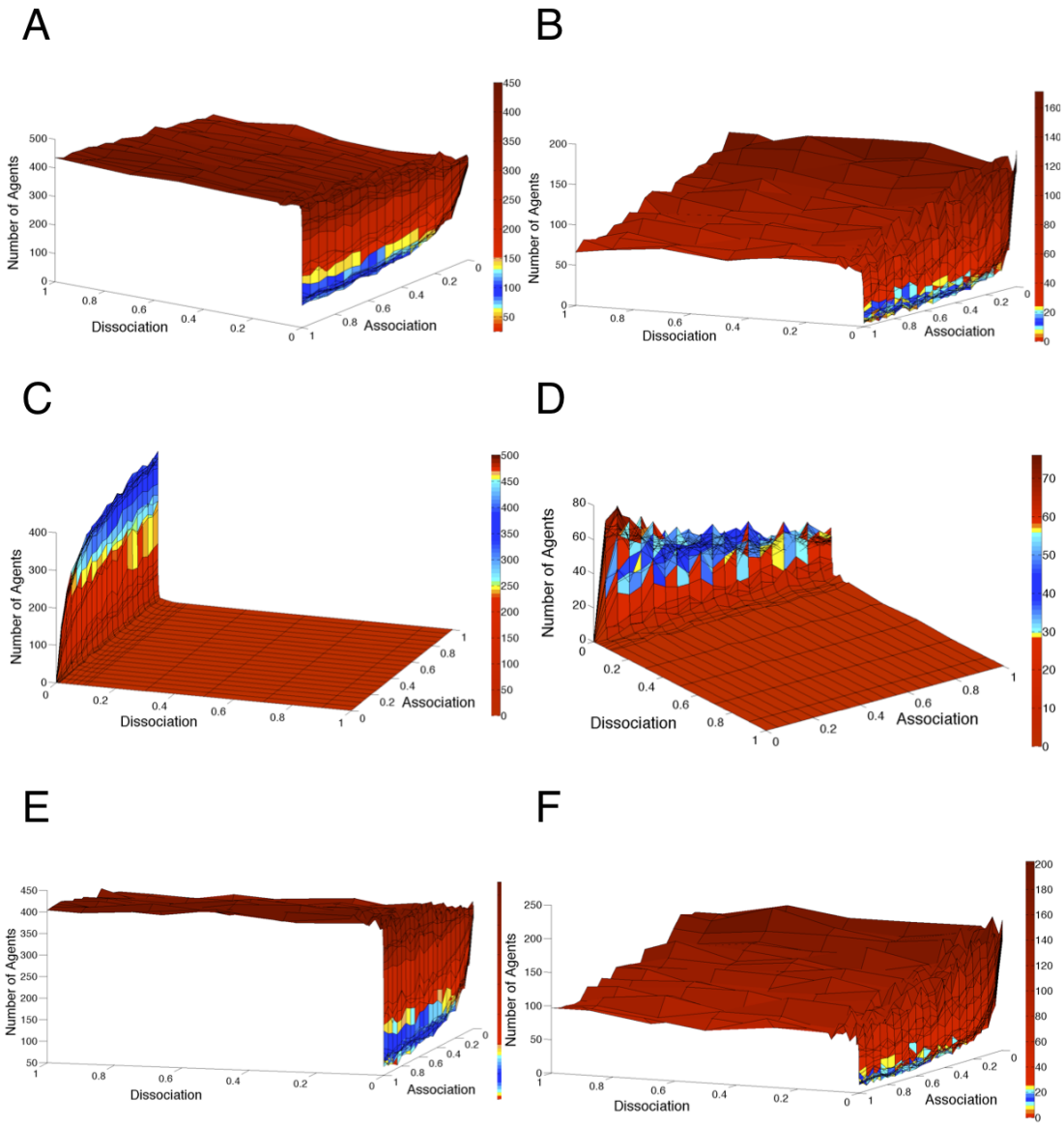

Figure 6.11: Sensitivity analysis of the association and dissociation parameters, and the effects of their co-depenence on the total number of NF-κB and IκBα agents and their respective states and locations after 25,000 iteration simulation runs. **A** shows the variation in the number of NF-κB agents within the cytoplasm. **B** shows the variation in the number of NF-κB agents within the nucleus. **C** shows the variation in the number of NF-κB-IκBα agents within the cytoplasm. **D** shows the variation in the number of NF-κB-IκBα agents within the nucleus. **E** shows the variation in the number of IκBα agents within the cytoplasm. **F** shows the variation in the number of IκBα agents within the nucleus. It can be seen that the computational model is relatively robust to association parameter value changes, but is extremely fragile for basal dissociation parameter values, with the computational model only staying within the calibrated range of agents with extremely low basal dissociation parameter values.

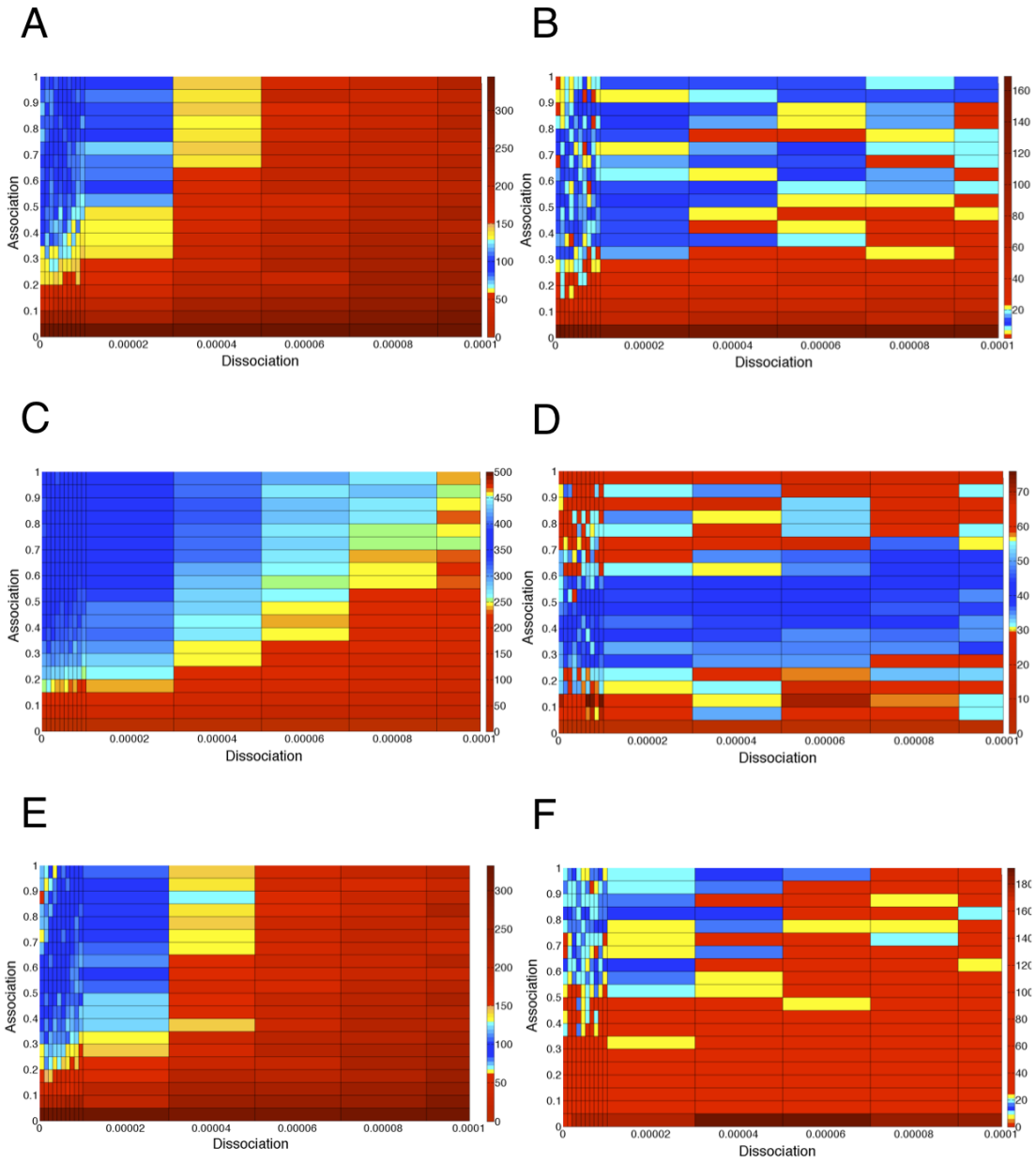

Figure 6.12: Sensitivity analysis of the association and dissociation parameters using a subset of the dissociation parameter values around the actual calibrated value. **A** shows the variation in the number of NF-κB agents within the cytoplasm. **B** shows the variation in the number of NF-κB agents within the nucleus. **C** shows the variation in the number of NF-κB-IκBα agents within the cytoplasm. **D** shows the variation in the number of NF-κB-IκBα agents within the nucleus. **E** shows the variation in the number of IκBα agents within the cytoplasm. **F** shows the variation in the number of IκBα agents within the nucleus. As per the calibration process, agents within the cytoplasm were found to be more robust to perturbations than those in the nucleus. It was discovered that an extremely narrow range of association and dissociation parameter values provided dynamics within the desired calibration range, comprising association between 0.5-0.65 and dissociation <0.00001.

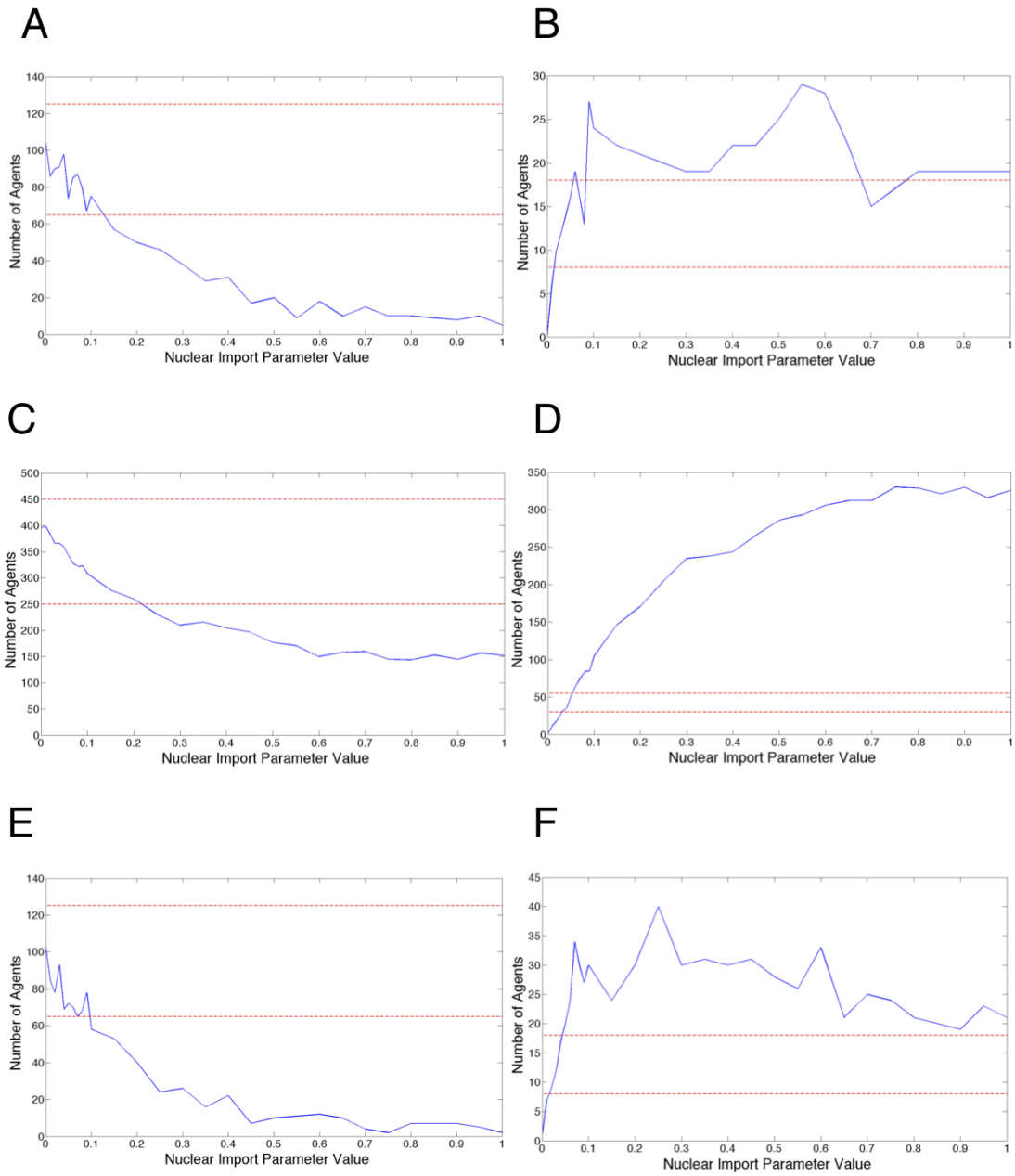

Figure 6.13: Sensitivity analysis of the nuclear import parameter, and its effects on the total number of NF-κB and IκBα agents and their respective states and locations after 25,000 iteration simulation runs. **A** shows the variation in the number of NF-κB agents within the cytoplasm. **B** shows the variation in the number of NF-κB agents within the nucleus. **C** shows the variation in the number of NF-κB-IκBα agents within the cytoplasm. **D** shows the variation in the number of NF-κB-IκBα agents within the nucleus. **E** shows the variation in the number of IκBα agents within the cytoplasm. **F** shows the variation in the number of IκBα agents within the nucleus. The dotted red lines show the calibration range for the agents. It can be seen that the computational model is sensitive to changes in the nuclear import parameter value, with only a small range (0.03-0.05) providing agent numbers within the desired calibration range.

6.3.2 Aleatory Uncertainty Analysis

Following the parameter estimation process (section 6.2), which generated a calibrated computational model for control (non-stimulated) dynamics, and the epistemic uncertainty analysis discussed above, it is essential that we investigate the *noise* (aleatory uncertainty) due to the stochastic nature of the computational model. Such aleatory uncertainty can severely effect our ability to compare experimental results against reference dynamics (e.g. the calibrated control dynamics or the calibrated IL-1 stimulated dynamics, which will be discussed in section 6.4), and thus interpret the simulation results with respect to the real-world domain (Harris et al., 1987). For example, the use of pseudo-random number generators are widely accepted as essential tools to generate heterogeneity within simulation results, even though the underlying parameter values of each simulation run are identical. This being the case, aleatory uncertainty analysis allows us to investigate the uncertainty that is introduced through the use of PRNGs (with different seed values), and to calculate a minimum number of replicate simulation runs to generate a stable median average of the simulation results (Read et al., 2012). The calculation of the minimum number of replicates, and its use within future *in silico* experimentation to generate a stable median average, allows us to mitigate stochastic effects and thus develop confidence that simulation results are representative of the condition(s) on which the simulation was run, and not an artefact of our computational model that is caused by the specific PRNG seed value.

Our aleatory uncertainty analysis closely followed the recent work of Alden et al. (2013), but instead of using their *Spartan* tool-chain, we developed our own analysis scripts using Matlab. Our analysis involved the running of 300 replicate simulations with different PRNG seed values to generate stochastic variation between the simulation runs. The stochastic nature of the computational model is clear to see (figure 6.14), with variation evident in the numbers of each agent state and location combination. It can be seen that the variation in free NF-κB, free IκBα, and NF-κB-IκBα within the cytoplasm is relatively smooth, denoted by the smooth red median average line. The replicate distributions of agents within the nucleus appear at first inspection to be less smooth, however we believe that this may be a direct consequence of the low absolute numbers of agents within the nucleus at calibrated dynamics. Thus, even small differences in absolute numbers over the 300 replicate simulation runs, equate to large percentage differences, and amplify the aleatory uncertainty for agents located within the nucleus.

The aleatory uncertainty was analysed by calculating median average distributions from subsets (increasing by 10 replicates) of these 300 replicate simulation runs. Briefly, we calculated median average distributions from 10, 20, 30, 40, ..., 300 replicate simulation run subsets, and utilised the Kolmogorov-Smirnov (KS) test (Massey, 1951) and the Vargha and Delaney non-parametric A-Test (Vargha and Delaney, 2000) to investigate the differences between these median average simulation results (see Appendix B for supporting material on these statistical tests). The KS-Test is used to understand the *statistical significance* of differences between two distributions, and the A-Test is used to understand the *scientific significance* (or effect magnitude) of differences between two distributions. In our analysis, we used the KS-Test and A-Test to interpolate when a stable median average was gained, to predict the minimum number of replicates

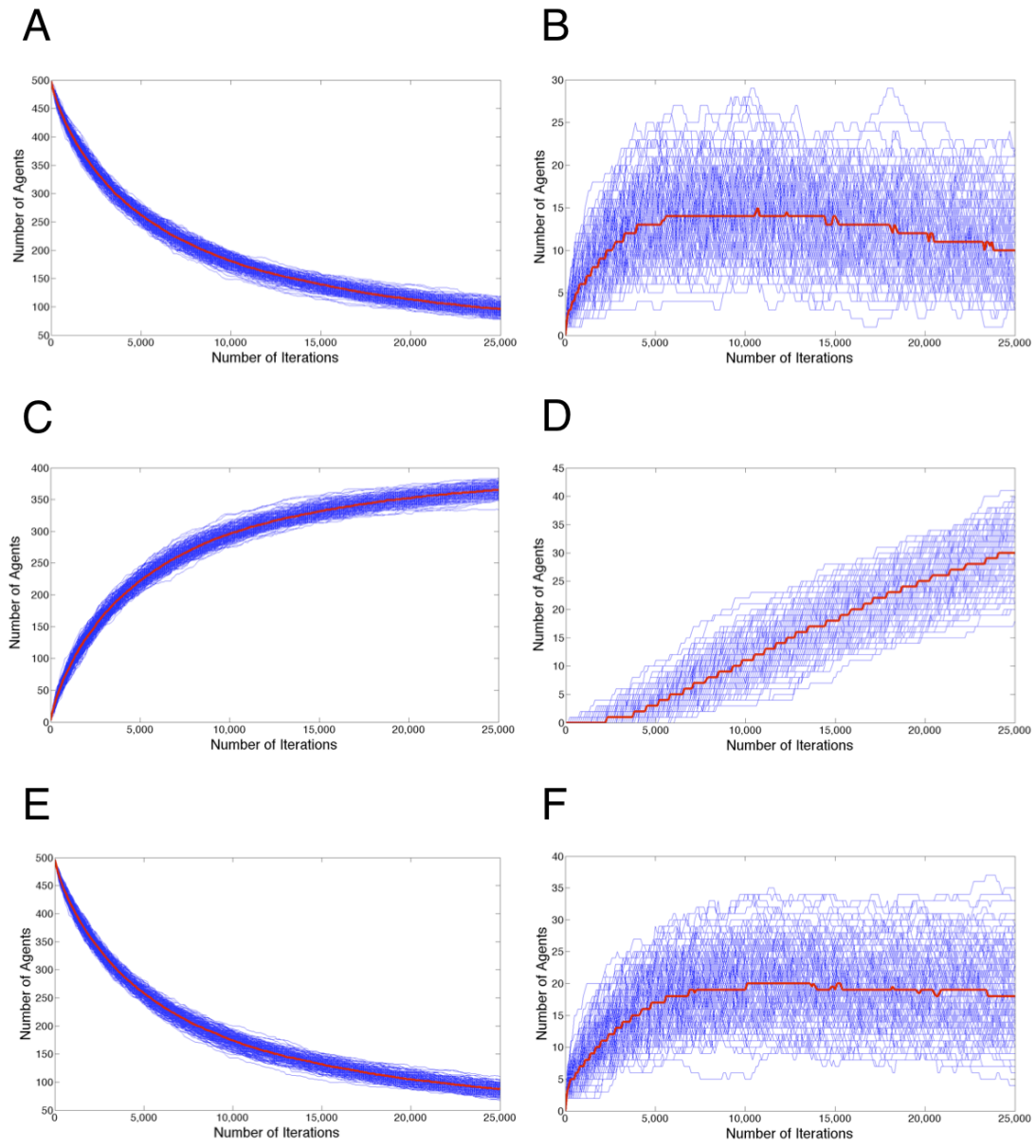

Figure 6.14: The computational model, using different PRNG seed values with the parameters estimated in sections 6.2.2 to 6.2.5, shows stochastic behaviour of the calibrated control dynamics over 300 replicate simulations. Replicate simulation runs are shown individually as blue curves, with the corresponding median average shown as a red curve. **A** shows the variation in the number of NF-κB agents within the cytoplasm. **B** shows the variation in the number of NF-κB agents within the nucleus. **C** shows the variation in the number of NF-κB-IκBα agents within the cytoplasm. **D** shows the variation in the number of NF-κB-IκBα agents within the nucleus. **E** shows the variation in the number of IκBα agents within the cytoplasm. **F** shows the variation in the number of IκBα agents within the nucleus. It can be seen that the variation within the cytoplasm for all agents is smooth. Whereas the variation in the nucleus appears to be very noisy. We believe that this is due to the low absolute calibrated values within the nucleus, so that even small variations around this value yield a large variance, thus acting to amplify the aleatory uncertainty.

required for *in silico* experimentation. This was performed by testing the median average distributions against each other, for example, by comparing the 10- and 20-replicate median average distributions, then the 20- and 30-replicate median average distributions, and continuing up to the 290- and 300-replicate median average distributions. Figure 6.15 shows a plot of the resulting p-values from the KS-Test, which shows that all agent states of interest stabilise at approximately 175 replicates, apart from the IκBα agents located in the nucleus. As before, we believe that this is due to the small absolute numbers involved, having an amplifiying effect, which in this case is on the p-value scores, even though the difference in absolute numbers may be small.

As stated above, the KS-Test provides an indication of statistical significance of the goodness of fit between two distributions. In our case, results of the KS-Test suggest that the aleatory uncertainty can be minimised, with a 7.5% level of significance, by generating a median average distribution from 175 replicate simulation runs. To provide additional confidence that 175 replicate simulation runs is the minimum number of replicates to achieve a stable median average distribution, we also analysed the data from the 300 replicates using the Vargha and Delaney A-Test. This test is usually performed to understand the scientifically significant (effect magnitude) difference between two median average distributions, and therefore normally uses A-Test scores <0.29 and >0.71 (indicating a large effect magnitude difference) in order to determine significant difference.

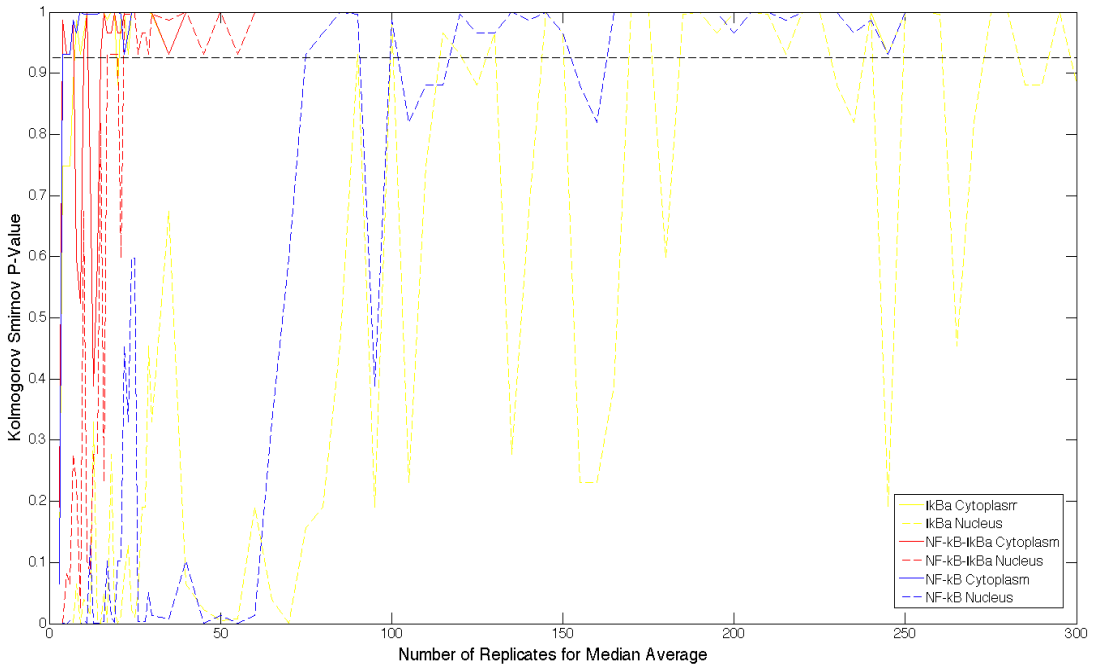

Figure 6.15: Graph of the Kolmogorov-Smirnov p-values, which were calculated from the median average subsets of 300 replicate simulation runs of the calibrated control dynamics. Median average distributions were generated, which increased by 10 replicates (i.e. 10, 20, 30, ..., 300) and compared using the KS-Test. It can be seen that all agent states stabilise at approximately 175 replicates, apart from the IκBα agents located in the nucleus. The dotted black line represents a p-value of 0.925, which we have used to signify stability.

We are using the A-Test in the opposite way here however, as we wish to ascertain the number of replicates required to provide a stable median average distribution, through the A-Test scores progressing towards 0.50, which signifies that the two distributions are the same. A cautionary note is required here however, as previous research (Read et al., 2012; Alden et al., 2013) has advised that median average distributions can always be forced to converge if enough replicates are run, therefore we need to choose a replicate number large enough for medians to stabilise, but small enough to ensure stochasticity continues. We therefore believe that the minimum number of replicates can be extrapolated through the use of A-Test scores between 0.47 and 0.53. We believe that this will minimise the aleatory uncertainty inherent to our computational model, but will also allow stochasticity, so will reduce the risk of overtuning the computational model by running large numbers of replicates that *force* the A-Test score to 0.50.

Figure 6.16 shows the corresponding A-Test score for each of the replicate subsets that were tested (cumulatively increasing subset of replicates, as per the KS-Test). In a similar manner to the KS-Test, it can be seen that agents within the cytoplasm stabilise relatively quickly (approximately 30 replicates), with those in the nucleus requiring more replicates, and continue to suffer from noise throughout throughout the full range of 300 replicates. For example, the NF-κB-IκBα complex within the nucleus seems to stabilise at approximately 40 replicates; the NF-κB agents within the nucleus also begin to stabilise by 110 replicates, and have stabilised by 150 replicates; with the IκBα agents within the nucleus showing the highest variability, by beginning to stabilise after approximately 100 replicates, but having an additional two spikes just outside of our *stable* A-Test score range of 0.47 to 0.53 at 150 and 250 replicates, respectively. As per our previous analysis, we again believe that this is due to the amplifying effect of small differences in absolute numbers at the desired calibrated range for control dynamics. As such, we believe that the A-Test has confirmed the results of the previous KS-Test, and conclude that the minimum number of replicates required to generate a stable median average distribution is 175 replicates.

Figure 6.17 shows the calibrated control dynamics that relate to the median average generated from the 175 replicate simulation runs. As expected, the curves closely follow the single simulation run dynamics displayed in figure 6.7, however closer inspection shows that the curves are considerably smoother. The actual median average simulation output values (s_o) are shown in the updated simulation platform translation table (see table 6.2), and can be seen to fall within the desired simulation results range (s_r Range).

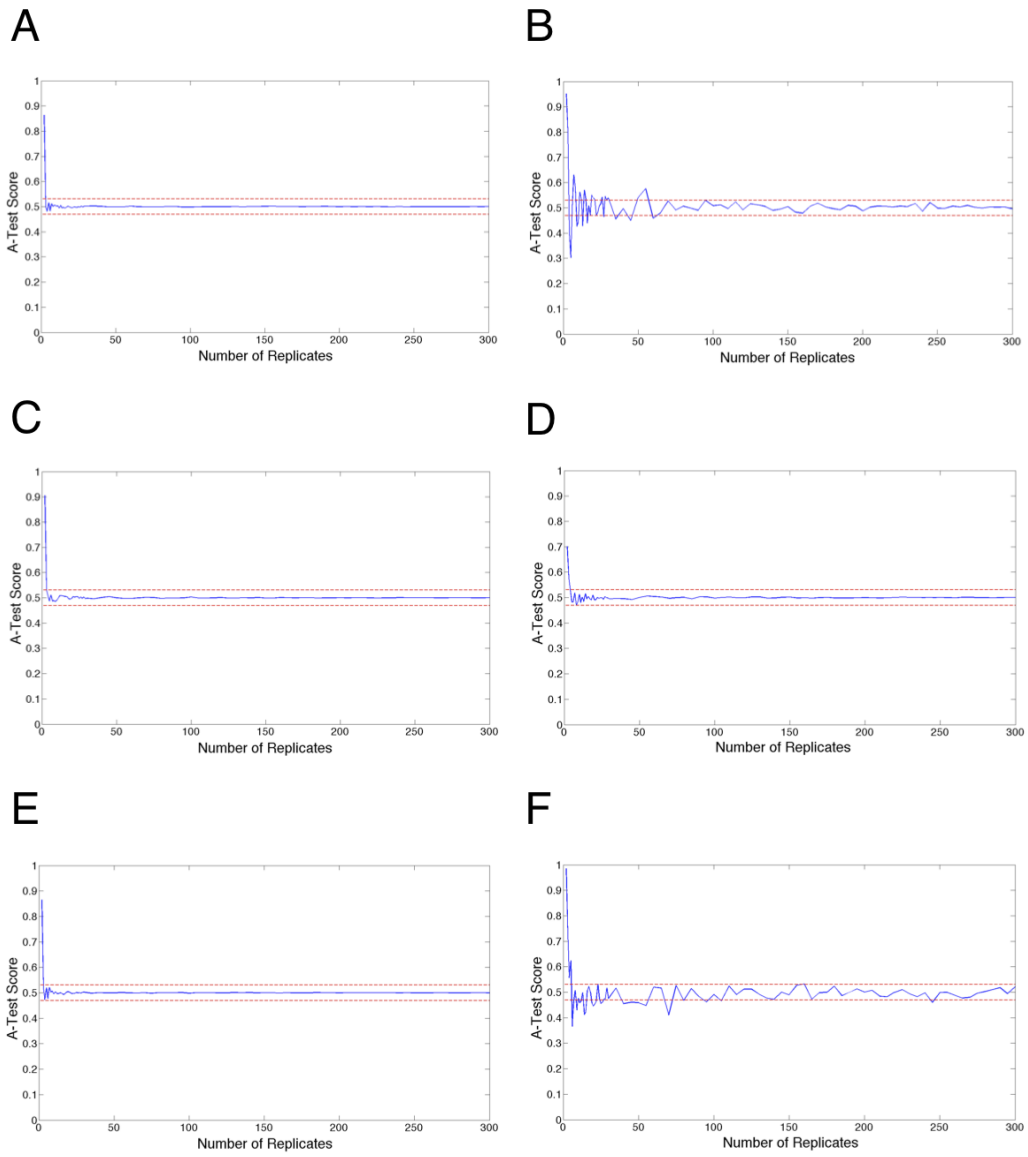

Figure 6.16: A-Test scores, which were calculated from the median average subsets of 300 replicate simulation runs using calibrated control dynamics. Median average distributions were generated, which increased by 10 replicates (i.e. 10, 20, 30, ..., 300) and compared using the A-Test. It can be seen that as per the KS-Test, all agent states stabilise at approximately 175 replicates, apart from the IκBα agents located in the nucleus. The dotted red lines represent the A-Test scores of 0.47 and 0.53, which we have used as the boundaries to signify stability. **A** shows the variation in the number of NF-κB agents within the cytoplasm. **B** shows the variation in the number of NF-κB agents within the nucleus. **C** shows the variation in the number of NF-κB-IκBα agents within the cytoplasm. **D** shows the variation in the number of NF-κB-IκBα agents within the nucleus. **E** shows the variation in the number of IκBα agents within the cytoplasm. **F** shows the variation in the number of IκBα agents within the nucleus. It can be seen that the variation is consistent with that from the KS-Test, with the cytoplasm stabilising relatively quickly, and the nucleus taking longer, and continuing to appear noisy even after the 175 replicates has been run (although the noise is contained within the A-Test score boundaries). As before, we believe that this is due to the low absolute calibrated values within the nucleus, with even small variations having the effect of amplifying the aleatory uncertainty.

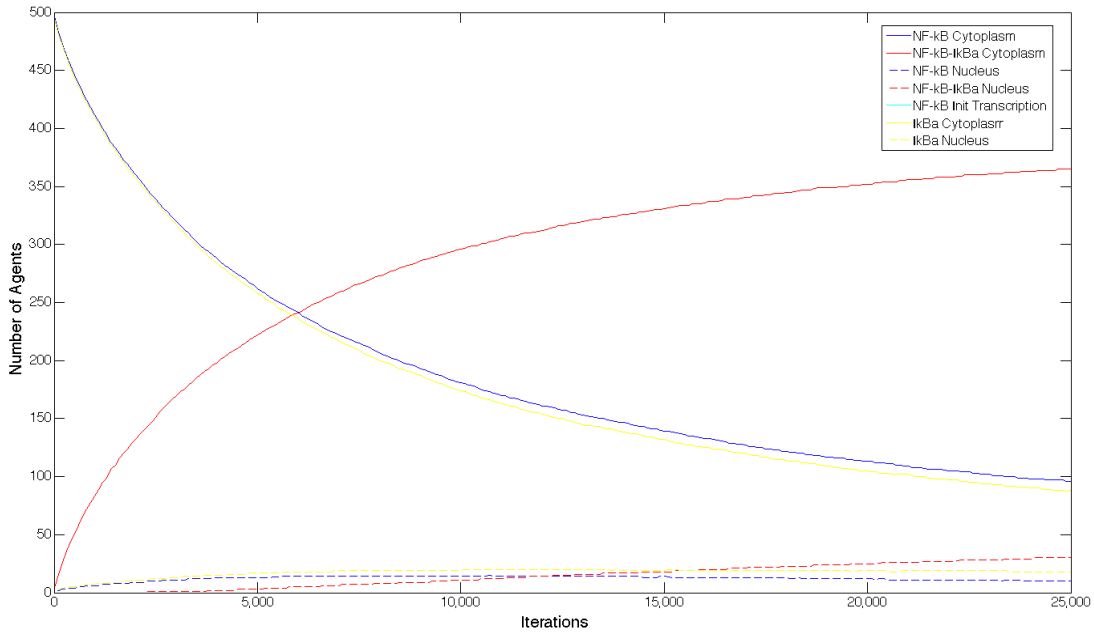

Figure 6.17: Calibrated control dynamics using the median average distributions of 175 replicate simulation runs using different PRNG seed values. It can be seen that although the curves closely follow the single simulation run curves from figure 6.7, the use of 175 replicates has generated smooth curves. We believe that this signifies a decrease in the aleatory uncertainty of our computational model.

Agent	Location	d_r	s_i	s_r Range	s_o
Free NF-κB	Cytoplasm	9,000	498	65 - 125	96
	Nucleus	1,000	1	5 - 20	11
Free IκBα	Cytoplasm	9,900	496	65 - 125	88
	Nucleus	1,100	2	5 - 20	17
NF-κB-IκBα	Cytoplasm	45,000	3	250 - 450	365
	Nucleus	5,000	0	30 - 55	30

Table 6.2: Calibration translation table as advocated by Stepney (2012), following aleatory uncertainty analysis, and the use of 175 replicate simulation runs to generate median average distributions. In a similar manner to table 6.1, this translates/maps: 1) the approximate numbers and locations of components from wet-lab experimentation, which forms the domain results data d_r; 2) the initial simulation platform values (s_i), which are used to commence simulation runs during the calibration process; 3) the desired range of simulation results data (s_r Range), which needs to be attained during the calibration process; and 4) the simulation output s_o, which represents the median average values generated through the use of 175 replicate simulation runs.

6.4 Calibration of Simulation Platform for IL-1 Stimulated Dynamics

The IL-1 stimulated dynamics within our computational model relate to the IKK-mediated release of NF-κB from the NF-κB-IκBα complex, the associated degradation of IκBα, and the translocation of the free NF-κB (that had been released via IKK) in to the nucleus. As iteration 1 relies on a countdown timer for activation of the cell membrane receptor, there are only three additional parameters that require calibration for IL-1 stimulated dynamics:

1. **IKK Rebind Delay** - The time delay associated with an IKK agent, following its interaction with an NF-κB-IκBα complex.

2. **Nuclear Import (IL-1 stimulated) Probability** - The probability of free NF-κB (released by IKK) binding to a nuclear import receptor to facilitate translocation from the cytoplasm into the nucleus. This is in addition to the standard parameter, which is used under control dynamics, and ensures that the nuclear translocation has a higher probability under IL-1 stimulated dynamics.

3. **IKK Binding Probability** - The probability of binding following the movement of an IKK agent into the interaction boundary of a NF-κB-IκBα complex.

As per the calibration process for control dynamics, we calibrated our computational model for IL-1 stimulated dynamics through varying parameter values, either one-at-a-time or two-at-a-time were there was co-dependence with another parameter. Again, this calibration process required consultation between the modeller and domain expert until a parameter space was found that yielded qualitative alignment to the underlying biological data on which the computational model was designed.

Carlotti et al. (1999) show a 40-fold increase in nuclear NF-κB following IL-1 stimulation. An increase of this level, using the 1:10 ratio of nuclear to cytoplasmic NF-κB from control dynamics and total of 500 NF-κB agents within our computational model, would push the ratio between the compartments to unnatural levels, i.e. 400 NF-κB agents in the nucleus and 100 NF-κB agents in the cytoplasm. As the transfected cells had 8 times more NF-κB than endogenous levels, we believe that it would be inappropriate to calibrate the IL-1 stimulated dynamics of our computational model for this level of increase, and have chosen to calibrate to a 5-fold increase in nuclear NF-κB numbers[5].

Due to the first iteration of our computational model focusing on the NF-κB signalling module, with a high-level abstraction of upstream events (e.g. cell membrane receptor activation, through to IKK activation), and no downstream events following NF-κB activation (e.g. transcription of gene products), we do not model the generation of new IκBα agents or the subsequent inhibition of active NF-κB in the nucleus, to facilitate negative feedback. This being the case, it would be

[5]Carlotti et al. (1999) have a 40-fold increase in nuclear NF-κB following IL-1 stimulation, with 8-fold higher numbers in transfected cells v endogenous. We have therefore chosen to use the quotient of 40/8, providing a 5-fold increase for our calibration requirements.

inappropriate to calibrate the model for IκBα numbers, and we therefore focus on calibrating the IL-1 stimulated dynamics for NF-κB agents only. Furthermore, as the Carlotti work looked at total NF-κB (both free and inhibited), we will calibrate the IL-1 stimulated dynamics to total NF-κB between the nucleus and cytoplasm. Table 6.3 represents the translation table for calibration of our computational model against IL-1 stimulated dynamics. The initial simulation platform values (s_i), which are used to commence simulation runs for the IL-1 stimulation calibration process are actually the median averaged simulation results (s_o) of the control calibration dynamics. The range of agent numbers from control dynamics calibration (s_r range) has also become the associated range for initial simulation platform values (s_i range), to facilitate a translation to the required simulation results range (s_r range) for calibration against IL-1 stimulated dynamics, incorporating the 5-fold increase in nuclear NF-κB, and the corresponding decrease in cytoplasmic numbers.

Location	State	s_i	s_i Range	s_r Range
Nucleus	NF-κB	11	35 - 75	175 - 375
	NF-κB-IκBα	30		
Cytoplasm	NF-κB	96	425 - 465	125 - 325
	NF-κB-IκBα	365		

Table 6.3: Translation table as advocated by Stepney (2012), for calibration of the computational model to IL-1 stimulation dynamics. This translates/maps the approximate numbers and locations of components from the calibrated control dynamics (s_i median average and range), which form the starting conditions of IL-1 stimulated simulation runs, against the required simulation results (s_r Range), which represent the calibrated dynamics of our computational model for IL-1 stimulation. Cytoplasmic values for s_i Range and s_r Range are the difference between the total number of NF-κB agents and the numbers extrapolated for the nucleus. Values for s_r Range are extrapolated from Carlotti et al. (1999) using a 5-fold increase (instead of a 40-fold) in nuclear numbers following IL-1 stimulation, as described in footnote 5 above.

6.4.1 IKK Rebind Delay

The IKK rebind delay parameter showed a remarkable degree of robustness during the calibration process. Figure 6.18 higlights that computational model dynamics are sensitive when simulations use IKK rebind delays between 1 to 4 time-steps, but very stable with values between 4 to 18 time-steps. We have therefore chosen to use a rebind delay of 10 time-steps, corresponding to the mid-point of this stable range.

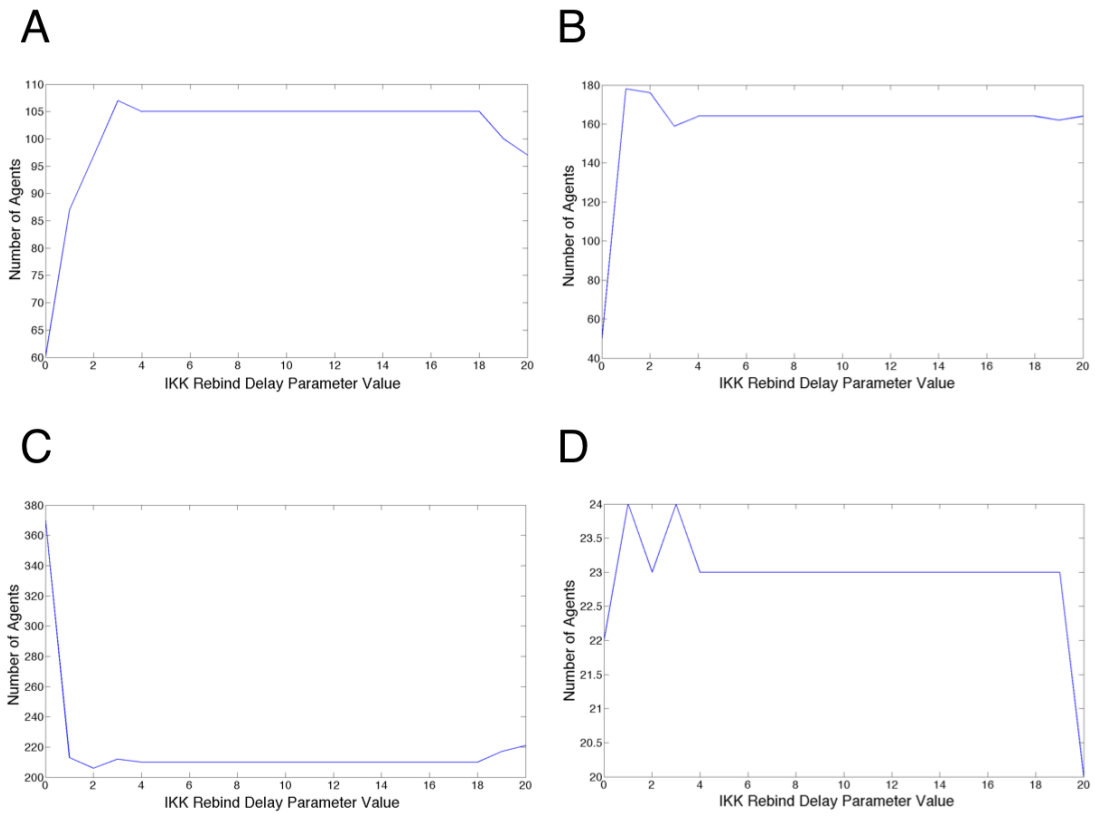

Figure 6.18: Calibration for IKK rebind delay during IL-1 stimulation. **A** shows the variation in the number of NF-κB agents within the cytoplasm. **B** shows the variation in the number of NF-κB agents within the nucleus. **C** shows the variation in the number of NF-κB-IκBα agents within the cytoplasm. **D** shows the variation in the number of NF-κB-IκBα agents within the nucleus. It can be seen that system dynamics are robust for IKK rebind delay parameter values between 4 to 18.

6.4.2 Nuclear Import Probability under IL-1 Stimulation

Initial calibration simulations focused on the nuclear import (under IL-1 stimulation) parameter value in isolation. This showed unstable dynamics as seen in figure 6.19, and we were unable to interpolate a calibrated parameter value. It can be seen that the nuclear import (under IL-1 stimulation) parameter value provides reasonably stable dynamics for NF-κB agents within the cytoplasm between 0.3 to 0.65; NF-κB-IκBα agents within the nucleus between 0.5 to 0.8; and NF-κB agents within the nucleus between 0.5 to 0.7. NF-κB-IκBα agents within the cytoplasm appear to be unstable. As Carlotti et al. (1999) looked at total NF-κB (both free and inhibited) however, graphs of these total NF-κB agents within the cytoplasm and nucleus (see figure 6.20) suggest a degree of stability, which we conjecture to represent non-linear dynamics of the nuclear import (under IL-1 stimulation) parameter, that may be co-dependent on IKK mediated dissociation through the IKK binding parameter.

