Modelling the Intracellular NF-κB Signalling Pathway

Mark Pogson

PhD Thesis
University of Sheffield
Department of Computer Science

December 2006

Abstract

NF-κB is a transcription factor which is central to the regulation of genes involved in inflammatory and immune responses. Understanding the operation of NF-κB and its associated intracellular signalling pathway is essential in order to control a wide range of chronic inflammatory diseases, including asthma and rheumatoid arthritis. Abnormalities in the pathway are present in a variety of human cancers and may also affect the pathogenesis of Alzheimer's disease.

Computational modelling of the signalling pathway is necessary to overcome the practical limitations of biological experiments and to facilitate a more comprehensive understanding of the system. The thesis begins by outlining existing understanding of the NF-κB signalling pathway, which in conjunction with a review of modelling methods is used to inform a different approach to model the pathway, using computational agents to represent individual molecules and receptors. The agent-based model is tested with well-understood chemical reactions before being used to describe the pathway. This provides a good appreciation of the system as a whole, offering a detailed description of events at every step in the pathway and allowing investigation of stochastic, spatial and structural parameters.

The modelling process and simulation help to provide a prediction about the role of actin filaments of the cytoskeleton in regulating the unstimulated pathway; this is quantitatively validated by biological experiment. The effect of cell shape on the pathway is also investigated.

Acknowledgements

Thank you foremost to my supervisors Mike Holcombe, Rod Smallwood and Eva Qwarnstrom for their continued advice, support and guidance.

Thank you also to many others, including Simon Coakley, Phil McMinn and Alan Waterworth for helping me with my computer; Dawn Walker for her modelling advice and for reading my draft thesis; Natasha Savage for also reading my draft thesis; Simon Scarle for his technical advice and suggestions; Ian Palmer and David Anderson for their experimental work and biological discussions; Steven Wood for his work on cell morphing; and everyone else who has helped over the last 3 years. Thanks again.

Declaration

The work presented in this thesis is original work undertaken between October 2003 and August 2006. Part of this work has been published in:

- M. Pogson, R. Smallwood, E. Qwarnstrom, M. Holcombe. Formal agent-based modelling of intracellular chemical interactions. *Biosystems* 85:37-45, 2006.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

Understanding the organisation and dynamics of the components of living systems is a fundamental challenge for the life sciences [CW03]. Systems biology seeks to address this by explaining biological processes in terms of the interactions of the components of the system, from molecules to cells, tissues, organs and organisms [Liu05, MN04].

Central to the operation of living systems are the complex interactions of genes, proteins and other molecules within the cell that regulate and determine cell behaviour [AJL+02, Nel04, LBZ+00]. A primary aim of systems biology is to understand better these interactions and consequently the mechanisms that determine the function and behaviour of cells.

Rather than simply catalogue and characterise the physical components of the cell, systems biology aims to gain a better understanding of cellular functional activity by collating this information in mathematical and computational models based on experimental data. This creates the possibility of investigating the cell in far greater detail than is achievable by practical experiment alone, providing a means to test and predict possible mechanisms that underlie observed behaviours.

Systems biology is a burgeoning research field which seeks to piece together biological understanding into a functional whole [Kit02a, Kit02b, Nob02]. Although system-level thinking about biological processes is not in itself new [HN63], only recently has increased computing power enabled realistically detailed simulation of large systems. The physiome project is a systems biology-based attempt to provide a framework to computationally model the human body, from proteins to organs [HRN02, HB03, Bas00]. Molecular in-

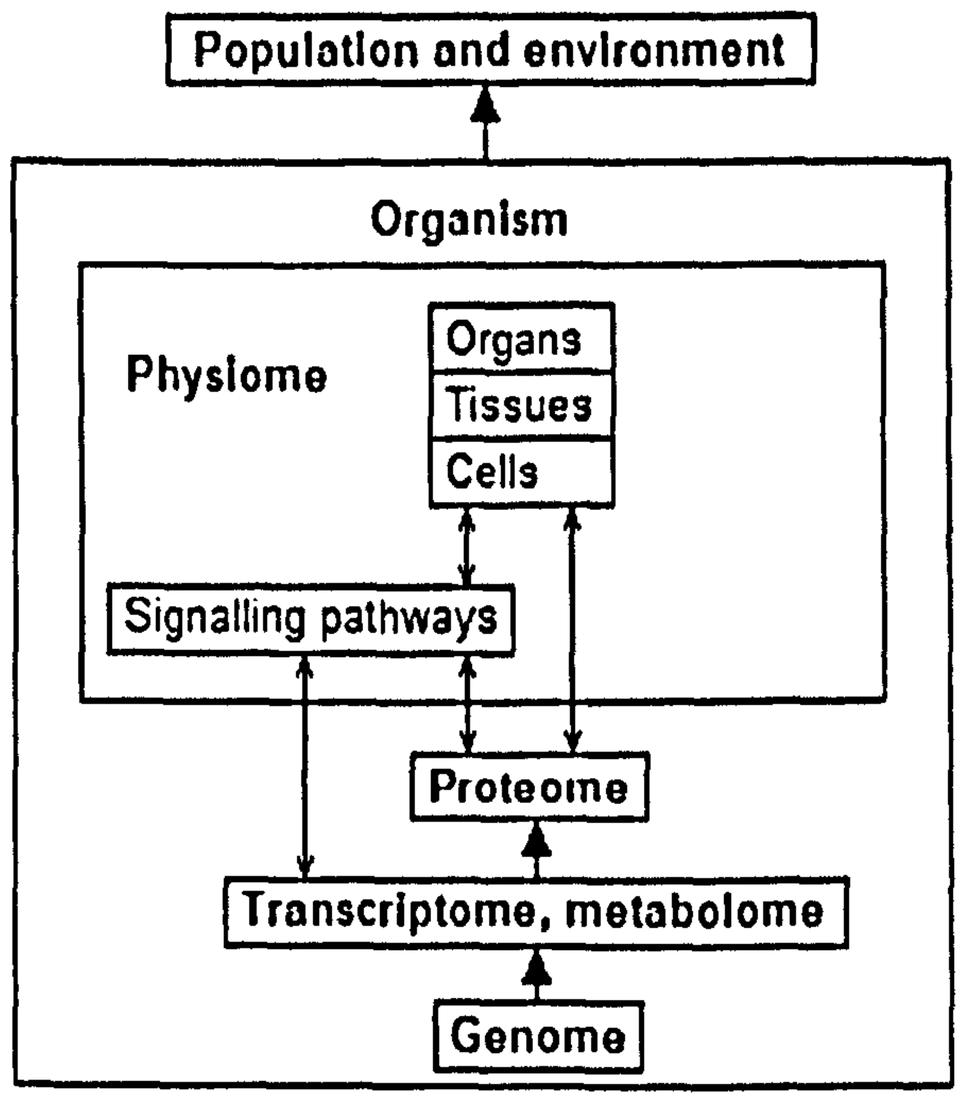

Figure 1.1: **Biological organisation and the physiome.** The physiome aims to model the organism at various levels, from molecular interactions to organ function. Other areas of biological organisation shown are the genome (genes encoded in DNA), transcriptome (messenger RNA produced by gene expression), metabolome (metabolites that are present) and proteome (proteins produced and where they reside). Above the organism level are population organisation and environmental interaction. Many processes are omitted. Figure adapted from Hunter and Borg [HB03] and Bassingthwaighte [Bas00].

teractions within the cell form an important foundation for such modelling, combining proteins into living systems, and tying together gene expression with cell, tissue and organ behaviour, as illustrated in Figure 1.1.

Intracellular signalling pathways are chains of molecular interactions that occur within the cell. Signals received from outside the cell are transmitted by these pathways to the nucleus to provide specific genetic responses. Signalling pathways are thus vital to the regulation of cellular behaviour, and are therefore central to the challenge of understanding the organisation and dynamics of the cell.

In the past, the biochemistry of intracellular signalling pathways has been oversimplified [TRRG02]. Signals do not propagate through simple linear chains of proteins, and proteins do not work individually but instead assemble fleetingly into molecular machines. Furthermore, signalling pathways can interact with each other and with the structure of the cell itself. To try to understand all these factors requires suitably detailed modelling techniques that account for both spatial and temporal considerations.

1.2 Background to the NF-κB Signalling Pathway

Nuclear factor κB (NF-κB) is a transcription factor, which in combination with its signalling pathway is responsible for regulating genes involved in inflammation and immune responses [KG95, BB96, CLB98]. The NF-κB signalling pathway is also central to the regulation of genes involved in cellular proliferation and apoptosis (cell death). As such it is associated with the pathogenesis of many chronic inflammatory diseases, including asthma and rheumatoid arthritis [TF01]. Abnormalities in the NF-κB signalling pathway are often present in a variety of human cancers and may also be involved in the pathogenesis of Alzheimer's disease [YG01].

The intracellular NF-κB signalling pathway is a crucial survival capability present in many mammalian cells, with related pathways present in various other species and plants. It is part of the innate immune system, which attacks generic classes of foreign molecules, as well as facilitating the adaptive immune system for the targeted attack of recognised pathogens [O'N05]. The pathway is also involved in responses to physical injury. Its modelling is not only important with regard to higher level cell and tissue modelling, but is also essential in order to gain a better understanding of how it can be controlled, helping to overcome the practical limitations of biological experiments [Bha03b].