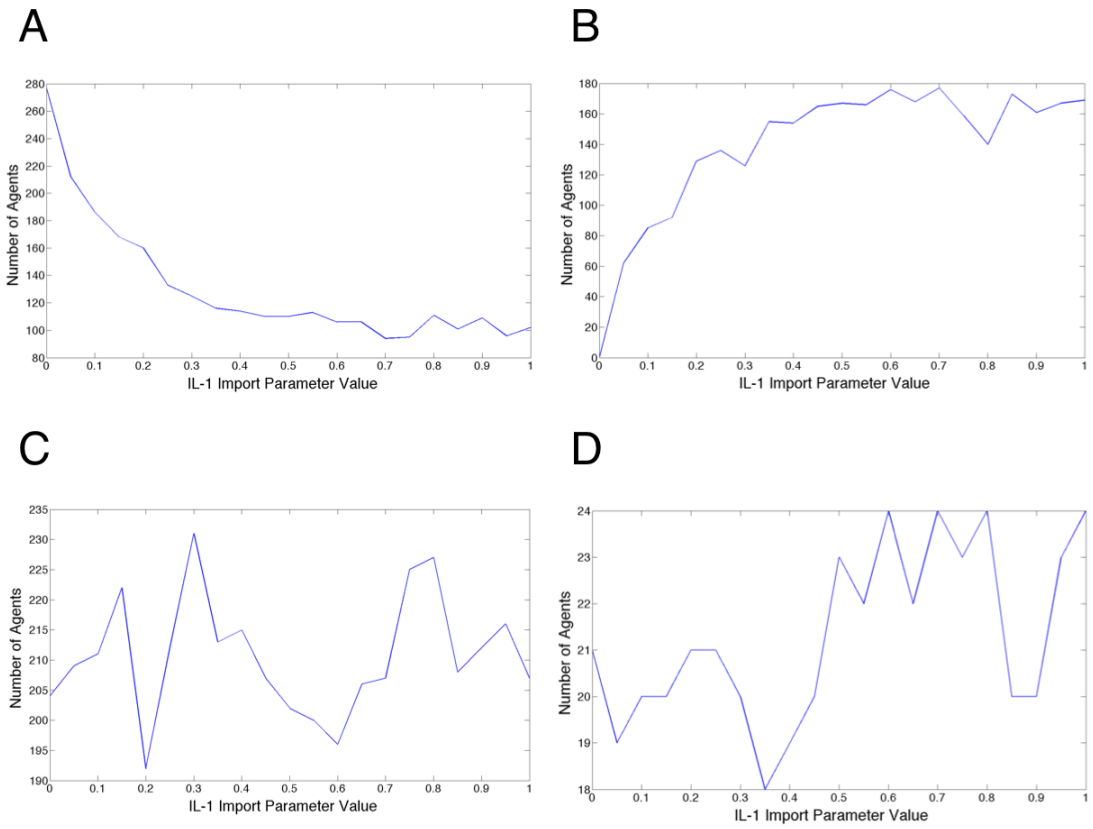

Figure 6.19: Calibration for nuclear import during IL-1 stimulation. **A** shows the variation in the number of NF-κB agents within the cytoplasm. **B** shows the variation in the number of NF-κB agents within the nucleus. **C** shows the variation in the number of NF-κB-IκBα agents within the cytoplasm. **D** shows the variation in the number of NF-κB-IκBα agents within the nucleus. It can be seen that system dynamics are relatively stable for NF-κB agents within the cytoplasm, NF-κB-IκBα agents within the nucleus, and NF-κB agents within the nucleus, but are unstable for NF-κB-IκBα agents within the cytoplasm.

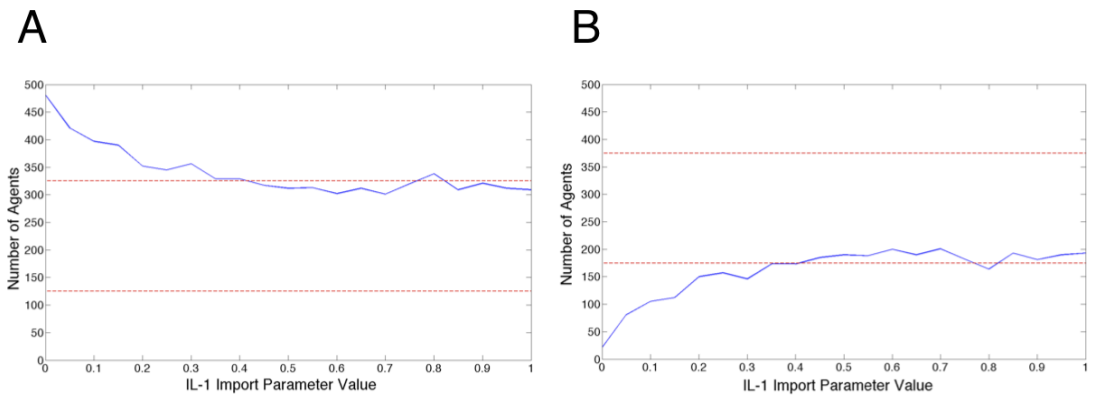

Figure 6.20: Calibration of nuclear import during IL-1 stimulation, for total NF-κB agents. **A** shows the variation in the total number of NF-κB agents within the cytoplasm. **B** shows the variation in the total number of NF-κB agents within the nucleus. The red dotted lines represent the desired calibration range (s_r Range) from table 6.3. It can be seen that total NF-κB agent numbers are within the desired calibration range for only a small section of parameter values.

6.4.3 IKK Binding Probability

To confirm any co-dependence between nuclear import (under IL-1 stimulation) and IKK binding, we performed two-at-a-time analysis, and focused on total NF-κB agent numbers across the nucleus and cytoplasm. The resulting 2D heatmaps (figure 6.21) suggest that desired total NF-κB agent numbers (with respect to the translation table 6.3) within the nucleus and cytoplasm are generated when nuclear import (under IL-1 stimulation) parameter value is between 0.75 to 1.0, and when the IKK binding parameter value is between 0.75 to 1.0. We have therefore chosen to set both these parameters to 0.85 for the IL-1 stimulation calibrated dynamics, to provide a degree of robustness.

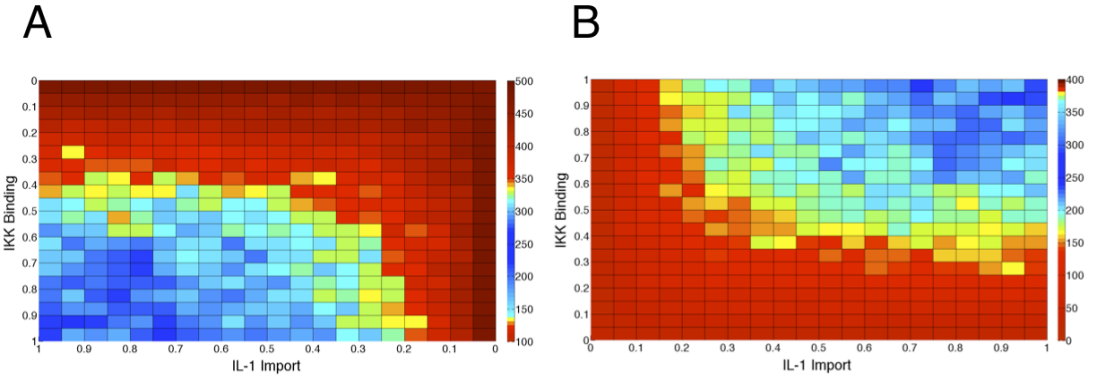

Figure 6.21: Calibration of nuclear import and IKK binding during IL-1 stimulation, for total NF-κB agents. **A** shows the variation in the total number of NF-κB agents within the cytoplasm. **B** shows the variation in the total number of NF-κB agents within the nucleus. It can be seen that total NF-κB agent numbers are within the desired calibration range when both parameter values fall between 0.75 to 1.0.

6.4.4 Calibrated IL-1 Stimulated Dynamics

Following parameter estimation experiments, the following parameter values were used to generate IL-1 stimulation dynamics of the NF-κB signalling pathway. DT = 2; Rebind Delay = 3; Basal Dissociation = 0.0000025; Association = 0.65; Import = 0.85; Export = 0.5; IKK Rebind Delay = 10; and IKK Binding Probability = 0.85. The median average simulation dynamics (using 175 replicates) are shown in figure 6.22 for a 50,000 iteration run, which was the time period used to reach calibration dynamics. See Appendix C for an example 0.xml parameters file depicting the calibrated parameter values and examples of the relevant agent definitions for iteration 1.

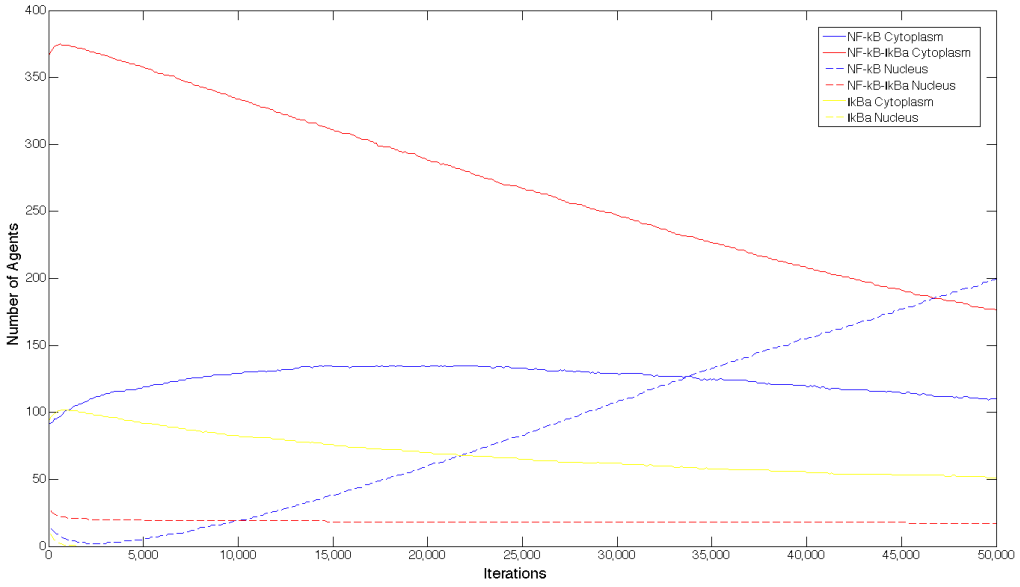

Figure 6.22: Calibrated IL-1 stimulation dynamics for the NF-κB signalling pathway. The computational model, using the parameters estimated in sections 6.2.2 to 6.2.5 for control dynamics, and 6.4.1 to 6.4.3 for IL-1 stimulation dynamics generates emergent behaviour over a 50,000 iteration simulation that corresponds to the desired calibration range of agent states and locations (s_r range with respect to table 6.3).

6.4.5 Calibration for Physical Time

The key agent for calibrating the simulation with respect to physical time is the increase in nuclear NF-κB, as its curve in the calibration graph (figure 6.22) bisects both the cytoplasmic NF-κB agents and the cytoplasmic NF-κB-IκBα agents. The wet-lab data from Yang et al. (2003) has a corresponding bisection of the curves for nuclear NF-κB and cytoplasmic NF-κB at approximately 1,070 seconds. Within our calibrated computational model, this occurs at approximately 33,750 iterations, which provides a ratio of 31.5 iterations per second. However, as the computational model will be used to convey qualitative differences in experimental dynamics, we have chosen to approximate the calibrated ratio to 30 iterations per second.

The 50,000 iteration graph (see figure 6.22) allowed calibration to desired ratios of agent states, however in order to provide a reference point or *baseline* behaviour for future experiments, we believe that the simulation dynamics should be allowed to progress for a longer period, which nears completion of the chemical reactions.

Using the calibrated time interval of 30 iterations per second, this equates to 1,800 iterations per minute, or 108,000 iterations per hour. As such, 432,000 iterations (4 hours) will be used to illustrate the running of the system to completion, as there is no negative feedback through IκBα transcription and translation (figure 6.23). Finally, the 3D visualisation of the calibrated computational model is shown in figure 6.24, and highlights the two concentric circles (cytoplasm and nucleus) that represent the cell environment, along with the various membrane-bound and nuclear/cytoplasmic agents.

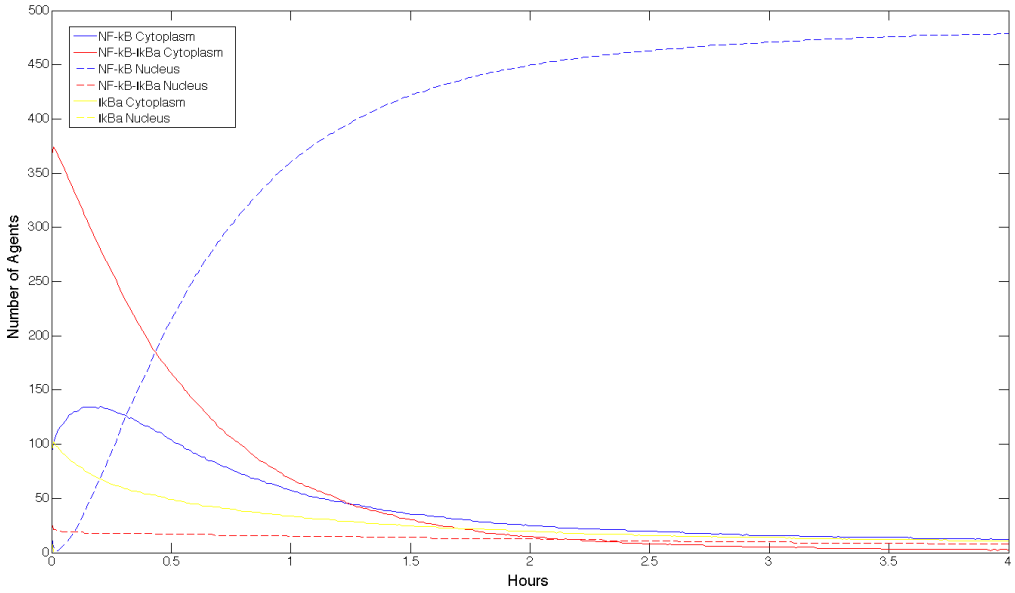

Figure 6.23: IL-1 stimulation dynamics for the NF-κB signalling pathway calibrated against real-world time. The computational model, using the calibrated parameter values generates emergent behaviour over a 432,000 iteration simulation. Through extrapolation of the dynamics from Yang et al. (2003), time has been calibrated to 108,000 iterations approximating to 1 hour of real-world time.

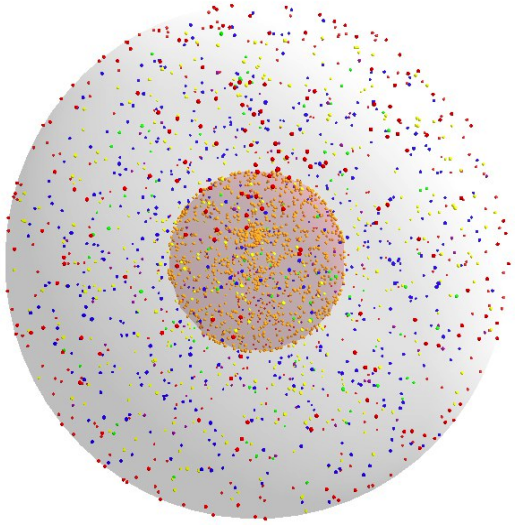

Figure 6.24: 3D visualisation of an IL-1 stimulated simulation. The cytoplasmic and nuclear locations of the cell envrionment are represented through the use of two concentric circles, with NF-κB and IκBα agents located throughout.

6.5 Implications of Hardware on Performance of Simulations

The initial simulations that were run as part of the calibration process, were run on a Windows desktop PC, which had quad-core processor and 8GB RAM. Even though three simulations could be run in parallel using a distributed mode, it quickly became evident that although the FLAME framework had been written in C to optimise performance, it still required a prohibitive amount of resources. For example, the 50,000 iteration simulations (for the IL-1 stimulated dynamics) took ∼8hr to run, with an additional 10hr to transform and process the XML results data to more usable CSV files through Python scripts. Another observation is that the simulator generates vast amounts of XML output data, with the raw XML files from FLAME amounting to 30GB and the Python scripts generating another 10GB per 50,000 iteration simulation run. This is because FLAME utilises the concept of communicating stream X-Machines, and generates a separate XML file (containing the parameter values and states for every instantiated agent) for each time-step. Therefore a 50,000 iteration run with 2,000 agents, generates 50,000 separate XML output files, each containing the detailed XML-tagged information relating to every single one of the 2,000 agents at that moment in time.

Diagnostic tests regarding the performance of FLAME indicated the length of time required to complete single simulation runs increases linearly with respect to the number of iterations in the simulation. It also became evident that the rate-limiting step of simulations was the Input/Output (I/O) speed, which was a direct consequence of the need to create a separate XML output file for each iteration. As the output XML file for a given iteration, becomes the input XML file for the next iteration, a performance bottleneck is formed, relating to the speed of writing to, and reading from the storage disk. To confrm this, we performed replicate simulations using the same Windows desktop PC, but this time writing output files to an external USB hard drive. As suspected, the performance was much worse, due to the flow of data across the USB port. Additional tests were performed using an Apple Mac mini with Solid State Disk (SSD) and the Northern Eight Universities Consortium (N8) High-Performance Computing (HPC) architecture. The Apple Mac mini with Solid State Disk (SSD), was found to be 28% faster than the Windows PC with internal disk (or 38% faster than the Windows PC with external USB hard drive), and reduced the running time of a 50,000 iteration simulation by two hours. Similarly, the N8 HPC was found to be 85% faster than the Mac mini with SSD, and reduced the running time down to just over 2 hours.

6.6 Summary

We have encoded the technical specification, as defined by the platform model, into a simulation platform that is suitable for *in silico* experimentation. The Flexible Large-scale Agent-based Modelling Environment (FLAME) was used, as this allows the modeller to focus on the agent and environment definitions, as the simulation framework takes care of the necessary parsing of the C and XML code

into executable code, and scheduling of the X-Machine functions throughout each iteration of a simulation run. Although the developmental approach advocated by the FLAME simulation framework is relatively intuitive to modellers familiar with the object-oriented and agent-based paradigms, it does suffer from a number of limitations, which resulted in technical issues during the development of the simulation platform. The two main limitations were the absence of functionality to utilise global mutable parameters, and the inability to set the PRNG seed value at the simulation level. Unfortunately, these limitations are direct consequences of the underlying design of FLAME, because it was developed to run individual simulations across multiple clusters using parallel processing functionality. Two critical issues resulted from these limitations, the first resulting in a loss of IκBα agents during basal dissociation of NF-κB-IκBα complexes (the conservation of mass issue), and the second resulting in simulation runs not fully benefitting from the stochastic behaviour that should emerge through the use of pseudo-random number generators. Fortunately, a number of workarounds were found to resolve these technical issues, however we believe that future users of the FLAME simulation framework should be mindful of these limitations.

Following development of these workarounds, the simulation platform was calibrated using one-at-a-time and two-at-a-time parameter estimation. Our computational model was calibrated against wet-lab data of Carlotti et al. (2000) and Yang et al. (2003). The calibration process was iterative, with each parameter being the focus of investigation, until a multidimensional parameter space was found that yielded qualitative alignment to the underlying biological data on which the computational model was designed. Parameter estimation initially focused on differential time, IκBα rebind delay, agent association, basal dissociation of the NF-κB-IκBα complex, and nuclear translocation. Once the computational model was calibrated for control dynamics, the parameter estimation process was repeated for IL-1 stimulation dynamics, by focusing on IKK binding, nuclear import under IL-1 stimulated conditions, and IKK rebind delay. Through the calibration process, it became evident that the computational model suffers from a conservation of mass issue for particular regions of parameter space where basal dissociation is >0.1 and IκBα rebind delay is <3.

Epistemic and aleatory analysis were performed to understand the uncertainty arising from our incomplete knowledge of the real-world system that we are modelling, and the uncertainty arising from the stochastic nature of systems. The computational model was found to be very sensitive to the distance that agents move per time-step (differential time parameter), nuclear import under control conditions, the basal dissociation of NF-κB-IκBα complexes, IKK binding and nuclear import (under IL-1 stimulation) parameters. Conversely, the computational model was found to be robust to perturbations of IκBα rebind delay (when >3), and IKK rebind delay. This sensitivity analysis also highlighted the co-dependence between: association and basal dissociation; association and IκBα rebind delay; and IKK binding and nuclear import under IL-1 stimulation. We believe the results of this epistemic uncertainty analysis are consistent with the findings of Kitano (2004a), who conjectures there to be a delicate balance between robustness and fragility within complex dynamical biological systems. Furthermore, aleatory analysis not only confirmed the stochastic nature of our computational model, but was used to interpolate that a minimum of 175 replicate

simulation runs are required to mitigate the uncertainty within our simulations that is introduced through the use of PRNGs.

These preliminary simulations, which have formed the basis of development, calibration and uncertainty analysis have also provided an insight to the resource intensive nature of FLAME, and the implications for future *in silico* experimentation. Due to FLAME utilising the concept of communicating stream X-Machines, even relatively small simulations can generate a disproportionate amount of data. Diagnostic tests regarding the performance of FLAME indicated that the rate-determinant step for the total time taken to run a simulation was the I/O speed of the stroage medium. This, in conjunction with the requirement to run 175 replicates in order to gain stable median averages of simulation dynamics, means that desktop hardware is unsuitable for running simulations of our computational model. Access to high-performance computing architecture is therefore essential to ensure that our computational model can be used for *in silico* experimentation, with a view to furthering our understanding of the IL-1 stimulated NF-κB signalling pathway.

This chapter has addressed reasearch objective 2: Create an agent-based computational model of the core intracellular components of the IL-1 stimulated NF-κB signalling pathway; research objective 3: Investigate techniques for calibrating agent-based computational models that have been developed using the FLAME simulation framework; and research objective 4: Perform novel *in silico* experimentation using the agent-based model. Furthermore, it has also contributed to research objective 7: Investigate the suitability of using the FLAME simulation framework for developing computational models of complex biological systems.

7 Experimentation using the Baseline Simulator

The fourth and final product of a CoSMoS project, is the *results model*, which Andrews et al. (2010) state should *"encapsulate the understanding that results from simulation: the simulation platform behaviour, results of data collection and observations of simulation runs."* Whereas the simulation platform is the computational implementation of the platform model, which in our case was developed using FLAME (C and XML) and various python, ruby, and Matlab scripts (to pre-process, transform, and analyse data), the results model is the understanding of the system that has been gained through actual simulation runs, and the analysis of system dynamics[1].

Following the development and calibration of the simulation platform, along with the epistemic and aleatory uncertainty analysis, we are now in a position to perform *in silico* experimentation to increase our understanding of the underlying mechanistic behaviours of the NF-κB signalling pathway. As discussed in chapter 2, agent-based models lend themselves to investigating the temporal and spatial dynamics of systems, allowing us to investigate the consequences of perturbations to component interactions. This chapter therefore represents further novel *in silico* experimentation in to the IL-1 stimulated NF-κB signalling pathway (research objective 4).

Section 7.1 defines the motivation for the chapter. There are four areas for experimentation reported here for iteration 1: section 7.2 discusses the first experiment, which investigates the effects of varying basal dissociation of the NF-κB-IκBα complex; section 7.3 discusses the second experiment, which investigates the effects on system dynamics of varying the number of IKK agents; section 7.4 discusses the third experiment, which investigates the effects on system dynamics of varying the number of IκBα agents; and section 7.5 discusses the fourth experiment, which investigates the effects of varying the lag-time between activation of IL-1 cell membrane receptors and the subsequent activation of IKK. Finally, section 7.6 concludes this chapter.

[1]Note there is ongoing debate on whether the results of uncertainty analysis can be considered part of the *results model*. We have decided to decouple the results of uncertainty analysis (which was used to understand the effects of our incomplete knowledge of the real-world system and the stochastic nature of our computational model) from the results of more formal *in silico* experiments. This was achieved by merging the aleatory and epistemic uncertainty analysis into chapter 6, which focused on *simulation platform* development and calibration, and using chapter 7 as a standalone chapter to look at the experimentation involved with iteration 1.

7.1 Motivation behind In Silico Experimentation for Iteration 1

This chapter directly addresses research objective 4: perform novel *in silico* experimentation using the agent-based computational model of NF-κB. In addition to providing a contribution to the field of NF-κB computational research, these *in silico* experiments allow us to consider the extent to which our principled approach to development, and the agent-based modelling technique, explored in this thesis, provide confidence in simulation results being representative of the real-world domain.

There are four sets of *in silico* experiments reported in this chapter. Section 7.2 investigates the effects of varying basal dissociation of the NF-κB-IκBα complex, through four different experimental setups that varied the basal dissociation parameter value away from the calibrated value. The basal dissociation parameter was identified as a point of fragility within the signalling pathway during epistemic uncertainty analysis, and this experiment allows us to qualify the extent of this fragility using tests for statistical significance (KS-Test) and effect magnitude (A-Test).

The second theme of experimentation, presented in section 7.3, examines the effects of varying IKK numbers on the signalling pathway dynamics. An arbitrary number of IKK agents (50 agents) was used during simulation development, with the subsequent calibration exercise that focused on the six key parameters (differential time, rebind delay, basal dissociation, association, nuclear import, and nuclear export) compensating for this assumption. This experiment therefore allows us to qualify the change in system dynamics along with the extent of robustness (or indeed fragility) within the system that is directly attributable to IKK number.

The third theme of experimentation, presented in section 7.4, examines the effects of varying IκBα numbers on the signalling pathway dynamics. This is of particular importance due to IκBα's role in masking the nuclear localisation sequence of NF-κB dimers, and thus restricting them to the cytoplasmic compartment of cells when bound to form NF-κB-IκBα complexes. Furthermore, Pogson et al. (2008) showed through computational studies that excess free IκBα perturbs system dynamics, and that normal dynamics can be regained if we assume the excess IκBα is sequestered to cytoskeleton within the cell. This experiment therefore allows us to qualify the change in system dynamics with the incorporation of 3-fold IκBα that is either free within the cytoplasm, or sequestered to the cytoskeleton.

The fourth theme of experimentation, presented in section 7.5, examines the effects of varying the lag-time between cell membrane receptor activation and the subsequent IKK activation within the cytoplasm. An arbitrary count of 500 iterations (or 17 seconds) was used as the lag-time between the commencement of IL-1 stimulated simulations and the activation of IKK agents during simulation development. This experiment therefore allows us to qualify the change in system dynamics, along with the robustness of the system, that is directly attributable to the lag-time between the commencement of IL-1 stimulated simulation runs and the activation of IKK agents.

7.2 Basal Dissociation of the NF-κB-IκBα Complex

Sensitivity analysis of control dynamics has highlighted that the system is very sensitive to changes in the basal dissociation parameter value. It has been shown that free cytoplasmic NF-κB (following basal dissociation) does not become activated to transcribe the relevant genes, but instead requires IKK-mediated release from the inhibited complex, and following translocation to the nucleus (Carlotti et al., 1999) may then be activated. Furthermore, this IKK-mediated release of NF-κB, is also linked to the phosphorylation and subsequent polyubiquitination and proteasomal degradation of the associated IκBα molecule (Alkalay et al., 1995b).

Carlotti et al. (2000) investigated the shuttling of NF-κB (RelA) and IκBα between the cytoplasm and nucleus. They postulated that the shuttling is a consequence of cytoplasmic dissociation of the NF-κB-IκBα complex, rather than the direct nuclear import of the NF-κB-IκBα complex, degradation of IκBα, and subsequent resynthesis of IκBα.

With increased basal dissociation, you would expect to see less degradation of IκBα within the system, and less NF-κB capable of promoting the transcription of genes within the nucleus. Similarly, you would expect to see a higher level of free NF-κB and IκBα within the nucleus due to the increased degree of shuttling between the cytoplasm and nucleus. This increase in nuclear NF-κB would support the mathematical model predictions of Carlotti et al. (2000), who suggested changes in association-dissociation would have a pronounced impact on nuclear NF-κB levels, and we believe may be useful for stimuli that modify IκBα inhibitory dynamics without resulting in its degradation. The Carlotti model was based around control (non-stimulated) dynamics of NF-κB and IκBα shuttling between the cytoplasm and nucleus, this allows us to perform *in silico* experimentation using our computational model to further their predictions to IL-1 stimulated dynamics. Our two working hypotheses are that an increase in basal dissociation of the inhibited NF-κB-IκBα complex to yield its constituent components will: 1) reduce the likelihood of activated NF-κB accumulating in the nucleus; and 2) increase the levels of free IκBα and NF-κB in the cytoplasm. Converting these to null hypotheses, we have:

H1$_0$ An increase in basal dissociation of the inhibited NF-κB-IκBα complex to yield its constituent components will not reduce the likelihood of activated NF-κB accumulating in the nucleus.

H2$_0$ An increase in basal dissociation of the inhibited NF-κB-IκBα complex to yield its constituent components will not increase the levels of free IκBα and NF-κB in the cytoplasm.

7.2.1 Experimental Procedure

Experimentation into the effects of varying basal dissociation of the NF-κB-IκBα complex, is conducted through perturbation of the *basal_dissociation_prob* simulation parameter. By default, this parameter is set to 0.000003, representing a probability of 0.0003% that an individual NF-κB-IκBα agent will basally dissociate at any given time-step within a simulation run. To test the two null hypotheses above, we ran four sets of *in silico* experiments using 10x, 100x, 1,000x

and 10,000x calibrated dissociation values (i.e. 0.00003, 0.0003, 0.003, and 0.03, respectively). For each set of 175 simulation replicates, the median distributions of cytoplasmic NF-κB, nuclear NF-κB, cytoplasmic NF-κB-IκBα, nuclear NF-κB-IκBα, cytoplasmic IκBα and nuclear IκBα over the lifetime of the simulation runs are interpolated. These distributions are contrasted with the baseline behaviour that results from simulations using the default (calibrated) parameter value. KS-Tests are then performed to understand whether there are any statistically significant differences from baseline behaviour (requiring p-values below 0.05), and A-Tests are also performed to understand the effect magnitude of these differences, assuming 'large' differences of < 0.29 and > 0.71 to be scientifically significant.

7.2.2 Results

Results indicate that the system is fragile to perturbations involving basal dissociation of the NF-κB-IκBα complex. Figure 7.1 depicts the effects on cytoplasmic IκBα numbers when the *basal_dissociation_prob* parameter value is set to: 10x, 100x, 1,000x and 10,000x with respect to the default parameter value. As the parameter value increases up the orders of magnitude, the system appears to spontaneously dissociate large numbers of NF-κB-IκBα agents to their respective free agents immediately after simulation runs begin. Indeed, with 1,000x and 10,000x default, the system appears to dissociate virtually all of the NF-κB-IκBα complexes into their respective NF-κB and IκBα agents within a minute.

Figure 7.2 depicts the effects on nuclear NF-κB numbers when the parameter value for *basal_dissociation_prob* is perturbed. It is clear that increases in basal dissociation probabilities, result in a reduction of NF-κB agents within the nucleus. Similarly, figure 7.3 depicts the effects on cytoplasmic NF-κB-IκBα numbers when the parameter value is perturbed. Here it is clear that cytoplasmic numbers markedly reduce, with increasing basal dissociation.

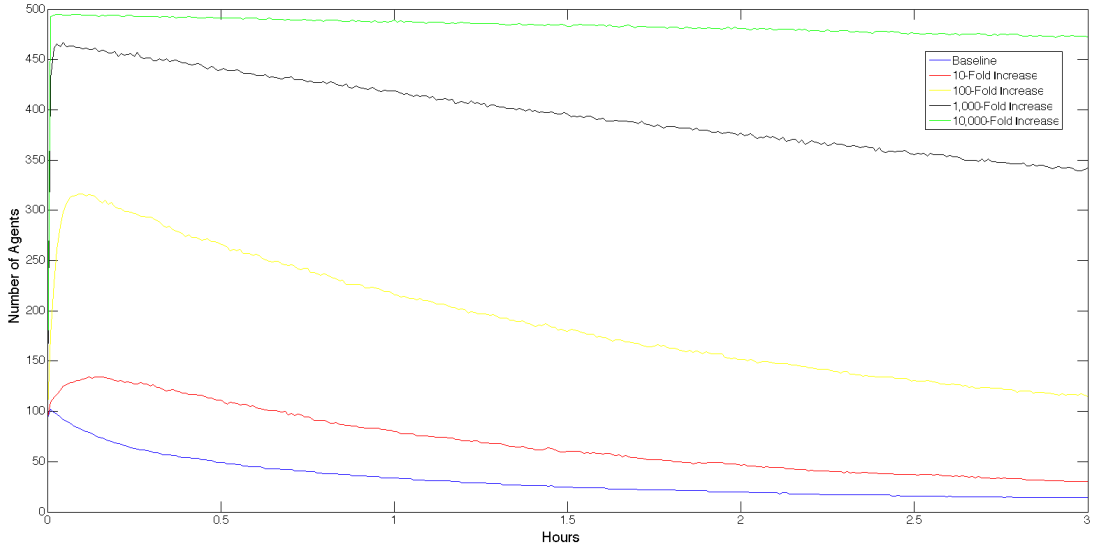

Figure 7.1: Cytoplasmic IκBα dynamics for the four sets of experiments into basal dissociation of the NF-κB-IκBα complex, which have been compared against baseline IL-1 stimulated dynamics.

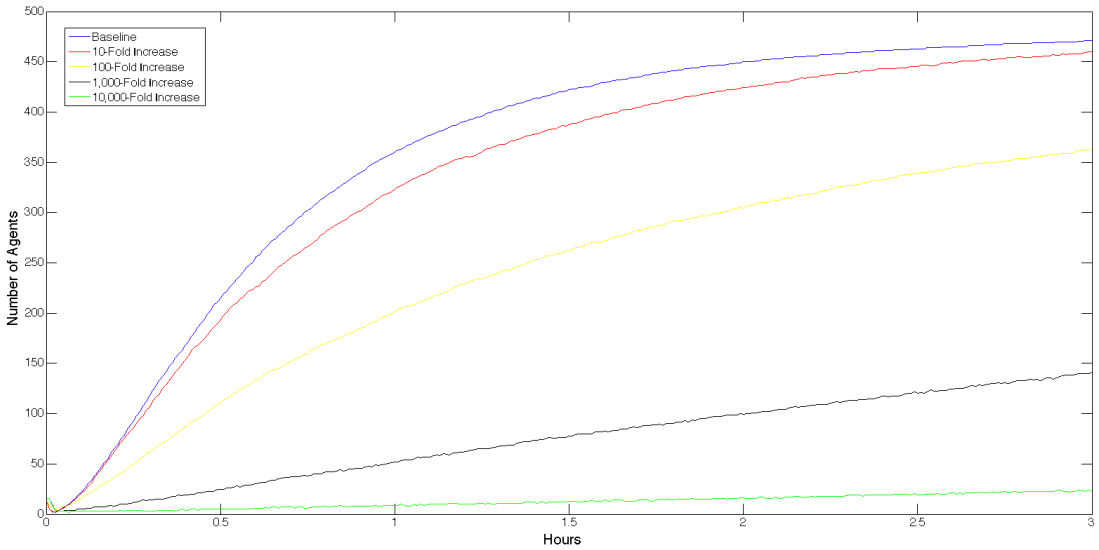

Figure 7.2: Nuclear NF-κB dynamics for the four sets of experiments into basal dissociation of the NF-κB-IκBα complex, which have been compared against baseline IL-1 stimulated dynamics.

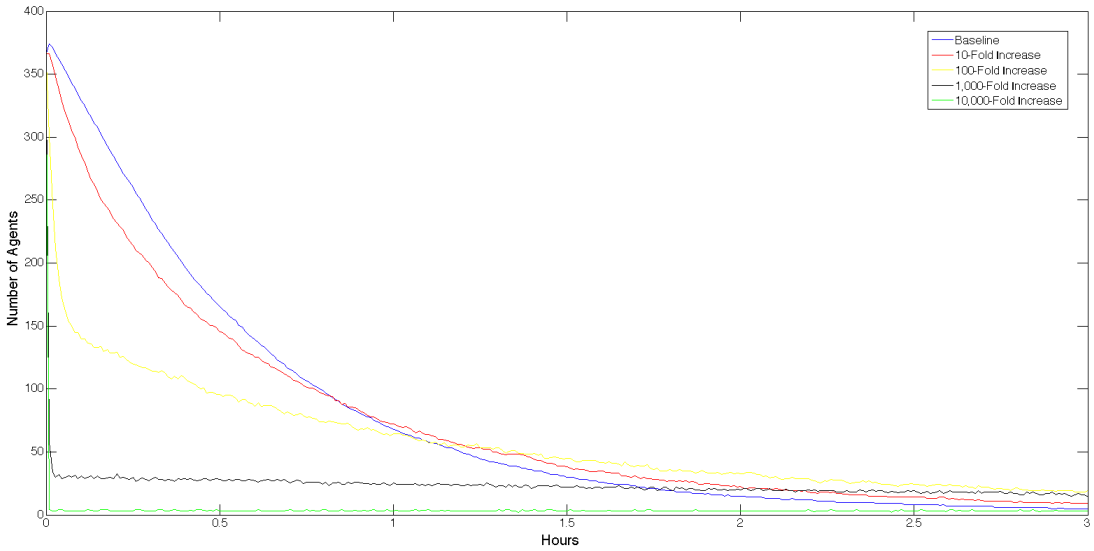

Figure 7.3: Cytoplasmic NF-κB-IκBα dynamics for the four sets of experiments into basal dissociation of the NF-κB-IκBα complex, which have been compared against baseline IL-1 stimulated dynamics.

Tables 7.1 and 7.2 depict the associated KS-Test p-values and A-Test scores resulting from the perturbations to the basal dissociation parameter with respect to the default value. It can be seen that all perturbations to the basal dissociation probability provide statistically significant differences (with respect to baseline dynamics) across all agent states, apart from nuclear IκBα for the 10x default basal dissociation probability. This is due to the KS-Tests generating p-values less than 0.05 for all agent states apart from IκBα for the 10x default in basal dissociation probability, which has p-value 0.4931. Furthermore, although KS-Tests with p-values below 0.05 indicate a statistically significant difference, the A-Test is used to calculate whether these differences yield a significant effect magnitude. Scientifically significant differences exist for: cytoplasmic NF-κB in all experiments; cytoplasmic NF-κB-IκBα in 10,000x default; nuclear NF-κB-IκBα in all experiments; nuclear NF-κB and activated NF-κB in all experiments apart from 10x default; cytoplasmic IκBα in all experiments; and nuclear IκBα in all experiments apart from 10x default.

Agent State	**10x**	**100x**	**1,000x**	**10,000x**
NF-κB Cyto	0.0	0.0	0.0	0.0
NF-κB-IκBα Cyto	0.0035	0.0008	0.0	0.0
NF-κB-IκBα Nuc	0.0	0.0	0.0	0.0
NF-κB Nuc	0.0	0.0	0.0	0.0
Active NF-κB Nuc	0.0	0.0	0.0	0.0
IκBα Cyto	0.0	0.0	0.0	0.0
IκBα Nuc	0.4931	0.0	0.0	0.0

Table 7.1: KS-Test scores for the four sets of *in silico* experiments, which have been compared against baseline IL-1 stimulated dynamics. All scores apart from IκBα Nucleus for the 10x default in basal dissociation probability show statistically significant differences, as they represent p-values below 0.05.

Agent State	**10x**	**100x**	**1,000x**	**10,000x**
NF-κB Cyto	0.2482	0.0101	0.0006	0.0006
NF-κB-IκBα Cyto	0.4401	0.4111	0.5653	0.9969
NF-κB-IκBα Nuc	0.9419	0.9919	0.9969	0.9969
NF-κB Nuc	0.6072	0.8010	0.9297	0.9774
Active NF-κB Nuc	0.6074	0.8019	0.9308	0.9781
IκBα Cyto	0.1343	0.0	0.0	0.0
IκBα Nuc	0.4678	0.1841	0.0036	0.0031

Table 7.2: A-Test scores for the four sets of *in silico* experiments, which have been compared against baseline IL-1 stimulated dynamics. A-Test scores show scientifically significant differences for: cytoplasmic NF-κB in all experiments; cytoplasmic NF-κB-IκBα using 10,000x default; nuclear NF-κB-IκBα in all experiments; nuclear NF-κB and activated NF-κB in all experiments apart from 10x default; cytoplasmic IκBα in all experiments; and nuclear IκBα in all experiments apart from 10x default.

7.3 Effect of IKK Numbers

Following IL-1 stimulation and activation of receptor complex components, the signal transduction reaches the NF-κB-IκBα signalling module (Hoffmann et al., 2002), with the key step being the IKK-mediated dissociation (and resultant degradation) of IκBα (Chen et al., 1996). This causes release of NF-κB, and facilitates its translocation into the nucleus and subsequent activation, to initiate transcription of genes for the inflammatory response.

Simulation dynamics have been calibrated using 501 IκBα and 502 NF-κB (both in various states of free, bound, cytoplasmic and nuclear), along with 500 IL-1R, and 50 IKK agents. This provides an opportunity for us to investigate the effects of varying IKK numbers (akin to varying concentration in wet-lab experiments), and whether the system is robust to such perturbations. As there is an amplification step, i.e. one IKK agent can facilitate dissociation of many NF-κB-IκBα complexes over the lifetime of a simulation, it is expected that an increase in IKK numbers would increase the rate at which active NF-κB accumulates in the nucleus. Converting this to null hypotheses, we have:

H3$_0$ An increase in IKK concentration will not increase the rate of dissociation of NF-κB-IκBα complexes, and resultant degradation of IκBα.

H4$_0$ An increase in IKK concentration will not increase the rate at which activated NF-κB accumulates in the nucleus.

7.3.1 Experimental Procedure

Experimentation into the effects of varying IKK numbers, is conducted through perturbation of the total number of IKK agents. By default, there were 50 IKK agents within the calibrated simulation platform. To test the two null hypotheses above, we ran six sets of *in silico* experiments using a tenth, a fifth, a half, 2x, 5x, and 10x default numbers (i.e. 5, 10, 25, 100, 250 and 500 IKK agents, respectively). For each set of 175 simulation replicates, the median distributions of cytoplasmic NF-κB, nuclear NF-κB, cytoplasmic NF-κB-IκBα, nuclear NF-κB-IκBα, cytoplasmic IκBα and nuclear IκBα over the lifetime of the simulation runs are interpolated. These distributions are contrasted with the baseline behaviour that results from simulations using the default number of IKK agents. KS-Tests are then performed to understand whether there are any statistically significant differences from baseline behaviour (requiring p-values below 0.05), and A-Tests are also performed to understand the effect magnitude of these differences, assuming 'large' differences of < 0.29 and > 0.71 to be scientifically significant.

7.3.2 Results

Results indicate that the system is sensitive to perturbations involving the total number of IKK agents, which when using stable numbers of NF-κB and IκBα agents, has the effect of altering the ratio of IKK to NF-κB and the ratio of IKK to IκBα. Figure 7.4 depicts the effects on cytoplasmic IκBα numbers when the total number of IKK agents is set to: 10%, 20%, 50%, 200%, 500% and 1,000% of the

default number. It can be seen that increasing the number of IKK agents, has the effect of decreasing the number of IκBα agents within the cytoplasm. Similarly, figure 7.5 depicts the effects on cytoplasmic NF-κB-IκBα numbers, and figure 7.6 depicts the effects on cytoplasmic NF-κB numbers when the total number of IKK agents is perturbed. Here it can be seen that an increase in IKK numbers results in both a decrease in cytoplasmic NF-κB-IκBα numbers, but also an increase in the rate of dissociation of the NF-κB-IκBα complexes. The effects on cytoplasmic NF-κB numbers are not as straight forward to interpret, as there appears to be a non-linear temporal component at play, whereby an increase in IKK numbers results in an initial increase in cytoplasmic NF-κB numbers upto approximately 15min, after which there is a sharp decrease, resulting in a negative relationship between the number of IKK agents and the number of cytoplasmic NF-κB agents at the end of simulation runs.

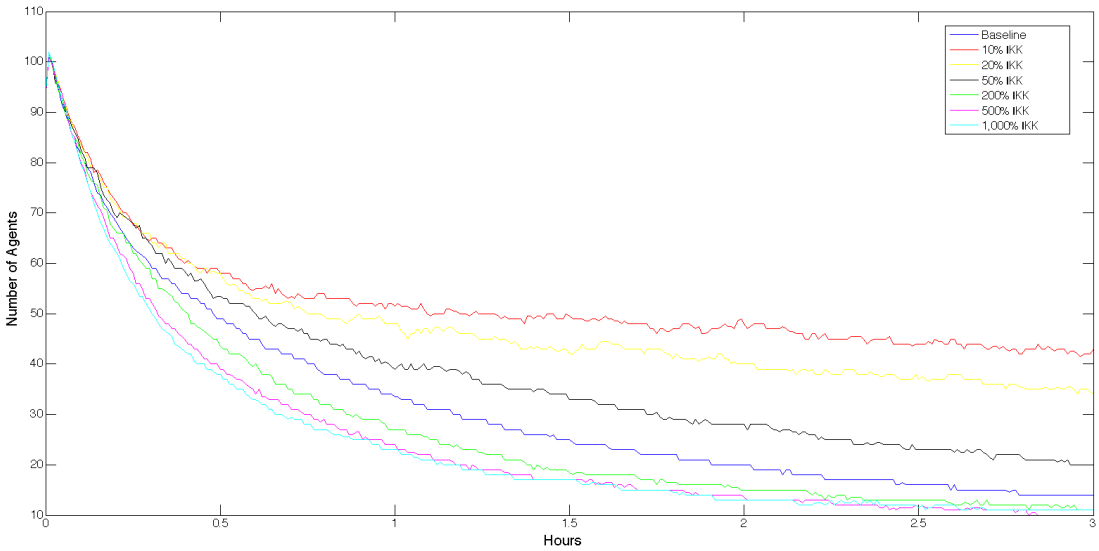

Figure 7.4: Cytoplasmic IκBα dynamics for the six sets of experiments into IKK concentration, which have been compared against baseline IL-1 stimulated dynamics.