The intracellular signalling pathway can be triggered by various extra-cellular chemical signals as well as by mechanical stress. In its inactive state, NF-κB exists in the cytoplasm bound by its inhibitor protein IκB, as in Figure 1.2. Following 'classical' activation of the signalling pathway by one or more of a range of extra-cellular stimuli, IκB is phosphorylated (gains a phosphate group) and then ubiquitinated (tagged for destruction), leaving NF-κB free to move (or 'translocate') to the nucleus and induce transcription of pro-inflammatory genes [YRQ03, YCQ01, CKTC$^+$01, CDQ01]. Activation also causes the expression of IκB, thus regulating the signalling pathway through negative feedback [SFL$^+$93].

1.3 Background to Modelling the Pathway

Models are simplified representations of reality that are used to enable or improve understanding; these include linguistic, graphical, experimental, statistical and mathematical descriptions. Different modelling methods are useful for different purposes, from a simple verbal explanation of events to a quantitative mathematical description and analysis.

Various modelling approaches have been used to investigate different features of signalling pathways [Bha03a, Bha03b]. The choice of modelling approach depends on the type

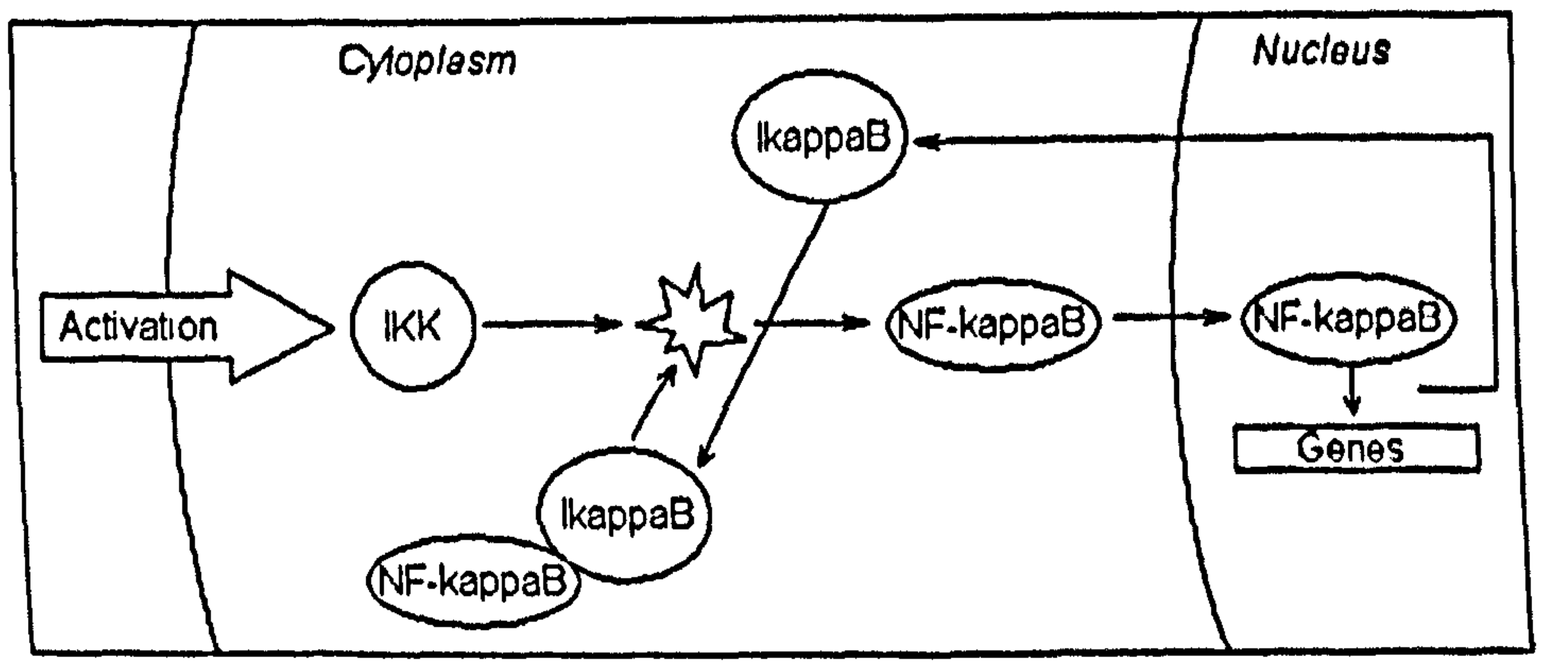

Figure 1.2: **Diagram of key interactions in the negative feedback of the NF-κB signalling pathway.** NF-κB is held inactive in the cytoplasm by its inhibitor IκB. Activation of the pathway causes IκB-kinase (IKK) to initiate the destruction of IκB, freeing NF-κB to translocate to the nucleus and activate appropriate genes. The resulting expression of IκB returns the system to its inactive state.

and detail of the required results, in addition to what information and data are available to construct the model; this depends both on the complexity and scope of the pathway as well as on the quality and quantity of experimental observations. Modelling is not only useful as a means of making sense of observations, but also in raising detailed questions that are necessary for the purposes of constructing the model—and are key to the operation of the system—but which might otherwise be overlooked [JHR04b, RJHR05].