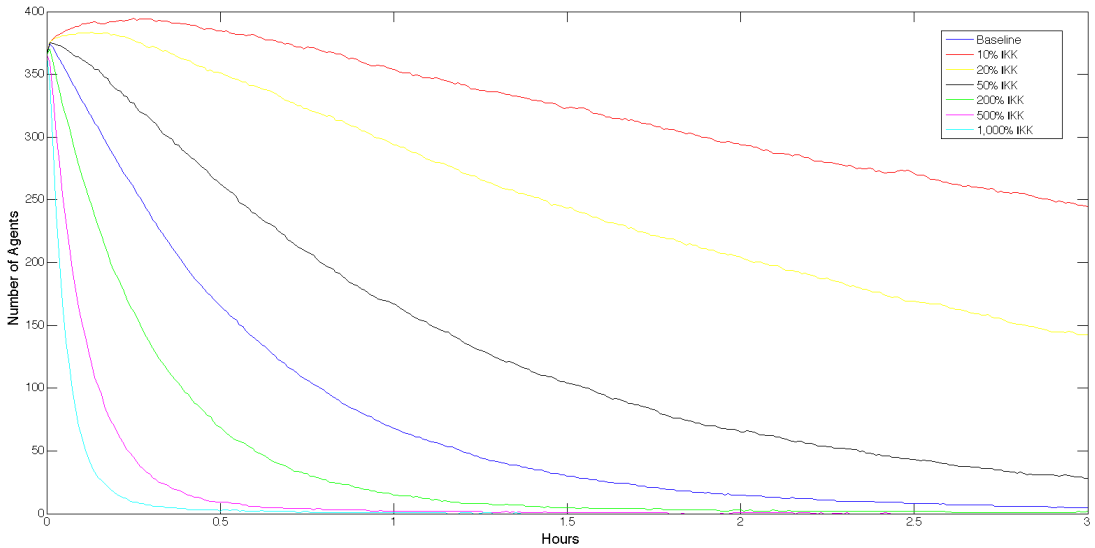

Figure 7.5: Cytoplasmic NF-κB-IκBα dynamics for the six sets of experiments into IKK concentration, which have been compared against baseline IL-1 stimulated dynamics.

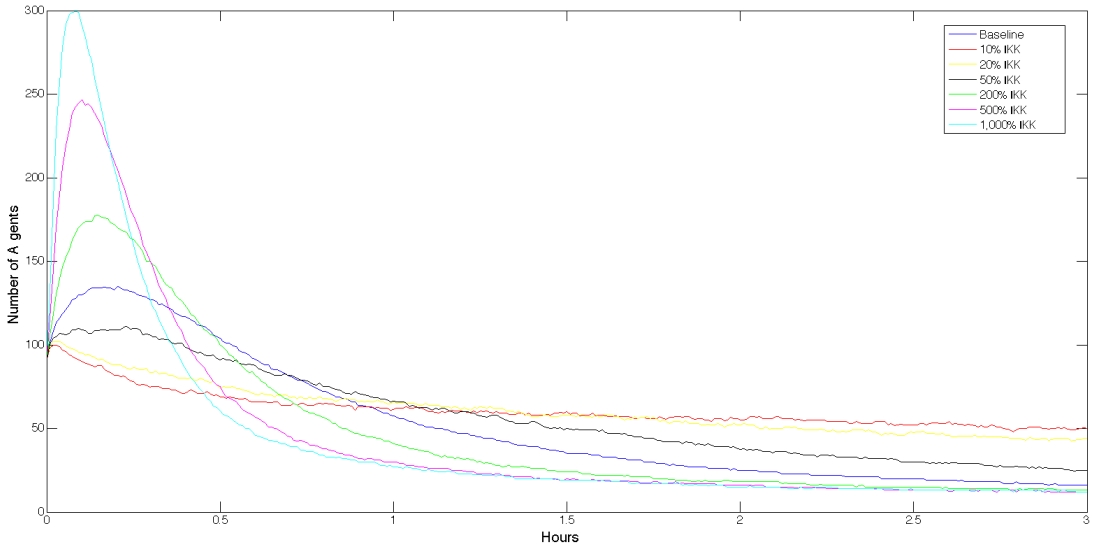

Figure 7.6: Cytoplasmic NF-κB dynamics for the six sets of experiments into IKK concentration, which have been compared against baseline IL-1 stimulated dynamics.

Tables 7.3 and 7.4 depict the associated KS-Test p-values and A-Test scores resulting from the perturbations to IKK numbers with respect to the default number. It can be seen that all perturbations to IKK numbers provide statistically significant differences (with respect to baseline dynamics) across all agent states, apart from nuclear IκBα, and the cytoplasmic and nuclear NF-κB-IκBα agent states when 50% of default IKK numbers are used (i.e. 25 IKK agents). This is due to the KS-Tests generating p-values less than 0.05 for all agent states apart from the cytoplasmic NF-κB-IκBα agents, which have a p-value of 0.0803, and nuclear NF-κB-IκBα agents, which have a p-value of 0.0979, when 50% of default IKK numbers are used. Furthermore, the A-Test was used to calculate the corresponding effect magnitude due to the perturbations, and whether scientifically significant differences result. Scientifically significant differences exist for: cytoplasmic NF-κB-IκBα in all perturbations, nuclear NF-κB and active NF-κB for perturbations that used 10%, 20%, 500% and 1,000% IKK (with respect to default); and cytoplasmic IκBα for perturbations that used 10% and 20% of IKK (with respect to default).

Agent State	**10%**	**20%**	**50%**	**200%**	**500%**	**1,000%**
NF-κB Cyto	0.0	0.0	0.0	0.0	0.0	0.0
NF-κB-IκBα Cyto	0.0	0.0	0.0803	0.0001	0.0	0.0
NF-κB-IκBα Nuc	0.0342	0.0102	0.0979	0.0008	0.0005	0.0001
NF-κB Nuc	0.0	0.0	0.0	0.0	0.0	0.0
Active NF-κB Nuc	0.0	0.0	0.0	0.0	0.0	0.0
IκBα Cyto	0.0	0.0	0.0	0.0	0.0	0.0
IκBα Nuc	1.0	1.0	1.0	1.0	1.0	1.0

Table 7.3: KS-Test scores for the six sets of *in silico* experiments, which have been compared against baseline IL-1 stimulated dynamics. All scores apart from IκBα Nucleus (for all experiments) and NF-κB-IκBα in the cytoplasm and nucleus for the perturbation that used 50% IKK (with respect to default) show statistically significant differences.

Agent State	10%	20%	50%	200%	500%	1,000%
NF-κB Cyto	0.319	0.3263	0.3857	0.617	0.6873	0.7077
NF-κB-IκBα Cyto	0.0402	0.0939	0.2550	0.7492	0.8961	0.9465
NF-κB-IκBα Nuc	0.4599	0.5713	0.5342	0.5581	0.5025	0.4763
NF-κB Nuc	0.9120	0.8583	0.7059	0.3284	0.2364	0.2133
Active NF-κB Nuc	0.9117	0.8581	0.7059	0.3283	0.2364	0.2133
IκBα Cyto	0.1582	0.2022	0.3376	0.6229	0.6786	0.6922
IκBα Nuc	0.5	0.5	0.4985	0.5	0.5	0.5

Table 7.4: A-Test scores for the six sets of *in silico* experiments, which have been compared against baseline IL-1 stimulated dynamics. These scores suggest there to be scientifically significant differences for cytoplasmic NF-κB-IκBα in all perturbations, nuclear NF-κB and active NF-κB for perturbations that used 10%, 20%, 500% and 1,000% IKK (with respect to default), and cytoplasmic IκBα for perturbations that used 10% and 20% of IKK (with respect to default).

7.4 Effect of IκBα Numbers

IκBα molecules inhibit NF-κB by masking their nuclear localisation sequence (NLS), thus restricting the NF-κB dimers to the cytoplasm. Carlotti et al. (2000) have shown that free IκBα competes with free NF-κB for translocation to the nucleus, and that approximately 17% of NF-κB remains free within resting cells, and does not form an inhibited complex with IκBα. Similarly, Yang et al. (2003) and Carlotti et al. (1999) predicted a larger total number of IκBα molecules than NF-κB molecules within cells (approximately 3:1 ratio). Furthermore, Pogson et al. (2008) have shown through computational modelling, that the system is more sensitive to changes in IκBα concentration than NF-κB concentration. They assumed unchanged levels of the inhibited complex, so linked excess IκBα to cytoskeletal sequestration in order to maintain the 17% free NF-κB predicted by Carlotti et al. (2000). They found maximal inhibited complex formation was reached at a 3:1 ratio of IκBα to NF-κB, confirming the earlier predictions of Yang et al. (2003) and Carlotti et al. (1999), and concluded that the excess IκBα was sequestered to the actin cytoskeleton.

Our calibrated simulation platform contains similar numbers of IκBα and NF-κB agents (501 v 502). This provides an opportunity to investigate the effects on control and IL-1 stimulated dynamics of perturbing the system through a 2-fold increase of free and sequestered IκBα agents. We believe the addition of excess free IκBα will generate catastrophic failures to system dynamics, but that addition of excess IκBα that is sequestered within the cell, will result in behaviour similar to that of the calibrated system. Converting this to null hypotheses, we have:

H5$_0$ The addition of excess free IκBα within the cytoplasm will not perturb system dynamics into a fragile state.

H6$_0$ The addition of excess IκBα sequestered within the cell will not result in stable system dynamics.

7.4.1 Experimental Procedure

Experimentation into the effects of varying IκBα numbers, is conducted through perturbation of the total number of IκBα agents. By default, there were 501 IκBα agents within the calibrated simulation platform. To test the two null hypotheses we ran four sets of experiments, two under control (non-stimulated) conditions and two under IL-1 stimulated conditions, using: 1,500 IκBα all free within the cell for control dynamics; 1,500 IκBα agents, corresponding to 500 free and 1,000 sequestered within the cell for control dynamics; 1,500 IκBα agents using the new pseudo steady-state dynamics from the control experiment using excess free IκBα, for IL-1 stimulated dynamics; and 1,500 IκBα agents, using the new pseudo steady-state dynamics from the control experiment using excess sequestered IκBα, for IL-1 stimulated dynamics. For each set of 175 simulation replicates, the median distributions are interpolated. These distributions are contrasted with the control behaviour (for non-stimulated conditions) and baseline behaviour (for IL-1 stimulated conditions) that result from simulations using the default number of IκBα agents. KS-Tests and A-Tests are then performed to investigate significant differences from control or baseline behaviours.

7.4.2 Results

Results indicate that the system is fragile to perturbations involving the numbers of free IκBα. Figure 7.7 depicts the effects on cytoplasmic NF-κB numbers, and figure 7.8 depicts the effects on the number of cytoplasmic NF-κB-IκBα complexes, when perturbations for 3-fold free IκBα and 3-fold with 66% actin-bound IκBα are applied. It can be seen that the addition of 1,000 actin-bound IκBα agents, results in the system approximating to baseline IL-1 stimulated dynamics. Conversely, with the addition of 1,000 free IκBα agents, the system is significantly perturbed, resulting in an initial spike of free cytoplasmic NF-κB between 5 - 10min, a corresponding decrease in cytoplasmic NF-κB-IκBα, then a rapid decrease in cytoplasmic numbers due to nuclear translocation.

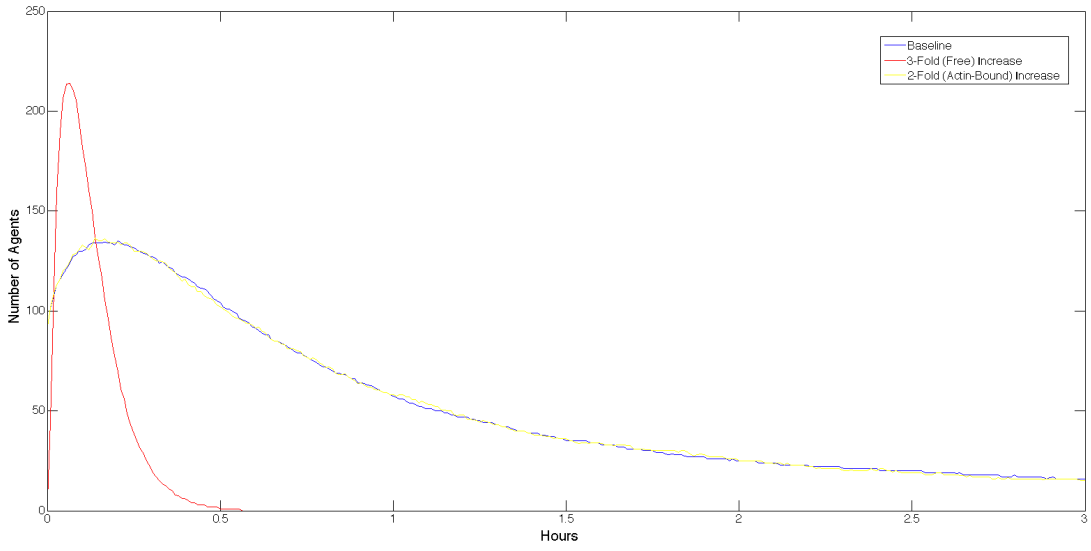

Figure 7.7: Cytoplasmic NF-κB dynamics for perturbations that used 3x free IκBα and perturbations that used default levels of free IκBα with the additional being sequestered by the actin cytoskeleton.

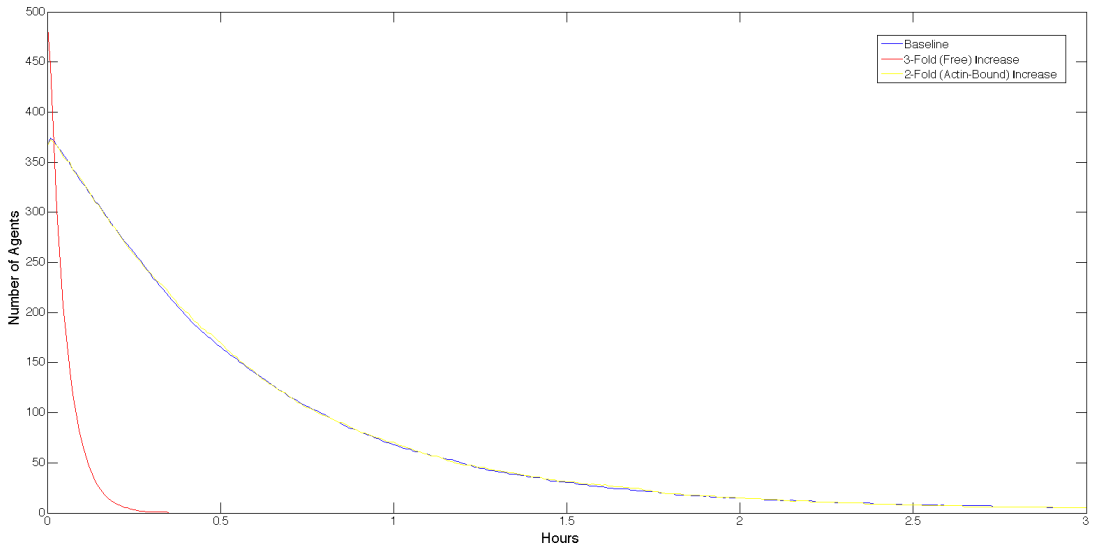

Figure 7.8: Cytoplasmic NF-κB-IκBα dynamics for perturbations that used 3x free IκBα and perturbations that used default levels of free IκBα with the additional being sequestered by the actin cytoskeleton.

Tables 7.5 and 7.6 depict the associated KS-Test p-values and A-Test scores resulting from the perturbations to IκBα numbers with respect to the default number. It can be seen that all perturbations using 3-fold IκBα numbers result in statistically and scientifically significant differences for cytoplasmic NF-κB. Additionally, the KS-Test and A-Test results indicate there to be both statistical and scientific significant differences for cytoplasmic and nuclear NF-κB-IκBα, when perturbations of 3-fold free IκBα under non-stimulated conditions are applied, and for all agent states, apart from cytoplasmic IκBα, when 3-fold free IκBα under IL-1 stimulation is applied.

Agent State	3x Free	3x Seq	3x Free (IL-1)	3x Seq (IL-1)
NF-κB Cyto	0.0	0.0	0.0	0.0
NF-κB-IκBα Cyto	0.0	0.9999	0.0	0.9913
NF-κB-IκBα Nuc	0.0	1.0	0.0	0.9999
NF-κB Nuc	1.0	0.9869	0.0	0.3264
Active NF-κB Nuc	1.0	1.0	0.0	0.9999
IκBα Cyto	1.0	1.0	0.0	0.3774
IκBα Nuc	1.0	1.0	0.0	1.0

Table 7.5: KS-Test scores for the four sets of *in silico* experiments, which have been compared against control (non-stimulated) and baseline IL-1 stimulated dynamics. Statistically significant differences are shown for: cytoplasmic NF-κB under all conditions; cytoplasmic and nuclear NF-κB-IκBα for perturbations that applied 3-fold free IκBα under non-stimulated conditions; and all agent states for perturbations that applied 3-fold free IκBα under IL1-stimulated conditions.

Agent State	3x Free	3x Seq	3x Free (IL-1)	3x Seq (IL-1)
NF-κB Cyto	0.9022	0.5022	0.9221	0.5054
NF-κB-IκBα Cyto	0.0829	0.4935	0.9541	0.4999
NF-κB-IκBα Nuc	0.6536	0.5108	0.9992	0.5306
NF-κB Nuc	0.9907	0.5346	0.0645	0.4976
Active NF-κB Nuc	0.5	0.5	0.0644	0.4978
IκBα Cyto	0.0	0.4987	0.3889	0.5099
IκBα Nuc	0.0369	0.493	0.0	0.5

Table 7.6: A-Test scores for the four sets of *in silico* experiments, which have been compared against baseline IL-1 stimulated dynamics. These scores indicate scientifically significant differences for: all agent states with perturbations that applied 3-fold free IκBα under non-stimulated conditions, apart from nuclear NF-κB-IκBα and active NF-κB; and all agent states with perturbations that applied 3-fold free IκBα under IL1-stimulated conditions apart from cytoplasmic IκBα.

7.5 Effect of Lag-Time before IKK Activation

Our calibrated simulation platform for the first iteration, contains a lag-time (countdown timer) between the physical commencement of simulation runs and activation of the IKK agents within these simulations, due to the activation of the cell membrane receptor complexes being abstracted away (although the focus of iteration 2 in chapter 8). Currently, iteration 1 has a 250 iteration (approximately 8.5 seconds) lag-time before all IL-1R agents are activated, and then a further 250 iteration lag-time (i.e. 500 iterations in total or 17 seconds) before IKK agents are activated. This provides an opportunity for us to investigate whether varying the lag-time before IKK activation has any effect on overall system dynamics, or whether it simply delays activation of the pathway, using the following null hypothesis:

H7$_0$ An increase in lag-time for IKK activation will only act to delay pathway dynamics, and will not affect the overall dynamics regarding IκBα degradation and NF-κB activation.

7.5.1 Experimental Procedure

Experimentation into the effects of varying the lag-time before IKK activation, is conducted through perturbation of the IKK *stimulation_delay* simulation parameter. By default, this parameter is set to 500, which approximated to 17s of physical time. To test the null hypothesis, we ran five sets of *in silico* experiments under IL-1 stimulated conditions, using the following *real-world* lag-times for IKK activation: 15s ($\sim$ default), 1min ($\sim$4x default), 2.5min ($\sim$10x default), 5min ($\sim$20x default), and 10min ($\sim$40x default). For each set of 175 simulation replicates, the median distributions of cytoplasmic NF-κB, nuclear NF-κB, cytoplasmic NF-κB-IκBα, nuclear NF-κB-IκBα, cytoplasmic IκBα and nuclear IκBα

over the lifetime of the simulation runs are interpolated. These distributions are contrasted with the baseline behaviour that results from simulations using the default parameter value. KS-Tests and A-Tests are then performed to investigate significant differences from baseline behaviour.

7.5.2 Results

Results unanimously indicate the system is robust to perturbations involving the lag-time delay (upto 40x default) between activation of the cell membrane receptors and the resultant activation of IKK. Figure 7.9 depicts the effects on cytoplasmic IκBα numbers when the IKK *stimulation_delay* simulation parameter value is set to: approximately equivalent, 4x, 10x, 20x and 40x default parameter value. No appreciable difference can be seen following increases (upto 40x) in the lag-time before IKK activation. Similarly, figure 7.10 depicting the effects on nuclear NF-κB numbers, and figure 7.11 depicting the effects on nuclear NF-κB-IκBα numbers, provide further support that there is no appreciable difference on system dynamics following perturbations (upto 40x) to the lag-time between cell membrane receptor activation and subsequent IKK activation.

Tables 7.7 and 7.8 depict the associated KS-Test p-values and A-Test scores resulting from the perturbations to the lag-time before IKK activation. It can be seen that the only statistically significant difference occurs to nuclear NF-κB-IκBα under a 40x default value in lag-time, due to the p-value of 0.0427. The corresponding A-Test score indicates no scientifically significant difference, and indeed there are no significant A-Test scores for any of the perturbations.

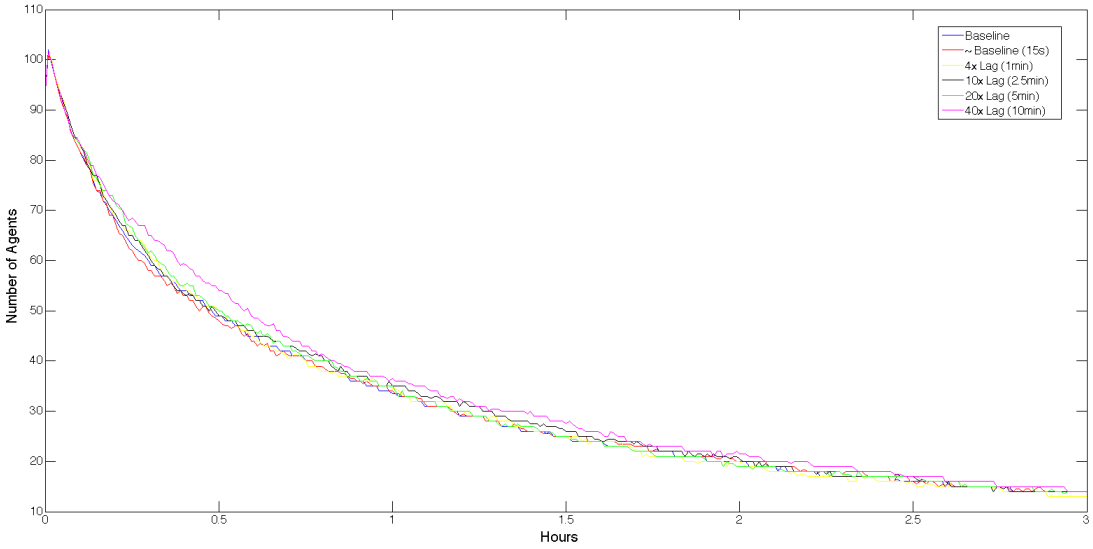

Figure 7.9: Cytoplasmic IκBα dynamics for the five sets of experiments varying the lag-time before IKK activation, which have been compared against baseline IL-1 stimulated dynamics.

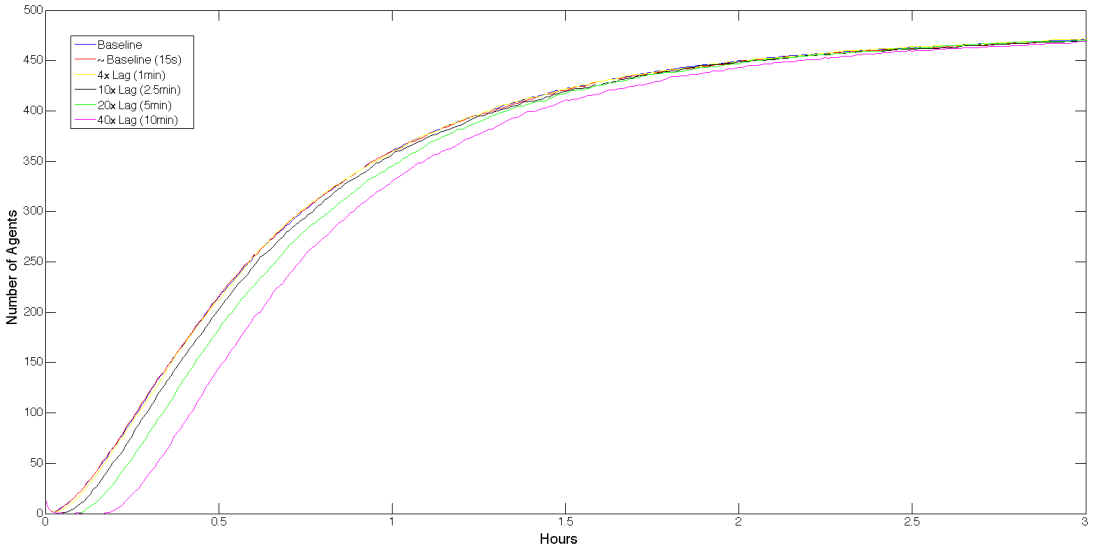

Figure 7.10: Nuclear NF-κB dynamics for the five sets of experiments varying the lag-time before IKK activation, which have been compared against baseline IL-1 stimulated dynamics.

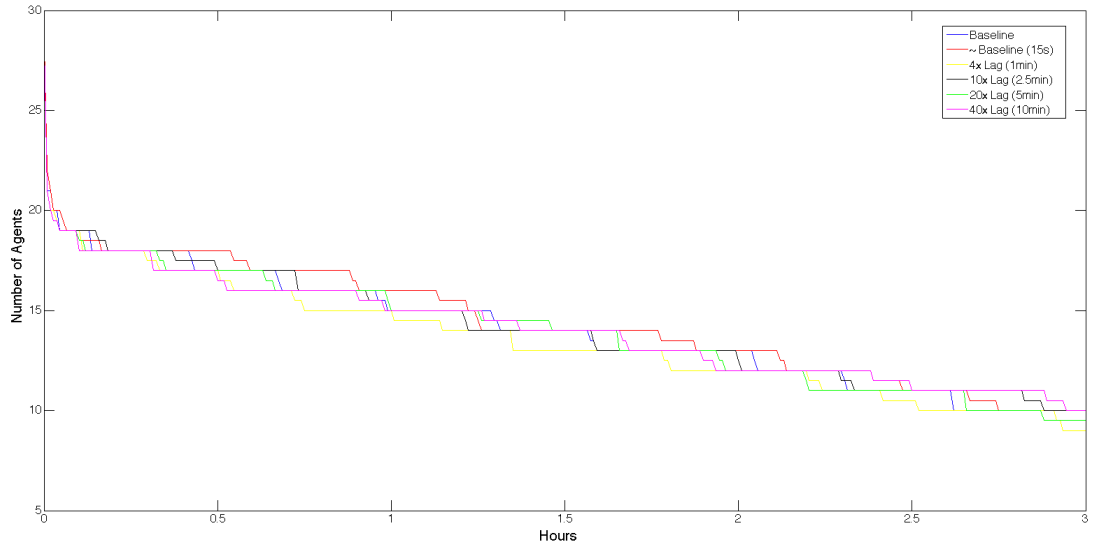

Figure 7.11: Nuclear NF-κB-IκBα dynamics for the five sets of experiments varying the lag-time before IKK activation, which have been compared against baseline IL-1 stimulated dynamics.

Agent State	**∼ default**	**∼ 4x**	**∼ 10x**	**∼ 20x**	**∼ 40x**
NF-κB Cyto	0.9976	0.9976	0.9535	0.9181	0.0979
NF-κB-IκBα Cyto	0.9976	0.9775	0.9181	0.9535	0.5566
NF-κB-IκBα Nuc	0.0979	0.4931	0.9535	0.9976	0.0427
NF-κB Nuc	0.9913	0.9999	0.1706	0.1185	0.3264
Active NF-κB Nuc	0.9181	0.9999	0.9775	0.9976	0.3264
IκBα Cyto	0.8718	0.9181	0.9181	0.9976	0.2391
IκBα Nuc	1.0	1.0	1.0	1.0	1.0

Table 7.7: KS-Test scores for the five sets of *in silico* experiments, which have been compared against baseline IL-1 stimulated dynamics. The scores indicate that only nuclear NF-κB-IκBα using 40x default value has a statistically significant difference from baseline IL-1 stimulated dynamics.

Agent State	**∼ default**	**∼ 4x**	**∼ 10x**	**∼ 20x**	**∼ 40x**
NF-κB Cyto	0.4988	0.5037	0.4945	0.4956	0.4684
NF-κB-IκBα Cyto	0.5022	0.4999	0.4877	0.4777	0.4527
NF-κB-IκBα Nuc	0.4623	0.5551	0.4936	0.5076	0.4957
NF-κB Nuc	0.5067	0.4992	0.5144	0.5182	0.551
Active NF-κB Nuc	0.5066	0.4988	0.5143	0.518	0.5506
IκBα Cyto	0.4933	0.505	0.4852	0.4918	0.4565
IκBα Nuc	0.5	0.5	0.5	0.5	0.5

Table 7.8: A-Test scores for the five sets of *in silico* experiments, which have been compared against baseline IL-1 stimulated dynamics. The scores indicate that there is no scientifically significant difference between any of the experiments or the baseline IL-1 stimulated dynamics.

7.6 Discussion

This chapter has reported the four *in silico* experiments that were performed using the simulation platform developed for iteration 1. The motivation behind these four experiments was presented in section 7.1, with the actual experimental procedures and simulation results being presented in sections 7.2 to 7.5.

Section 7.2 presented the background, experimental procedure, and results of *in silico* experimentation into basal dissociation of the NF-κB-IκBα complex. As the work of Carlotti et al. (2000) focused on control (i.e. non-stimulated) dynamics, and due to our epistemic uncertainty analysis on control dynamics (see section 6.3.1) indicating that the system is very sensitive to changes in the basal dissociation parameter value, this provided an ideal opportunity to test whether perturbations to basal dissociation of the NF-κB-IκBα complex also have an effect on system dynamics under IL-1 stimulation. We hypothesised that an increase in basal dissociation of the inhibited NF-κB-IκBα complex would reduce the likelihood of activated NF-κB accumulating in the nucleus, and also increase the levels of free NF-κB and IκBα in the cytoplasm (these were converted to their respective null hypotheses).

Results indicate that an increase in basal dissociation leads to: a decrease in both cytoplasmic and nuclear NF-κB-IκBα complexes; an increase in cytoplasmic NF-κB, with an associated decrease in nuclear NF-κB; and an increase in cytoplasmic IκBα, with a small increase in nuclear IκBα. From a biological perspective the reduction in nuclear NF-κB and cytoplasmic NF-κB-IκBα complexes, will be because an increase in basal dissociation results in a reduction of the total numbers of NF-κB-IκBα complexes in the cytoplasm, and thus a reduction in the IKK-mediated release of NF-κB, which further causes a reduction in the number of NF-κB agents that become activated following their translocation to the nucleus.

KS-Tests and subsequent A-Tests indicated scientifically significant differences exist for all scenarios, apart from: nuclear IκBα under 10x default value for basal dissociation, which we believe may be due to the low absolute numbers of agents located within the nucleus; and cytoplasmic NF-κB-IκBα using 10x, 100x, and 1,000x default values, which we believe may be due to an inability of the A-Test to account for the convergence of baseline, 10x and 100x data curves at approximately 1hr, and their convergence with the 1,000x data curve at approximately 2hr (see figure 7.3). We therefore believe that the A-Test scores, when viewed in isolation would yield a false negative in relation to the cytoplasmic NF-κB-IκBα complex. Through analysing the full suite of results (individual figures, KS-Test p-values, and A-Test scores), we believe there is a significant difference across all agent states under the basal dissociation experiments. As such, we reject both $H1_0$, that an increase in basal dissociation of the inhibited NF-κB-IκBα complex to yield its constituent components will not reduce the likelihood of activated NF-κB accumulating in the nucleus; and $H2_0$, that an increase in basal dissociation of the inhibited NF-κB-IκBα complex to yield its constituent components will not increase the levels of free IκBα and NF-κB in the cytoplasm. Furthermore, we also conclude that the system is fragile to perturbations involving basal dissociation for both control conditions, as shown through epistemic uncertainty analysis (see section 6.3.1) and IL-1 stimulation conditions.

Section 7.3 presented the background, experimental procedure and results of *in silico* experimentation into the effect of varying IKK numbers. Following the discovery of Chen et al. (1996), that the signal-induced activation of NF-κB requires phosphorylation of the inhibitory IκBα molecule (within the inhibited NF-κB-IκBα complex), and its subsequent proteolytic degradation, we now know that the initial phosphorylation to dissociate the IκBα molecule and release *free* NF-κB is performed by IKK. This has provided us with an opportunity to investigate the effects of varying IKK number within the calibrated simulation platform for iteration 1. As there is an amplification step in the signalling pathway (i.e. one IKK agent can facilitate dissociation of many NF-κB-IκBα complex over time), we hypothesised that an increase in IKK numbers would increase the numbers of free NF-κB and free IκBα (due to the increased dissociation of the inhibited complex), and increase the accumulation of active NF-κB in the nucleus (these were converted to their respective null hypotheses).

Results indicate that the system is stable to perturbations in IKK numbers within a small range either side of the default value (i.e. between 50% and 200%), but sensitive to more extreme perturbations (i.e. 10%, 20%, 500%, and 1,000%). Using these higher IKK numbers (i.e. 500% and 1,000% of the default value), results also indicate that an increase in the number of IKK agents, leads to: a decrease in cytoplasmic IκBα; a decrease in both cytoplasmic and nuclear NF-κB-IκBα complex, along with an associated increase in the dissociation of the complexes; and no significant change to either nuclear NF-κB-IκBα complexes or nulcear IκBα, although we suspect that this is due to the small absolute numbers that are associated with the calibrated simulation platform. Additionally, a small change in cytoplasmic IκBα was found when perturbing the system with low IKK numbers (i.e. 10% and 20% of the default value).

KS-Tests and subsequent A-Tests provide additional support to these interpretations, by indicating scientifically significant differences for cytoplasmic NF-κB-IκBα complexes under all perturbations to IKK number, and nuclear (both free and active) NF-κB under all perturbations to IKK number, apart from 50% and 200% of the default value for IKK numbers. Unfortunately, the effects on cytoplasmic NF-κB are not as straight forward to interpret, as non-linear temporal dynamics emerge, which we believe have yielded false negative A-Test scores in a similar manner to those seen under the previous *in silico* experimentation around basal dissociation of the inhibited complex. We believe that these non-linear dynamics of cytoplasmic NF-κB are due to the rapid increase in dissociation of NF-κB-IκBα complexes to yield free NF-κB, and then subsequent translocation to the nucleus whereby these IKK-mediated free NF-κB agents become activated and remain in the nuclear compartment. As such, we believe that an increase in IKK numbers generates an initial increase in cytoplasmic NF-κB, which then slowly reduces over time through translocation to the nucleus. Once in the nucleus, the NF-κB becomes activated and thus remains there until the end of the simulation, due to the abstractions that we have taken as part of our platform model. We therefore reject both $H3_0$, that an increase in IKK concentration will not increase the rate of dissociation of NF-κB-IκBα complexes, and resultant degradation of IκBα; and $H4_0$, that an increase in IKK concentration will not increase the rate of at which activated NF-κB accumulates in the nucleus.

Section 7.4 presented the background, experimental procedure and results of *in*

silico experimentation into the effect of varying IκBα numbers. As discussed in chapter 3 where we focused on the biological domain, IκBα is the key inhibitory molecule, which masks the nuclear localisation sequence of NF-κB molecules, to restrict the NF-κB-IκBα complexes within the cytoplasm, and thus mitigate the activation of NF-κB and its subsequent promotion of the transcription of inflammatory proteins. Yang et al. (2003) and Carlotti et al. (1999) predicted a larger total number of IκBα molecules than NF-κB molecules within cells (approximately 3:1 ratio). Pogson et al. (2008) have recently shown through computational modelling, that the excess IκBα was sequestered to the actin cytoskeleton. This has provided us with an opportunity to investigate the effects of increasing the levels of free IκBα to 3x the default value, and the effects of adding excess IκBα agents that are sequestered within the cell (i.e. to the cytoskeleton).

Results indicate that the system in both control (non-stimulated) and IL-1 stimulated conditions, is fragile to perturbations that increase the number of free IκBα agents to 3x the default value, but stable when the excess IκBα agents are sequestered to the cytoskeleton. Under control conditions, the excess free IκBα results in the system settling at a pseudo steady-state away from our calibrated dynamics, as indicated by statistically significant KS-Test p-values and scientifically significant A-Test scores for: cytoplasmic NF-κB, cytoplasmic NF-κB-IκBα complexes, and nuclear NF-κB-IκBα complexes. Conversely, the results of excess sequestered IκBα agents, show no scientifically significant differences, as indicated by a tendency towards A-Test scores of 0.5. We believe that these results for control dynamics are consistent with the findings of Pogson et al. (2008).

Results of excess free IκBα during IL-1 stimulation also show significant differences, and in fact generate an initial spike in free NF-κB between 5 - 10min, with a corresponding decrease in cytoplasmic NF-κB-IκBα complexes, before a rapid decrease in cytoplasmic numbers due to their translocation to the nucleus. We believe that these phenomena may be due to the extremely low numbers of free NF-κB at the beginning of simulations, as the pseudo steady-state dynamics following excess free IκBα under control conditions, had resulted in the overwhelming majority of NF-κB agents to be complexed with IκBα within the cytoplasm. KS-Test p-values and A-Test scores indicate scientifically significant differences for all agent states under IL-1 stimulated conditions with excess free IκBα, apart from cytoplasmic IκBα. Furthermore, as per the control conditions experiment, the addition of excess sequestered IκBα had no effect on the system. We therefore reject both $H5_0$, that the addition of excess free IκBα within the cytoplasm will not perturb system dynamics into a fragile state; and $H6_0$, that the addition of excess IκBα sequestered within the cell will not result in stable system dynamics.

Section 7.5 presented the background, experimental procedure and results of *in silico* experimentation into the effect of varying the lag-time before IKK activation. Simulation platform development and calibration has yielded a default lag-time for IKK activation within our computational model of 500 iterations, which approximates to 17 seconds. This has provided us with an opportunity to investigate the effects of varying the lag-time between cell membrane receptor activation and the subsequent activation of IKK within the cytoplasm. Results indicate that the system is robust to perturbations involving the lag-time before activation of IKK. Increases of upto 40x to the lag-time (with respect to the

default value) between cell membrane receptor activation and the subsequent activation of IKK within the cytoplasm, showed no appreciable difference for any of the seven agent states (i.e. cytoplasmic and nuclear NF-κB, cytoplasmic and nuclear NF-κB-IκBα, active NF-κB, cytoplasmic and nuclear IκBα).

KS-Test p-values and A-Test scores confirm these findings, as it can be seen that only nuclear NF-κB-IκBα under 40x default lag-time yields a statistically significant difference (p-value of 0.0427), but the associated A-Test score indicates that although statistically significant, the difference does not generate a sufficient effect magnitude for us to conclude a scientifically significant difference, i.e. the effect of the difference would not be appreciable in the *real world*. We are therefore unable to reject $H7_0$, and believe that the system is stable to perturbations extending the lag-time (upto 40x default or 10min) before IKK activation.

This chapter has addressed research objective 4: to perform novel *in silico* experimentation using the agent-based model. Through *in silico* experimentation, it provides contributions to the field of research into the IL-1 stimulated NF-κB signalling pathway, and additionally provides a platform for exploring how modelling techniques may assist wet-lab research in furthering our understanding of complex biological systems. This is the subject of the following chapter, where the simulation platform for iteration 1 is augmented to provide further granularity of components and interactions associated with the cell membrane receptor complex.

8 The Augmented Simulator

As discussed previously, the CoSMoS process was defined to be iterative in nature, and therefore allows (and indeed expects) multiple versions of computational models, which incrementally increase the functionality of the model with respect to the real-world system under investigation. The initial version of our computational model, focused on the NF-κB-IκBα-IKK signalling module, with minimal functionality for the cell membrane receptor, through use of a single generic IL-1R agent. Iteration 1 was defined from a technical perspective in the platform model of chapter 5; developed and calibrated into the simulation platform documented in chapter 6, where uncertainty analysis was also performed to ascertain the various uncertainties emerging from our understanding of the real-world system (epistemic uncertainty) and the stochastic processes within the computational model itself (aleatory uncertainty); and experimented with in chapter 7. The second version of the computational model will incorporate increased granularity of components at the cell membrane, and before activation of IKK, to extend the scope of *in silico* experimentation to components upstream of the NF-κB signalling module.

This chapter focuses on the second version of the computational model. Section 8.1 commences with the motivation for augmenting the model; section 8.2 defines the augmented platform model; section 8.3 presents the new simulation platform and associated dynamics (research objective 5); section 8.4 investigates the effects of perturbing the numbers of various cell membrane receptor components through *in silico* experimentation (research objective 6); whilst section 8.5 concludes this chapter.

8.1 Motivation behind Augmenting the Computational Model

As discussed in section 3.3, in addition to the various NF-κB dimers, IκB inhibitors, and IKK complexes, there are a large number of upstream components which facilitate signal transduction through the NF-κB signalling pathway. Perhaps the most important of these upstream components are those that make up the cell membrane receptor complex. With specific reference to the IL-1 stimulated NF-κB signalling pathway, these upstream components include the actual IL-1 receptors themselves, along with various co-receptors, adaptor proteins and kinases. Upon recognition of appropriate extracellular signal, the receptors undergo conformational changes at their intracellular surfaces to facilitate binding of proteins within the cytoplasm, and resultant propagation of signal inside the cell. We therefore have an opportunity to extend the computational model developed for iteration 1, with additional functionality for receptor complex formation at the cell membrane. As such, the simulation platform will be updated to incor-

porate dimerisation of the IL-1 receptors, binding of the adaptor protein MyD88, and binding of the IRAK and TRAF6 kinases, before the activation of IKK for propagation of the signal. Our approach to augmenting iteration 1 with additional functionality adheres to the recent work of Greaves et al. (2012, 2013), that discusses a process for augmenting existing computational models with additional pathway components.

8.2 The Augmented Platform Model

Due to the CoSMoS framework advocating an incremental approach to developing model functionality, we have attempted to capture within our domain model *all* of the components of the signalling pathway that will be relevant to us during the lifetime of the project. As such, the existing domain model presented in chapter 4, continues to be appropriate for our augmented computational model developed as part of iteration 2. This iterative approach to development, ensures that the underlying basis of iteration 2 is the technical infrastructure previously developed as part of the first iteration. We therefore believe it appropriate to utilise a number of UML diagrams previously developed for iteration 1 (defined in chapter 5), as the basis for the augmented platform model, and have highlighted the additional components, interactions, and activities for this second iteration in the colour red, as this ensures a quick and easy differentiation of the platforms for the two iterations.

Along with the five main agents from iteration 1 that consisted of the IL-1R cell membrane receptor, IKK, IκBα, NF-κB, and nuclear transporter; we have now added three additional agents relating to MyD88, IRAK, and TRAF6. Within our abstracted technical view of the system, these additional components are subtypes of the generic *Receptor* component, as highlighted in the updated inheritance class diagram of figure 8.1, and interact with the IL-1 receptor components to form the IL-1R receptor complex, as highlighted in the updated containment class diagram of figure 8.2.

The class association diagram (figure 8.3) specifies the additional high-level interactions for cell membrane receptor complex formation. It can be seen from this diagram and the updated sequence and communication diagrams (figures 8.4 and 8.5) that the additional functionality incorporated through the MyD88, IRAK and TRAF6 components is modular in nature, being a discrete additional step over-and-above that for iteration 1. As such, the incorporation of the additional components has no affect on the existing functionality once the IL-1R complex has become activated. To further support the idea that these additional components represent extra granularity within the receptor activation *module*, the activity diagram of figure 8.6 provides an updated partial view of the system, which focuses on the activities upto and including IKK activation, as the downstream activities remain consistent with figure 5.8 for iteration 1. Finally, the low-level detail of the system components for iteration 2 are specified in the state machine diagram (figure 8.7), which specifies the various agent states, and the corresponding X-Machine stategraph diagram (figure 8.8), which along with specifying the agent states, also defines the flow of methods within individual agents and the communication between agent types.

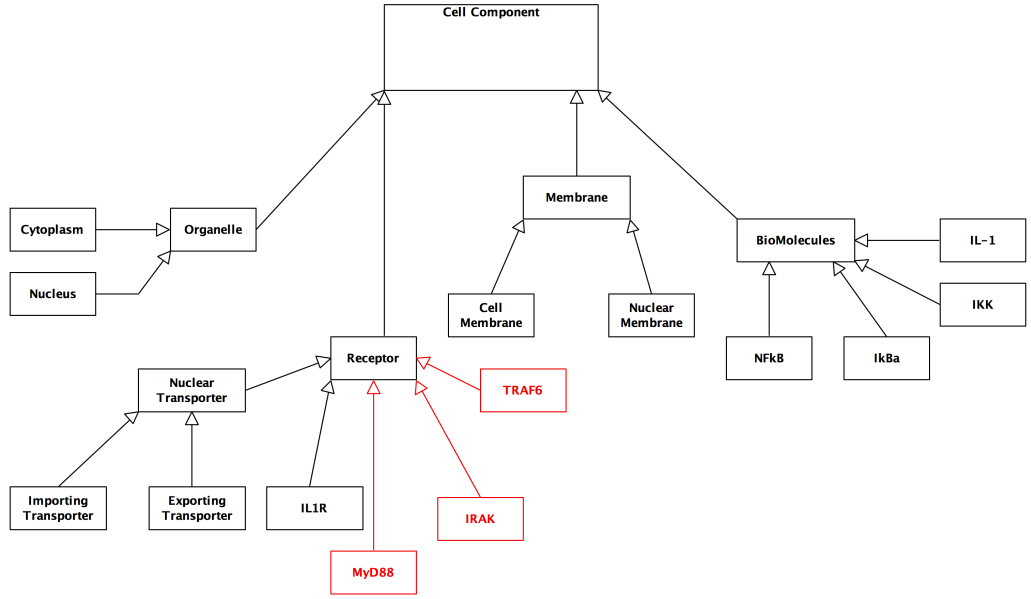

Figure 8.1: UML inheritance class diagram for the augmented platform model. The inheritance class diagram for iteration 1 has been augmented with the addition of MyD88, IRAK and TRAF6, which are subtypes of the generic receptor component.

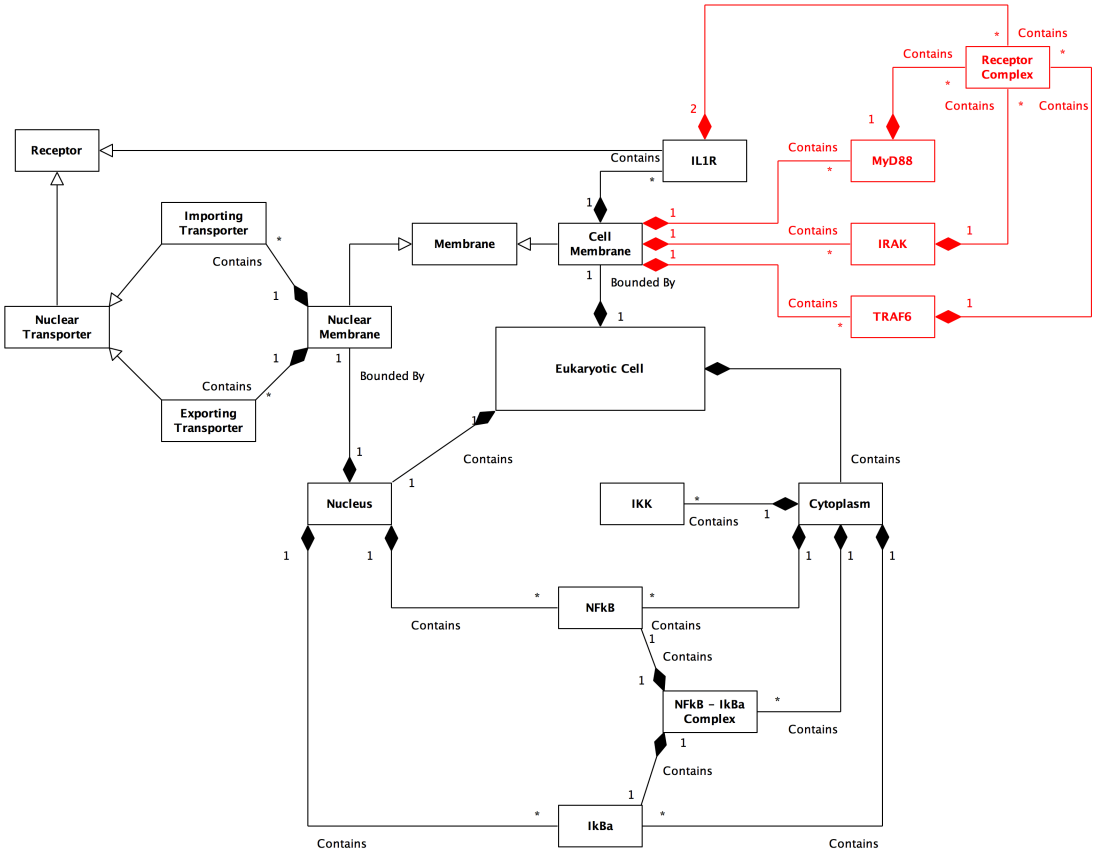

Figure 8.2: UML containment class diagram for the augmented platform model. Containment continues to follow that for iteration 1, apart from the addition of MyD88, IRAK and TRAF6. These additional components interact with IL-1R to form the IL-1R receptor complex at the cell membrane.

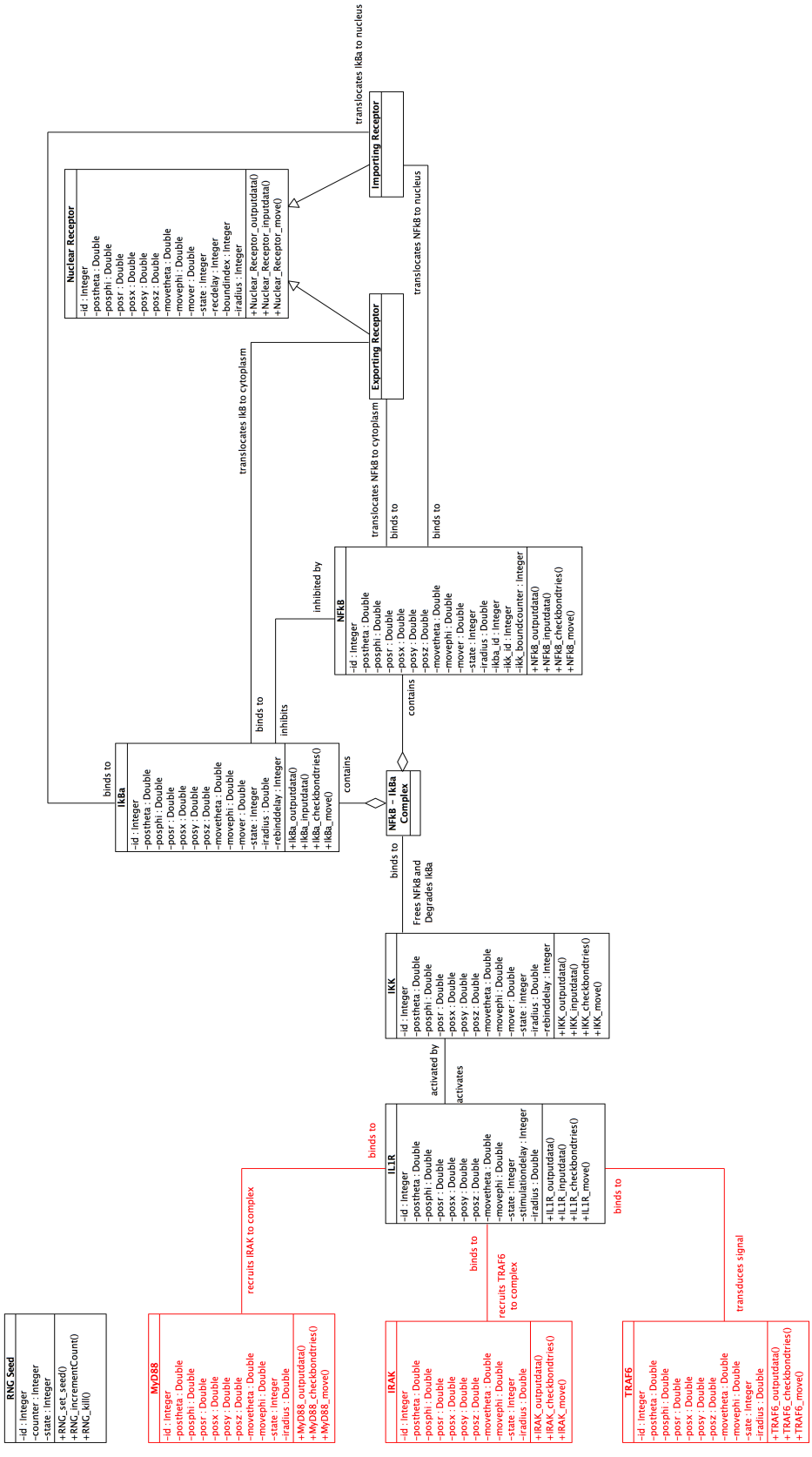

Figure 8.3: UML class association diagram for the augmented platform model. The associations between different agent types continues as per iteration 1, however the class diagram has been updated to incorporate the additional associations between the new MyD88, IRAK, and TRAF6 components, with the existing IL-1R component.

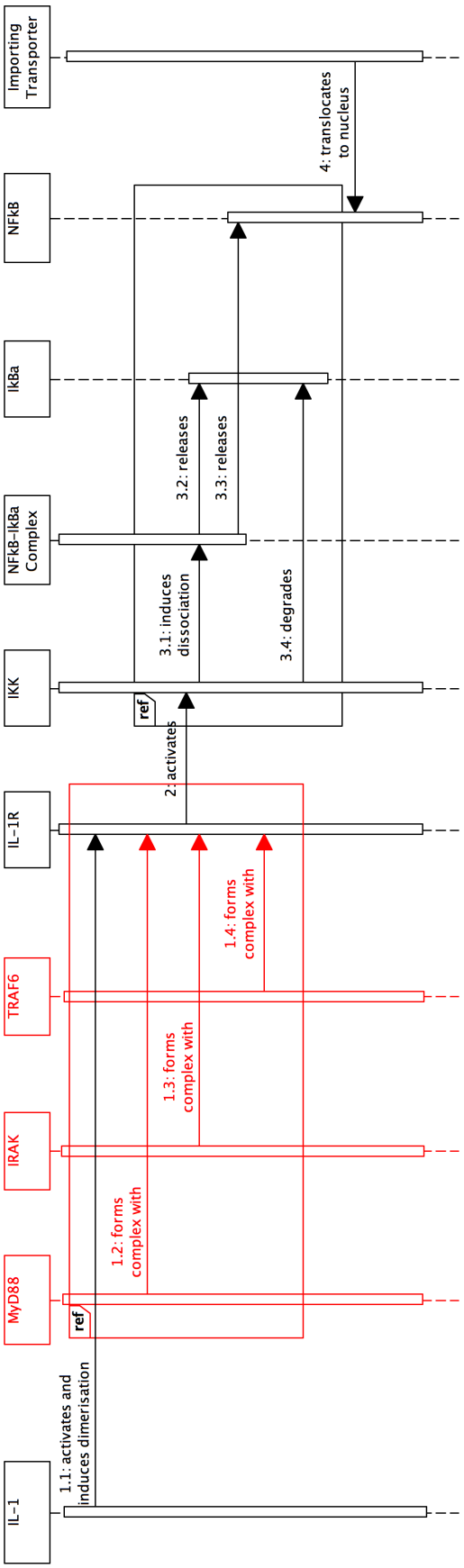

Figure 8.4: UML sequence diagram for the augmented platform model. The sequence diagram from iteration 1 has been updated to incorporate the additional events between dimerisation and activation of IL-1R and the subsequent activation of IKK.

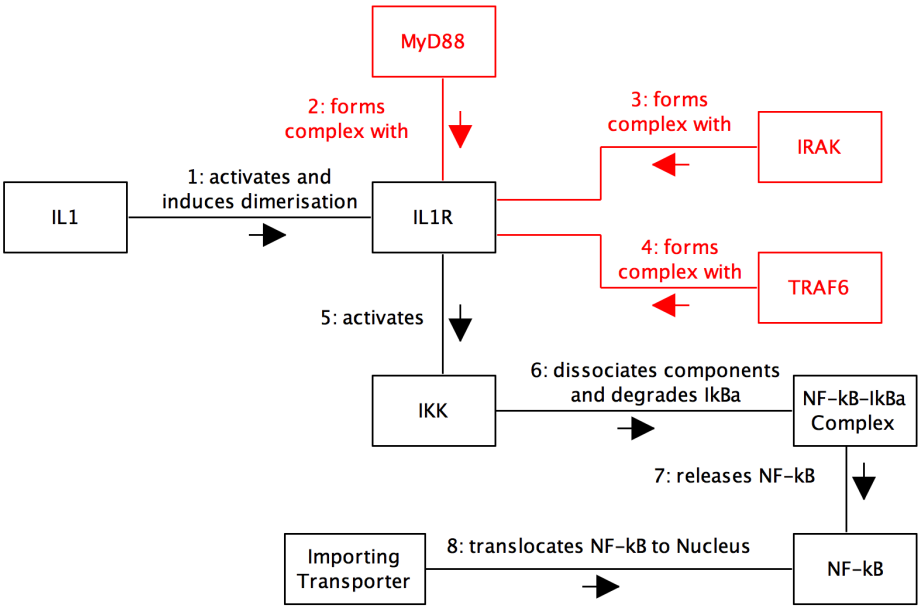

Figure 8.5: UML communication diagram for the augmented platform model. As per previous UML diagrams, the version from iteration 1 has been updated with the additional MyD88, IRAK and TRAF6 components, and their associated interactions.

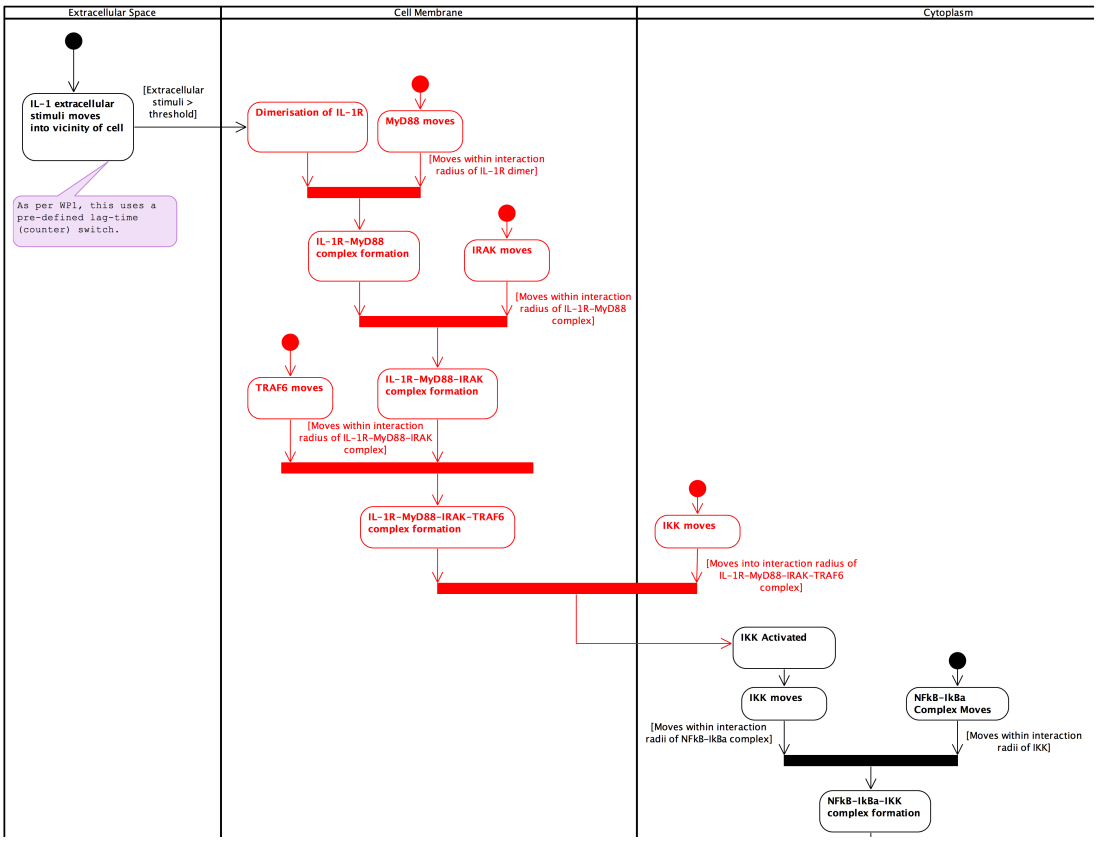

Figure 8.6: UML activity diagram for the augmented platform model. This UML diagram focuses explicitly on the activities associated with the additional MyD88, IRAK and TRAF6 components, as the other downstream activities remain the same as iteration 1.

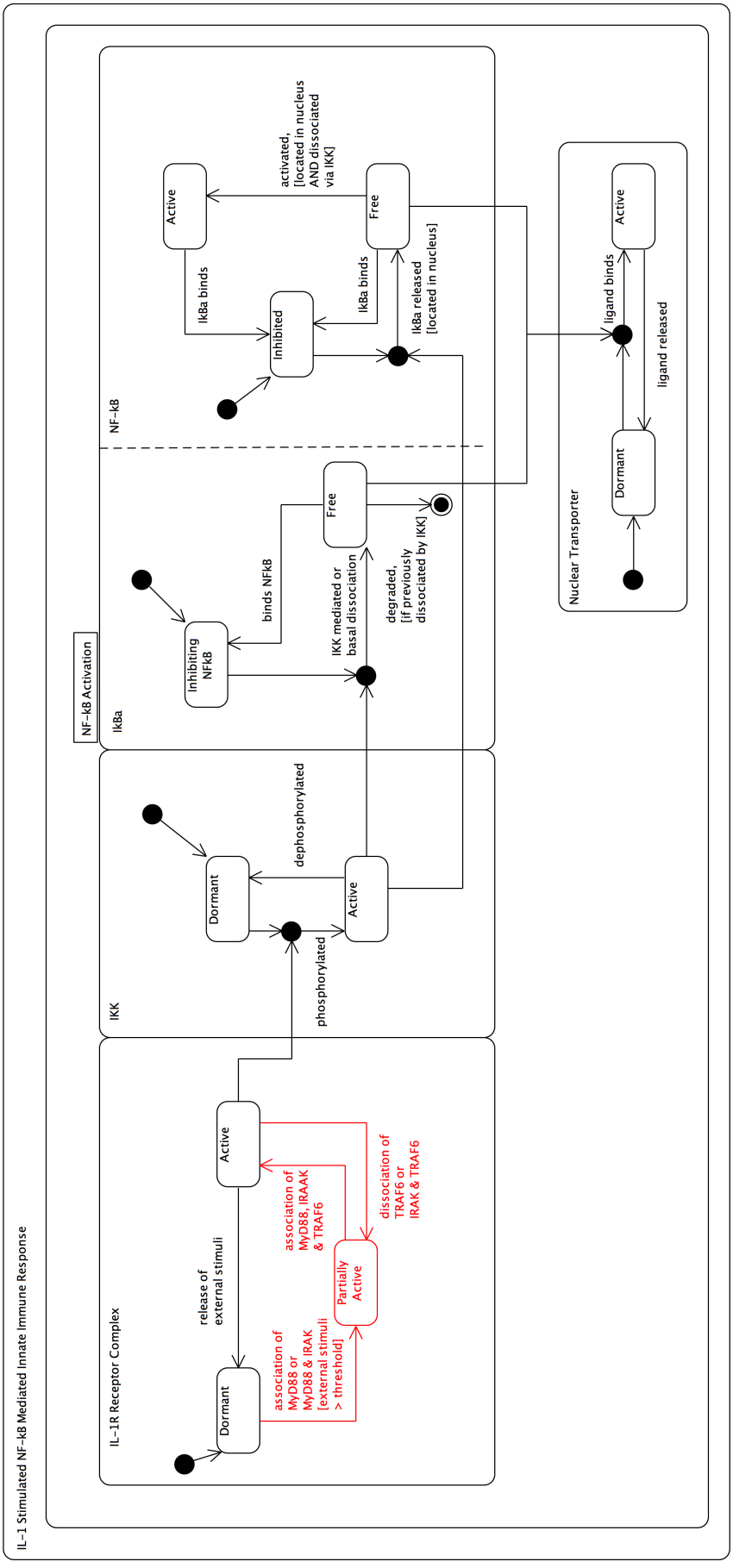

Figure 8.7: UML state machine diagram for the augmented platform model, depicting the full set of detailed state changes for each agent type within iteration 2.

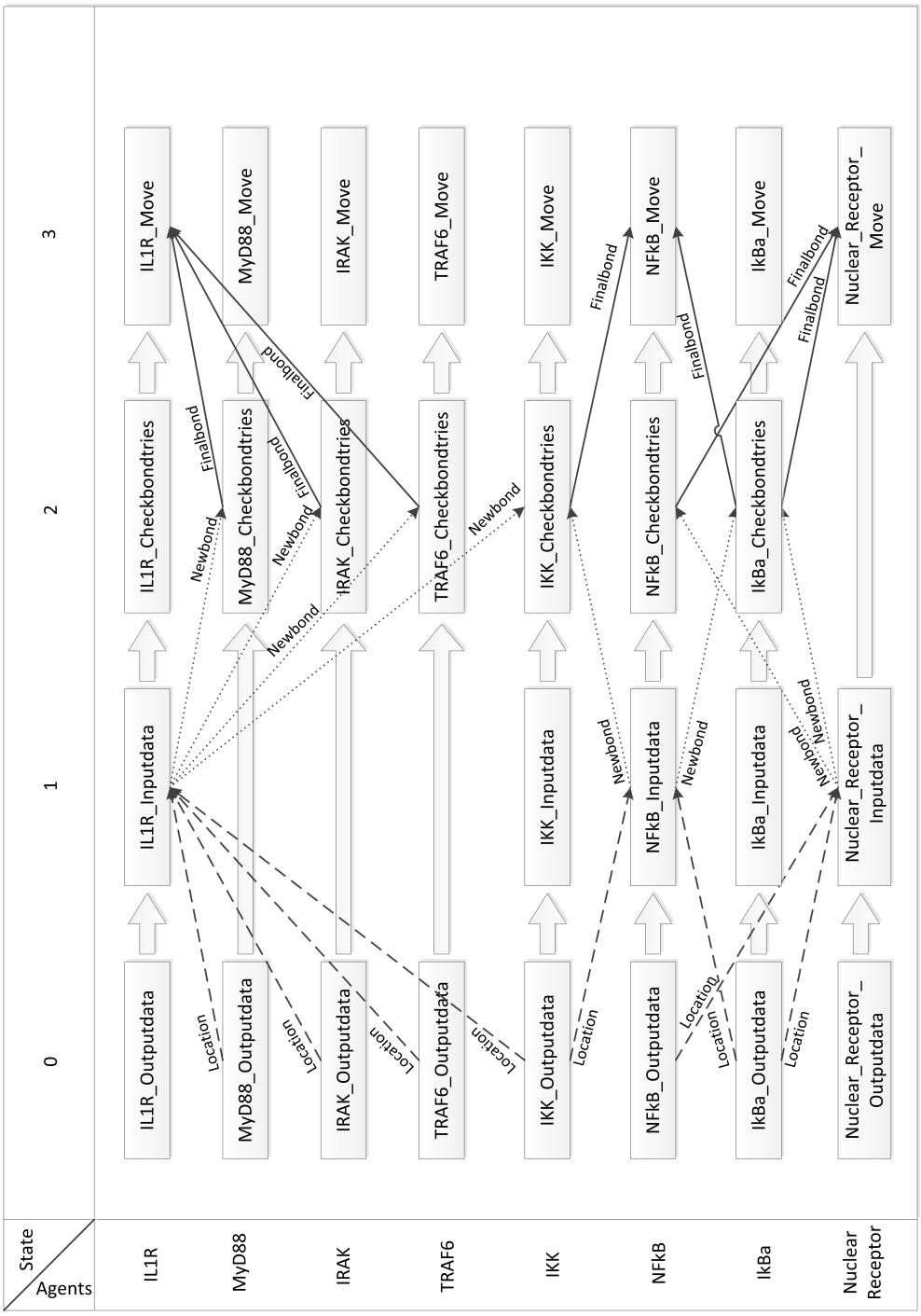

Figure 8.8: X-Machine stategraph diagram for the augmented platform model, depicting a matrix of the eight agent types (X-Machines) versus their internal states and associated message types for communication.

8.3 The Augmented Simulation Platform

As discussed in chapter 6, the initial agent-based computational model was developed using the FLAME simulation framework, against the technical specification defined in the platform model from iteration 1 (see chapter 5). This platform model was intentionally designed to be modular in nature, which has allowed the augmentation for additional components (MyD88, IRAK and TRAF6) to be a relatively straightforward process, as defined in the previous section. Similarly, as the simulation platform is the physical embodiment of the platform model in computer code, this too followed a modular architecture, which has again facilitated a relatively straightforward process for augmenting the code with the additional functionality for the formation of the cell membrane receptor complexes. Furthermore, as the augmentation of the simulation platform required the addition of extra granularity with an existing functional module (i.e. of the process of cell receptor complex activation, and the subsequent activation of IKK within the cytoplasm), we did not need to recalibrate the simulation platform for iteration 2, as the downstream mechanisms following IKK activation remained the same as the first iteration. Unfortunately, two major technical issues were encountered during the augmentation of the simulation platform for iteration 2 functionality, which related to the requirement for agents to only interact with their nearest neighbour, and poor performance of the hardware architecture when running simulations, due to the resource intensive nature of the FLAME simulation framework.

The first issue relates to the fact that Communicating X-Machines interact through the generation of messages by agents, which are posted to a centralised message board. Agent behaviour (state changes) are then generated through filtering these messages and acting accordingly. Unfortunately, the agents within a biological simulation probabilistically bind to their nearest neighbours, and it was the calculation of nearest neighbours that caused issues in the system. During cell receptor complex formation, a defined sequence of bindings needs to occur: 1) IL-1R dimerisation, 2) IL1R-MyD88, 3) IL1R-MyD88-IRAK, and 4) IL1R-MyD88-IRAK-TRAF6. It became evident that agents within the cytoplasm (IKK, NF-κB and IκBα) were consistently within the interaction radius of cell receptor components, and were affecting the dynamics of cell receptor complex formation. We mitigated this issue through introducing additional message types for interactions between cell receptor complex components and for interactions between the activated cell receptor complex and dormant IKK. This allowed an intrinsic filtering system, as only cell receptor components could read their message type, so mitigating any issues arising from agents in the cytoplasm being closer in 3D space. We therefore required a message type for cell receptor components (Membrane Messages), a message type for IKK activation by the activated cell receptor complex (IKK Activation Messages), and a third message type for the cytoplasm and nuclear components. Figure 8.9 depicts a 3D visualisation of the additional components IL-1R, MyD88, IRAK and TRAF6 orbitting within the vicinity of the cell membrane.

The second issue relates to the high-volume of data produced by the FLAME simulation framework, and the resultant resource intensive nature of running replicate sets of simulations. The 0.xml starting parameters file for iteration 2

has grown in size to approximately 1.4MB, due to defining separate instantiations of the 500 MyD88, 500 IRAK and 500 TRAF6 agents within simulation runs. An *in silico* experiment from iteration 1 ran simulations for 324,000 iterations (∼3hr biological time), which generated approximately 454GB of cumulative XML output files. The issue was then further compounded by our requirement to run 175 replicates to minimize aleatory uncertainty (see section 6.3.2), which is predicted to generate approximately 80TB of data. We brought down the Northern 8 Universities Consortium (N8) High Performance Computing Cluster (HPC) on a number of occasions during our initial development of the simulation platform, which we were later informed was due to running out of storage space as the N8 HPC had 100TB of fast Lustre storage. Due to the fact that other users required access to fast storage space, we were forced to run a maximum of 50 replicates at a time to mitigate the risk of the N8 HPC running out of fast storage space. Unfortunately, our performance issue was further compounded by the fact that the 324,000 time-step simulations took over 48hr to run, which exposed an additional issue with the setup and configuration of the N8 HPC. The N8 HPC technical design committee had previously decided that all jobs would be stopped upon reaching 48hr of wall clock (physical) time, to ensure that no single user was disproortionately using this shared computing resource. We were unfortunately unable to overturn this decision, and therefore had to reduce the number of time-steps for individual simulation runs, from 324,000 (∼3hr biological time) to 270,000 (∼2.5hr biological time), to ensure they could run to completion.

Figure 8.10 represents the baseline IL-1 stimulated dynamics of the augmented simulation platform, whilst table 8.1 provides the results of the associated statistical tests (KS-Test and A-Test) that were performed against the IL-1 stimulated baseline of iteration 1. It can be seen by the KS-Test p-values that there are no statistically significant differences, and by the A-Test scores that there are no scientifically significant differences, between the two sets of simulation data distributions. This confirms that the augmentation of the simulation platform for additional functionality at the cell membrane receptor, has not perturbed simulation dynamics with respect to the calibrated simulation platform of the first iteration.

Agent State	**KS-Test p-value**	**A-Test Score**
NF-κB Cyto	0.1891	0.4943
NF-κB-IκBα Cyto	0.9997	0.4741
NF-κB-IκBα Nuc	0.4375	0.5305
NF-κB Nuc	0.9999	0.5228
Active NF-κB Nuc	0.9999	0.5228
IκBα Cyto	0.998	0.4899
IκBα Nuc	1.0	0.5

Table 8.1: KS-Test p-values and A-Test scores for the baseline IL-1 stimulated dynamics of iteration 2 against the baseline IL-1 stimulated dynamics of iteration 2. The scores indicate no statistically or scientifically significant difference between the two distributions of simulation data.

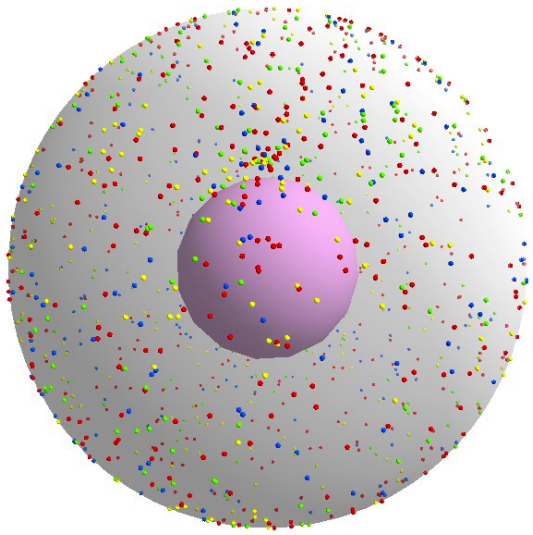

Figure 8.9: 3D visualisation of the spherical orbitting of IL-1R and the additional MyD88, IRAK and TRAF6 components for iteration 2. The cytoplasmic and nuclear locations of the cell envrionment are represented through the use of two concentric circles. The IKK, NF-κB and IκBα agents have been omitted to reduce confusion and ensure focus is applied to the vicinity of the cell membrane.

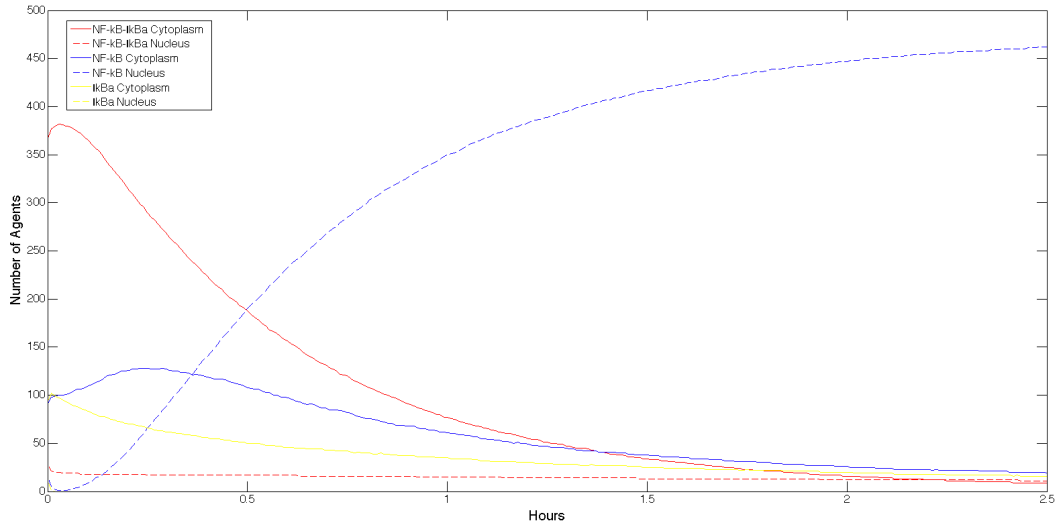

Figure 8.10: Median average IL-1 stimulation baseline dynamics for the augmented simulation platform, following 175 replicates. As discussed above, simulations could only run for a maximum of 270,000 iterations, which equates to approximately 2.5 hour of real-world time.

8.4 Effect of Cell Receptor Complex Component Numbers

As previously discussed in chapter 3 (The Domain), the extracellular signal is propagated through the signalling pathway via the cell membrane receptor complex, intermediate components, the NF-κB signalling module, and the transcription and translation machinery for generation of inflammatory response proteins. Whereas experiments 1-4 investigated the effects of perturbing the system with respect to the NF-κB signalling module, experiment 5 focuses on investigating IKK activation dynamics through perturbations of cell membrane receptor complex formation. Investigations will therefore focus on varying the number of adaptor protein (MyD88) and kinase (IRAK and TRAF6) agents within the system.

Simulation dynamics have been calibrated using 500 agents for each of the IL-1R, MyD88, IRAK and TRAF6 components. This provides an opportunity for us to investigate the effects of varying the numbers of these cell receptor complex components (akin to varying the concentration in wet-lab experiments), and whether the system is robust to such perturbations. As there is a direct sequence of binding events within our computational model for formation of the active cell membrane receptor, it is expected that an increase in numbers for each of the individual components will increase the speed at which IKK is activated, and thus affect the IκBα degradation and NF-κB activation dynamics. It is also expected that an increase in adaptor protein and kinase numbers (MyD88, IRAK and TRAF6), whilst maintaining the numbers of IL-1R agents, will also increase the rate of activated cell membrane receptor complexes, and thus the IκBα degradation and NF-κB activation dynamics. Converting these to null hypotheses, we have:

H8$_0$ An increase in the numbers of each cell membrane receptor complex component (e.g. IL-1R, MyD88, IRAK and TRAF6) will not affect the overall dynamics regarding IκBα degradation and NF-κB activation.

H9$_0$ An increase in the numbers of adaptor protein (MyD88) and kinase (IRAK and TRAF6) components will not affect the overall dynamics regarding IκBα degradation and NF-κB activation.

8.4.1 Experimental Procedure

Experimentation into the effects of varying cell membrane receptor complex component numbers, is conducted through perturbation of the total numbers of individual components. By default, there were 500 IL-1R agents, 500 MyD88 agents, 500 IRAK agents, and 500 TRAF6 agents, within the calibrated simulation platform. To test the two null hypotheses above, we ran four sets of *in silico* experiments using half of each component (i.e. 250 of each type of agent), double each component (i.e. 1,000 of each type of agent), half adaptor protein and kinase components (i.e. 500 IL-1R, but 250 MyD88, IRAK and TRAF6 agents), and double adaptor protein and kinase components (i.e. 500 IL-1R, but 1,000 MyD88, IRAK and TRAF6 agents). For each set of 175 simulation replicates, the median distributions of cytoplasmic NF-κB, nuclear NF-κB, cytoplasmic NF-κB-IκBα,

nuclear NF-κB-IκBα, cytoplasmic IκBα and nuclear IκBα over the lifetime of the simulation runs are interpolated. These distributions are constrasted with the baseline behaviour that results from simulations using the default number of the cell membrane receptor complex agents. KS-Tests are then performed to understand whether there are any statistically significant differences from baseline behaviour (requiring p-values below 0.05), and A-Tests are also performed to understand the effect magnitude of these differences, assuming 'large' differences of < 0.29 and > 0.71 to be scientifically significant.

8.4.2 Results

Results indicate that the system is relatively robust to perturbations involving the total numbers of individual cell membrane receptor complex components. Figure 8.11 depicts the effects on cytoplasmic IκBα numbers when the component numbers are: all halved, all doubled, constant IL-1R with halved adaptor protein and kinases, and constant IL-1R with doubled adaptor protein and kinases. It can be seen that none of these perturbations has an effect on the overall temporal profile of IκBα agents within the cytoplasm. Similarly, figure 8.12 depicts the effects on cytoplasmic NF-κB-IκBα numbers, and figure 8.13 depicts the effects on cytoplasmic NF-κB numbers when perturbing the system. Here it can be seen that the experimental setups for baseline and doubling of non-IL-1R agents, and for halving all agents and halving the non-IL-1R agents, show very similar temporal profiles, with the only appreciable difference resulting from the experimental setup that utilised a doubling of all agents.

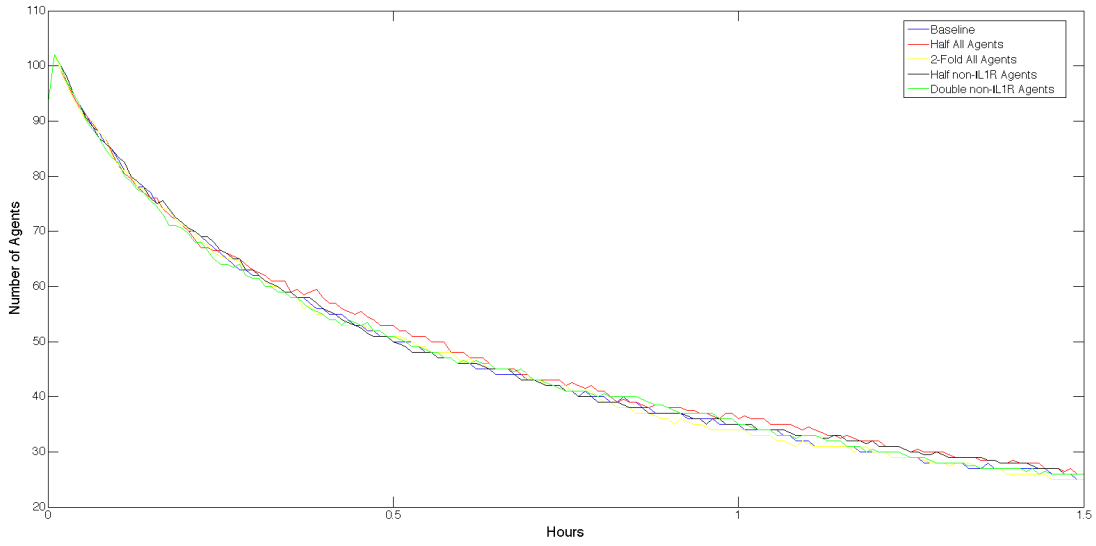

Figure 8.11: Cytoplasmic IκBα dynamics for the four sets of experiments into receptor complex component numbers, which have been compared against the baseline IL-1 stimulated dynamics of iteration 2.

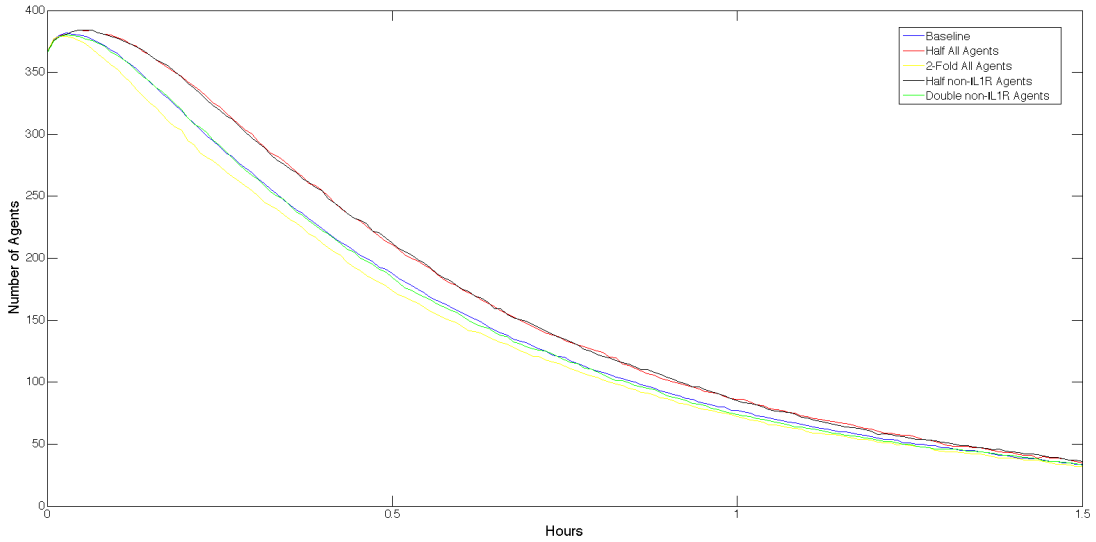

Figure 8.12: Cytoplasmic NF-κB-IκBα dynamics for the four sets of experiments into receptor complex component numbers, which have been compared against the baseline IL-1 stimulated dynamics of iteration 2.

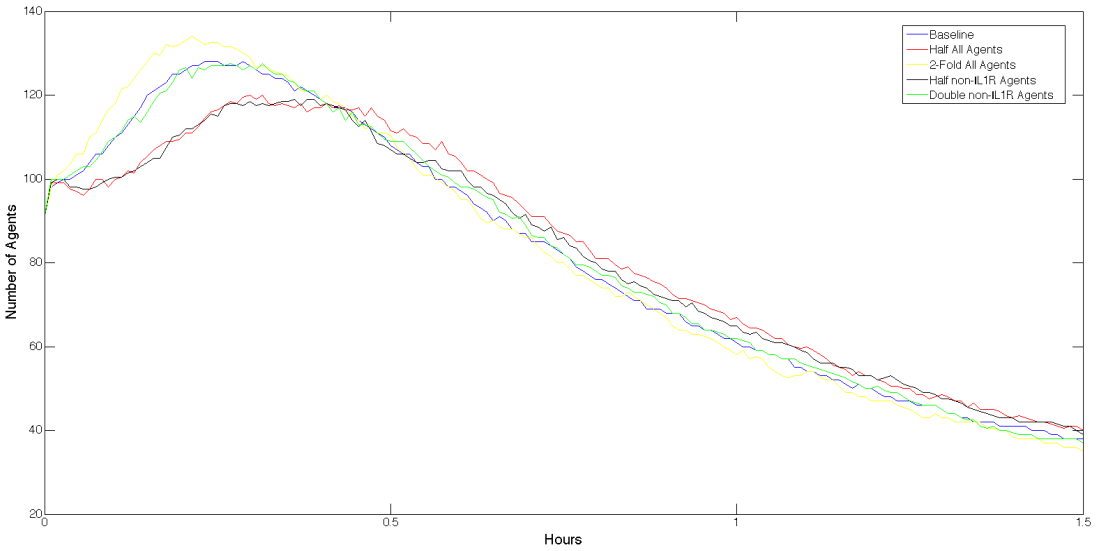

Figure 8.13: Cytoplasmic NF-κB dynamics for the four sets of experiments into receptor complex component numbers, which have been compared against the baseline IL-1 stimulated dynamics of iteration 2.

Tables 8.2 and 8.3 depict the associated KS-Test p-values and A-Test scores resulting from the perturbations to cell membrane receptor complex component numbers. It can be seen that only the cytoplasmic NF-κB when perturbed with halving of all agents and halving of non-IL-1R agents; and additionally the nuclear NF-κB-IκBα when perturbed with halving of all agents, doubling of all agents, and doubling of non-IL-1R agents; show statistically significant differences from the baseline dynamics. This does not translate to a scientifically significant difference however, as none of the A-Test scores yield a high difference from baseline dynamics.

Agent State	Half all agents	Double all agents	Half non-IL-1R agents	Double non-IL-1R agents
NF-κB Cyto	0.0279	0.3939	0.0386	0.9601
NF-κB-IκBα Cyto	1.0	1.0	1.0	1.0
NF-κB-IκBα Nuc	0.0000	0.0001	0.3939	0.0031
NF-κB Nuc	0.9875	1.0	0.9999	1.0
Active NF-κB Nuc	0.9978	0.9999	0.9999	1.0
IκBα Cyto	0.6599	0.9978	0.8394	1.0
IκBα Nuc	1.0	1.0	1.0	1.0

Table 8.2: KS-Test scores for the four sets of *in silico* experiments, which have been compared against baseline IL-1 stimulated dynamics. The scores indicate statistically significant differences for cytoplasmic NF-κB when perturbed with halving of all agents and halving of non-IL-1R agents; and additionally nuclear NF-κB-IκBα when perturbed with halving of all agents, doubling of all agents, and doubling of non-IL-1R agents.

Agent State	Half all agents	Double all agents	Half non-IL-1R agents	Double non-IL-1R agents
NF-κB Cyto	0.4938	0.4979	0.5042	0.4978
NF-κB-IκBα Cyto	0.4623	0.522	0.4619	0.5058
NF-κB-IκBα Nuc	0.3599	0.4099	0.5103	0.3995
NF-κB Nuc	0.5384	0.4849	0.5343	0.5001
Active NF-κB Nuc	0.5381	0.4843	0.5339	0.4999
IκBα Cyto	0.4715	0.5065	0.4873	0.4946
IκBα Nuc	0.5	0.5	0.5	0.497

Table 8.3: A-Test scores for the four sets of *in silico* experiments, which have been compared against baseline IL-1 stimulated dynamics. The scores indicate that there is no scientifically significant difference between any of the experiments or the baseline IL-1 stimulated dynamics.

8.5 Discussion

This chapter has reported the augmentation of the previously developed computational model. The motivation behind augmenting the computational model with additional granularity of components at the cell membrane was presented in section 8.1, with the actual technical design (augmented platform model) and augmented simulation platform being presented in section 8.2 and 8.3. Furthermore, section 8.4 presented the background, experimental procedure, and results, of *in silico* experimentation into the effects of perturbing component numbers within the cell membrane receptor complex.

As discussed in section 8.2, our domain model (see chapter 4) was developed to capture the overall functional aspects of this project in its entirety, and therefore did not need to be augmented, as it defined the underlying biological behaviour for both iterations, and therefore continues to be appropriate. We were however required to develop a new platform model, and due to the modular nature of our design, we were able to utilise the UML diagrams that were previously developed for iteration 1 (see chapter 5) as the basis for our augmented platform model. This was achieved through the additional technical aspects for the computational model being added to the existing UML notations, and using a red colour-coding to provide an easy differentiation between the two platforms. The three additional agents relating to MyD88, IRAK and TRAF6, are all subtypes of the generic *Receptor* component (as defined in the updated inheritance class diagram of figure 8.1), and confined to the vicinity of the cell membrane as defined in the updated containment diagram (see figure 8.2). The resultant high-level interactions between these three additional agents and the existing IL-1R agent have also been specified in the updated class association, sequence, and communication diagrams (see figures 8.3-8.5 respectively). It is clear from these three UML notations that the additional functionality for iteration 2 is modular in nature, which has also been reinforced through the updated activity diagram (see figure 8.6), which shows the additional activities upto and including IKK activation. Finally, the low-level dynamics for the additional agents were defined in the end-to-end state machine diagram (see figure 8.7) and updated X-Machine stategraph diagram (see figure 8.8).

Due to this modular nature, the simulation platform was relatively straightforward to augment for the additional functionality of cell membrane receptor complex formation. Furthermore, as the augmentation did not alter any of the logic or functionality following activation of IKK, the simulation platform did not require recalibration. Unfortunately, two additional issues with the FLAME simulation framework were identified during this iteration of the CoSMoS process. These were similar to those identified during simulation platform development in iteration 1, with the first relating to probabilistic binding of agents through the use of the central message board to communicate the location of nearest neighbours, and the second relating to the resource intensive nature of running simulations. The first issue, once identified was relatively straight forward to resolve, and required the use of a new message type, that was specific to the IL-1R, MyD88, IRAK and TRAF6 agents. The second issue could not be fully resolved due to the design and policy decisions taken by the N8 HPC technical committee, and as such, our only option was to reduce the size of our simulation

runs from 324,000 iterations to 270,000 iterations, and to reduce the number of parallel jobs that could be submitted to the HPC at any given time. Although this facilitated our simulations to run to completion (and therefore not fail), it incurred considerable time delays on running experiments.

Section 8.4 presented the background, experimental procedure, and results of *in silico* experimentation into the effect of varying the numbers of cell membrane receptor complex components. Simulation platform development and calibration utilised a default number of 500 agents for each receptor complex component (IL-1R, MyD88, IRAK and TRAF6). This has provided us with an opportunity to investigate the effects of varying the numbers of all components, and indeed of varying the numbers of adaptor protein and kinase components, whilst maintaining a stable number of IL-1R agents. We hypothesised that an increase in the number of each individual type of receptor complex component will increase the speed at which IKK is activated, and thus increase the rate of IκBα degradation and NF-κB activation; and that an increase in adaptor protein and kinase numbers, whilst maintaining IL-1R numbers, would also increase these rate dynamics, As per iteration 1 *in silico* experimentation, these were also converted to their respective null hypotheses.

Results indicate that the system is robust to perturbations involving the total numbers of receptor complex components. Although correlations were found in the temporal profiles (with respect to cytoplasmic NF-κB-IκBα and cytoplasmic NF-κB) for baseline and doubling of non-IL-1R agents conditions, and for the halving of all agents and the halving of non-IL-1R agents, no scientifically significant difference was found, following the use of the Kolmogorov-Smirnov and Vargha-Delaney A-Tests. We are therefore unable to reject either $H8_0$, that an increase to receptor complex component numbers will not affect system dynamics, or $H9_0$, that an increase in adaptor protein and kinase numbers will not affect system dynamics.

This chapter has addressed research objective 5: Augment the agent-based model with additional upstream signalling components related to the cell membrane receptor complex; and research objective 6: Perform novel *in silico* experimentation using the augmented agent-based model. This augmented computational model provides additional contributions, which are over and above those from chapter 7, to the field of research into the IL-1 stimulated NF-κB signalling pathway. It provides a useful platform for the research community to explore the mechanistic rules that underpin cell receptor complex formation, in particular through the rule-based interactions and associated ratios of the respective receptor complex components (IL-1R, MyD88, IRAK, and TRAF6). Furthermore, this chapter also makes an additional contribution to research objective 7: Investigate the suitability of using the FLAME simulation framework for developing computational models of complex biological systems, which are over and above those previously made within chapter 6.

Part III

Discussion, Conclusions and Further Work

9 Discussion, Conclusions and Further Work

This final chapter concludes the work of this doctoral thesis, which had the overall aim:

> *To develop an agent-based model of the IL-1 stimulated NF-κB signalling pathway in a quality assured manner, which uses leading practices for software engineering, calibration, verification and validation. Furthermore, once developed, the agent-based model will be used to perform novel in silico experimentation to extend our knowledge of the signalling pathway.*

The chapter commences in section 9.1, with discussion of the contributions made to computational biology, that include: reflections on the completeness of UML for modelling complex biological systems; reflections on the suitability of FLAME for modelling complex biological systems; reflections on the CoSMoS process as a project lifecycle for modelling complex biological systems; and reflections on the necessity for statistical rigour. Section 9.2 discusses the contributions to NF-κB modelling, whilst section 9.3 provides a chapter-by-chapter summary of the thesis content and contributions. Section 9.4 presents three broad areas for further investigation, including: additional modelling and simulation of NF-κB signalling; further investigation into the use of UML and statistics to define complex biological systems; and investigations into the effects that different modelling paradigms (e.g. equation-based versus agent-based), or modelling frameworks (e.g. FLAME versus FLAME GPU versus Java MASON) may have on simulation results, and whether the results from *in silico* experiments that utilise different approaches are comparable. Finally, section 9.5 presents the concluding remarks of this doctoral thesis.

9.1 Contributions to Computational Biology

Biological systems are complex, with behaviours and characteristics that result from a highly connected set of interaction networks that function through time and space. One of the main strengths of the systems biology approach, is that it focuses on three key properties of complex systems: 1) system structures, 2) system dynamics, and 3) system control (Ideker et al., 2001). Simulation attempts to predict the dynamics of systems, so that the validity of the underlying assumptions behind the models can be tested. As such, computational modelling and simulation can be useful tools for exploring the behaviours and dynamics of biological systems. Through *in silico* experimentation, these models provide a

relatively easy mechanism for testing complex hypotheses of how complex low-level system dynamics, result in the myriad of system-wide behaviours.

For these models to be successful in fulfilling their role as computational abstractions of real-world biology, it is essential that the chosen modelling system is able: to effectively capture system structure and dynamics; is scalable with respect to the hierarchical-level of biology, e.g. molecules, to cells, to tissues/organs; is modular, so that additional real-world functionality can be applied in an incremental basis, without having to re-engineer the existing models; and lends itself to rigourous interrogation through statistical analysis of simulation results. We will reflect on these various aspects for the remainder of this section, with initial focus applied to the use of UML for defining system structure and dynamics, before turning to the suitability of FLAME for modelling complex biological systems. We then complete this section through an evaluation of the CoSMoS process as a project lifecycle for design and development of computational models, and reflections on the necessity for statistical rigour when calibrating computational models and interpreting simulation results.

9.1.1 Reflections on the Completeness of UML for Domain Modelling of Complex Biological Systems

As discussed in chapter 4, we have developed a domain model of the IL-1 stimulated NF-κB signalling pathway, for use as functional specification for our agent-based model. As per Read et al. (2009b) and Bersini (2012), we agree that a subset of UML notations are able to efficiently represent elements of the domain model of biological systems. We have found activity diagrams and communication diagrams particularly effective at depicting system-wide behaviours; communication diagrams to be effective at depicting relationships between components; and state machine diagrams effective at depicting low-level dynamics within individual components. Furthermore, we have found that activity diagrams are particularly effective when used in conjunction with swim-lanes to convey the location (e.g. cytoplasm or nucleus) of activities, and sequentially-linked state machine diagrams are particularly effective at depicting the end-to-end state changes within a system.

Although we have found UML to be particularly useful in these cases, it does have a number of deficiencies however. Along with the issues found by Read et al. (2009b, 2014), we have discovered a number of additional areas where the current UML standards have deficiencies in modelling biology. For example, although UML facilitates detailed information to be depicted as attributes of individual components, it relies on the reader to unpick the multitude of diagrams to collate all of the information, for example parameter values (such as size of cell, and speed of movement of intracellular components) and rate constants (such as degradation of IκBα, and translocation of components across the nuclear membrane). We believe that a table of such information would provide a more effective mechanism to convey this information, than to over-engineer a UML diagram. Furthermore, although UML allows the range of individual objects to be depicted through multiplicity, in the form of a zero to many '0..*' association, this does not effectively convey the degree of simultaneous interactions between agents. Similarly, it is well understood that observations of genetically identical,

individual cells in a standardised environment often display significant differences in their response to perturbations (Tijskens et al., 2003), thus leading to the large degree of *inherent* variation within biological populations (be they cells, organisms, or communities). At a molecular level, this may be due to the varying numbers of particular proteins within a population of cells. UML does not have the ability to depict this variation, and nor was it designed to.

We agree with Cook (2012) that "UML is likely to influence model-driven development for the foreseeable future", but counter that UML should not be seen as the only tool to be used in the domain modelling process. Our domain modelling work, found a number of linear and multivariate statistical techniques to be key to gaining a fuller understanding of the domain, and for scoping the abstraction of the domain to be taken forward into the simulation. Likewise, UML does not currently have the ability to depict patterns within wet-lab data, which we believe is an essential component of the domain model for complex biological systems. We therefore believe that domain models developed using UML, benefit from the complementary views that emerge from statistical analysis. Furthermore, Rumpe and France (2011) advise that different stakeholders and modellers from different domains have varying interpretations of what constitutes an appropriate UML diagram. They further advise that as the UML specification allows for a wide variety of semantic variations[1], diagrams can be tailored to better support the varied requirements of individual modellers, stakeholders, and their respective domains. We therefore suggest that the statistical techniques used within this case study, along with the various cartoon-like diagrams for modelling the expected behaviours of the system (see figure 4.1) and physical containment of components (see figure 4.2) represent an example semantic variation point for modelling complex intracellular signalling pathways.

Furthermore, we believe that community and industry standards are important for improving the communication between developers and domain experts. The use of these standards, should make the reimplementation of models by different researchers (and labs) easier, and indeed should reduce the duplication of work, and more importantly reduce implementation errors, which may become introduced through reverse engineering of existing models and manual walkthroughs of published papers. The use of cartoon and UML diagrams are an essential first step towards development of a domain model, which may be published alongside the results of *in silico* experimental papers; however in isolation they are not enough to provide a comprehensive model, which other researchers and labs may use to reproduce computational models. In particular, cartoon and UML diagrams have been unable to convey the dynamics of IκBα degradation (along with the associated NF-κB release and subsequent activation), or indeed model the quantitative aspects of the signalling pathway. We therefore conclude that the use of linear and multivariate statistical techniques to complement the UML diagrams, can aid the development of more comprehensive domain models of complex biological systems, such as the IL-1 stimulated NF-κB signalling pathway, and that on its own, UML is not enough for developing domain models of complex biological systems.

[1]Semantic variation points are where part of the semantics of individual UML notations are not specified in detail, thus allowing the user to augment with additional semantics to tailor for their domain.

9.1.2 Reflections on the Suitability of FLAME for Modelling Complex Biological Systems

As discussed in chapter 2, FLAME is a discrete-event, agent-based, modelling and simulation framework, whose underlying design principles are based on the communicating X-Machine architecture. Kiran et al. (2008) and Coakley et al. (2012) advise that FLAME was designed and developed from the outset to be able to deal with massive simulations, incorporating a large scope with respect to the underlying real-world domain, and very large numbers of individual agents - in the order of hundreds of thousands, to millions, of agents. Through interfacing to the Message Passing Interface (MPI) communication framework, the simulation code is also fully compliant with parallel hardware platforms, enabling efficient communication between agents, and ensuring that concurrently executing agents remain synchronised with each other (Foster, 1995). One of the main advantages of FLAME is that it allows the modeller to focus on agent and environment definitions, through C and XML templates, and therefore removes the distractions associated with writing parsers and scheduling routines. Given a set of predefined code templates for agent definitions and their associated behaviours, the template engine generates custom simulation code which can be compiled and executed (Kiran et al., 2010).

Although the developmental approach advocated by FLAME is relatively intuitive to modellers familiar with the object-oriented and agent-based paradigms, the development and *in silico* experimental work performed within this doctoral thesis, has unfortunately shown that FLAME suffers from a number of serious limitations. The first of which, is that the requirement to utilise nearest-neighbour functionality for agent-binding, means that we are unable to harness the ability to run our simulations across parallel hardware (even though the framework has been developed to mitigate routing deadlock); as we would encounter message-dependent deadlocks across nodes due to the necessity to update various internal agent variables, and generate various binding related messages, during the interactions between complementary agents (e.g. NF-κB and IκBα to form the inhibited complex).

The second major limitation is linked to the first, in that the ability to run simulations across parallel architectures has forced the designers to *ban* the use of global mutable parameters in models. This limitation generated a number of issues in the development of our simulation platform, which are evidenced through: our inability to easily set the pseudo-random number generator seed value; and our inability to easily set the ID of new IκBα agents upon basal dissociation of the NF-κB-IκBα complex, which led to the conservation of mass issue due to our loss of IκBα agents throughout the timecourse of simulations. As discussed in chapter 6, workarounds were developed to resolve these technical issues, however we believe that future users of FLAME should be mindful of these limitations, in particular because the simulation framework has been positioned by a number of the original development team as being suitable for domain experts who have very little computational or programming competence. We believe that this message should be tempered, as domain experts will in all probability not have the technical experience or competence to diagnose and resolve the kinds of issues that we have experienced within this case study. As an alternative, we believe that

adherence to the CoSMoS process, where a close collaborative environment exists between modeller and domain expert, would mitigate propagation of such technical issues into the production code and associated publication to the scientific community.

The third major limitation relates to the extremely resource intensive nature of FLAME. As discussed in chapter 6, a number of performance issues were encountered during development, calibration and *in silico* experimentation. Stochastic agent-based models can be computationally expensive, with our model taking in excess of 20 hours for each 324,000 time-step simulation run on the N8 High Performance Computer (N8 HPC). Diagnostic tests on the N8 HPC indicated the length of time required to complete single simulation runs increases linearly with respect to the number of iterations (time-steps) in the simulation. They also indicated that FLAME simulations were Input-Output (IO) rate-limited, and not Central Processing Unit (CPU) rate-limited, which was a direct consequence of the need to create a separate XML output file for each iteration, due to it being underpinned by communicating X-Machine architecture. As such, an XML output file needs to be produced at the end of each time-step to record the ending states (and internal memory) of each agent, which is then read in at the beginning of the next time-step to set the starting states (and internal memory) of each agent. A performance bottleneck is therefore formed, relating to the speed of writing to, and reading from, the storage disk.

This issue was compounded following aleatory uncertainty analysis, as the minimum number of replicates required to gain stable median dynamics was 175 replicates, which generated over 56 million read-write accesses in total. There exists a trade-off during aleatory uncertainty between selecting high numbers of replicates to ensure the most stable meduan average simulation results, and the computational resources required to consistently run these high numbers of replicate experimental simulations. The number of replicate simulations must therefore represent a balance between minimising the effects of aleatory uncertainty, and the computational expense that can be accepted for the project. As such, completion of each set of 175 replicates within this project required a considerable amount of computational resources indeed, and was found to be a pragmatic balance between minimising aleatory uncertainty and being able to complete *in silico* experiments in a timely manner.

The fourth major limitation relates to the predefined sequence of events of the FLAME scheduler, which is not accessible to the modeller, and controls the iteration through the sequence of internal X-Machine states and the addition/removal of agents from the system. Through the calibration process, it became evident that our computational model suffered from a conservation of mass issue. Simulation dynamics when using rebind delays of 1 or 2 time-steps (for the newly re-introduced IκBα following basal dissociation of the inhibted complex), or basal dissociation rates between 0.000001 - 0.0001, suffered from an issue with the conservation of IκBα agents, with simulation runs temporarily losing these newly re-introduced agents. Investigations indicated that the issue was due to the underlying scheduler for FLAME when adding new agents to a simulation. It appears that a lag period of one time-step is introduced by the FLAME simulation framework between functions being called to generate a new IκBα agent (following basal dissociation), and its actual incorporation into the simulation.

The fifth major limitation once again links to the resource intensive nature of FLAME, however this time relates to epistemic uncertainty analysis during simulation platform development. As discussed in chapter 6, our epistemic uncertainty analysis followed the recent work of Read et al. (2012), but instead of using a global sensitivity technique, such as the latin hypercube (McKay et al., 1979), we were forced to use a mixture of robustness analysis through the perturbation of parameter values for a single parameter (one-at-a-time analysis), and local sensitivity analysis for co-dependent parameters through a two-at-a-time approach (Saltelli et al., 2009). A global sensitivity approach, such as the latin hypercube, would have been beneficial as there is currently a risk that we may have calibrated our simulation platform to a local minima within parameter space, however with single simulations taking approximately 20 hours to run; the 175 replicates generating approximately 80TB of data; and the need to ensure other users of the N8 HPC are able to run their code; it would have been intractable in the lifetime of a 3 year doctoral research project to have used such a technique.

Finally, we would like to state that although a number of the original development team have positioned FLAME as running efficiently across all architectures; from laptop, to desktop, to HPC; we believe that its resource intensive nature for complex biological systems that use nearest-neighbour functionality (as evidenced through this case study), makes large-scale simulations intractable on anything other than HPCs, and even these need vast amounts of fast solid state disk storage in order to cope with the IO issue - our minimal model generated 80TB of data per 175 replicate simulation run, and we were forced to schedule a maximum of 50 replicates at a time to ensure other users were also able to utilise the N8 HPC.

9.1.3 Reflections on the CoSMoS Process as a Project Lifecyle for Modelling Complex Biological Systems

"Scientists are taught the scientific method from the time they perform their first experiments. Similarly, software engineers are taught about the software development lifecycle before they write their first 'if' statement" (Baxter et al., 2006).

The CoSMoS process, reviewed in chapter 2, and used as the project lifecycle for this doctoral thesis, merges these two approaches, by advocating a principled approach to design and development of computational models and their subsequent use for *in silico* experimentation. We believe that such a merging of approaches is paramount for the successful management and implementation of computational biology projects - it is all to easy for modellers (software developers) to get carried away with the technology and technical aspects of the software development, and deviate away from the overall objectives of using the computational model to investigate various scientific questons of interest. We have found the CoSMoS project lifecycle to be extremely useful in ensuring the modeller is always cognisant of the real-world context behind the computational modelling work, and therefore focused on the objective of investigating the real-world domain, and not just developing the most elegant, efficient, and technically proficient code. This finding agrees with Read (2011), who previously reported that the creation of separate artefacts for the functional specification (domain

model), technical specification (platform model), and physical software (simulation platform), provides confidence that the computational model's mechanics are a faithful representation of the target domain.

One of the major strengths of the CoSMoS process is its advocation of separating the abstracted view of biology (domain model) from the technical specification of the computational model (platform model). This separation ensures the abstracted view of biology and the technical specifications of the system remain discrete models, and thus aims to minimise confusion during the development of the computational model around what aspects of the programming code relate to biological requirements, and what aspects are necessary as technical workarounds due to constraints of the specific modelling frameworks being used (e.g. communicating X-Machines and FLAME).

The CoSMoS process was defined to be iterative in nature, and therefore expects multiple iterations of the domain model, platform model and simulation platform before the final versions are agreed. We believe this to be valuable, in particular because the iterative process of domain modelling, helps the modeller to explore the biological domain in conjunction with one or more domain experts, in order to ensure the functional requirements of the computational model are captured, and formally agreed to define the scope of the project, before development begins. Similarly, once the domain model has been completed, the CoSMoS process advocates an incremental approach to development of the platform model and simulation platform over the lifetime of a CoSMoS project. The use of modularisation within our platform model and simulation platform, which were underpinned by an object-oriented approach to design and development, ensured that development activities were performed in a quality assured manner, and that the various increments of our simulation platform were easily verified against the associated platform model, and validated against the domain model.

Although we have found the CoSMoS project lifecycle to be suitable for modelling and simulation projects within computational biology, we believe that it could be extended further, through augmentation with additional details relating to the various activities that are performed, and processes that are followed, during the three CoSMoS phases (discovery, development, and exploration). For example, the use of UML to document part (or indeed all) of a domain model is now becoming normalised, following the recent work of Read et al. (2009a, 2014) on Experimental Autoimmune Encephalomyelitis, Alden et al. (2012) on Peyer's patches, and chapter 4 within this doctoral thesis, which advocates the use of statistics to complement UML when documenting the multiple hierarchical abstractions of domain models. Similarly, recent work by Ghetiu et al. (2010) advocates the use of argumentation-driven approaches (using Goal Structured Notation) for the validation process when developing computational models of complex systems. Additionally, Stepney (2012) has advocated a process for mapping the real-world data to the various artefacts developed within a CoSMoS project. For complex biological systems, this would require mappings between the *in vitro* or *in vivo* experimental data and the domain model; the domain model and the platform model; the platform model and the simulation platform; and the results model back to the real-world biological domain. Furthermore, there has recently been considerable work from the York Computational Immunology Lab (YCIL) on cal-

ibration and uncertainty analysis for agent-based modelling, and the need for a solid statistical underpinning to ensure the results of *in silico* experimentation can be translated back to the real-world domain in order to ensure predictions are based on the underlying mechanisms of the real-world system, and are not simply due to the random nature of stochastic simulations. We believe that our revised CoSMoS project lifecycle diagram (figure 2.6) provides an indication of where additional details relating to activities and processes should be aligned.

Along with the technical activities and processes for design and development of computational models, there are also essential project management related activities and processes. We therefore believe that any updates and enhancements to the CoSMoS process should also incorporate elements of project management in order to ensure that CoSMoS projects are not only quality assured in relation to development of artefacts, but also adhere to leading practices in project management, to mitigate negative impacts from non-technical areas of the project. The project management approach does not need to be overly administrative, bureaucratic or burdensome, but may in fact consist of a light-touch approach, utilising concepts from the Project Management Institute (2004) and the Unified Software Development Process (Arlow and Neustadt, 2009), which is closely aligned to UML, and would lend itself to the development of the domain model, platform model, simulation platform, and results model, which are the four artefacts required from a CoSMoS project.

We believe that the nine knowledge areas from the PMI PMBoK along with two processes from the Oracle Unified Method (OUM), could be used to augment the current CoSMoS process, and facilitate its tranformation from a high-level project lifecycle, into a development method for the modelling and simulation of complex systems. The nine knowledge areas from the PMI PMBoK cover: Integration Management, Scope Management, Time Management, Cost Management, Quality Management, Staff Management, Communication Management, Risk Management, and Procurement Management. Similarly, the two processes that we believe may be useful from the OUM, relate to Configuration Management and Infrastructure Management. These additional project management knowledge areas and processes, can be grouped into three high-level phases, which relate to Project Startup, Project Execution and Control, and Project Closure.

The Project Startup phase precedes the Discovery phase from the CoSMoS process, and is where the project planning activities take place, and where the objectives of the project are defined. The Project Execution and Control phase covers the project management activities and processes that are involved in the full CoSMoS project lifecycle as it currently stands (Discovery, Development and Exploration phases, along with development of the domain model, platform model, simulation platform and results model). Finally, the Project Closure phase occurs after completion of all *in silico* experimentation and development of the final version of the results model. During this phase, the CoSMoS project will be formally *closed* from an administrative perspective. This will include ensuring the various increments of the project artefacts are archived within an institutional repository, publishing any outstanding results in relevant academic outlets (peer-reviewed journal articles, or presentation at conferences), uploading of code to institutional and preferably public repositories, and capturing of lessons learned from the project, for example what aspects of the project worked well, what could

have been improved, and where did any limitations and/or constraints arise - the previous subsection of this discussion provided an example of such a reflective lessons learned commentary on the FLAME simulation framework. In order to provide a single diagrammatic representation of how a future CoSMoS project lifecycle may look, we have overlaid these three high-level project management phases over our CoSMoS process diagram from chapter 2 (see figure 9.1).

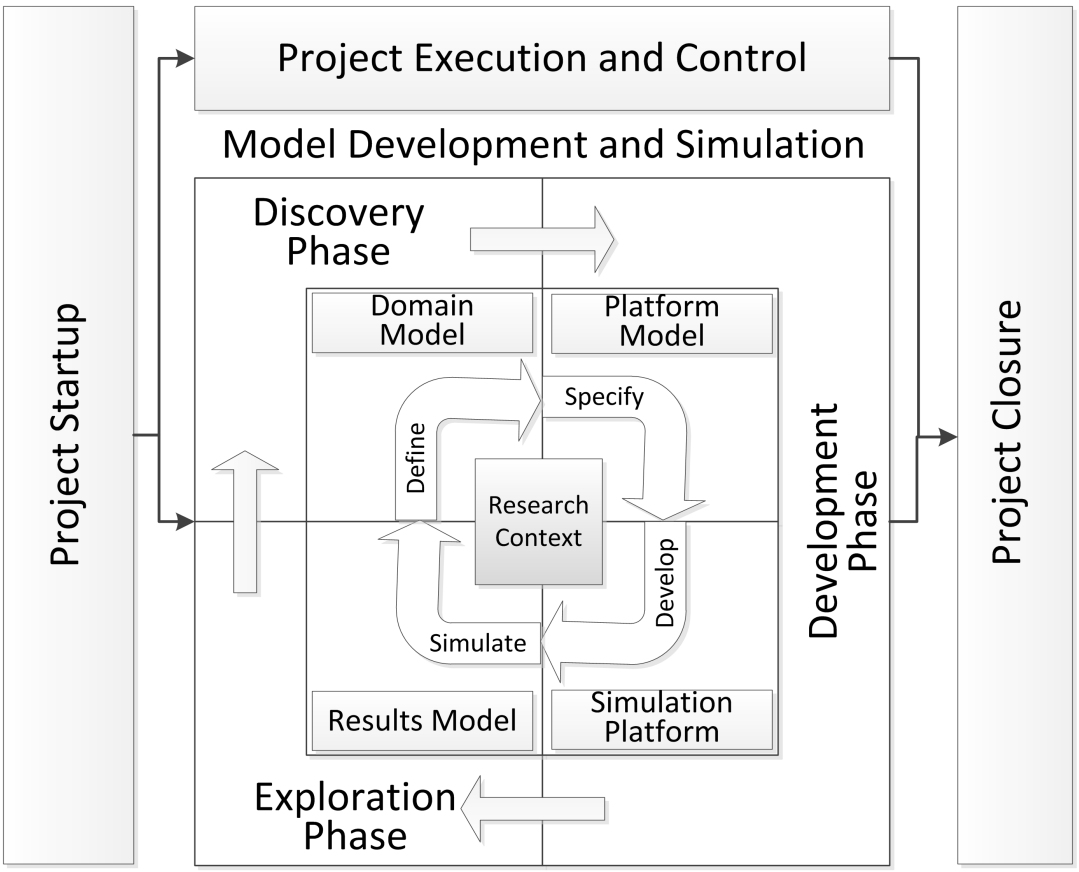

Figure 9.1: The proposed augmented CoSMoS process continues to advocate an iterative lifecycle, but this time also contains three project management phases, along with the existing three implementation phases.

9.1.4 Reflections on the Necessity for Statistical Rigour

As discussed in chapter 2, the choice of which modelling paradigm to employ is guided by multiple factors, including the real-world questions that you wish to ask, the assumptions that can be made when abstracting the real-world domain into the domain model, the availability of data relating to individual model parameters, and the computational expense of running the model. Agent-based models, being based at the level of the individual component and stochastic in nature, are able to reveal emergent behaviours resulting from the interactions of agents within time and space. Unfortunately, ABMs require a larger number of parameters than both ODE and PDE models, and because these are often estimated within ABMs during the calibration process, it is not often possible to generate exact matches between parameters from the real-world system (from *in vitro* and *in vivo* experimentation) and those defined within computational models.

Our agent-based model takes an abstracted view of the real-world biology (see chapters 5 and 6) in terms of both the agent types, and indeed the numbers of the respective agents. As such, we cannot directly compare an executing model to the wet-lab data, and utilised the process advocated by Stepney (2012) in order to translate the real-world biological data into an appropriate range of numbers for calibration of the simulation platform. We therefore believe that a rigourous framework for calibration, uncertainty analysis, and interpretation of simulation results, is therefore essential to ensure that the model adequately relates back to the underlying real-world domain, and furthermore that we can have confidence that the results of *in silico* experimentation are reflective of the experimental conditions and not a mere random occurence due to the stochastic nature of the model.

As with all agent-based models that utilise stochastic behaviours and set different pseudo-random number generator seed values for each replicate, there is significant aleatory uncertainty within our model. Such aleatory uncertainty can severely effect our ability to compare experimental results against reference dynamics. This being the case, aleatory uncertainty analysis allows us to investigate the uncertainty that is introduced through the use of PRNGs, and to calculate a minimum number of replicate simulations to generate a stable median average of the simulation results. The calculation of the minimum number of replicates, and its use within *in silico* experimentation to generate a stable median, allows us to mitigate stochastic effects and thus develop confidence that our simulation results are representative of the condition(s) on which the simulation was run, and not an artefact of our computational model that is caused by any specific PRNG seed value.

Our aleatory uncertainty analysis followed the recent work of Alden et al. (2013), but instead of using their Spartan tool-chain, we developed a novel approach using the Kolmogorov-Smirnov two-sample test (for statistical significance) and the Vargha-Delaney A-Test (for effect magnitude), for calculating the minimum number of replicates require to ensure a stable median average of simulation results. This was required, because the Spartan tool-chain, and previous work by Read et al. (2012), upon which Spartan was based, use global sensitivity and uncertainty analysis techniques, which would be intractable for this doctoral thesis, due to the resource intensive nature of FLAME. As discussed in chapter

6, the KS-Test indicated that all agent states of interest stabilised at approximately 175 replicates (with p-values >92.5), apart from the IκBα agents located in the nucleus, which we believe is due to the small numbers involved, and resultant small *absolute* differences in agent numbers having a disproportionately large effect on the p-value. The A-Test was therefore used to provide additional confidence that 175 replicate simulation runs is the minimum number of replicates to achieve a stable median average distribution, with stabilisation of the A-Test scores between 0.47 and 0.53.

Along with aleatory analysis to understand the uncertainty due to the stochastic nature of the agent-based model, we also performed epistemic uncertainty analysis to look at the uncertainty due to the model's design, which is based on our abstracted view of the real-world system. Results indicate that the system is robust to perturbations involving Rebind Delay, Nuclear Import and Nuclear Export parameters, sensitive to the Differential Time parameter, and displayed fragility with respect to the Basal Dissociation parameter. We believe that the effect of this uncertainty must be considered when interpreting the results of *in silico* experimentation back to the real-world system, as predictions may only hold for a small area of parameter space, which could be smaller than the range of possible values reported in the biological literature. Finally, due to the stochastic nature of agent-based models, a rigourous statistical underpinning is required in order to interpret the results of *in silico* experimentation against reference data, such as baseline or control dynamics. As discussed in the empirical chapters for the baseline simulator (chapter 7) and the augmented simulator (chapter 8), and the supplementary material (Appendix B), we have also used the KS-Test and A-Test to ascertain statistical significance and scientific significance.

We believe that rigourous statistical underpinnings, such as those used within this doctoral thesis, are not an *added extra*, but are in fact an essential part of simulation platform development and analysis of *in silico* experimental results. We also believe that without such rigour, it will be hard for the niche area of computational biology to gain credibility with the wider scientific community.

9.2 Contributions to NF-κB Modelling

As discussed in chapter 4, the IL-1 stimulated NF-κB signalling pathway is a complex dynamical biological system, which activates transcription of a large number of inflammatory response genes. The system is multi-dimensional, being influenced by numerous intermediate components within the network, being regulated through temporal and spatial interactions of components, and being regulated by both positive and negative feedback loops. Through design and development of our computational model, and subsequent *in silico* experimentation, we have investigated both the dynamics of the system, and tested the validity of our assumptions behind the model. We will reflect on how these various aspects have contributed to the body of knowledge around modelling the NF-κB signalling pathway, within this section. Initial focus will be applied to domain modelling and simulation platform development, before discussing how the results of sensitivity and uncertainty analysis (of the model) may be used to make predictions on real-world system dynamics, and finally discussing how the results of our *in silico* experimentation can be translated back to real-world biology.

9.2.1 Reflections on the Domain Model

We found the iterative process of domain modelling using the deep-curation approach to be extremely helpful in allowing the modeller to explore the biological domain (in conjunction with the *domain expert*) before development of the computational model. There exists substantial quantity of literature on the NF-κB signalling pathway, and various aspects of the pathway are independently studied by a wide variety of labs. It is generally understood that representing every aspect of a real-world system in models and simulations is computationally intractable. As such, a subset of the properties and behaviours from the real-world system need to be defined for subsequent investigation. One of the primary purposes of the domain model is to capture this subset of real-world system properties, and therefore provide a definition of the abstraction level taken for the modelling project. The present domain model represents the most recent iteration, and has focused on the subset of signalling components investigated by the Qwarnstrom lab; with focus applied to the observations of NF-κB and IκBα dynamics from Carlotti et al. (1999, 2000) and Yang et al. (2001, 2003) using single-cell analysis.

The use of cartoon and UML diagrams were an essential first step towards development of our domain model, however in isolation they were not enough to provide a comprehensive model. In particular, they were unable to convey the dynamics of IκBα degradation (along with the associated NF-κB release and subsequent activation), or indeed model the quantitative aspects of the signalling pathway. We therefore used a number of univariate and multivariate statistical techniques to complement the UML diagrams, in order to develop a more comprehensive domain model of the NF-κB signalling pathway. Following completion (by the modeller), and validation (by the domain expert), the domain model acted as functional specification for our agent-based model, and furthermore provided a comprehensive and transparent understanding of the domain that underpins the *in silico* experimentation performed as part of the exploration phase of this CoSMoS project. As such, we believe that the domain model is an essential project deliverable that provides an audit trail on how the real-world biology is linked, through abstractions, assumptions, and constraints, to the functionality of the computational model. We also believe that our domain model can be utilised by other modellers from the Qwarnstron group, and indeed modellers from other research groups, in order to ensure a consistent abstraction level (with respect to real-world biology) for future work around the IL-1 stimulated NF-κB signalling pathway at the single-cell level.

Aside from its use as functional specification, the domain model in conjunction with published literature, has also indicated the NF-κB signalling pathway to be a good candidate for forming a scale-free network (Kitano, 2004a; Albert, 2005). Scale-free networks are identified by the fact that most nodes have relatively low numbers of connections, but that some nodes have very high numbers of connections, and are termed hubs. These network topologies are able to tolerate random removal of nodes, such as the individual extracellular signals or inflammatory gene products, but are vulnerable to loss of the interactive hubs, which in this case is the NF-κB-IκBα signalling module. As such, the inflammatory gene networks linked through the NF-κB signalling pathway can be considered highly connected hubs of signal transduction, and if perturbed through mutation or disease, may lead to eventual collapse of the system-wide inflammatory response.

9.2.2 Reflections on the Simulation Platform

> "The scientific goal of systems biology is not merely to create precision models of cells and organs, but also to discover fundamental and structural principles behind biological systems that define the possible design space of life. The value of understanding fundamental and structural theories is that they provide deeper insights into the governing principles that complex evolvable systems including biological systems follow." (Kitano, 2007)

As elegantly quoted above by Kitano, the goal of systems biology is not merely to develop a precision model, but also to utilise the model to investigate the underlying principles of the real-world system. As such, computational models provide not only a means to rigorously think about and describe complex dynamical systems, but also provide an opportunity for us to extend our knowledge of system dynamics through performing novel *in silico* experimentation, and extrapolating these results back to the real-world system.

Our agent-based simulation platform contains additional granularity at the cell membrane receptor complex with respect to existing models, through explicit modelling of the IL-1R, MyD88, TRAF, and IRAK components. It has been developed using the concept of communicating X-Machines, which should ensure the model can be run across multiple computing architectures (e.g. laptop, desktop, cluster and grid), and calibrated against wet-lab data of Carlotti et al. (2000) and Yang et al. (2003), to ensure that the model can be used for predictive purposes against the IL-1 stimulated NF-κB signalling pathway, at the single-cell level. As such, we believe that our simulation platform can be used by other modellers from the Qwarnstrom group, and indeed modellers from other research groups, to perform additional *in silico* experimentation around the IL-1 stimulated NF-κB signalling pathway.

One of the key strengths of the FLAME simulation framework is its flexibility, in that once the highly technical underpinning of a model has been developed (e.g. the 3D Brownian movement of agents, the mirroring effect to contain agents within the correct compartment, and PRNG seed setting functionality), the addition of new functionality and agent types is relatively straightforward due to its modular agent architecture (XML agent definition and C agent functions). Furthermore, due to our principled approach to design and development, which followed the CoSMoS process, and our rigourous calibration phases and uncertainty analysis, we believe that the model will provide a useful way for wet-lab biologists and computational biologists (who do not have the depth of technical expertise to develop their own models) to perform novel *in silico* experimentation in a predictive capacity and for the purpose of hypothesis generation. In this way, the biologists and computational biologists who subsequently use our model will be able to *close* the loop, with respect to the hypothesis-driven systems biology lifecycle advocated by Kitano (see figure 2.2).

9.2.3 Reflections on Sensitivity and Uncertainty Analysis

As discussed in chapter 6, there are two main types of uncertainty within computational models: epistemic uncertainty, which arises from our incomplete knowledge of the real-world system that we are modelling; and aleatory uncertainty, which arises from the stochasticity within both the real-world system, and indeed our computational model due to the use of probabilistic interactions and pseudo-random number generators.

Our epistemic uncertainty analysis followed the recent work of Read et al. (2012), but instead of using a global sensitivity analysis, such as the latin hypercube (McKay et al., 1979), we were forced to use a mixture of robustness analysis, through the perturbation of parameter values for a single parameter (one-at-a-time approach), and local sensitivity analysis for co-dependent parameters through a two-at-a-time approach (Saltelli et al., 2009); due to the extremely resource intensive nature of FLAME making a global approach intractable. Our analysis identified that: 1) the profiles of agents within the nucleus are very noisy, which we believe is due to the small *absolute* numbers following calibration, and that even small changes in absolute numbers (of agents in the nucleus) promote amplification of this noise; 2) the system is robust to perturbations of the rebind delay value with respect to the cytoplasmic agents, but sensitive with respect to the nuclear agents; 3) there is co-dependence between the rebind delay parameter value and the association parameter value; 4) the system is extremely fragile to perturbations of basal dissociation between the NF-κB and IκBα agents in the inhibited complex; 5) the nuclear export parameter appeared to be redundant under control (non-stimulated) conditions, with the export dynamics being dependent on the nuclear import parameter value, although we suspect that this may be a system artefact of our abstraction level (i.e. reduction of $\sim$60,000 NF-κB agents down to $\sim$500, and the small *absolute* number of agents in the nucleus); and 6) the system is sensitive to changes in the import parameter value. We believe the results of this epistemic uncertainty analysis are consistent with the findings of Kitano (2004a), who conjectures there to be a delicate balance between robustness and fragility within complex dynamical biological systems.

Our aleatory uncertainty analysis closely followed the recent work of Alden et al. (2013), but instead of using their *Spartan* tool-chain, we developed our own analysis scripts using Matlab. This involved the running of 300 replicate simulations with different PRNG seed values to generate stochastic variation between the simulation runs, and utilised the Kolmogorov-Smirnov test to understand statistical significance and the Vargha-Delaney A-Test to understand effect magnitude difference. Our analysis indicates that 175 replicate simulations are required when using our computational model for *in silico* experimentation. We believe that this will minimise the aleatory uncertainty inherent to our computational model, but will also allow stochasticity, so will reduce the risk of overtuning the computational model by running large numbers of replicates.

The calculation of the minimum number of replicates, and its use within future *in silico* experimentation to generate a stable median average, allows us to mitigate stochastic effects and thus develop confidence that simulation results are representative of the condition(s) on which the simulation was run, and not an artefact of our computational model that is caused by the specific PRNG seed value.

9.2.4 Reflections on In Silico Experimentation

Following the development and calibration of the simulation platform, and subsequent epistemic and aleatory uncertainty analysis, we were able to perform *in silico* experimentation to increase our understanding of the underlying mechanistic behaviours of the NF-κB signalling pathway. As discussed in chapter 2, agent-based models lend themselves to investigating the temporal and spatial dynamics of systems, and this allowed us to investigate the consequences of perturbations to component interactions.

Four sets of *in silico* experiments were performed on the initial version of the simulation platform (discussed in chapter 7). These investigated: the effects of varying basal dissociation of the NF-κB-IκBα complex; the effects of varying IKK numbers on the signalling pathway dynamics; the effects of varying IκBα numbers on the signalling pathway dynamics; and the effects of varying the lag-time between cell membrane receptor activation and the subsequent IKK activation within the cytoplasm.

The first experiment was important because the work of Carlotti et al. (2000), which focused on control (i.e. non-stimulated) dynamics, postulated that shuttling of free NF-κB and IκBα was a consequence of basal (i.e. non-IKK-mediated) dissociation of the NF-κB-IκBα complex within the cytoplasm, rather than the direct nuclear import of the NF-κB-IκBα complex, degradation of IκBα, and subsequent resynthesis of IκBα. Our results under IL-1 stimulated conditions, indicate that an increase in basal dissociation leads to: a decrease in both cytoplasmic and nuclear NF-κB-IκBα complexes; an increase in cytoplasmic NF-κB, with an associated decrease in nuclear NF-κB; and an increase in cytoplasmic IκBα, with a small increase in nuclear IκBα. From a biological perspective the reduction in nuclear NF-κB and cytoplasmic NF-κB-IκBα complexes, will be because an increase in basal dissociation results in a reduction of the total numbers of NF-κB-IκBα complexes in the cytoplasm, and thus a reduction in the IKK-mediated release of NF-κB, which further causes a reduction in the number of NF-κB agents that become activated following their translocation to the nucleus. We also conclude that the system is fragile to perturbations involving basal dissociation for both control conditions, as shown through epistemic uncertainty analysis (see section 6.3.1) and IL-1 stimulation conditions.

Results of the second experiment indicate that the system is stable to perturbations in IKK numbers within a small range either side of the default value (i.e. between 50% and 200%), but sensitive to more extreme perturbations (i.e. 10%, 20%, 500%, and 1,000%). Using these higher IKK numbers (i.e. 500% and 1,000% of the default value), results also indicate that an increase in the number of IKK agents, leads to: a decrease in cytoplasmic IκBα; a decrease in both cytoplasmic and nuclear NF-κB-IκBα complex, along with an associated increase in the dissociation of the complexes; and no significant change to either nuclear NF-κB-IκBα complexes or nulcear IκBα, although we suspect that this is due to the small absolute numbers that are associated with the calibrated simulation platform. Non-linear dynamics of cytoplasmic NF-κB were also identified, which we believe are due to the rapid increase in dissociation of NF-κB-IκBα complexes to yield free NF-κB, and then subsequent translocation to the nucleus whereby these IKK-mediated free NF-κB agents become activated and remain in the nuclear compartment. As such, we conjecture that an increase in IKK numbers

generates an initial increase in cytoplasmic NF-κB, which then slowly reduces over time through translocation to the nucleus.

Results of the third experiment indicate that the system in both control (non-stimulated) and IL-1 stimulated conditions, is fragile to perturbations that increase the number of free IκBα agents to 3x the default value, but stable when the excess IκBα agents are sequestered to the cytoskeleton. Under control conditions, the excess free IκBα results in the system settling at a pseudo steady-state away from our calibrated dynamics, whereas excess sequestered IκBα agents, conversely show no scientifically significant differences. We believe that these results for control dynamics are consistent with the findings of Pogson et al. (2008). Results of excess free IκBα during IL-1 stimulation also show significant differences, generating an initial spike in free NF-κB (with a corresponding decrease in cytoplasmic NF-κB-IκBα complexes), before a rapid decrease in cytoplasmic numbers due to their translocation to the nucleus. We conjecture that these phenomena may be due to the extremely low numbers of free NF-κB at the beginning of simulations, as the pseudo steady-state dynamics following excess free IκBα under control conditions, had resulted in the overwhelming majority of NF-κB agents to be complexed with IκBα within the cytoplasm. Furthermore, as per the control conditions experiment, the addition of excess sequestered IκBα had no effect on the system.

Results of the fourth experiment indicate that the system is robust to perturbations involving the lag-time before activation of IKK. Increases of upto 40x to the lag-time (with respect to the default value) between cell membrane receptor activation and the subsequent activation of IKK within the cytoplasm.

Finally, the second iteration of the computational model (as discussed in chapter 8) incorporated increased granularity of components at the cell membrane, to extend the scope of agent interactions upstream of the NF-κB signalling module. Along with the five main agents from iteration 1 that consisted of the IL-1R cell membrane receptor, IKK, IκBα, NF-κB, and nuclear transporter; the second iteration included three additional agents relating to MyD88, IRAK, and TRAF6. This allowed us to investigate through *in silico* experimentation, the effects of varying the numbers of these cell receptor complex components (akin to varying the concentration in wet-lab experiments), and whether the system is robust to such perturbations. Results of this fifth (and final) experiment indicate that the system is relatively robust to perturbations involving the total numbers of individual cell membrane receptor complex components.

These *in silico* experiments demonstrate the robust, yet fragile, nature of the intracellular pathway. Furthermore, along with the results from epistemic uncertainty analysis, they also demonstrate the importance of appreciating the effect(s) that arbitrarily assigned parameter values (defined during platform model development) may have on results of subsequent simulations using the calibrated simulation platform. We therefore agree with Read (2011) that "*in silico* predictions that rest on arbitrarily assigned parameter values should be challenged."

9.3 Thesis Summary and Contribution

This section returns to the overall purpose and specific objectives of this doctoral thesis, and provides a summary of how the individual objectives were met, and the contributions that were made in doing so. Seven research objectives were identified (see chapter 1) to guide the work in this thesis towards the development of an agent-based model of the IL-1 stimulated NF-κB signalling pathway, along with its use to perform novel *in silico* experimentation. These were as follows:

Obj 1: Explore the role of diagrammatic and statistical techniques for developing a domain model of the NF-κB case study.

Obj 2: Create an agent-based computational model of the core intracellular components of the IL-1 stimulated NF-κB signalling pathway.

Obj 3: Investigate techniques for calibrating agent-based computational models that have been developed using the FLAME simulation framework.

Obj 4: Perform novel *in silico* experimentation using the agent-based model.

Obj 5: Augment the agent-based model with additional upstream signalling components related to the cell membrane receptor complex.

Obj 6: Perform novel *in silico* experimentation using the augmented simulator.

Obj 7: Investigate the suitability of using the FLAME simulation framework for developing computational models of complex biological systems.

The remainder of this section provides a chapter-by-chapter summary of this doctoral thesis, which maps the specific research objectives and contributions against the chapter where they were achieved.

Chapter 1: The motivation for investigating the IL-1 stimulated NF-κB signalling pathway through agent-based modelling is provided, along with the seven research objectives, and an overview of the thesis structure.

Chapter 2: A review of the theoretical underpinning to the computational aspects of this work is provided. Emphasis was placed on General Systems Theory, Network Theory, Systems Biology, Computational Biology and Computational Immunology. The chapter also discussed the equation-based and agent-based approaches for modelling complex biological systems, and gave an overview of the CoSMoS process.

Chapter 3: A review of the cell biology at the level of detail required to understand the IL-1 stimulated NF-κB intracellular signalling pathway, and the potential consequences from dysregulation. The single-cell experimentation performed by the Qwarnstrom lab is described. The complexity of the signalling pathway is used to motivate the benefits of computational modelling, and indeed a thorough review of the existing equation-based and agent-based models was provided. This chapter forms the domain of the CoSMoS process.
Contributions:

- Review the existing literature around the complexity that is inherent to the NF-κB signalling pathway; the way that computational approaches have facilitated our increased understanding of the pathway; and the need for a new computational model to extend our understanding even further, with particular reference to the IL-1 stimulated pathway.

Chapter 4: Defines a domain model of the IL-1 stimulated NF-κB signalling pathway. This domain model was developed using cartoon and UML diagrammatic notations, and complemented through univariate and multivariate statistical techniques. Critical reflections were provided on the appropriateness of UML for modelling complex biological systems, and where deficiencies were perceived, discussion was made on how statistical techniques may complement UML, to develop a more complete domain model. In terms of the CoSMoS process, this chapter translates the real-world domain into our abstracted view, which will be used to define the scope of our agent-based model. This chapter addressed research objective 1.
Contributions:

- A complete and comprehensive domain model of the IL-1 stimulated NF-κB signalling pathway. No such model existed prior to the work detailed in this doctoral thesis.
- A detailed analysis of UML's ability to define the complex structural, temporal, and interaction dynamics, which are inherent to complex dynamical biological systems.
- An investigation into the ability of statistical techniques to complement UML where deficiencies have been found.

Chapter 5: A platform model for the first iteration of our agent-based model, which is implementation platform specific (i.e. the FLAME simulation framework). Along with various UML notations, the platform model also utilises communicating X-Machine diagrams for the individual agents, and a state-graph for the communication between agents. In terms of the CoSMoS process, this chapter translates the abstracted view of the underlying biological functionality into a technical specification.
Contributions:

- Design of an agent-based model at the single-cell level, through which *in silico* experimentation may be conducted.

Chapter 6: A calibrated simulation platform for iteration 1, which has undergone aleatory and epistemic uncertainty analysis. In terms of the CoSMoS process, this chapter defines the translation of the platform model into our actual agent-based model, which has been calibrated to wet-lab dynamics, and is therefore ready for *in silico* experimentation. The chapter reports the novel calibration technique, which uses the mapping approach advocated by Stepney (2012) to translate the results of wet-lab biology (from the domain model) into a desired range of agent numbers (median and interquartile range) for our calibrated simulation platform. This chapter addresses research objective 2 and 3, and contributes to research objective 7.
Contributions:

- An interdisciplinary and collaborative calibration procedure, which was underpinned by the transformation of single-cell analysis wet-lab data (from the domain model) into desired ranges for output data (from the simulation platform) of individual agent types.
- Design and development of an agent-based model, calibrated to wet-lab data at the single-cell level, through which *in silico* experimentation may be conducted.
- Build on the discussion by Read et al. (2012) of the necessity to establish a calibrated *baseline* for the simulation platform, which has taken account of the epistemic uncertainty of the underlying real-world biology, and the aleatory uncertainty that arises through the stochasticity of the computational model.
- Development of a novel approach for using the Kolmogorov-Smirnov two-sample test and the Vargha-Delaney A-Test for calculating the minimum number of replicates required to ensure a stable median average of simulation results. This was required because the previously published *consistency analysis* technique by Read et al. (2012), could not be used due to the resource intensive nature of the FLAME simulation framework, making global sensitivity and uncertainty analysis techniques intractable.
- Initiate a debate on the necessity for the computational biology community to ensure a rigourous and robust statistical underpinning to simulation platform development, in order to develop credibility with the wider scientific community.
- Assessment of the suitability of the FLAME simulation framework to model complex biological systems in general, and the IL-1 stimulated NF-κB signalling pathway in particular.

Chapter 7: The simulation platform developed for the first iteration of our agent-based model has been used for novel *in silico* experimentation, which forms the basis of the results model - the fourth and final artefact from a CoSMoS project. The first experiment into basal dissociation of the NF-κB-IκBα complex, identified the fragility of the system with respect to perturbations involving basal dissociation in both control and IL-1 stimulated conditions. The second experiment into varying IKK numbers, indicated that the system is stable to perturbations within a small range either side of the default value, but sensitive to more extreme perturbations. The third experiment into the effect of varying IκBα numbers, identified that the system is fragile to perturbations that increase the number of free IκBα agents, but stable when the excess agents are sequestered within the cell, through for example the cytoskeleton. Finally, the fourth experiment into the effects of varying the lag-time between cell membrane receptor activation and the subsequent IKK activation, indicated that the system is robust to increases in lag time. This chapter addresses research objective 4.

Contributions:

- Elucidation of the robust, yet fragile, nature of the intracellular pathway, with particular emphasis on the probabilistic nature of association

and basal-dissociation of the NF-κB-IκBα complex.

- Identification of the substantial robustness within the signalling pathway with respect to the ratios between individual component numbers and their associated interaction dynamics.
- Assessment of the suitability of the FLAME simulation framework to model complex biological systems in general, and the IL-1 stimulated NF-κB signalling pathway in particular.

Chapter 8: Discusses the second iteration of the CoSMoS project, through development of the augmented platform model and simulation platform, which added increased granularity of components at the cell mebrane receptor complex. *In silico* experimentation into the effects of varying component numbers within the receptor complex, indicate the system to be robust to such perturbations. This chapter addresses research objectives 5 and 6, and makes an additional contribution to research objective 7.
Contributions:

- Design and development of an agent-based model, calibrated to wet-lab data at the single-cell level, through which *in silico* experimentation may be conducted.
- Identification of the substantial robustness within the signalling pathway with respect to the ratios between individual component numbers and their associated interaction dynamics.
- Assessment of the suitability of the FLAME simulation framework to model complex biological systems in general, and the IL-1 stimulated NF-κB signalling pathway in particular.

Chapter 9: The concluding chapter of this doctoral thesis. Discussion has been made on: the completeness of UML for domain modelling of complex biological systems; the suitability of FLAME for modelling complex biological systems; the appropriateness of the CoSMoS process as a project lifecycle for modelling complex biological systems; and the necessity for statistical rigour when interpreting simulation results. Similarly, the contributions that *in silico* experimentation has made to the *real-world* NF-κB signalling pathway, have been discussed. The contributions of this doctoral thesis are summarised, specific areas for further work are presented, and final concluding remarks are made. This chapter makes additional contributions to research objectives 1, 3 and 7.
Contributions:

- Assessment of the CoSMoS approach as a project lifecycle for computational biology.
- Discussion of the necessity for the computational biology community to ensure a rigourous and robust statistical analysis is performed on simulation results, in order to develop credibility with the wider scientific community.

9.4 Further Work

The contributions of this doctoral thesis have identified a number of areas for further work, which are discussed in this section. Three broad areas of further work are identified, relating to: further investigations of the NF-κB signalling pathway; additional work qualifying the advantages and disadvantages of using UML to develop domain models of complex biological systems, and how alternative techniques such as statistics may provide a complementary approach; and investigations into the effects that the underlying modelling paradigm (e.g. equation-based versus agent-based) and modelling framework (e.g. FLAME versus Java MASON) may have on simulation results.

9.4.1 Further Investigation of NF-κB

This thesis has described a principled approach to design, development and calibration of an agent-based model of the IL-1 stimulated NF-κB signalling pathway; along with subsequent *in silico* experimentation to investigate various aspects of the pathway at the component-level, to provide insight to the nature of the underlying mechanisms and dynamics. Although we are confident that the simulation platform has adhered to good software engineering practices, the modelling activities were not the end goal itself, but were instead incremental activities for the development of a *tool* to increase our understanding of the signalling pathway. As such, it is hoped that future researchers will use our agent-based model to perform additional, novel, *in silico* experiments around the mechanisms and dynamics of the NF-κB-IκBα signalling module in particular, and the wider pathway more generally. We believe that an increased understanding of the pathway under normal physiology and disease, will elucidate the underlying causes of inflammatory diseases, and lead to new treatment strategies.

The most significant finding from our *in silico* experimentation, is confirmation that the system is fragile to perturbations around basal dissociation dynamics, as conjectured by Carlotti et al. (2000). We believe further work is required to fully define the dynamics of this reversible reaction; with initial focus applied to global sensitivity analysis techniques (such as the latin hypercube), before subsequent confirmation in the wet-lab.

The second significant finding was the robustness of system dynamics following perturbations to the numbers of IL-1R, MyD88, TRAF and IRAK agents. We believe that additional *in silico* experimentation around cell membrane receptor complex formation are required, in order to further investigate the mechanisms and dynamics behind receptor complex formation, receptor activation, and propagation of the signal from extracellular space into the cytoplasm. Our augmented simulation platform contained the receptor agent IL-1R, the adaptor protein MyD88, and the *generic* kinases TRAF and IRAK. This provides a basis for augmentation with additional granularity for the kinases IRAK1, IRAK4, TRAF6 and TAK1; the recently discovered co-receptor TILRR; and the inhibitory adaptor protein Tollip. Furthermore, there is also potential to provide additional granularity at the sub-component level through modelling the individual NF-κB units (e.g. RelA and p50) and the individual IKK units (e.g. NEMO, IKKα and IKKβ).

Expansion of the computational model (in this way) would then allow *in silico* experimentation into potential targets for therapeutic agents to treat diseases and disorders due to pathway dysregulation. Such targets could be the transcription factor itself, or any of the associated components within the pathway, that lead to induction of gene activation. Given that the signalling pathway is a multi-component pathway, there are numerous targets for therapeutic interventions, including cell membrane receptors (e.g. anti-IL-1R antibodies), adaptor proteins (e.g. ubiquitinase inhibitors), IKK, IκB stabilization, cytoplasmic retention of NF-κB, or direct inhibition of NF-κB (Gilmore and Garbati, 2011). Due to the complex nature of the signalling pathway in time and space, and its cross-talk with other pathways, it is unclear whether targeting of specific individual pathway components, or inhibition by broader activity against multiple components in the pathway, will show the greatest degree of clincial efficacy and safety. Gilmore and Herscovitch (2006) reviewed the known inhibitors of NF-κB, which may provide a basis for research regarding pharmacological intervention.

Given that our simulation platform is based at the single-cell level, we believe there is merit in extending the scope of the model to simulate responses at the tissue and organ levels. We note however that multiscale models are difficult to develop as they require the integration of molecular, cellular and organ level data, within a single model. In addition, due to the limitations that we have encountered with the FLAME simulation framework, which requires very large amounts of computational resources, this may be intractable with our current simulation platform. As such, we believe that the domain model (in chapter 4) should be used as the basis for our model to be ported to the GPU-based version of FLAME (currently in beta version), which has been reported to improve performance over the CPU-based version that we have used, and also has the capability of real-time visualisation as the data from each iteration is persistent in the GPU memory (Richmond et al., 2010).

To complete this subsection, we believe there are a number of key questions that may be answered through development of large-scale computational models of NF-κB. From a cell biology perspective, these are: 1) what can *in silico* experimentation tell us about the relative roles of the intermediate components within the signal transduction events; 2) what can *in silico* experimentation tell us about the various receptors, co-receptors, and adaptor proteins, and their role(s) in signal transduction events; 3) what can *in silico* experimentation tell us about the dysregulation that can occur (regarding NF-κB activation and IκBα degradation and resynthesis) when cells are in diseased states, and how can we perturb the system back to a healthy state; and 4) what can *in silico* experimentation tell us about the cross-talk that occurs when various extracellular signals (e.g. LPS, TNFα, IL-1) and other environmental stimuli (e.g. bacteria, UV radiation, physical stress) converge on the NF-κB signalling module.

9.4.2 Defining Complex Biological Systems with UML and Statistics

The Unified Modelling Language (UML) has become one of the *de facto* notations in industry for documenting the functional requirements and technical specifications of software systems. Biologists have also attempted to use UML to model biological systems, and to this end, we have developed a domain model of the NF-κB signalling pathway, for use as functional specification for our agent-based computational model. As discussed in the reflective section of chapter 4, we believe that UML is a useful tool for defining the static characteristics and linear interactions of components within a complex biological system, but that it suffers from a number of serious deficiencies when trying to define the non-linear dynamics and heterogeneity inherent to such systems.

We have previously discussed our reflections on UML within this chapter, however would like to add here that we agree with Chaudron et al. (2012) that "even if the UML syntax may not be followed very precisely, the concepts of UML form the *de facto* language for discussing model designs", and reiterate that on its own, UML is not enough for developing domain models of complex biological systems. We therefore believe that further work should be performed into the suitability of UML for defining domain models of complex biological systems, which builds on this thesis and the previous work of Read et al. (2009a,b, 2014) and Bersini et al. (2012). Where deficiencies are exposed, further investigation will then be needed to define workarounds using semantic variation points, and the use of complementary approaches (such as statistical techniques), to facilitate a more complete diagrammatic approach for developing domain models.

9.4.3 Investigating the Effects of Modelling Paradigm on Simulation Results

Over the last two decades there has been a significant expansion in computational biology research, primarily led by academia, but also in the pharmaceutical industry for the purposes of drug target discovery and drug design. The computational biology community has spoken at length on the need to standardize tools, in particular simulation platforms such as Systems Biology Workbench and CellDesigner, and especially the output format of results (e.g. a variant of XML, such as the Systems Biology Markup Language). However, just because the community may use the same toolset, it does not mean that research is comparable. There are various reasons for this inability to compare results, including differences in the design approaches taken, differences in the approaches taken for calibration and optimization, differences in the level and approach taken for quality assurance (validation, verification and testing), and more interestingly, differences in the modelling paradigm used (i.e. equation-based versus agent-based) as noted by Ray et al. (2009) who gained subtley different results from their agent-based model when replicating their previous equation-based work.

We believe that this is symptomatic of the underlying mechanistic differences between modelling paradigms, and differences between the architectures of modelling frameworks. This is exemplified by the fact that functionality within computational models does not always match biological reality due to technological

constraints, and thus requires a computational workaround, which is based on the specific architecture and programming approach of the modelling paradigm and framework being used. For example, the development of an equation-based model using Matlab or CellDesigner would require a different approach to handling computational workarounds than an agent-based model using Java-MASON or FLAME.

We believe that this area of computational biology requires urgent attention in order to identify and qualify the effects that specific modelling paradigms and frameworks, may have on the results of *in silico* experimentation, so that the community may better appreciate the *hidden* aleatory uncertainty inherent to different modelling approaches. We therefore advocate that our domain model be used to develop three new platform models and associated simulation platforms for: 1) differential equation-based model using CellDesigner; 2) Java-MASON agent-based model, which will be simulated in distributed mode; and 3) FLAME GPU agent-based model, which will be simulated in parallel mode. We suggest that focus should be applied to investigating the similarities and differences between the three platform models, and compare the simulation results and associated predictions from the three different software frameworks. With specific reference to agent-based modelling, further work into the impact of different pseudo-random number generators (PRNGs) and associated seed values on the simulation results and associated predictions, would also be beneficial to the community. Following the development of these three additional models, we believe that it would also be beneficial to develop interfaces for Java-MASON and FLAME GPU software frameworks into the Systems Biology Workbench platform, so that all three computational models may be interfaced to a common analysis and results platform.

As computational models are increasingly being used to generate predictions that have clinical implications, through for example predictions on the robustness and fragility of disease networks, we urgently require a new framework that incorporates the effects that different software platforms have on simulation results, and that will instill trust in the associated simulation-based predictions. We believe that development of such a framework, and its wide-spread adoption by the community, will instill trust in predictions from computational models, through providing the ability to interpret simulation results not only in the context of the model's design (i.e. domain model), but also in the context of the specific software platform used. Additionally, we believe the following key questions should also be investigated to instill trust in simulation results (by the wider scientific community): 1) does the abstraction level affect the accuracy of simulation-driven predictions; 2) does the resolution level (e.g. number of agents) affect the accuracy of simulation-driven predictions; 3) what are the relative merits of averaged population data versus single-cell data for calibration and validation of computational models; and 4) what are the advantages and limitations of using massively parallel computing architectures, for reproducing the large-scale variation (as seen in biological systems) into computational simulations.

9.5 Concluding Remarks

Although the relationship between mathematics and biology has been established for a considerable period of time, early work was limited to the power of the calculating machines available at the time. The application of mathematics and computer science to develop complex models of real-world biology has only recently gained acceptance in the scientific community at-large, following the significant advancements to computer hardware and software technologies over the past two decades.

There are many advantages to using theoretical models over wet-lab experimental studies in the initial phases of scientific studies. For example, mathematical and computational models are generally far less expensive to setup, calibrate and run, are less time consuming (with respect to the scientists time for each experimental run), and are more flexible to changes in environmental conditions and parameter ranges. As such, these models provide a cost effective mechanism to perform *in silico* investigations with a view to acting as a plausibility filter, for the generation of new hypotheses for future *in vitro* or *in vivo* experimentation. The major drawback however, is that theoretical models can only be as good as the data that they are designed and calibrated against, and the theory that they are based on (Woelke et al., 2010).

It is important to emphasize that a *good* model does not have to be complete (with respect to the real-world domain), as arguably all of the models that are currently accepted by the scientific community researching NF-κB are incomplete to some degree or other (see chapter 3). We therefore believe that a *good* model should incorporate the minimum set of components that are sufficient to reproduce the emergent behaviour of interest, and be designed and developed to high software engineering standards, in order to ensure rigourous calibration, validation, and verification, so that we may understand how the simulation outputs relate to the real-world system under study (Bown et al., 2012). We also believe that no computational model should be deemed to be conclusive, as at best it is an approximation to the real-world system, which as shown through the two iterations of the simulation platform in this thesis, needs to be progressively updated through additional functionality and *potentially* corrections to parameter values and rate constants.

The use of computational modelling and simulation has begun to pose new challenges to scientific research. Traditionally, scientific journals required a complete and rigorous *Materials and Methods* section to all manuscripts, however with the rapid increase in theoretical and computational studies over the past decade, this requirement appears to have been dropped, which has only recently been recognised as cause for concern. As discussed in this thesis (chapters 5 and 6), replication and reproduction of computational work can become extremely difficult due to reliance on simulation code, analysis scripts, datasets, pre-processing and transformation scripts, and initialisation parameter files. Without access to the full suite of simulation code and associated *instrumentation* tools used by the original researchers who performed the initial model development and *in silico* experimentation, it is almost impossible to replicate the work, and requires a disproportionate amount of effort (with respect to the rewards generated) to reproduce the computational model, which generally requires the design and development

of the model from scratch, relying on the discursive sections of published material to reverse engineer the design, and the published figures to retrofit simulation dynamics through the process of qualitative curve-fitting. We therefore applaud the growing movement within the computational biology community who advocate not only the need for open access publications, but also the need for open access code and the requirement to fully document our complete computational methods within the supplementary material sections of journal manuscripts (Mesirov, 2010).

Finally, we agree with Kitano (2004b) and Slepchenko et al. (2002) that the theoretical investigations (through *in silico* experimentation) into the underlying cellular dynamics of biological systems, followed up by verification in actual biological systems, needs to be promoted as a new aspect of *The Scientific Method*, and that the CoSMoS process can be used to aid in this promotion.

Appendices

A χ^2 Goodness of Fit Tests

Given two sets of data, we can test to ascertain whether they come from the same population distribution using a number of statistical techniques. The accepted test for differences between non-continuous (binned) distributions is the chi-square (χ^2) test, whereas the *Kolmogorov-Smirnov* test is used for continuous data (Press et al., 2007). The single-cell analysis fluorescence data of Yang et al. (2003) contains 36 observations for *control* (unstimulated) dynamics and 52 observations for *IL-1 stimulated* dynamics. These observations can be grouped into regular intervals (binned), and are therefore amenable to the two-tailed χ^2 goodness of fit test to ascertain whether the data (control and IL-1 stimulated) approximates to known mathematical distributions, which in this case was revealed to be either a *Normal* or *Negative Binomial* distribution under visual inspection.

Initial χ^2 tests were performed on the data with integer binning intervals. Figure A.1 represents the control observations with integer binning intervals, along with a superimposed line to represent a Normal distribution using the mean and standard deviation calculated from the data. The associated χ^2 calculations for this integer binning interval are displayed in table A.1. The hypothesis $H1_0$ is that the control observations from the full dataset approximate to a Normal distribution. The ${\chi_7}^2$ threshold for 97.5% is 16.01 and 99% is 18.48. As the χ^2 score for the data is 19.35, we reject $H1_0$. Similarly, figure A.2 and table A.2 represent the IL-1 stimulated observations with integer binning intervals. The hypothesis $H2_0$ is that the IL-1 stimulated observations from the full dataset approximate to a Normal distribution. The ${\chi_{16}}^2$ threshold for 97.5% is 28.85 and 99.9% is 39.25. As the χ^2 score for the data is 94.16, we again reject $H2_0$.

Due to the first phase of χ^2 tests rejecting an approximation (by the full dataset) to a Normal distribution, the second phase of tests were performed on a subset of the data, using initial fluorescence from 0 to 3.0 fluorescence units. Figure A.3 represents the control observations with integer binning, that have initial fluorescence between 0 and 3.0 units. It is evident that an integer interval is too high a binning interval, and in effect *hides* detail. An alternative representation is given in figure A.4, which uses a binning interval of 0.5 fluorescence units. This provides a greater spread of frequencies, and the associated χ^2 calculations are displayed in table A.3. The hypothesis $H3_0$ is that the control observations from the partial dataset approximate to a Normal distribution. The ${\chi_4}^2$ threshold for 97.5% is 11.483, and as the χ^2 score for the data is 3.099, we accept $H3_0$. This tested whether the subset of data approximated to a Normal distribution using positive values of fluorescence only. Although not physically possible, i.e. you cannot gain a negative fluorescence value using microscopy, we also believed that it was pertintent to test against the wider Normal distribution that also included negative fluorescence values within its tail. Figures A.5 and A.6 utilise the integer and 0.5 interval binning as per figures A.3 and A.4, but this time incorporating the wider tails of the Normal distribution. Similarly, table A.4 provides the asso-

ciated χ^2 calculations. The hypothesis $H4_0$ is again that the control observations from the partial dataset approximate to a Normal distribution. The ${\chi_6}^2$ threshold for 97.5% is 14.45, and as the χ^2 score for the data is 7.512, we accept $H4_0$. Figures A.7 and A.8 represent the IL-1 stimulated partial dataset superimposed with the wider normal distribution, using integer and 0.5 interval binning respectively. The associated χ^2 calculations are displayed in table A.5. The hypothesis $H5_0$ is that the IL-1 stimulated observations from the partial dataset approximates to a Normal distribution. The ${\chi_6}^2$ threshold for 97.5% is 14.45, and as the χ^2 score for the data is 14.29, we accept $H5_0$.

Although tables A.4 and A.5 provide the χ^2 scores that allow us to accept $H4_0$ and $H5_0$ that the control and IL-1 stimulated observations from the partial dataset follow a Normal distribution, the superimposed Normal distribution curves on figures A.5 to A.8 raise the concern that the requirement to include negative fluorescence values within the test does not reflect the underlying biological data. As such, the third phase of χ^2 tests were performed on a mathematical distribution (the negative binomial distribution) that mandates positive x-axis values, and thus will better reflect the underlying biological data. Figure A.9 represents the control data with integer binning, with a superimposed curve that follows the negative binomial distribution with median average of 1.947153. The associated χ^2 calculations for this integer binning interval are displayed in table A.6. The hypothesis $H6_0$ is that the control observations from the full dataset approximate to a negative binomial distribution. The ${\chi_4}^2$ threshold for 97.5% is 11.483, and as the χ^2 score for the data is 0.84, we accept $H6_0$. Similarly, figure A.10 represents the IL-1 stimulated data with integer binning, with a superimposed curve that follows the negative binomial distribution with median average of 1.729876. The associated χ^2 calculations for this integer binning interval are displayed in table A.7. The hypothesis $H7_0$ is that the IL-1 stimulated observations from the full dataset approximate to a negative binomial distribution. The ${\chi_{10}}^2$ threshold for 97.5% is 20.483, and as the χ^2 score for the data is 20.07, we accept $H7_0$.

With the above χ^2 tests in mind, we conclude that the full wet-lab dataset of Yang et al. (2003) may be deemed to approximate to a negative binomial distribution.

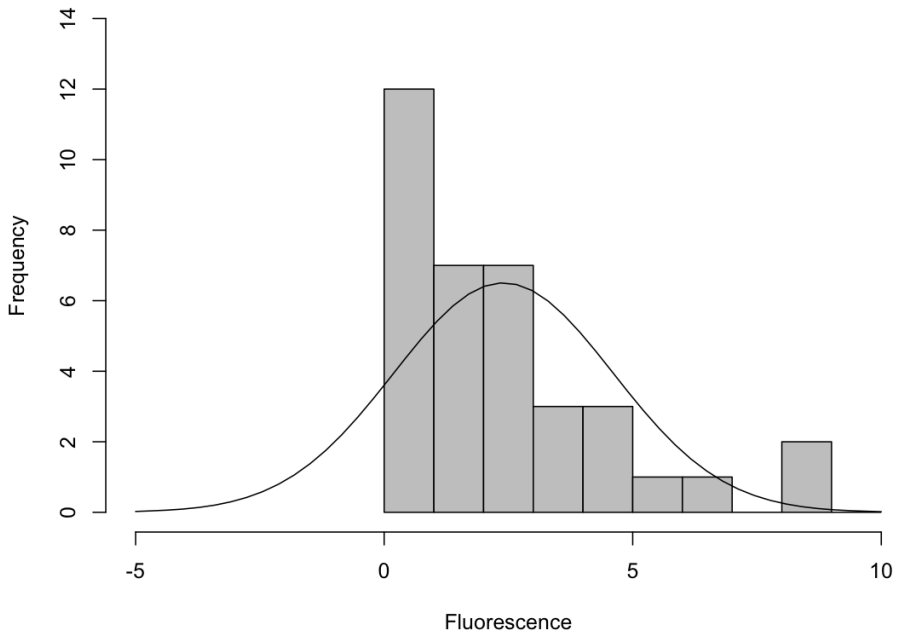

Figure A.1: Histogram of control observations that have been binned (grouped) to integer values of fluorescence. The superimposed line represents a Normal distribution, using the mean and standard deviation calculated from the raw data. The mean has been calculated as 2.3995.

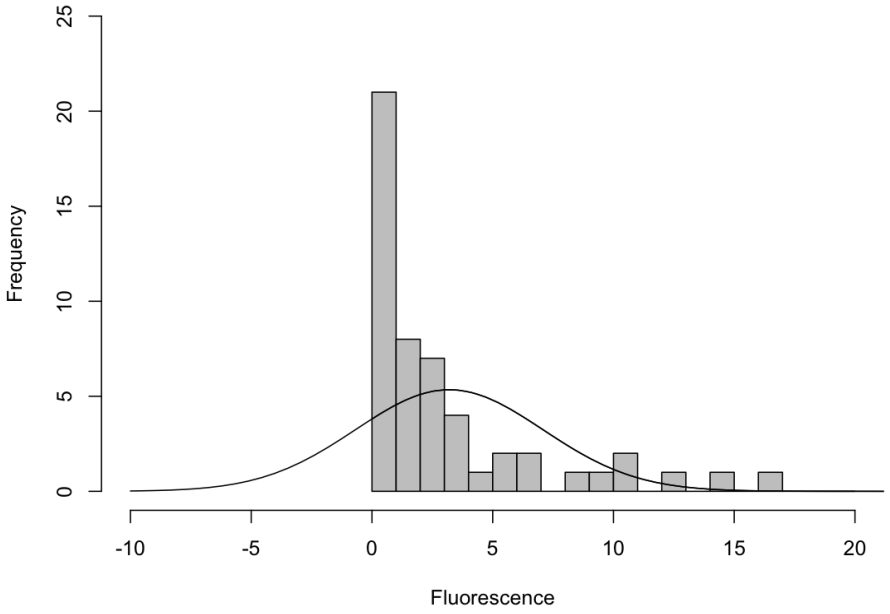

Figure A.2: Histogram of IL-1 stimulated observations that have been binned (grouped) to integer values of fluorescence. The superimposed line represents a Normal distribution, using the mean and standard deviation calculated from the raw data. The mean has been calculated as 3.1987.

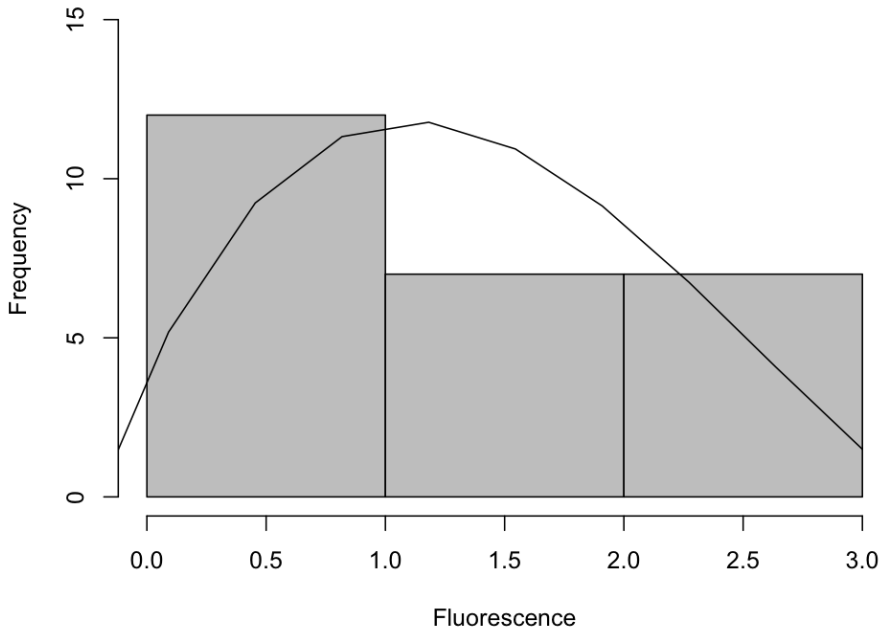

Figure A.3: Histogram of control observations from the partial dataset (with initial fluorescence 0 to 3.0) that have been binned (grouped) to integer values of fluorescence. The superimposed line represents a Normal distribution, using the mean and standard deviation calculated from the raw data. The mean has been calculated as 1.282673.

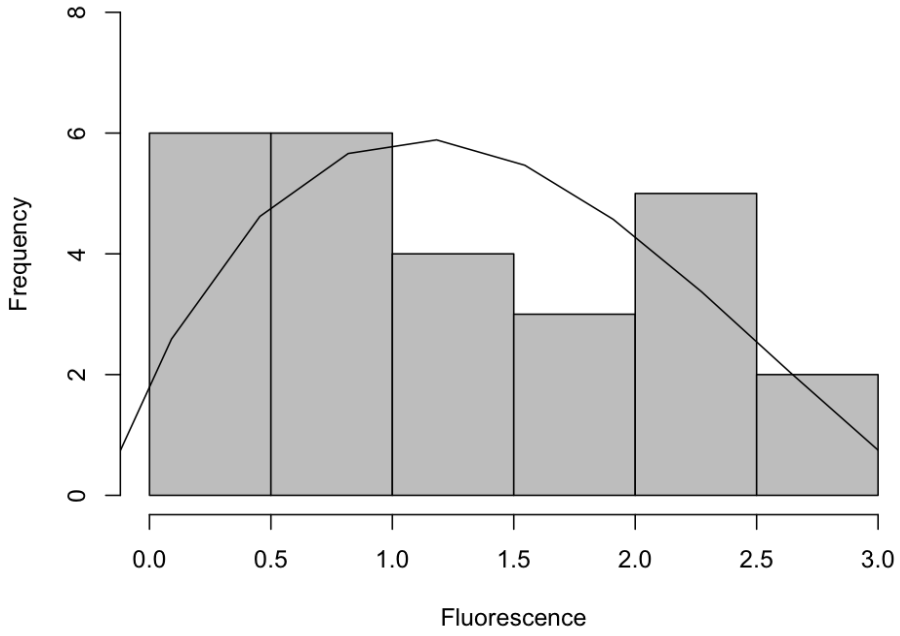

Figure A.4: Histogram of control observations from the partial dataset (with initial fluorescence 0 to 3.0) that have been binned (grouped) using a 0.5 interval of fluorescence. The superimposed line represents a Normal distribution, using the mean and standard deviation calculated from the raw data. The mean has been calculated as 1.282673.

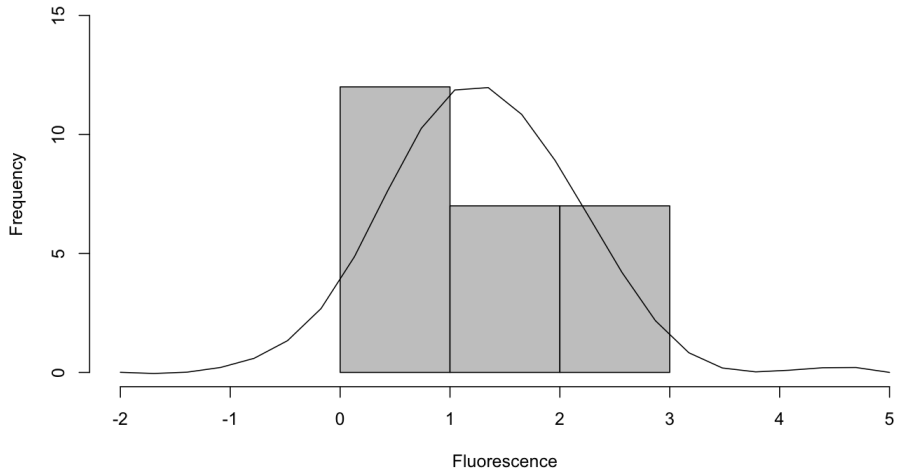

Figure A.5: Histogram of control observations from the partial dataset (with initial fluorescence 0 to 3.0) that have been binned (grouped) to integer values of fluorescence. The superimposed line represents a Normal distribution with a negative fluorescence tail, using the mean and standard deviation calculated from the raw data. The mean has been calculated as 1.282673.

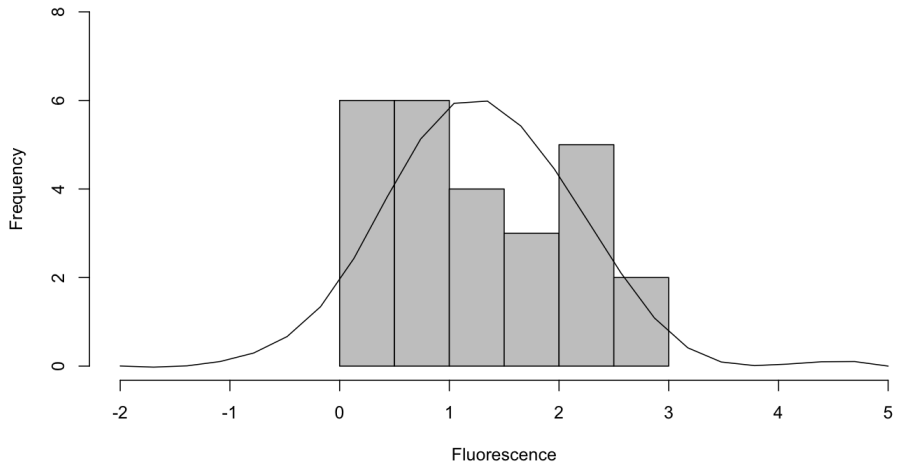

Figure A.6: Histogram of control observations from the partial dataset (with initial fluorescence 0 to 3.0) that have been binned (grouped) using a 0.5 interval of fluorescence. The superimposed line represents a Normal distribution with a negative fluorescence tail, using the mean and standard deviation calculated from the raw data. The mean has been calculated as 1.282673.

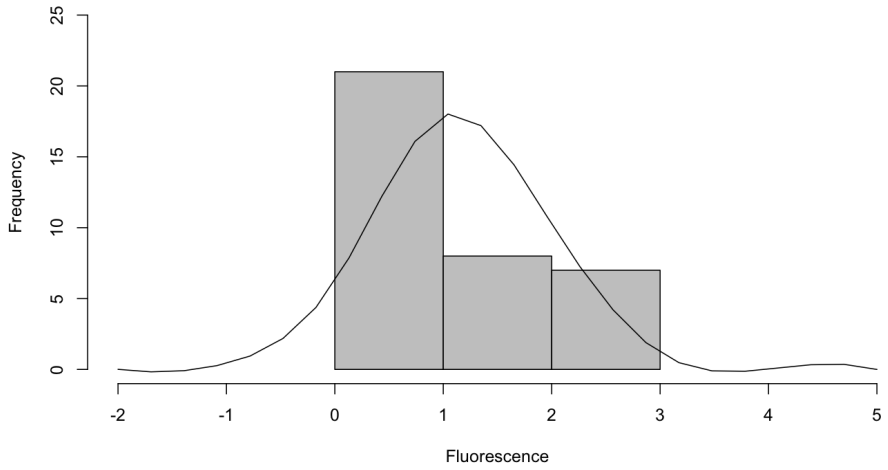

Figure A.7: Histogram of IL-1 stimulated observations from the partial dataset (with initial fluorescence 0 to 3.0) that have been binned (grouped) to integer values of fluorescence. The superimposed line represents a Normal distribution with a negative fluorescence tail, using the mean and standard deviation calculated from the raw data. The mean has been calculated as 1.156009.

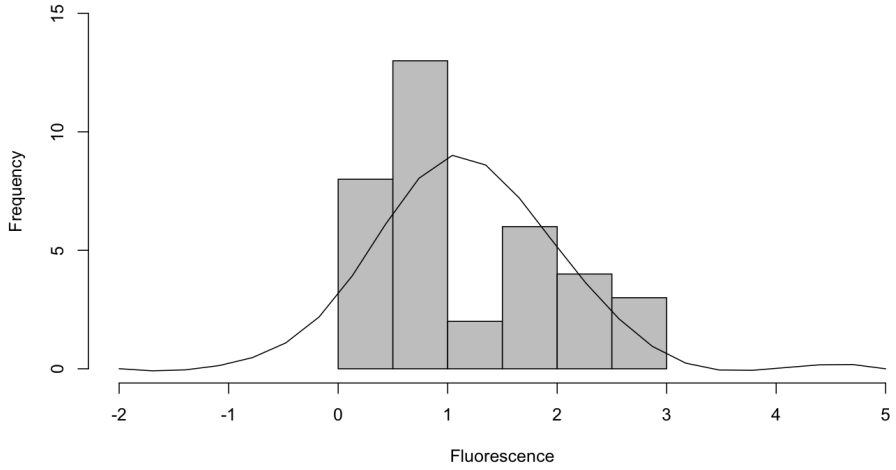

Figure A.8: Histogram of IL-1 stimulated observations from the partial dataset (with initial fluorescence 0 to 3.0) that have been binned (grouped) using a 0.5 interval of fluorescence. The superimposed line represents a Normal distribution with a negative fluorescence tail, using the mean and standard deviation calculated from the raw data. The mean has been calculated as 1.156009.

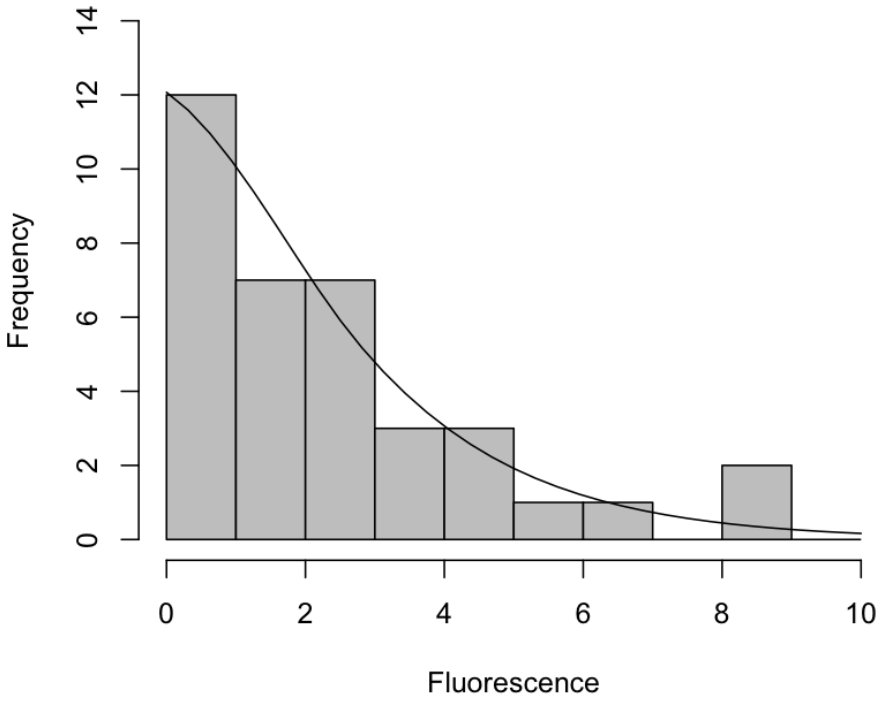

Figure A.9: Histogram of control observations from the full dataset that have been binned (grouped) using an integer interval of fluorescence. The superimposed line represents a negative binomial distribution, using the median calculated from the raw data. The median average has been calculated as 1.947153.

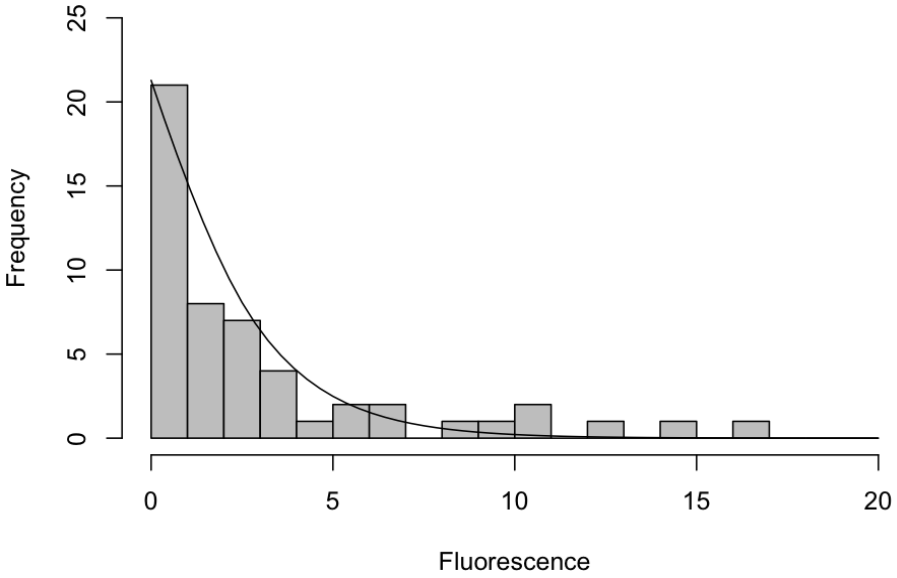

Figure A.10: Histogram of IL-1 stimulated observations from the full dataset that have been binned (grouped) using an integer interval of fluorescence. The superimposed line represents a negative binomial distribution, using the median calculated from the raw data. The median average has been calculated as 1.729876.

Value	Observed (O_i)	Expected (E_i)	O_i-E_i	$(O_i\text{-}E_i)_2$ / E_i
< -2	0	0.83206980	-0.83206980	0.83206980
-2 to -1	0	1.39094496	-1.39094496	1.39094496
-1 to 0	0	2.76233004	-2.76233004	2.76233004
0 - 1	12	4.48308720	7.51691280	12.60380970
1 - 2	7	5.94606600	1.05399400	0.18682990
2 - 3	7	6.44527800	0.55472200	0.04774290
3 - 4	3	5.70973320	-2.70973320	1.28598900
> 4	7	8.43049080	-1.43049080	0.24272650
Total	**36**	**36**	**0**	**19.35244280**

Table A.1: χ^2 test for full dataset control observations approximating to a Normal distribution

Value	Observed (O_i)	Expected (E_i)	O_i-E_i	$(O_i\text{-}E_i)_2$ / E_i
< -5	0	0.89847576	-0.89847576	0.89847576
-5 to -4	0	0.75232248	-0.75232248	0.75232248
-4 to -3	0	1.21064944	-1.21064944	1.21064944
-3 to -2	0	1.82360256	-1.82360256	1.82360256
-2 to -1	0	2.57122216	-2.57122216	2.57122216
-1 to 0	0	3.39350440	-3.39350440	3.39350440
0 - 1	21	4.19232320	16.80767680	67.38459460
1 - 2	8	4.84802760	3.15197240	2.04927260
2 - 3	7	5.24764760	1.75235240	0.58516490
3 - 4	4	5.31702600	-1.31702600	0.32622700
4 - 5	1	5.04279360	-4.04279360	3.24109640
5 - 6	2	4.47685160	-2.47685160	1.37033670
6 - 7	2	3.72026200	-1.72026200	0.79545510
7 - 8	0	2.89383120	-2.89383120	2.89383120
8 - 9	1	2.10702960	-1.10702960	0.58163140
9 - 10	1	1.43604240	-0.43604240	0.13240070
> 10	5	2.06844040	2.93155960	4.1548130
Total	**52**	**52**	**0**	**94.1646287**

Table A.2: χ^2 test for full dataset IL-1 stimulated observations approximating to a Normal distribution

Value	Observed (O_i)	Expected (E_i)	O_i-E_i	$(O_i\text{-}E_i)_2$ / E_i
< 0.5	6	4.5310330	1.4689670	0.4762411
0.5 - 1.0	6	5.0231740	0.9768260	0.1899574
1.0 - 1.5	4	6.1155432	-2.1155432	0.7318276
1.5 - 2.0	3	5.2569530	-2.2569530	0.9689713
> 2.0	7	5.0732968	1.9267032	0.7317106
Total	**26**	**26**	**0**	**3.0987080**

Table A.3: χ^2 test for partial dataset control observations approximating to a Normal distribution

Value	Observed (O_i)	Expected (E_i)	O_i-E_i	$(O_i\text{-}E_i)_2$ / E_i
< -0.5	0	0.42573856	-0.42573856	0.42573856
-0.5 to 0	0	1.19235714	-1.19235714	1.19235714
0 - 0.5	6	2.91293730	3.08706270	3.27159670
0.5 - 1.0	6	5.0231740	0.9768260	0.1899574
1.0 - 1.5	4	6.1155432	-2.1155432	0.7318276
1.5 - 2.0	3	5.2569530	-2.2569530	0.9689713
> 2.0	7	5.0732968	1.9267032	0.7317106
Total	**26**	**26**	**0**	**7.51215930**

Table A.4: χ^2 test for partial dataset control observations approximating to a Normal distribution with negative fluorescence tail

Value	Observed (O_i)	Expected (E_i)	O_i-E_i	$(O_i\text{-}E_i)_2$ / E_i
< -0.5	0	0.63275940	-0.63275940	0.63275940
-0.5 to 0	0	1.91269908	-1.91269908	1.91269908
0 - 0.5	8	4.72651992	3.27348008	2.26713780
0.5 - 1.0	13	7.89639840	5.10360160	3.29856070
1.0 - 1.5	2	8.92127520	-6.92127520	5.36964160
1.5 - 2.0	6	6.81655680	-0.81655680	0.09781550
> 2.0	7	5.09379120	1.90620880	0.71334530
Total	**36**	**36**	**0**	**14.29196**

Table A.5: χ^2 test for partial dataset IL-1 stimulated observations approximating to a normal distribution with negative fluorescence tail.

Value	Observed (O_i)	Expected (E_i)	O_i-E_i	$(O_i\text{-}E_i)_2$ / E_i
< 1	12	11.225772	0.774228	0.0533976
1 - 2	7	8.7013764	-1.7013764	0.3326694
2 - 3	7	5.917482	1.082518	0.1980311
3 - 4	3	3.8369432	-0.8367432	0.1824827
> 4	7	6.3186264	0.6813736	0.0734764
Total	**36**	**36**	**0**	**0.8400572**

Table A.6: χ^2 test for full dataset control observations approximating to a negative binomial distribution.

Value	Observed (O_i)	Expected (E_i)	O_i-E_i	$(O_i\text{-}E_i)_2$ / E_i
< 1	21	15.9764696	5.0235304	1.5795641
1 - 2	8	11.3233276	-3.3233276	0.9753764
2 - 3	7	7.8343252	-0.8343252	0.0888524
3 - 4	4	5.3764620	-1.3764620	0.3523967
4 - 5	1	3.6742108	-2.6742108	1.9463781
5 - 6	2	2.5050012	-0.5050012	0.1018068
6 - 7	2	1.7049656	0.2950344	0.0510540
7 - 8	0	1.1590748	-1.1590748	1.1590748
8 - 9	1	0.7872644	0.2127356	0.0574857
9 - 10	1	0.5343520	0.4656480	0.4057776
> 10	5	1.1245468	3.8754532	13.3557247
Total	**52**	**52**	**0**	**20.0734913**

Table A.7: χ^2 test for full dataset IL-1 stimulated observations approximating to a negative binomial distribution.

B Supporting Material for Statistical Techniques

Statistics does not tell us whether we are right about our interpetations of empirical data, but instead tells us the probability of being wrong. For example, replicates of experiments that are based on stochastic systems (*in vitro*, *in vivo* or *in silico*), will never yield exactly the same results. Instead, the repeated measurements will span a range of values, due to the inherent heterogeneity of the biological system(s) under investigation, and also the precision limits of our measuring equipment (Krzywinski and Altman, 2013). With such uncertainty within the data, it is important to use appropriate statistical techniques in order to capture the role of chance within our experiments, and to assign confidence levels to the experimental data, so that we may determine whether our measurements are compatible with our experimental hypotheses. For the purposes of this thesis, we have chosen to use the Kolmogorov-Smirnov test (Massey, 1951) to investigate the statistical significance of our experimental data against baseline/control data, and the Vargha-Delaney A-Test (Vargha and Delaney, 2000) to investigate the effect magnitude (size difference) between the data. Through this dual-test approach, we are able to confer *scientifically significant* difference on experimental data that has been found to contain significant differences under both tests.

B.1 The Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test (KS-Test) is a non-parametric test of the equality of continuous, one-dimensional probability distributions that can be used to either compare a sample against a reference probability distribution (one-sample KS-Test), or to compare two samples (two-sample KS-Test) against each other. The two-sample KS-Test is considered to be one of the most useful non-parametric methods for comparing two continuous datasets, because it is sensitive to differences in both location and shape of the *cumulative distribution* functions of the two samples.

As the data generated through simulation is continuous (with respect to time) and one-dimensional in nature (as each agent state can be separated out from the data), we have used the two-sample KS-Test that was pre-built into Matlab, in order to investigate whether there is any significant difference between our experimental and baseline/control datasets. This is performed through generation of a cumulative distribution for each dataset of interest (i.e. different cumulative distributions for each agent type), and testing whether the underlying one-dimensional probability distributions differ between experimental conditions and baseline/control conditions. The KS-Test generates a p-value, and we have taken a p-value <0.05 to indicate a statistically significant difference between the two distributions of experimental data.

B.2 The Vargha-Delaney A-Test

The Vargha-Delaney A-Test, is a non-parametric test that indicates the significance of any differences in size (termed effect magnitude) between two populations of data (Vargha and Delaney, 2000). Through comparison of the two distributions (A and B), for example by comparing an experimental dataset against a reference (baseline or control sample) dataset, the A-Test investigates whether a randomly chosen data point from distribution A is larger than a randomly chosen data point from distribution B. The resulting A-Test score, has a value between 0.0 and 1.0, and indicates the probability that the two distributions of experimental data are taken from different populations. Vargha and Delaney (2000) suggest a number of threshold values (for the A-Test scores) to indicate significant differences between the two distributions (see table B.1). A value of 0.5 is used to indicate no difference between the distributions, whereas values < 0.29 and > 0.71 are used to indicate *large* differences.

Difference	**Large**	**Medium**	**Small**	**None**	**Small**	**Medium**	**Large**
A-Test Score	<0.29	<0.36	<0.44	0.50	>0.56	>0.64	>0.71

Table B.1: The effect magnitude (A-Test score) thresholds, as indicated by Vargha and Delaney (2000). We take the *large* difference, with A-Test scores <0.29 and >0.71 for samples that have already yielded a statistically significant difference from the KS-Test, to indicate scientifically significant differences.

The following Matlab code is used to generate the Vargha-Delaney A-Test score. It was generously provided by Dr Mark Read, who was a fellow researcher within the York Computational Immunology Lab. A caveat needs to be applied, in that the code only works if the two distributions being compared contain exactly the same number of data points.

$$
\begin{aligned}
functionA &= Atest(X, Y) \\
[p, h, st] &= ranksum(X, Y, \text{`}alpha', 0.05); \\
N &= size(X, 1); \\
M &= size(Y, 1); \\
A &= (st.ranksum/N - (N+1)/2)/M;
\end{aligned}
$$

C Example 0.XML Parameters File


```
<states>
<itno>0</itno>
<environment>
  <simulation_seed>1</simulation_seed>
  <stimulation_status>1</stimulation_status>
  <recdelay_constant>10</recdelay_constant>
  <biomolecule_speed_ave>20</biomolecule_speed_ave>
  <biomolecule_speed_range>10</biomolecule_speed_range>
  <biomolecule_angle_range>0.314159</biomolecule_angle_range>
  <nuclear_receptor_speed_ave>20</nuclear_receptor_speed_ave>
  <nuclear_receptor_speed_range>5</nuclear_receptor_speed_range>
  <nuclear_receptor_angle_range>0.314159</nuclear_receptor_angle_range>
  <dt>2</dt>
  <nuclear_radius>3750</nuclear_radius>
  <cell_radius>10000</cell_radius>
  <basal_dissociation_prob>0.000003</basal_dissociation_prob>
  <binding_prob_biomolecules>0.65</binding_prob_biomolecules>
  <binding_prob_importreceptor>0.03</binding_prob_importreceptor>
  <binding_prob_exportreceptor>0.5</binding_prob_exportreceptor>
  <binding_prob_ikk>0.75</binding_prob_ikk>
  <binding_prob_il_importreceptor>0.85</binding_prob_il_importreceptor>
  <binding_prob_il_exportreceptor>0.5</binding_prob_il_exportreceptor>
  <default_ikk_rebinddelay>10</default_ikk_rebinddelay>
  <rebind_delay>3</rebind_delay>
</environment>
<xagent>
  <name>NFkB</name>
  <id>1</id>
  <postheta>0.519620</postheta>
  <posphi>-1.164042</posphi>
  <posr>4992.761417</posr>
  <posx>1714.567424</posx>
  <posy>980.830804</posy>
  <posz>-4585.400315</posz>
  <movetheta>-1.928947</movetheta>
  <movephi>-0.748157</movephi>
  <mover>10.174385</mover>
  <state>1</state>
  <ikk_id>0</ikk_id>
  <ikk_boundcounter>5</ikk_boundcounter>
  <iradius>100</iradius>
```

```
  <ikba_id>560</ikba_id>
</xagent>

  .... Additional NF-κB Agents ....

<xagent>
  <name>IkBa</name>
  <id>500</id>
  <postheta>1.165670</postheta>
  <posphi>1.108712</posphi>
  <posr>9685.570796</posr>
  <posx>1701.861737</posx>
  <posy>3968.444144</posy>
  <posz>8669.798114</posz>
  <movetheta>-0.270213</movetheta>
  <movephi>-6.707881</movephi>
  <mover>24.655802</mover>
  <state>100</state>
  <iradius>100</iradius>
  <rebinddelay>0</rebinddelay>
</xagent>

  .... Additional IκBα Agents ....

<xagent>
  <name>IKK</name>
  <id>2000</id>
  <postheta>0.612705</postheta>
  <posphi>0.072128</posphi>
  <posr>6235.363700</posr>
  <posx>5087.860270</posx>
  <posy>3576.523137</posy>
  <posz>449.355767</posz>
  <movetheta>9.514695</movetheta>
  <movephi>2.001282</movephi>
  <mover>10.900060</mover>
  <state>200</state>
  <iradius>100</iradius>
  <stimulationdelay>500</stimulationdelay>
  <ikkrebinddelay>10</ikkrebinddelay>
</xagent>

  .... Additional IKK Agents ....

<xagent>
  <name>Nuclear_Receptor</name>
  <id>1000</id>
  <postheta>183.606789</postheta>
```

```
   <posphi>-180.958286</posphi>
   <posr>3750.000000</posr>
   <posx>204.929407</posx>
   <posy>1149.782568</posy>
   <posz>3563.496034</posz>
   <movetheta>0.000096</movetheta>
   <movephi>0.000024</movephi>
   <mover>0.000000</mover>
   <state>400</state>
   <recdelay>0</recdelay>
   <boundindex>587</boundindex>
   <iradius>100</iradius>
</xagent>

   .... Additional Nuclear Receptor Agents ....

<xagent>
   <name>IL1R</name>
   <id>2501</id>
   <postheta>186.459726</postheta>
   <posphi>-183.881180</posphi>
   <posr>10000.000000</posr>
   <posx>438.828713</posx>
   <posy>874.617350</posy>
   <posz>-9952.008534</posz>
   <movetheta>-0.000116</movetheta>
   <movephi>-0.000089</movephi>
   <state>300</state>
   <stimulation_delay>250</stimulation_delay>
</xagent>

   .... Additional IL1 Cell Membrane Receptor Agents ....

</states>
```

Abbreviations

ABMS	Agent-Based Modelling and Simulation
API	Application Programming Interface
CoSMoS	Complex Systems Modelling and Simulation
DNA	Deoxyribonucleic Acid
EAE	Experimental Autoimmune Encephalomyelitis
EGFP	Enhanced Green Fluorescent Protein
FLAME	Flexible Large-scale Agent-based Modelling Environment
GTP	Guanosine 5'-Triphosphate
HIV	Human Immunodeficiency Virus
IκB	Inhibitor of Nuclear Factor-kappa B
IκBα	Inhibitor of Nuclear Factor-kappa B alpha
IκBβ	Inhibitor of Nuclear Factor-kappa B beta
IκBδ	Inhibitor of Nuclear Factor-kappa B delta
IκBϵ	Inhibitor of Nuclear Factor-kappa B epsilon
IKK	Inhibitor of Nuclear Factor-kappa B Kinase
IL-1	Interleukin-1
IRAK	Interleukin-1 Receptor-Associated Kinase
LPS	Lipopolysaccharide
mRNA	Messenger Ribonucleic Acid
NEMO	NF-κB Essential Modulator
NF-κB	Nuclear Factor-kappa B
ODE	Ordinary Differential Equation
PC	Principal Component
PCA	Principal Component Analysis
PDE	Partial Differential Equation
RNA	Ribonucleic Acid
SBGN	Systems Biology Graphical Notation
SDE	Stochastic Differential Equation
TAK	Transforming Growth Factor-β-Activated Protein Kinase
TILRR	Toll-like/IL-1 Receptor Regulator
TLR	Toll-Like Receptor
TNF	Tumour Necrosis Factor
TRAF	Tumour Necrosis Factor Receptor-Associated Factor
UML	Unified Modelling Language

Glossary

Adaptive Immune System The second line of defence within the immune system, which unlike the innate immune system that acts generically, provides a specific immune response. The cells and tissues of the adaptive immune system are capable of recognising individual pathogens, and are able to provide a targeted response. Furthermore, following the successful defence against a pathogen, cells within the adaptive immune system that provided specificity to the pathogen are maintained, in order to provide a *memory* of the pathogenic attack, and thus provide long lasting immunity against the pathogen.

Agent-Based Modelling A computational modelling approach, in which each individual component (such as individual biologocal molecules) are explicitly represented. The emergent behaviour that is generated through agent-based simulations is the result of prespecified rules, that define the possible interactions between components and state changes of components.

Cytokine A broad category of small-protein signalling molecules secreted by cells. They generally migrate away from the secreting cell to act on other cells, however in some circumstances, may remain in the vicinity of the secreting cell and act upon it, thus forming a feedback loop.

Domain The real-world system of interest that is the subject of our investigation, which in this case is the IL-1 stimulated NF-κB Signalling Pathway.

Domain Model	The abstracted view of our understanding of the domain. The concept of the domain model was defined within the CoSMoS approach, and may represent both the current view of the underlying real-world system, and our hypotheses on the mechanistic interactions and processes within the system. The CoSMoS approach does not specify what techniques should be used for developing the domain model, and therefore the modeller and subject matter expert(s) are free to utilise the best tools available, such as diagrammatic modelling using cartoon and UML notations, statistical modelling, and kinetic modelling using the various types of equations.
Epithelial Cell	A cell of the epithelium, which is a tissue that lines the surfaces of structures throughout the body, and also forms many glands. Epithelial cells have many functions, including the secretion and selective absorption of extracellular signals, protection against infection, and transcellular transport.
Fibroblast Cell	A cell common to the connective tissue in animals, due to their synthesis of extracellular matrix and collagen - the structural framework for animal tissues. They also play a critical role in wound healing.
IL-1	Interleukin-1. A family of cytokines, involved in the regulation of immune and inflammatory responses due to infection. IL-1 is produced by a large number of cells related to the immune response, such as tissue macrophages, monocytes, fibroblasts, and dendritic cells, but is also produced by epithelial cells.
Immune System	The term given to the internal system that ensures organisms can defend themselves against attack from bacteria, virus' and toxic chemicals. The immune system comprises specific cells, tissues and organs, that produce immune response molecules for the maintainence of a healthy state for the organism as a whole.
Inflammatory Response	The inflammatory response is initiated by the innate immune system, and is a protective tissue response to injury or destruction of tissue-cells. Its aim is to destroy, dilute, or quarantine both the pathogenic/toxic agent and the injured tissues.

Innate Immune System The first line of defence within the immune system, which acts through a system-wide response, and thus non-specific to the pathogenic threat. The cells of the innate immune system recognise and respond to pathogenic threats in a generic way, and therefore unlike the adaptive immune system, it does not confer long-lasting or protective immunity to the host.

In Silico An approach that uses computers to perform experimentation (i.e. in silicon chip), rather than the more traditional laboratory-based experimental approaches, which are performed in glass or in living organisms.

In Vitro An approach that performs experimentation outside of a living organism, although it may utilise cells and tissues harvested from a living organism. Such experiments are traditionally performed in a laboratory environment, and may be confined to a test tube, petri dish, or other suitable container.

In Vivo An approach that performs experimentation within a living organism.

Kolmogorov-Smirnov Test A non-parametric statistical test, which is used on continuous, one-dimensional distributions. It can be used to compare a sample with a reference probability distribution (one-sample KS test), which was the application for our work, or to compare two samples (two-sample KS test).

Membrane Receptor A protein complex located in a cellular membrane, such as the cell surface membrane, which binds signal molecules (e.g. cytokine), and once activated may propagate the signal from the outside of the cell to the inside of the cell for propagation of the signal.

Model An abstracted view of a real-world system, theory or phenomenon. There are many different types of models, however for the purposes of this thesis, they may be thought of as conceptual, computational, mathematical and statistical.

ODE Abbreviation for Ordinary Differential Equation. This mathematical approach is usually used to investigate the temporal dynamics of population-level phenomena, using a series of linked differential equations.

Pathogen An infectious agent, such as a virus, bacterium, or parasite, that may cause disease in the host.

Platform Model The technical specification from a CoSMoS project, which translates the functional requirements of the system (as documented in the domain model), into a software platform specific design, for subsequent coding into the actual computational model (the simulation platform).

Sensitivity Analysis The investigation using statistical techniques, of how a system (e.g. a computational model) responds to perturbations of input parameter values away from their calibrated values.

Simulation An instantiation (i.e. exeception with specific parameter values) of a computational model.

TNFα Tumour Necrosis Factor alpha. A type of cytokine, secreted by a broad range of cells, such as cells from the innate and adaptive immune system, along with epithelial and fibroblast cells. It primarily acts to regulate the action of immune cells, by inducing apoptotic cell death, inflammation, and the inhibition of viral replication. It may also propagate the immune response through inducing the production of further cytokines, such as IL-1.

UML The Unified Modelling Language. A diagrammatic notation, which is maintained and administered by the Object Management Group. Although originally developed for software engineering, where it was used to define the functional, technical and process related aspects of computer systems, it has recently been applied within the complex systems analysis community.

Vargha-Delaney A-Test A non-parametric statistical technique to test the effect magnitude of the difference between two continuous, one dimensional distributions. On its own, the test does not provide an indication of significant difference, however when used on distributions that have been shown to contain statistically significant differences (through for example p-values < 0.05 from the KS-Test), a large effect magnitude is deemed indicative of a scientifically significant difference.

Bibliography

Acerbi, E; Decraene, J, and Gouaillard, A. Computational reconstruction of biochemical networks. In *15th International Conference on Information Fusion*, pages 1134–1141. IEEE, 2012.

Ahn, A C; Tewari, M; Poon, C S, and Phillips, R S. The clinical applications of a systems approach. *PLoS Medicine*, 3:e209, 2006.

Akira, S and Takeda, K. Toll-like receptor signalling. *Nature Reviews Immunology*, 4: 499–511, 2004.

Aksenov, S V; Church, B; Dhiman, A; Georgieva, A; Sarangapani, R; Helminger, G, and Khalil, I G. An integrated approach for inference and mechanistic modelling for advancing drug development. *FEBS Letters*, 579:1878–1883, 2005.

Albert, R. Scale-free networks in cell biology. *Journal of Cell Science*, 118:4947–4957, 2005.

Albert, R; Jeong, H, and Barabasi, A L. Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000.

Alberts, B; Bray, D; Lewis, J; Raff, M; Roberts, K, and Watson, J D. *Molecular Biology of the Cell*. Garland Publishing Inc, 1994.

Alden, K; Timmis, J; Andrews, P S; Veiga-Fernandes, H, and Coles, M. Pairing experimentation and computational modelling to understand the role of tissue inducer cells in the development of lymphoid organs. *Frontiers in Inflammation*, 3:172, 2012.

Alden, K; Read, M; Timmis, J; Andrews, P S; Veiga-Fernandes, H, and Coles, M C. Spartan: A comprehensive tool for understanding uncertainty in simulations of biological systems. *PLoS Computational Biology*, 9(2):e1002916, 2013.

Alkalay, I; Yaron, A; Hatzubai, A; Jung, S; Avraham, A; Gerlitz, O; Pashut-Lavon, I, and Ben-Neriah, Y. In vivo stimulation of IκB phosphorylation is not sufficient to activate NF-κB. *Molecular and Cellular Biology*, 15:1294–1301, 1995a.

Alkalay, I; Yaron, A; Hatzubai, A; Orian, A; Ciechanover, A, and Ben-Neriah, Y. Stimulation-dependent IκBα phosphorylation marks the NF-κB inhibitor for degradation via the ubiquitin-proteasome pathway. *Proceedings of the National Academy of Sciences*, 92:10599–10603, 1995b.

Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman and Hall/CRC, Boca Raton, Florida, 2006.

Altman, R B. Bioinformatics in support of molecular medicine. In *Proceedings of the American Medical Informatics Association Symposium*, pages 53–61, 1998.

Alves, R; Antunes, F, and Salvador, A. Tools for kinetic modelling of biochemical networks. *Nature Biotechnology*, 24:667–672, 2006.

An, G. Introduction of an agent-based multi-scale modular architecture for dynamic knowledge representation of acute inflammation. *Theoretical Biology and Medical Modelling*, 5:11, 2008.

Anderson, K V and Nusslein-Volhard, C. Information for the dorsal-ventral pattern of Drosophila embryo is stored as maternal mRNA. *Nature*, 311:223–227, 1984.

Andrews, P S; Polack, F A C; Sampson, A T; Timmis, J; Scott, L, and Coles, M C. Simulating biology: Towards understanding what the simulation shows. In Stepney, S; Polack, F A C, and Welch, P, editors, *Proceedings of the 2008 Workshop on Complex Systems Modelling and Simulation*, pages 93–123. Luniver Press, 2008.

Andrews, P S; Polack, F A C; Sampson, A T; Stepney, S, and Timmis, J. The CoSMoS process, version 0.1: A process for the modelling and simulation of complex systems. Technical Report YCS-2010-453, University of York, 2010.

Andrews, P S; Stepney, S; Hoverd, T; Polack, F A C; Sampson, A T, and Timmis, J. CoSMoS process, models, and metamodels. In Stepney, S; Welch, P H; Andrews, P S, and Ritson, C G, editors, *Proceedings of the 2011 Workshop on Complex Systems Modelling and Simulation*, CoSMoS, pages 1–14. Luniver Press, 2011.

Andrews, S S and Bray, D. Stochastic simulation of chemical reactions with spacial resolution and single molecule detail. *Physical Biology*, 1:137–151, 2004.

Arenzana-Seisdedos, F; Thompson, J; Rodriguez, M S; Bachelerie, F; Thomas, D, and Hay, R T. Inducible nuclear expression of newly synthesised IκBα negatively regulates DNA-binding and transcriptional activities of NF-κB. *Molecular and Cellular Biology*, 15:2689–2696, 1995.

Arlow, J and Neustadt, I. *UML2 and The Unified Process: Practical Object-Oriented Analysis and Design*. Addison Wesley, New Jersey, USA, 2nd edition, 2009.

Ashall, L; Horton, C A; Nelson, D E; Paszek, P; Harper, C V; Sillitoe, K; Ryan, S; Spiller, D G; Unitt, J F; Broomhead, D S; Kell, D B; Rand, D A; See, V, and White, M R H. Pulsatile stimulation determines timing and specificity of NF-κB-dependent transcription. *Science*, 324:242–246, 2009.

Babaoglu, O; Canright, G; Deutsch, A; Caro, G A Di; Ducatelle, F; Gambardella, L M; Ganguly, N; Jelasity, M; Montemanni, R; Montresor, A, and Urnes, T. Design patterns from biology for distributed computing. *ACM Transactions on Autonomous Adaptive Systems*, 1:26–66, 2006.

Baeuerle, P A and Baltimore, D. IκB a specific inhibitor of the NF-κB transcription factor. *Science*, 242:540–546, 1988a.

Baeuerle, P A and Baltimore, D. Activation of DNA-binding activity in an apparently cytoplasmic precursor of the NF-κB transcription factor. *Cell*, 53(211-217), 1988b.

Baeuerle, P A and Henkel, T. Function and activation of NF-κB in the immune system. *Annual Review of Immunology*, 12:141–179, 1994.

Bagnoli, F; Lio, P, and Sguanci, L. Modeling viral coevolution: HIV multi-clonal persistence and competition dynamics. *Physica Acta*, 366:333–346, 2006.

Baichwal, V R and Baeuerle, P A. Apoptosis: Activate NF-κB or die? *Current Biology*, 7:R94–R96, 1997.

Balanescu, T; Cowling, A J; Georgescu, H; Gheorghe, M; Holcombe, M, and Vertan, C. Communicating stream X-machine systems are no more than X-machines. *Journal of Universal Computer Science*, 5(9):494–507, 1999.

Baldwin, A S. The NF-κB and IκB proteins: New discoveries and insights. *Annual Review of Immunology*, 14:649–681, 1996.

Banga, J R. Optimization in computational systems biology. *BMC Systems Biology*, 2:47, 2008.

Barabasi, A L. *Linked*. Plume, New York, 2003.

Barabasi, A L. Scale-free networks: A decade and beyond. *Science*, 325:412–413, 2009.

Barabasi, A L and Bonabeau, E. Scale-free networks. *Scientific American*, 288:50–59, 2003.

Barabasi, A L and Oltvai, Z N. Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5:101–113, 2004.

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

Barken, D; Wang, C J; Kearns, J; Cheong, R; Hoffmann, A, and Levchenko, A. Comment of"oscillations in NF-κB signaling control the dynamics of gene expression". *Science*, 308:52, 2005.

Barnard, J; Whitworth, J, and Woodward, M. Communicating X-machines. *Information and Software Technology*, 38(6):401–407, 1996.

Barry, S C. How much impact does the choice of a random number generator really have? *International Journal of Geographical Information Science*, 25(4):523–530, 2011.

Basak, S; Kim, H; Kearns, J D; Tergaonkar, V; O'Dea, E; Werner, S L; Benedict, C A; Ware, C F; Ghosh, G; Verma, I M, and Hoffmann, A. A fourth IkappaB protein within the NF-kappaB signaling module. *Cell*, 128:369–381, 2007.

Bassingthwaighte, J B. Strategies for the physiome project. *Annals of Biomedical Engineering*, 28(8):1043–1058, 2000.

Bauer, A L; Beauchemin, C A A, and Perelson, A S. Agent-based modeling of host-pathogen systems: The successes and challenges. *Information Sciences*, 179:1379–1389, 2009.

Baxter, S M; Day, S W; Fetrow, J S, and Reisinger, S J. Scientific software development is not an oxymoron. *PLoS Computational Biology*, 2:e87, 2006.

Beard, D A; Babson, E; Curtis, E, and Qian, H. Thermodynamic constraints for biochemical networks. *Journal of Theoretical Biology*, 228:327–333, 2004.

Beg, AA; Ruben, S M; Scheinman, R I; Haskill, S; Rosen, C A, and Baldwin, A S. IκB interacts with the nuclear localization sequences of the subunits of NF-κB: A mechanism for cytoplasmic retention. *Genes & Development*, 6:1899–1913, 1992.

Behn, U. Idiotypic networks: Toward a renaissance? *Immunological Reviews*, 216: 142–152, 2007.

Bender, K; Gottlicher, M; Whiteside, S; Rahmsdorf, H, and Herrlich, P. Sequential DNA damage-independent and -dependent activation of NF-κB by UV. *The EMBO Journal*, 17:5170–5181, 1998.

Benoist, C; Germain, R N, and Mathis, D. A plaidoyer for 'systems immunology'. *Immunological Reviews*, 210:229–234, 2006.

Bernaschi, M and Castiglione, F. Design and implementation of an immune system simulator. *Computers in Biology and Medicine*, 31:303–331, 2001.

Bersini, H. UML for ABM. *Journal of Artificial Societies and Social Simulation*, 15:9, 2012.

Bersini, H; Klatzmann, D; Six, A, and Thomas-Vaslin, V. State-transition diagrams for biologists. *PLoS ONE*, 7:e41165, 2012.

Bhalla, U S and Iyengar, R. Emergent properties of networks of biological signaling pathways. *Science*, 283:381–388, 1999.

Bivona, T G; Hieronymus, H; Parker, J; Chang, K; Taron, M; Rosell, R; Moonsamy, P; Dahlman, K; Miller, V A; Costa, C; Hannon, G, and Sawyers, C L. FAS and NF-κB signalling modulate dependence of lung cancers on mutant EGFR. *Nature*, 471:523–526, 2011.

Blair, R H; Kliebstein, D J, and Churchill, G A. What can causal networks tell us about metabolic pathways? *PLoS Computational Biology*, 8:e1002458, 2012.

Bliss, C I and Fisher, R A. Fitting the negative binomial distribution to biological data. *Biometrics*, 9(2):176–200, 1953.

Boehm, U; Klamp, T; Groot, M, and Howard, J C. Cellular responses to interferon-g. *Annual Review of Immunology*, 15:749–795, 1997.

Bonabeau, E. Agent-based modelling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(s3):7280–7287, 2002.

Bonetti, B; Stegagno, C; Cannella, B; Rizzuto, N; Moretto, G, and Raines, C S. Activation of NF-κB and c-jun transcription factors in multiple sclerosis lesions: Implications for oligodendrocyte pathology. *American Journal of Pathology*, 155:1433–1438, 1999.

Booch, G; Rumbaugh, J, and Jacobson, I. *The Unified Modeling Language User Guide*. Addison Wesley Longman, Reading, Massachusetts, 1998.

Boulding, K E. General systems theory - the skeleton of science. *Management Science*, 2:197–208, 1956.

Bown, J; Andrews, P S; Deeni, Y; Goltsov, A; Idowu, M; Polack, F A C; Sampson, A T; Shovman, M, and Stepney, S. Engineering simulations for cancer systems biology. *Current Drug Targets*, 13(12):1560–1574, 2012.

Brasier, A R. The NF-κB regulatory network. *Cardiovascular Toxicology*, 6:111–130, 2006.

Bridges, C C. Hierarchical cluster analysis. *Psychological Reports*, 18:851–854, 1966.

Brown, K; Gerstberger, S; Carlson, L; Franzoso, G, and Siebenlist, U. Control of IκBα proteolysis by site-specific, signal-induced phosphorylation. *Science*, 267:1485–1488, 1995.

Bruggeman, F J; Westerhoff, H V; Hoek, J B, and Kholodenko, B N. Modular response analysis of cellular regulatory networks. *Journal of Theoretical Biology*, 218:507–520, 2002.

Bubici, C; Papa, S; Dean, K, and Franzoso, G. Mutual cross-talk between reactice oxygen species and nuclear factor-kappa B: Molecular basis and biological basis. *Oncogene*, 25:6731–6748, 2006.

Burns, K; Clatworthy, J; Martin, L; Martinon, F; Plumpton, C; Maschera, B; Lewis, A; Ray, K; Tschopp, J, and Volpe, F. Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor. *Nature Cell Biology*, 2:346–351, 2000.

Butcher, E C; Berg, E L, and Kunkel, E J. Systems biology in drug discovery. *Nature Biotechnology*, 22:1253–1259, 2004.

Cao, Z; Henzel, W J, and Gao, X. IRAK: A kinase associated with the interleukin-1 receptor. *Science*, 271:1128–1131, 1996a.

Cao, Z; Xiong, J; Takeuchi, M; Kurama, T, and Goeddel, D V. TRAF6 is a signal transducer for interelukin-1. *Nature*, 383:443–446, 1996b.

Carlotti, F; Chapman, R; Dower, S K, and Qwarnstrom, E E. Activation of nuclear factor κB in single living cells. *Journal of Biological Chemistry*, 274:37941–37949, 1999.

Carlotti, F; Dower, S K, and Qwarnstrom, E E. Dynamic shuttling of nuclear factor κB between the nucleus and cytoplasm as a consequence of inhibitor dissociation. *Journal of Biological Chemistry*, 275:41028–41034, 2000.

Carlson, J M and Doyle, J C. Highly optimized tolerance: A mechanism for power laws in designed systems. *Physical Review E*, 60(2):1412–1427, 1999.

Carlson, J M and Doyle, J C. Highly optimized tolerance: Robustness and design in complex systems. *Physical Review Letters*, 84(11):2529–2532, 2000.

Carlson, J M and Doyle, J C. Complexity and robustness. *Proceedings of the National Academy of Sciences*, 99 (suppl 1):2538–2545, 2002.

Caron, E; Ghosh, S; Matsuoka, Y; Ashton-Beaucage, D; Therrien, M; Lemieux, S; Perreault, C; Roux, P P, and Kitano, H. A comprehensive map of the mTOR signaling network. *Molecular Systems Biology*, 6:453, 2010.

Chakraborty, A K and Das, J. Pairing computation with experimentation: A powerful coupling for understanding T cell signalling. *Nature Reviews Immunology*, 10:59–71, 2010.

Chakraborty, A K; Dustin, M L, and Shaw, A S. In silico models for cellular and molecular immunology: Successes, promises and challenges. *Nature Immunology*, 4: 933–936, 2003.

Chaudron, M R V; Heijstek, W, and Nugroho, A. How effective is UML modelling? an empirical perspective on costs and benefits. *Software and Systems Modelling*, 12: 571–580, 2012.

Chen, B S; Yang, S K; Lan, C Y, and Chuang, Y J. A systems biology approach to construct the gene regulatory network of systemic inflammation via microarray and databases mining. *BMC Medical Genomics*, 1:46, 2008.

Chen, Z J; Parent, L, and Maniatis, T. Site-specific phosphorylation of IκBα by a novel ubiquitination-dependent protein kinase activity. *Cell*, 84:853–862, 1996.

Cheong, R; Bergmann, A; Werner, S L; Regal, J; Hoffmann, A, and Levchenko, A. Transient IκB kinase activity mediates temporal NF-κB dynamics in response to wide range of tumour necrosis factor-α doses. *Journal of Biological Chemistry*, 281: 2945–2950, 2006.

Cheong, R; Hoffmann, A, and Levchenko, A. Understanding NF-κB signalling via mathematical modeling. *Molecular Systems Biology*, 4:192, 2008.

Cho, K H and Wolkenhauer, O. Analysis and modelling of signal transduction pathways in systems biology. *Biochemical Society Transactions*, 31:1503–1509, 2003.

Choudhary, S; Kalita, M; Fang, L; Patel, K; Tian, B; Zhao, Y; Edeh, C B, and Brasier, A R. Inducible TNF receptor associated factor-1 expression coules the canonical to the non-canonical NF-κB pathway in TNF stimulation. *Journal of Biological Chemistry*, 288:14612–14623, 2013.

Coakley, S. *Formal Software Architecture for Agent-Based Modelling in Biology*. PhD thesis, University of Sheffield, 2007.

Coakley, S and Kiran, M. *FLAME User Manual*. University of Sheffield, Sheffield, UK, October 2009.

Coakley, S; Smallwood, R, and Holcombe, M. From molecules to insect communities - how formal agent-based computational modelling is uncovering new biological facts. *Scientiae Mathematicae Japonicae Online*, e-2006:765–778, 2006a.

Coakley, S; Smallwood, R, and Holcombe, M. Using X-Machines as a formal basis for describing agents in agent-based modelling. *Simulation Series*, 38(2):33–40, 2006b.

Coakley, S; Gheorghe, M; Holcombe, M; Chin, S; Worth, D, and Greenough, C. Exploitation of high-performance computing in the FLAME agent-based simulation framework. In *14th IEEE International Conference on High-Performance Computing and Communications (HPCC)*, Liverpool, UK, 2012. IEEE Computer Society.

Cohen, I R. Modelling immune behaviour for experimentalists. *Immunological Reviews*, 216:232–236, 2007.

Cohn, M and Langman, R E. The protection: The unit of humoral immunity selected by evolution. *Immunological Reviews*, 115(1):7–147, 1990.

Cohn, M and Mata, J. Quantitative modeling of immune responses. *Immunological Reviews*, 216:5–8, 2007.

Cook, S. Looking back at UML. *Software and Systems Modelling*, 11:471–480, 2012.

Corradini, F; Merelli, E, and Vita, M. A multi-agent system for modelling carbohydrate oxidation in cell. In Gervasi, O; Gavrilova, M L; Kumar, V; Lagana, A; Lee, H P; Mun, Y; Taniar, D, and Tan, C J K, editors, *International Conference on Computational Science and its Applications*, volume 3481 of *Lecture Notes in Computer Science*, pages 1264–1273, Singapore, 2005. Springer.

Creighton, T E. *Encyclopedia of Molecular Biology*, volume 1-4. John Wiley & Sons Inc, New York, 1999.

Csete, M E and Doyle, J C. Reverse engineering of biological complexity. *Science*, 295: 1664–1669, 2002.

Csete, M E and Doyle, J C. Bow ties, metabolism and disease. *Trends in Biotechnology*, 22(9):446–450, 2004.

Cuthbert, D. One-at-a-time plans. *Journal of the American Statistical Association*, 68 (342):353–360, 1973.

Dancik, G M; Jones, D E, and Dorman, K S. Parameter estimation and sensitivity analysis in an agent-based model of Leishmania major infection. *Journal of Theoretical Biology*, 262:398–412, 2010.

DiDonato, J A; Mercurio, F, and Karin, M. Phosphorylation of IκBα precedes but is not sufficient for its dissociation from NF-κB. *Molecular and Cellular Biology*, 15: 1302–1311, 1995.

DiDonato, J A; Hayakawa, M; Rothwarf, D M; Zandi, E, and Karin, M. A cytokine-responsive IκB kinase that activates the transcription factor NF-κB. *Nature*, 388: 548–564, 1997.

Diestel, R. *Graph Theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

dos Santos, R M Z and Coutinho, S. Dynamics of HIV infection: A cellular automata approach. *Physical Reviews Letters*, 87:168102–1–168102–4, 2001.

Downward, J. The ins and outs of signalling. *Nature*, 411:759–762, 2001.

Doyle, S L and O'Neill, L A J. Toll-like receptors: From the discovery of NF-κB to new insights into transcriptional regulations in innate immunity. *Biochemical Pharmacology*, 72:1102–1113, 2006.

Duan, Z; Holcombe, M, and Bell, A. A logic for biological systems. *BioSystems*, 55: 93–105, 2000.

Dutta, J; Fan, Y; Gupta, N; Fan, G, and Gelinas, C. Current insights into the regulation of programmed cell death by NF-κB. *Oncogene*, 25:6800–6816, 2006.

Efroni, S; Harel, D, and Cohen, I R. Emergent dynamics of thymocyte development and lineage determination. *PLoS Computational Biology*, 3:e13, 2007.

Einstein, A. Uber die von der molekularkinetischen theorie der warme geforderte bewegung von ib ruhenden flussigkeiten suspendierten teilchen (Translated by A. D. Cowper, 1956 On the Theory of the Brownian Motion). *Annalen der Physik (Leipzig)*, 17:549–560, 1905.

Elowitz, M B; Levine, A J; Siggia, E D, and Swain, P S. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.

Fallahi-Sichani, M; Kirschner, D E, and Linderman, J J. NF-κB signaling dynamics play a key role in infection control in tuberculosis. *Frontiers in Physiology*, 3:170, 2012.

Ferrell, J E. Q&A: Systems biology. *Journal of Biology*, 8:2, 2009.

Fisher, J and Henzinger, T A. Executable cell biology. *Nature Biotechnology*, 25(11): 1239–1248, 2007.

Flavell, R A; Sanjabi, S; Wrzesinski, S H, and Lincona-Limon, P. The polarization of immune cells in the tumour environment by TGFβ. *Nature Reviews Immunology*, 10:554–567, 2010.

Foster, I. *Designing and Building Parallel Programs*, chapter Message Passing Interface. Addison Wesley, 1995.

Fowler, M. *UML Distilled: A Brief Guide to the Standard Object Modeling Language*. Addison Wesley, Boston, USA, 3rd edition, 2004.

Foxwell, B; Browne, K; Bondeson, J; Clarke, C; de Martin, R; Brennan, F, and Feldmann, M. Efficient adenoviral infection with IκBα reveals that macrophage tumor necrosis factor α production in rheumatoid arthritis is NF-κB dependent. *Proceedings of the National Academy of Sciences*, 95:8211–8215, 1998.

Fraley, C and Raftery, A E. How many clusters? which clustering method? answers via model-based cluster analysis. *The Computer Journal*, 41(8):578–588, 1998.

Frank, M B; Wang, S; Aggarwal, A; Knowlton, N; Jiang, K; Chen, Y; McKee, R; Chaser, B; McGhee, T; Osban, J, and Jarvis, J N. Disease-associated pathophysiologic structures in pediatric rheumatic diseases show characteristics of scale-free networks seen in physiologic systems: Implications for pathogenesis and treatment. *BMC Medical Genomics*, 2:9, 2009.

Gabay, C; Lamacchia, C, and Palmer, G. IL-1 pathways in inflammation and human diseases. *Nature Reviews Rheumatology*, 6:232–241, 2010.

Ganguli, A; Persson, L; Palmer, I R; Evans, I; Yang, L; Smallwood, R; Black, R, and Qwarnstrom, E E. Distinct NF-κB regulation by shear stress through Ras-dependent IκBα oscillations: Real time analysis of flow-mediated activation in live cells. *Circulation Research*, 96:626–634, 2005.

Garnett, P; Stepney, S; Day, F, and Leyser, O. Using the CoSMoS process to enhance an executable model of auxin transport canalisation. In Stepney, S; Welch, P H; Andrews, P S, and Sampson, A T, editors, *Proceedings of the 2010 Workshop on Complex Systems Modelling and Simulation*, CoSMoS, pages 9–31, Odense, Denmark, 2010. Luniver Press.

Garrett, R H and Grisham, C M. *Molecular Aspects of Cell Biology*. Saunders College Publishing, Orlando, USA, international edition, 1995.

Gay, N J and Keith, F J. Drosophila Toll and IL-1 receptor. *Nature*, 351:355–356, 1991.

Germain, R N. The art of the probable: System control in the adaptive immune system. *Science*, 293:240–245, 2001.

Ghetiu, T; Polack, F A C, and Bown, J. Argument-driven validation of computer simulations - a necessity rather than an option. In *Proceedings of the 2nd International Conference on Advances in System Testing and Validation Lifecycle*, pages 1–4, Nice, France, August 2010. IEEE Computer Society.

Ghosh, S; May, M J, and Kopp, E B. NF-κB and Rel proteins: Evolutionarily conserved mediators of immune response. *Annual Review of Immunology*, 16:225–260, 1998.

Ghosh, S; Matsuoka, Y; Asai, Y; Hsin, K Y, and Kitano, H. Software for systems biology: From tools to integrated platforms. *Nature Reviews Genetics*, 12:821–832, 2012.

Giersch, C. Mathematical modelling of metabolism. *Current Opinion in Plant Biology*, 3:249–253, 2000.

Gillespie, D T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81:2340–2361, 1977.

Gilmore, T D and Garbati, M R. Inhibition of NF-κB signalling as a strategy in disease therapy. *Current Topics in Microbiology and Immunology*, 349:245–263, 2011.

Gilmore, T D and Herscovitch, M. Inhibitors of NF-κB signalling: 785 and counting. *Oncogene*, 25:6887–6899, 2006.

Greaves, R B; Read, M; Timmis, J; Andrews, P S, and Kumar, V. Extending an established simulation: Exploration of the possible effects using a case study in experimental autoimmune encephalomyelitis. In Lones, M A; Smith, S L; Teichmann, S; Naef, F; Walker, J A, and Trefzer, M A, editors, *Proceedings of the 9th International Conference on Information Processing in Cells and Tissues (IPCAT)*, volume 7223 of *Lecture Notes in Computer Science*, pages 150–161, Cambridge, UK, 2012. Springer-Verlag.

Greaves, R B; Read, M; Timmis, J; Andrews, P S; Butler, J A; Gerckens, B O, and Kumar, V. In silico investigation of novel biological pathways: The role of CD200 in regulation of T-cell priming in experimental autoimmune encephalomyelitis. *BioSystems*, 112:107–121, 2013.

Grilli, M and Memo, M. Nuclear factor-κB/Rel proteins. *Biochemical Pharmacology*, 57:1–7, 1999.

Grizzi, F and Chiriva-Internati, M. The complexity of anatomical systems. *Theoretical Biology and Medical Modelling*, 2:26, 2005.

Hamahashi, S and Kitano, H. Parameter optimization in hierarchical structures. In Floreano, D; Nicoud, J D, and Mondada, F, editors, *Proceedings of the 5th European Conference in Artificial Life*, volume 1674 of *Lecture Notes in Artificial Intelligence*, pages 467–471, Lausanne, Switzerland, 1999. Springer.

Hancock, J T. *Cell Signalling*. Prentice Hall, 1st edition, 1997.

Harris, R B; Maguire, L A, and Shaffer, M L. Sample sizes for minimum viable population estimation. *Conservation Biology*, 1(1):72–76, 1987.

Hasan, U; Chaffois, C; Gaillard, C; Saulnier, V; Merck, E; Tancredi, S; Guiet, C; Briere, F; Vlach, J; Lebecque, S; Trinchieri, G, and Bates, E E M. Human TLR10 is a functional receptor, expressed by B cells and plasmacytoid dendritic cells, which activates gene transcription through MyD88. *The Journal of Immunology*, 174:2942–2950, 2005.

Hayden, M S; West, A P, and Ghosh, S. NF-κB and the immune response. *Oncogene*, 25:6758–6780, 2006.

Hayot, F and Jayaprakash, C. NF-κB oscillations and cell-to-cell variability. *Journal of Theoretical Biology*, 240:583–591, 2006.

Helton, J C. Uncertainty and sensitivity analysis for models of complex systems. In Barth, T J; Griebel, M; Keyes, D E; Nieminen, R M; Roose, D, and Schlick, T, editors, *Computational Models in Transport: Verification and Validation*, pages 207–228. Springer, Heidelberg, Germany, 2008.

Hiscott, J; Nguyen, T-L; Arguello, M; Nakhaei, P, and Paz, S. Manipulation of the nuclear factor-κB pathway and the innate immune response by viruses. *Oncogene*, 25:6844–6867, 2006.

Hoffmann, A and Baltimore, D. Circuitry of nuclear factor κB signaling. *Immunological Reviews*, 210:171–186, 2006.

Hoffmann, A; Levchenko, A; Scott, M L, and Baltimore, D. The IκB - NF-κB signaling module: Temporal control and selective gene activation. *Science*, 298:1241–1245, 2002.

Holcombe, M. X-machines as a basis for dynamic system specification. *Software Engineering Journal*, (69-76), 1988.

Holcombe, M; Coakley, S, and Smallwood, R. A general framework for agent-based modelling of complex systems. In *European Conference on Complex Systems (ECCS)*, Warwick, UK, September 2006.

Hultmark, D. Macrophage differentiation marker MyD88 is a member of the Toll/IL-1 receptor family. *Biochemical and Biophysical Research Communications*, 199:144–146, 1994.

Hunter, P J and Borg, T K. Integration from proteins to organs: The physiome project. *Nature Reviews Molecular Cell Biology*, 4:237–243, 2003.

Hunter, P J; Li, W W; McCulloch, A D, and Noble, D. Multiscale modeling: Physiome project standards, tools and databases. *Computer*, 39(11):48–54, 2006.

Ideker, T; Galitski, T, and Hood, L. A new approach to decoding life: Systems biology. *Annual Review of Genomics and Human Genetics*, 2:343–372, 2001.

Ihekwaba, A E C; Broomhead, D S; Grimley, R L; Benson, N, and Kell, D B. Sensitivity analysis of parameters controlling oscillatory signalling in the NF-κB pathway: The roles of IKK and IκBα. *Systems Biology*, 1:93–103, 2004.

Ihekwaba, A E C; Broomhead, D S; Grimley, R; Benson, N; White, M R, and Kell, D B. Synergistic control of oscillations in the NF-kappaB signalling pathway. *IEEE Proceedings Systems Biology*, 152:153–160, 2005.

Ipate, F and Holcombe, M. Specification and testing using generalised machines: A presentation and a case study. *Software Testing, Verification and Reliability*, 8(2): 61–81, 1998.

Israël, A. The IKK complex, a central regulator of NF-κB activation. In Karin, M and Staudt, L M, editors, *NF-κB A Network Hub Controlling Immunity, Inflammation, and Cancer*, Cold Spring Harbour Perspectives in Biology, pages 51–64. Cold Spring Harbour Press, New York, USA, 2009.

Janzen, D and Saiedian, H. Test-driven development: Concepts, taxonomy, and future direction. *Computer*, pages 43–50, September 2005.

Jennings, N R. On agent-based software engineering. *Artificial Intelligence*, 117:277–296, 2000.

Jeong, H; Tombor, B; Albert, R; Oltvai, Z N, and Barabasi, A L. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

Jones, D. All systems go: How might systems biology approaches be applied in drug discovery and development. *Nature Reviews Drug Discovery*, 7:278–279, 2008.

Joo, J; Plimpton, S; Martin, S; Swiler, L, and Faulon, J L. Sensitivity analysis of a computational model of the IKK-NF-κB-IκBα-A20 signal transduction network. In *Reverse Engineering Biological Networks: Opportunities and Challenges in Computational Methods for Pathway Inference*, volume 1115 of *Annals of the New York Academy of Sciences*, pages 221–239. Wiley, New York, USA, 2007.

Kaern, M; Elston, T C; Blake, W J, and Colins, J J. Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics*, 6:451–464, 2005.

Kam, N; Cohen, I R, and Harel, D. The immune system as a reactive system: Modeling T cell activation with statecharts. In *Symposium on Human Centric Computing Languages and Environments*, Stresa, Italy, 2001. IEEE.

Karin, M. The beginning of the end: IκB kinase (IKK) and NF-κB activation. *Journal of Biological Chemistry*, 274:27339–27342, 1999a.

Karin, M. How NF-κB is activated: the role of the IκB kinase (IKK) complex. *Oncogene*, 18:6867–6874, 1999b.

Karin, M and Ben-Neriah, Y. Phosphorylation meets ubiquitination: The control of NF-κB activity. *Annual Review of Immunology*, 18:621–663, 2000.

Karin, M and Lin, A. NF-κB at the crossroads of life and death. *Nature Immunology*, 3(3):221–227, 2002.

Kawai, T and Akira, S. TLR signaling. *Cell Death and Differentiation*, 13:816–825, 2006.

Kearns, J D and Hoffmann, A. Integrating computational and biochemical studies to explore mechanisms in NF-κB signaling. *Journal of Biological Chemistry*, 284: 5439–5443, 2009.

Kearns, J D; Basak, S; Werner, S L; Huang, C S, and Hoffmann, A. IκBϵ provides negative feedback to control NF-κB oscillations, signaling dynamics, and inflammatory gene expression. *Journal of Cell Biology*, 173:659–664, 2006.

Kefalas, P; Eleftherakis, G, and Kehris, E. Communicating X-machines: A practical approach for formal and modular specification of large systems. *Information and Software Technology*, 45:269–280, 2003.

Kehris, E; Eleftherakis, G, and Kefalas, P. Using X-machines to model and test discrete event simulation prograns. In Mastorakis, N, editor, *Proceedings of the 4th World MultiConference on Circuits, Systems, Communications & Computers*, pages 163–168. World Scientific and Engineering Society Press, 2000.

Khan, S; Makkena, R; Gillis, W, and Shmidt, C. A multi-agent system for the quantitative simulation of biological networks. In *Autonomous Agents and Multi-Agent Systems 2003 (AAMAS'03)*, pages 385–392, Melbourne, Australia, July 2003.

Kholodenko, B N. Cell-signalling dynamics in time and space. *Nature Reviews Molecular Cell Biology*, 7:165–176, 2006.

Kiran, M; Coakley, S; Walkinshaw, N; McMinn, P, and Holcombe, M. Validation and discovery from computational biology models. *BioSystems*, 93:141–150, 2008.

Kiran, M; Richmond, P; Holcombe, M; Chin, S; Worth, D, and Greenough, C. FLAME: simulating large populations of agents on parallel hardware architectures. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, volume 1, pages 1633–1636, Toronto, Canada, 2010. ACM.

Kirschner, D E and Linderman, J J. Mathematical and computational approaches can complement experimental studies of host-pathogen interactions. *Cellular Microbiology*, 11(4):531–539, 2009.

Kirschner, D E; Chang, S T; Riggs, T W; Perry, N, and Linderman, J J. Toward a multiscale model of antigen presentation in immunity. *Immunological Reviews*, 216: 93–118, 2007.

Kitano, H. Perspectives on systems biology. *New Generation Computing*, 18:199–216, 2000.

Kitano, H. *Foundations of Systems Biology*. MIT Press, Massachusetts, USA, 2001a.

Kitano, H. *Foundations of Systems Biology*, chapter Toward System-Level Understanding of Biological Systems, pages 1–36. MIT Press, 2001b.

Kitano, H. Computational systems biology. *Nature*, 420:206–210, 2002a.

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

Kitano, H. A graphical notation for biochemical networks. *BioSilico*, 1:169–176, 2003.

Kitano, H. Biological robustness. *Nature Reviews Genetics*, 5:826–837, 2004a.

Kitano, H. Cancer as a robust system: Implications for anticancer therapy. *Nature Reviews Cancer*, 4:227–235, 2004b.

Kitano, H. A robustness-based approach to system-oriented drug design. *Nature Reviews Drug Discovery*, 6:202–210, 2007.

Kitano, H. Grand challenges in systems physiology. *Frontiers in Physiology*, 1:3, 2010.

Kitano, H; Funahashi, A; Matsuoka, Y, and Oda, K. Using process diagrams for the graphical representation of biological networks. *Nature Biotechnology*, 23:961–966, 2005.

Kleinstein, S H. Getting started in computational immunology. *PLoS Computational Biology*, 4:e1000128, 2008.

Kleinstein, S H and Seiden, P E. Simulating the immune system. *Computing in Science and Engineering*, 2:69–77, 2000.

Kohn, K W; Aladjem, M I; Weinstein, J N, and Pommier, Y. Molecular interaction maps of bioregulatory networks: A general rubric for systems biology. *Molecular Biology of the Cell*, 17:1–13, 2006.

Kovacs, J A; Baseler, M; Dewar, R J; Vogel, S; Jr, R T Davey; Falloon, J; Polis, M A; Walker, R E; Stevens, R; Salzman, N P; Metcalf, J A; Masur, H, and Lane, H C. Increases in CD4 T lymphocytes with intermittent courses of interleukin-2 in patients with human immunodeficiency virus infection: A preliminary study. *New England Journal of Medicine*, 332(9):567–575, 1995.

Krauss, G. *Biochemistry of Signal Transduction and Regulation*. Wiley-VCH, Weinheim, Germany, 3rd edition, 2003.

Krebs, E G. Protein phosphorylation and cellular recognition, i. *Nobel Lectures in Physiology or Medicine 1991-1995*, pages 72–89, 1992.

Krebs, H A. The citric acid cycle and the Szent-Gyorgyi cycle in pigeon breast muscle. *Biochemical Journal*, 34:775–779, 1940.

Krikos, A; Laherty, C D, and Dixit, V M. Transcriptional activation of the tumor necrosis factor alpha-inducible zinc finger protein, A20, is mediated by kappa B elements. *Journal of Biological Chemistry*, 267:17971–17976, 1992.

Krishna, S; Jensen, M H, and Sneppen, K. Minimal model of spiky oscillations in NF-κB signaling. *Proceedings of the National Academy of Sciences*, 103(29):10840–10845, 2006.

Krzywinski, M and Altman, N. Importance of being uncertain. *Nature Methods*, 10(9): 809–810, 2013.

Kugler, H; Larjo, A, and Harel, D. Biocharts: A visual formalism for complex biological systems. *Journal of the Royal Society Interface*, doi:10.1098/rsif.2009.0457, 2009.

Kumar, N; Hendriks, B S; Janes, K A; de Graaf, D, and Lauffenburger, D A. Applying computational modelling to drug discovery and development. *Drug Discovery Today*, 11:806–811, 2006.

Laflamme, N and Rivest, S. Toll-like receptor 4: The missing link of cerebral innate immune response triggered by circulating gram-negative bacterial cell wall components. *The FASEB Journal*, 15:155–163, 2001.

Lazebnik, Y. Can a biologist fix a radio? – or, what i learned while studying apoptosis. *Cancer Cell*, 2:179–182, 2002.

le Novere, N; Hucka, M; Mi, H; Moodie, S; Schreiber, F; Sorokin, A; Demir, E; Wegner, K; Aladjem, M; Wimalaratne, S M; Bergman, F T; Gauges, R; Ghazal, P; Kawaji, H; Li, L; Matsuoka, Y; Villeger, A; Boyd, S E; Calzone, L; Courtot, M; Dogrusoz, U; Freeman, T C; Funahashi, A; Ghosh, S; Jouraku, A; Kim, S; Kolpakov, F; Luna, A; Sahle, S; Schmidt, E; Watterson, S; Wu, G; Goryanin, I; Kell, D B; Sander, C; Sauro, H; Snoep, J L; Kohn, K, and Kitano, H. The systems biology graphical notation. *Nature Biotechnology*, 27:735–741, 2009.

Lee, E G; Boone, D L; Chai, S; Libby, S L; Chien, M; Lodolce, J P, and Ma, A. Failure to regulate TNF-induced NF-κB and cell death responses in A20-deficient mice. *Science*, 289:2350–2354, 2000.

Lee, T K and Covert, M W. High-throughput, single-cell NF-κB dynamics. *Current Opinion in Genetics & Development*, 20:1–7, 2010.

Lemaitre, B; Nicholas, E; Michaut, L; Reichhart, J M, and Hoffmann, J A. The dorsoventral regulatory gene cassette spatzle/toll/cactus controls the potent antifungal response in Drosophila adults. *Cell*, 86:973–983, 1996.

Levine, B and Kroemer, G. Autophagy in the pathogenesis of disease. *Cell*, 132:27–42, 2007.

Levine, B; Mizushima, N, and Virgin, H W. Autophagy in immunity and inflammation. *Nature*, 469:323–335, 2011.

Lipniacki, T; Paszek, P; Brasier, A R; Luxon, B, and Kimmel, M. Mathematical model of NF-κB regulatory module. *Journal of Theoretical Biology*, 228:195–215, 2004.

Lipniacki, T; Paszek, P; Brasier, A R; Luxon, B A, and Kimmel, M. Stochastic regulation in early immune response. *Biophysical Journal*, 90:725–742, 2006.

Lipniacki, T; Puszynski, K; Paszek, P, and Brasier, A R. Single TNFα trimers mediating NF-κB activation: Stochastic robustness of NF-κB signaling. *BMC Bioinformatics*, 8:376, 2007.

Lodish, H; Berk, A; Zipursky, S L; Matsudaira, P; Baltimore, D, and Darnell, J. *Molecular Cell Biology*. W.H. Freeman & Co, New York, 4th edition, 2000.

Longo, D M; Selimkhanov, J; Kearns, J D; Hasty, J; Hoffmann, A, and Tsimring, L S. Dual delayed feedback provides sensitivity and robustness to the NF-κB signaling module. *PLoS Computational Biology*, 9(6):e1003112, 2013.

Lorenz, D M; Jeng, A, and Deem, M W. The emergence of modularity in biological systems. *Physics of Life Reviews*, 8:129–160, 2011.

Lucey, B P; Nelson-Rees, W A, and Hutchins, G M. Henrietta Lacks, HeLa cells, and cell culture contamination. *Archives of Pathology and Laboratory Medicine*, 133: 1463–1467, 2009.

Luke, S; Cioffi-Revilla, C; Panait, L; Sullivan, K, and Balan, G. MASON: A multi-agent simulation environment. *Simulation*, 81:517–527, 2005.

Luttge, U. Modularity and emergence: Biology's challenge in understanding life. *Plant Biology*, 14:865–871, 2012.

Macal, C M and North, M J. Tutorial on agent-based modeling and simulation. In Kuhl, M E; Steiger, N M; Armstrong, F B, and Jones, J A, editors, *Winter Simulation Conference*, pages 2–15, Orlando, Florida, 2005.

Maindonald, J and Braun, W J. *Data Analysis and Graphics using R*. Cambridge University Press, 3rd edition edition, 2010.

Mann, H B and Whitney, D R. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18: 50–60, 1947.

Mardinoglu, A and Nielsen, J. Systems medicine and metabolic modelling. *Journal of Internal Medicine*, 271:142–154, 2012.

Massey, F J. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

Mathes, E; O'Dea, E L; Hoffmann, A, and Ghosh, G. NF-κB dictates the degradation pathway of IκBα. *The EMBO Journal*, 27:1357–1367, 2008.

Mayr, E. Cause and effect in biology. *Science*, 134:1501–1506, 1961.

McKay, M D; Beckman, R J, and Conover, W J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Techometrics*, 21:239–245, 1979.

Mendes, P. GEPASI: A software package for modelling the dynamics, steady staes and control of biochemical and other systems. *Computer Applications in the Biosciences*, 9:563–571, 1993.

Mendes, P. Biochemistry by numbers: Simulation of biochemical pathways with Gepasi 3. *Trends in Biological Sciences*, 22:361–363, 1997.

Mercurio, F; Zhu, H; Murray, B W; Shevchenko, A; Bennett, B L; wu Li, J; Young, D B; Barbosa, M, and Mann, M. IKK-1 and IKK-2: Cytokine-activated IκB kinases essential for NF-κB activation. *Science*, 278:860–866, 1997.

Mesarovic, M D, editor. *Systems Theory and Biology: Proceedings of the 3rd Systems Symposium*, 1968. Case Institute of technology, Systems Research Center, Springer.

Mesirov, J P. Accessible reproducible research. *Science*, 327:415–416, 2010.

Mizuno, A; Lijima, R; Ogishima, S; Kikuchi, M; Matsuoka, Y; Ghosh, S; Miyamoto, T; Miyashita, A, and Tanaka, H. AlzPathway: A comprehensive map of signalling pathways of Alzheimer's disease. *BMC Systems Biology*, 6:52, 2012.

Morgan, M J and Liu, Z-G. Crosstalk of reactive oxygen species and NF-κB signaling. *Cell Research*, 21:103–115, 2011.

Morohashi, M; Winn, A E; Borisuk, M T; Bolouri, H; Doyle, J C, and Kitano, H. Robustness as a measure of plausibility in models of biochemical networks. *Journal of Theoretical Biology*, 216:19–30, 2002.

Murphy, K; Travers, P, and Walport, M. *Janeway's Immunobiology*. Garland Science, 7th edition, 2008.

Naumann, M and Scheidereit, C. Activation of NF-κB in vivo is regulated by multiple phosphorylations. *The EMBO Journal*, 13:4597–4607, 1994.

Navlahka, S and Bar-Joseph, Z. Algorithms in nature: The convergence of systems biology and computational thinking. *Molecular Systems Biology*, 7:546, 2011.

Nelson, D E; Ihekwaba, A E C; Elliott, M; Johnson, J R; Gibney, C A; Foreman, B E; Nelson, G; See, V; Horton, C A; Spiller, D G; Edwards, S W; McDowell, H P; Unitt, J F; Sullivan, E; Grimley, R; Benson, N; Broomhead, D; Kell, D B, and White, M R H. Oscillations in NF-κB signaling control the dynamics of gene expression. *Science*, 306:704–708, 2004.

Nemeth, Z H; Deitch, E A; Davidson, M T; Szabo, C; Vizi, E S, and Hasko, G. Disruption of the actin cytoskeleton results in Nuclear Factor-κB activation and inflammatory mediator production in cultured human epithelial cells. *Journal of Cellular Physiology*, 200:71–80, 2004.

Neves, S R and Iyengar, R. Modeling of signalling networks. *BioEssays*, 24:1110–1117, 2002.

Nicholson, D E. IUBMB-Nicholson metabolic pathways charts. *Biochemistry and Molecular Biology Education*, 29:42–44, 2001.

Noble, D. The rise of computational biology. *Molecular and Cellular Biology*, 3:460–463, 2002.

Noble, D. Systems biology and the heart. *BioSystems*, 83:75–80, 2006.

Novack, D V. Role of NF-κB in the skeleton. *Cell Research*, 21:169–182, 2011.

Nykamp, D Q. Spherical coordinates. *http://mathinsight.org/spherical_coordinates*, 2011.

Object Management Group, . Unified modeling language superstructure specification v2.4. *http://www.omg.org/spec/UML/2.4/Superstructure*, 2011.

Oda, K and Kitano, H. A comprehensive map of the toll-like receptor signaling network. *Molecular Systems Biology*, 2006.0015, 2006.

O'Dea, E L; Barken, D; Peralta, R Q; Tran, K T; Werner, S L; Kearns, J D; Levchenko, A, and Hoffmann, A. A homeostatic model of IκB metabolism to control constitutive NF-κB activity. *Molecular Systems Biology*, 3:111, 2007.

Oeckinghaus, A and Ghosh, S. The NF-κB family of transcription factors and its regulation. In Karin, M and Staudt, L M, editors, *NF-κB A Network Hub Controlling Immunity, Inflammation, and Cancer*, Cold Spring Harbour Perspectives in Biology, pages 5–18. Cold Spring Harbour Press, New York, USA, 2009.

Ohshima, D; Inoue, J I, and Ichikawa, K. Roles of spatial parameters on the oscillation of nuclear NF-κB: Computer simulations of a 3D spherica cell. *PLoS ONE*, 7(10): e46911, 2012.

Oltvai, Z N and Barabasi, A L. Life's complexity pyramid. *Science*, 298:763–764, 2002.

O'Neill, L A J and Dinarello, C A. The IL-1 receptor/toll-like receptor superfamily: Crucial receptors for inflammation and host defense. *Immunology Today*, 21:206–209, 2000.

Ootsuki, J T and Sekiguchi, T. Application of the immune system network concept to sequential control. In *IEEE International Conference on Systems, Man, and Cybernetics 1999*, volume 3, pages 869–874, Tokyo, Japan, 1999. Springer-Verlag.

Oracle Corporation, . Oracle Unified Method (OUM): Oracle's full lifecycle method for deploying Oracle-based business solutions. White paper, Redwood Shores, California, USA, November 2012.

Pahl, H L. Activators and target genes of Rel/NF-κB transcription factors. *Oncogene*, 18:6853–6866, 1999.

Palsson, B. The challenges of in silico biology. *Nature Biotechnology*, 18:1147–1150, 2000.

Palsson, B. *Systems Biology: Simulation of Dynamic Network States*. Cambridge University Press, Cambridge, UK, 2011.

Pappalardo, F; Musumeci, S, and Motta, S. Modeling immune system control of atherogenesis. *Bioinformatics*, 24:1715–1721, 2008.

Pappalardo, F; Forero, I M; Pennisi, M; Palazon, A; Melero, I, and Motta, S. SimB16: Modeling induced immune system response against B16-melanoma. *PLoS ONE*, 6 (10):e26523, 2011.

Paszek, P; Ryan, S; Ashall, L; Sillitoe, K; Harper, C V; Spiller, D G; Rand, D A, and White, M R H. Population robustness arising from cellular heterogeneity. *Proceedings of the National Academy of Sciences*, 107(25):11644–11649, 2010.

Patel, A; Harker, N; Moreira-Santos, L; Ferreira, M; Alden, K; Timmis, J; Foster, K; Garefalaki, A; Pachnis, P; Andrews, P; Enomoto, H; Milbrandt, J; Pachnic, V; Coles, M C; Kloussis, D, and Veiga-Fernandes, H. Differential RET signalling pathways drive development of the enteric lymphoid and nervous systems. *Science Signalling*, 5(235):ra55, 2012.

Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 6(2):559–572, 1901.

Pekalski, J; Zuk, P J; Kochanczyk, M; Junkin, M; Kellogg, R; Tay, S, and Lipniacki, T. Spontaneous NF-κB activation by autocrine TNFα signaling: A computational analysis. *PLoS ONE*, 8(11):e78887, 2013.

Perelson, A S and Nelson, P W. Mathematical analysis of HIV-1 dynamics in vivo. *Society for Industrial and Applied Mathematics Review*, 41:3–44, 1999.

Perelson, A S and Weisbuch, G. Immunology for physicists. *Review of Modern Physics*, 69:1219–1267, 1997.

Petrovsky, N and Brusic, V. Computational immunology: The coming of age. *Immunology and Cell Biology*, 80:248–254, 2002.

Pfeifer, A C; Timmer, J, and Klingmuller, U. Systems biology of JAKS/STAT signalling. In Wolkenhauer, O; Wellstead, P, and Cho, K H, editors, *Systems Biology*, volume 45 of *Essays in Biochemistry*, pages 109–120. Portland Press, London, UK, 2008.

Pogson, M; Smallwood, R; Qwarnstrom, E E, and Holcombe, M. Formal agent-based modelling of intracellular chemical interactions. *BioSystems*, 85:37–45, 2006.

Pogson, M; Holcombe, M; Smallwood, R, and Qwarnstrom, E E. Introducing spatial information into predictive NF-κB modelling - an agent-based approach. *PLoS ONE*, 3:e2367, 2008.

Poltorak, A; He, X; Smirnova, I; Liu, M Y; van Huffel, C; Du, X; Birdwell, D; Alejos, E; Silva, M; Galanos, C; Freudenberg, M; Ricciardi-Castagnoli, P I; Layton, B, and Beutler, B. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: Mutations in Tlr4 gene. *Science*, 282:2085–2088, 1998.

Press, W H; Teukolsky, S A; Vetterling, W T, and Flannery, B P. *Numerical Recipes: The Art of Scientific Computing*, chapter Statistical Description of Data, pages 720–772. Cambridge University Press, Cambridge, UK, 3rd edition, 2007.

Priami, C. Algorithmic systems biology: An opportunity for computer science. *Communication of the ACM*, 52(5):80–88, 2009.

Project Management Institute, . *A Guide to the Project Management Body of Knowledge*. Project Management Institute, Inc., Pennsylvania, USA, 3rd edition, 2004.

Rajapakse, I; Groudine, M, and Mesbahi, M. What can systems theory of networks offer to biology. *PLoS Computational Biology*, 8:e1002543, 2012.

Ravasz, E; Somera, A L; Mongru, D A; Oltvai, Z N, and Barabasi, A L. Hierarchical organization of modularity in metabolic networks. *Science*, 297:1551–1555, 2002.

Ray, J C J; Flynn, J L, and Kirschner, D E. Synergy between individual TBF-dependent functions determines granuloma performance for controlling Mycobacterium tuberculosis infection. *The Journal of Immunology*, 182:3706–3717, 2009.

Re, F and Strominger, J L. Monomeric recombinant MD-2 binds Toll-like receptor 4 tightly and confers lipopolysaccharide responsiveness. *Journal of Biological Chemistry*, 277:23427–23432, 2002.

Read, M; Timmis, J; Andrews, P S, and Kumar, V. A domain model of experimental autoimmune encephalomyelitis. In Stepney, S; Welch, P H; Andrews, P S, and Timmis, J, editors, *Proceedings of the 2009 Workshop on Complex Systems Modelling and Simulation*, CoSMoS, pages 9–44. Luniver Press, 2009a.

Read, M; Timmis, J; Andrews, P S, and Kumar, V. Using UML to model EAE and its regulatory network. In Andrews, P S; Timmis, J; Owens, N D L; Aickelin, U; Hart, E; Hone, A, and Tyrrell, A M, editors, *ICARIS*, volume 5666 of *LNCS*, pages 4–6. Springer, 2009b.

Read, M; Andrews, P S; Timmis, J, and Kumar, V. Techniques for grounding agent-based simulations in the real domain: A case study in experimental autoimmune encephalomyelitis. *Mathematical and Computer Modelling of Dynamical Systems*, 18 (1):67–86, 2012.

Read, M; Andrews, P S; Timmis, J, and Kumar, V. Modelling biological behaviours with the unified modelling language: An immunological case study and critique. *Journal of the Royal Society Interface*, 11:20140704, 2014.

Read, M N. *Statistical and Modelling Techniques to Build Confidence in the Investigation of Immunology through Agent-Based Simulation.* PhD thesis, University of York, September 2011.

Read, M N; Andrews, P S; Timmis, J; Williams, R A; Greaves, R B; Sheng, H; Coles, M, and Kumar, V. Determining disease intervention strategies usng spatially resolved simulations. *PLoS ONE*, 8(11):e80506, 2013.

Rice, N R and Ernst, M K. In vivo control of NF-κB activation by IκBα. *The EMBO Journal*, 12:4685–4695, 1993.

Richmond, P; Coakley, S, and Romano, D. Cellular level agent-based modelling on the graphics processing unit. In *International Workshop on High-Performance Computational Systems Biology (HiBi'09)*, pages 43–50, Trento, Italy, 2009. IEEE.

Richmond, P; Walker, D; Coakley, S, and Romano, D. High performance cellular level agent-based simulation with FLAME for the GPU. *Briefings in Bioinformatics*, 11 (3):334–347, 2010.

Royce, W W. Managing the development of large software systems. *Proceedings of IEEE WESCON*, 26:1–9, 1970.

Rumpe, B and France, R. Variability in UML language and semantics. *Software and Systems Modelling*, 10:439–440, 2011.

Ryan, G B. Acute inflammation. *American Journal of Pathology*, 86(1):185–276, 1977.

Saltelli, A; Chan, K, and Scott, E M. *Sensitivity Analysis.* Wiley Series in Probability and Statistics. John Wiley & Sons Inc, Chichester, UK, 2009.

Sauer, U. Metabolic networks in motion: 13C-based flux analysis. *Molecular Systems Biology*, 62, 2006.

Sauro, H M. Modularity defined. *Molecular Systems Biology*, 4:166, 2008.

Schadt, E E; Friend, S H, and Shaywitz, D A. A network view of disease and compound screening. *Nature Reviews Drug Discovery*, 8(286-295), 2009.

Schooley, K; Zhu, P; Dower, S K, and Qwarnstrom, E E. Regulation of nuclear translocation of nuclear factor-κB RelA: Evidence for complex dynamics at the single-cell level. *Biochemical Journal*, 369:331–339, 2003.

Seeley, T D. When is self-organization used in biological systems? *Biology Bulleton*, 202(3):314–318, 2002.

Segel, L A. Grappling with complexity. *Complexity*, 1:18–25, 1995.

Seiden, P E and Celada, F. A model for simulating cognate recognition and response in the immune system. *Journal of Theoretical Biology*, 158:329–357, 1992.

Sen, R and Baltimore, D. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell*, 46:705–716, 1986a.

Sen, R and Baltimore, D. Inducibility of κ immunoglobulin enhancer-binding protein NF-κB by a posttranslational mechanism. *Cell*, 47:921–928, 1986b.

Senftleben, U; Cao, Y; Xiao, G; Greten, F R; Krahn, G; Bonizzi, G; Chen, Y; Hu, Y; Fong, A; Sun, S C, and Karin, M. Activation by IKKα of a second, evolutionary conserved, NF-κB signaling pathway. *Science*, 293:1495–1499, 2001.

Sethi, G; Sung, B, and Aggarwal, B B. Nuclear factor-κB activation: From bench to bedside. *Experimental Biology and Medicine*, 233:21–31, 2008.

Sguanci, L; Lio, P, and Bagnoli, F. Modeling evolutionary dynamics of HIV infection. In Priami, C, editor, *International Conference on Computational Methods in Systems Biology*, volume 4210 of *Lecture Notes in Bioinformatics*, pages 196–211, Trento, Italy, October 2006. Springer-Verlag.

Sguanci, L; Bagnoli, F, and Lio, P. Modeling HIV quasispecies evolutionary dynamics. *BMC Evolutionary Biology*, 7(Suppl 2):S5, 2007.

Shapiro, B E; Levchenko, A; Meyerowitz, E M; Wold, B J, and Mjolsness, E D. Cellerator: Extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics*, 19:677–678, 2003.

Shih, V F S; Kearns, J D; Basak, S; Savinova, O V; Ghosh, G, and Hoffmann, A. Kinetic control of negative feedback regulators of NF-κB/RelA determines their pathogen- and cytokine-receptor signaling specificity. *Proceedings of the National Academy of Sciences*, 106:9619–9624, 2009.

Shih, V F S; Davis-Turak, J; Macal, M; Huang, J Q; Ponomarenko, J; Kearns, J D; Yu, T; Fagerlund, R; Asagiri, M; Zuniga, E I, and Hoffmann, A. Control of RelB during dendritic cell activation integrates canonica and noncanonical NF-κB pathways. *Nature Immunology*, 13(12):1162–1172, 2012.

Shimazu, R; Akashi, S; Ogata, H; Nagai, Y; Fukudome, K; Miyake, K, and Kimoto, M. MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *Journal of Experimental Medicine*, 189:1777–1782, 1999.

Shirakawa, F and Mizel, S B. In vitro activation and nuclear translocation of NF-κB catalyzed by cyclic AMP-dependent protein kinase and protein kinase C. *Molecular and Cellular Biology*, 9:2424–2430, 1989.

Shore, J and Warden, S. *The Art of Agile Development*. O'Reilly, Sebastopol, California, USA, 2008.

Siebenlist, U; Franzoso, G, and Brown, K. Structure, regulation and function of NF-κB. *Annual Review of Cell Biology*, 10:405–455, 1994.

Siegel, S. Nonparametric statistics. *The American Statistician*, 11:13–19, 1957.

Silver, P A. How proteins enter the nucleus. *Cell*, 64:489–497, 1991.

Slepchenko, B M; Schaff, J C; Carson, J H, and Loew, L M. Computational cell biology: Spatiotemporal simulation of cellular events. *Annual Review of Biophysics and Biomolecular Strcuture*, 31:423–441, 2002.

Sonenshein, G E. Rel/NF-κB transcription factors and the control of apoptosis. *Seminars in Cancer Biology*, 8:113–119, 1997.

Sontag, E D. Molecular systems biology and control. *European Journal of Control*, 11: 396–435, 2005.

Sreenath, S N; Cho, K H, and Wellstead, P. Modelling the dynamics of signalling pathways. In Wolkenhauer, O; Wellstead, P, and Cho, K H, editors, *Systems Biology*, volume 45 of *Essays in Biochemistry*, pages 1–28. Portland Press, London, UK, 2008.

Stamatopoulou, I; Kefalas, P, and Gheorghe, M. Modelling the dynamics structure of biological state-based systems. *BioSystems*, 87:142–149, 2007.

Stark, J; Chan, C, and George, A J. Oscillations in the immune system. *Immunological Reviews*, 216:213–231, 2007.

Stepney, S. A pattern language for scientific simulations. In Stepney, S; Andrews, P S, and Read, M, editors, *Proceedings of the 2012 Workshop on Complex Systems Modelling and Simulation*, pages 77–103, Orleans, France, September 2012. Luniver Press.

Stylianou, E; O'Neill, L A J; Rawlinson, L; Edbrooke, M R; Woo, P, and Saklatvala, J. Interleukin 1 induces NF-κB through its type I but not its type II receptor in lymphocytes. *Journal of Biological Chemistry*, 267:15836–15841, 1992.

Sun, S C and Liu, Z G. A special issue on NF-κB signaling and function. *Cell Research*, 21:1–2, 2011.

Sun, S C; Ganchi, P A; Ballard, D W, and Greene, W C. NF-κB controls expression of inhibtor IκBα: Evidence for an inducible autoregulatory pathway. *Science*, 259: 1912–1915, 1993.

Sun, T; Adra, S; Smallwood, R; Holcombe, M, and MacNeil, S. Exploring hypotheses of the actions of TGF-β1 in epidermal wound healing using 3D computational multiscale model of the human epidermis. *PLoS ONE*, 4(12):e8515, 2009.

Sung, M H and Simon, R. In silico simulation of inhibitor drug effects on nuclear factor-κb pathway dynamics. *Molecular Pharmacology*, 66(1):70–75, 2004.

Sung, M H; Salvatore, L; de Lorenzi, R; Indrawan, A; Pasparakis, M; Hager, G L; Bianchi, M E, and Agresti, A. Sustained oscillations of NF-κB produce distinct genome scanning and gene expression profiles. *PLoS ONE*, 4(9):e7163, 2009.

Sung, M H; Li, N; Lao, Q; Gottschalk, R A; Hager, G L, and Fraser, I D C. Switching of the relative dominance between feedback mechanisms in lipopolysaccharide-induced NF-κB signaling. *Science Signalling*, 7(308):ra6, 2014.

Tay, S; Hughey, J J; Lee, T K; Lipniacki, T; Quake, S R, and Covert, M W. Single-cell NF-κB dynamics reveal digital activation and analogue information processing. *Nature*, 466:267–272, 2010.

Terry, A J and Chaplain, M A J. Spatio-temporal modelling of the NF-κB intracellular signalling pathway: The roles of diffusion, active transport, and cell geometry. *Journal of Theoretical Biology*, 290:7–26, 2011.

Tian, B and Brasier, A R. Identification of a nuclear factor kappa B-dependent gene network. *Recent Progress in Hormone Research*, 58:95–130, 2003.

Tieri, P; Valensin, S; Latora, V; Castellani, G C; Marchiori, M; Remondini, D, and Franceschi, C. Quantifying the relevance of different mediators in the human immune cell network. *Bioinformatics*, 21:1639–1643, 2005.

Tiger, C F; Krause, F; Cedersund, G; Palmer, R; Klipp, E; Hohmann, S; Kitano, H, and Krantz, M. A framework for mapping, visualisation and automatic model creation of signal transduction networks. *Molecular Systems Biology*, 8:578, 2012.

Tijskens, L M M; Konopacki, P, and Simcic, M. Biological variance, burden or benefit? *Postharvest Biology and Technology*, 27:15–25, 2003.

Tong, JC and Ren, E C. Immunoinformatics: Current trends and future directions. *Drug Discovery Today*, 14:684–689, 2009.

Turner, D A; Paszek, P; Woodcock, D J; Nelson, D E; Horton, C A; Wang, Y; Spiller, D G; Rand, D A; White, M R H, and Harper, C V. Physiological levels of TNFα stimulation induce stochastic dynamics of NF-κB responses in single living cells. *Journal of Cell Science*, 123(16):2834–2843, 2010.

Upton, G and Cook, I. *Understanding Statistics*. Oxford University Press, 1996.

van Iersel, M P; Villeger, A C; Czauderna, T; Boyd, S E; Bergman, F T; Luna, A; Demir, E; Sorokin, A; Dogrusoz, U; Matsuoka, Y; Funahashi, A; Aladjem, M; Mi, H; Moodie, S L; Kitano, H; Novere, N Le, and Schreiber, F. Software support for SBGN maps: SBGN-ML and LibSBGN. *Bioinformatics*, 28(15):2016–2021, 2012.

van Niel, K P and Laffan, S W. Gambling with randomness: The use of pseudo-random number generators in GIS. *International Journal of Geographical Information Science*, 17(1):49–68, 2003.

van Niel, K P and Laffan, S W. There is no good excuse for a bad random number generator: A reply to Barry. *International Journal of Geographical Information Science*, 25(4):531–539, 2011.

Vargha, A and Delaney, H D. A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioural Statistics*, 25:101–132, 2000.

Voet, D and Voet, J. *Biochemistry*. John Wiley & Sons Inc, New York, 2nd edition, 1995.

von Bertalanffy, L. An outline of general systems theory. *The British Journal for the Philosophy of Science*, 1:134–165, 1950.

von Bertalanffy, L. *General Systems Theory: Foundations, Development, Applications*. George Braziller Inc, New York, 17th edition, 1969.

von Bertalanffy, L. The history and status of general systems theory. *The Academy of Management Journal*, 15:407–426, 1972.

von Dassow, G and Munro, E. Modularity in animal development and evolution: Elements of a conceptual frameowkr for EvoDeveo. *Journal of Experimental Zoology*, 285:307–325, 1999.

Wada, A. Bioinformatics - the necessity of the quest for 'first principles' in life. *Bioinformatics*, 16:663–664, 2000.

Walker, D; Wood, S; Southgate, J; Holcombe, M, and Smallwood, R. An integrated agent-mathematical model of the effect of intercellular signalling via the epidermal growth factor receptor on cell proliferation. *Journal of Theoretical Biology*, 242: 774–789, 2006.

Walker, D C and Southgate, J. The virtual cell - a candidate co-ordinator for 'middle-out' modelling of biological systems. *Briefings in Bioinformatics*, 10(4):450–461, 2009.

Walker, D C; Hill, G; Wood, S M; Smallwood, R H, and Southgate, J. Agent-based computational modelling of wounded epithelial cell monolayers. *IEEE Transactions on Nanobioscience*, 3:153–163, 2004.

Wang, Y; Paszek, P; Horton, C A; Kell, D B; White, M R H; Broomhead, D S, and Muldoon, M R. Interactions among oscillatory pathways in NF-κB signalling. *BMC Systems Biology*, 5:23, 2011.

Webb, K. UML modeling of finite state machines and molecular machines. In *Symposium on Complex Systems Engineering*, Santa Monica, CA, USA, January 2007. RAND Corporation.

Webb, K and White, T. Cell modeling with reusable agent-based formalisms. In Orchard, B; Yang, C, and Moonis, A, editors, *17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE)*, volume 3029 of *Lecture Notes in Computer Science*, pages 128–137. Springer, 2004.

Webb, K and White, T. UML as a cell and biochemistry modeling language. *BioSystems*, 80:283–302, 2005.

Weiner, N. *Cybernetics of Control and Communication in the Animal and the Machine*. MIT Press, 1948.

Weng, G; Bhalla, U S, and Iyengar, R. Complexity in biological signalling systems. *Science*, 284:92–96, 1999.

Werner, S L; Barken, D, and Hoffmann, A. Stimulus specificity of gene expression programs determined by temporal control of IKK activity. *Science*, 309:1857–1861, 2005.

Wesche, H; Henzel, W J; Shillinglaw, W; Li, S, and Cao, Z. MyD88: An adapter that recruits IRAK to the IL-1 receptor complex. *Immunity*, 7:837–847, 1997.

West, S; Bridge, L J; White, M R H; Paszek, P, and Biktashev, V N. A method of 'speed coefficients' for biochemical model reduction applied to the NF-κB system. *Journal of Mathematical Biology*, DOI 10.1007/s00285-014-0775-x, 2014.

Weston, A D and Hood, L. Systems biology, proteomics, and the future of health care: Toward predictive, preventative, and personalized medicine. *Journal of Proteome Research*, 3:179–196, 2004.

White, G C and Bennetts, R E. Analysis of frequency count data using the negative binomial distribution. *Ecology*, 77:2549–2557, 1996.

Wildermuth, M C. Metabolic control analysis: Biological applications and insights. *Genome Biology*, 1:103.1–103.5, 2000.

Williams, R A. Blurring the boundaries. *ITNow*, 53(5):16–17, 2011a.

Williams, R A. Agent-based modelling & simulation of the NF-κB intracellular signalling pathway. In *Proceedings of the 4th York Doctoral Symposium*, pages 83–87, http://www.cs.york.ac.uk/ftpdir/reports/2011/YCS/468/YCS–2011–468.pdf. University of York, 2011b.

Williams, R A. Statistical techniques complement UML when modelling domains of complex systems. In *Proceedings of the 5th York Doctoral Symposium*, pages 69–73, http://www.cs.york.ac.uk/ftpdir/reports/2011/YCS/480/YCS–2012–480.pdf. University of York, 2012.

Williams, R A. Spinning plates and juggling balls: Project managing your PhD. *EMBO Reports*, 14(4):305–309, 2013.

Williams, R A; Read, M; Timmis, J; Andrews, P S, and Kumar, V. In silico investigation into CD8Treg mediated recovery in murine experimental autoimmune encephalomyelitis. In Lio, P; Nicosia, G, and Stibor, T, editors, *International Conference on Artificial Immune Systems (ICARIS)*, volume 6825 of *Lecture Notes in Computer Science*, pages 51–54, 2011.

Williams, R A; Greaves, R; Read, M; Timmis, J; Andrews, P S, and Kumar, V. In silico investigation into dendritic cell regulation of CD8Treg mediated killing of Th1 cells in murine experimental autoimmune encephalomyelitis. *BMC Bioinformatics*, 14(Suppl 6):S9, 2013.

Williams, R A; Timmis, J, and Qwarnstrom, E E. A domain model of the IL-1 stimulated NF-κB signalling pathway. *PLoS ONE*, In Preparation, 2014a.

Williams, R A; Timmis, J, and Qwarnstrom, E E. Computational models of the NF-κB signalling pathway. *Computation*, 2:131–158 (In Press), 2014b.

Wirth, T and Baltimore, D. Nuclear factor NF-κB can interact functionally with its cognate binding site to provide lymphoid-specific promoter function. *The EMBO Journal*, 7:3109–3113, 1988.

Woelke, A L; Murgueitio, M S, and Preissner, R. Theoretical modeling techniques and their impact on tumor immunology. *Clinical and Developmental Immunology*, page Article ID 271794, 2010.

Wold, S; Esbensen, K, and Geladi, P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2:37–52, 1987.

Wolkenhauer, O. Systems biology: The reincarnation of systems theory applied in biology? *Briefings in Bioinformatics*, 2:258–270, 2001.

Wolkenhauer, O and Mesarovic, M. Feedback dynamics and cell function: Why systems biology is called systems biology. *Molecular BioSystems*, 1:14–16, 2005.

Wolkenhauer, O; Sreenath, S N; Wellstead, P; Ullah, M, and Cho, K J. A systems- and signal-oriented approach to intracellular dynamics. *Biochemical Society Transactions*, 33:507–513, 2005.

Wooldridge, M. Agent-based software engineering. *IEE Proceedings Software Engineering*, 144:26–37, 1997.

Wright, S D; Ramos, R A; Tobias, P S; Ulevitch, R J, and Mathison, J C. CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science*, 249:1431–1433, 1990.

Xiao, G; Fong, A, and Sun, S C. Induction of p100 processing by NF-κB-inducing kinase involves docking IκB kinase α (IKKα) to p100 and IKKα-mediated phosphorylation. *Journal of Biological Chemistry*, 279:30099–30105, 2004.

Yang, L; Chen, H, and Qwarnstrom, E. Degradation of IκBα is limited by a postphosphorylation/ubiquitination event. *Biochemical and Biophysical Research Communications*, 285:603–608, 2001.

Yang, L; Ross, K, and Qwarnstrom, E E. RelA control of IκBα phosphorylation. *Journal of Biological Chemistry*, 278:30881–30888, 2003.

Yde, P; Mengel, B; Jensen, M H; Krishna, S, and Trusina, A. Modeling the NF-κB mediated inflammatory response predicts cytokine waves in tissue. *BMC Systems Biology*, 5:115, 2011.

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*, 97:4649–4653, 2000.

Yook, S H; Oltvai, Z N, and Barabasi, A L. Functional and topological characterization of protein interaction networks. *Proteomics*, 4:928–942, 2004.

Young, D; Stark, J, and Kirschner, D. Systems biology of persistent infection: Tuberculosis as a case study. *Nature Reviews Microbiology*, 6:520–528, 2008.

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. *Molecular BioSystems*, 2:640–649, 2006.

Zambrano, S; Bianchi, M E, and Agresti, A. High-throughput analysis of NF-κB dynamics in single cells reveal basal nuclear localization of NF-κB and spontaneous activation of oscillations. *PLoS ONE*, 9(3):e90104, 2014a.

Zambrano, S; Bianchi, M E, and Agresti, A. A simple model of NF-κB reproduces experimental observations. *Journal of Theoretical Biology*, 347:44–53, 2014b.

Zandi, E; Rothwarf, D M; Delhase, M; Hayakawa, M, and Karin, M. The IκB kinase complex (IKK) contains two kinase subunits, IKKα and IKKβ, necessary for IκB phosphorylation and NF-κB activation. *Cell*, 91:243–252, 1997.

Zhang, L; Athale, C A, and Deisboeck, T S. Development of a three-dimensional multiscale agent-based tumor model: Simulating gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer. *Journal of Theoretical Biology*, 244:96–107, 2007.

Zhang, X; Shephard, F; Kim, H B; Palmer, I R; McHarg, S; Fowler, G J S; O'Neill, L A J; Kiss-Toth, E, and Qwarnstrom, E E. TILRR, a novel IL-1RI co-receptor, potentiates MyD88 recruitment to control Ras-dependent amplification of NF-κB. *Journal of Biological Chemistry*, 285:7222–7232, 2010.

Zhang, X; Pino, G M; Shephard, F; Kiss-Toth, E, and Qwarnstrom, E E. Distinct control of MyD88 adapter-dependent and Akt kinase-regulated responses by the interleukin (IL)-1RI co-receptor TILRR. *Journal of Biological Chemistry*, 287(15): 12348–12352, 2012.