Pathway modelling can be categorised into static and dynamic methods. Static models are concerned with the topology of interaction networks, investigating how signals propagate as a sequence of interactions [Bra03, PHPS05, BCT+04a]. Such models are useful for providing an overview of the behaviour of a pathway rather than investigating specific details of its operation. To understand the functional behaviour of pathways, dynamic modelling is necessary.

Previous dynamic pathway modelling has often relied on differential equations derived from reaction kinetics to describe chemical concentrations with time, treating the cell and its constituents as a continuum [CW03, HLSB02]. Although such models are able to provide useful results, they are limited in their scope for a variety of reasons.

A major concern is that the cell is not a well-mixed bag of chemicals, instead containing internal structure and low numbers and non-uniform distributions of certain key molecules [BT03, BBJ+03]. Ordinary differential equations (ODEs) do not account for this inhomogeneity; although temporal modelling is clearly essential in order to understand the

dynamics of signalling pathways, it may be insufficient in isolation, and the inclusion of spatial information could be equally vital in certain situations due to the nature of pathway operation. Although ODE models can be decomposed into separate compartments, such as a distinct cytoplasm and nucleus, spatial resolution is still evidently limited [EI04]. Greater spatial detail can be incorporated by the use of partial differential equations (PDEs), which are functions of both time and space [MSSL02]. However, PDE models can be hard to construct and solve.

Signalling pathways are composed of sequences of discrete stochastic interactions [EL00], as opposed to the mass-action determinism of a differential equation model. Due to the low numbers of certain critical molecules in many pathways, this can be of great significance [ELSS02, ARM98, MA97]. The issue has been addressed by the use of stochastic differential equations [BT03, BBJ+03, LPB+06], though this does not account for spatial concerns. Other modelling techniques, including petri nets [GP98, HKV01], process algebras [Mil99, RSS01, CDPT03, CVGO, CGH] and cellular automata [Wei02a, Wei02b], have been used to describe signalling pathways, each with their own advantages and disadvantages.

Models of signalling pathways are strongly informed by insights obtained from biological experiments, which themselves are models of the biological systems in question [VAB04]. While this is unavoidable, different modelling techniques depend to different degrees on experimental data, so comparison of models not only with biological results but also with other models can minimise over-reliance on any one source of information.

Most modelling methods that are used to investigate signalling pathways are abstract mathematical constructions with little visual output other than graphical representations of results. As models are simply a means of making information more easily understandable, a visual representation of the pathway in action is a basic but effective feature of a model, requiring not just temporal but also spatial graphing of chemical concentrations to form a movie. Although it is very much a secondary requirement, a clearer picture of pathway operation is a desirable aspect of any model, providing both a simple means to debug the model (since much incorrect behaviour will be immediately obvious) as well as an instantly understandable, if not quantifiable, representation of pathway dynamics.

1.4 Aims of Thesis

The overall objective of the thesis is to gain an insight into the intracellular NF-κB signalling pathway. More precisely, the thesis will investigate the pathway by using mathematical and computational modelling, focussing on spatial aspects of the pathway, with the following specific aims:

1. To identify the advantages and disadvantages of different modelling approaches, and consequently select the most appropriate for the purposes of modelling the dynamics of the NF-κB signalling pathway. The model must incorporate spatial detail and include scope for modelling biomechanical stimulation, as these are both vital to the operation of the pathway but are neither well understood nor modelled.

2. To verify the model with existing models and subsequently test it against biological results for well understood operation of the pathway.

3. To use the model to investigate new areas of pathway operation, with the intention of forming testable predictions.

In order to achieve the major modelling aims, the model must fulfil the following requirements:

- To be extensible in its formulation, so that extra detail can be added without changing the rest of the model.

- To be sufficiently flexible in its description so that abstractions can be made about areas of the pathway that are less fully understood or are of lower importance.

Although not essential for successfully modelling the pathway, further desirable features of the model are:

- To present a clear and immediately understandable picture of the pathway in operation, in the form of a movie.

- To be composed from detailed mathematical structures that can be investigated by techniques developed in computer science.

- To be in a format that can be used by other modellers, with scope for manipulation of the model by biologists.

1.5 Contributions of Thesis

The thesis contributes the following:

1. The main biological features of the NF-κB signalling pathway are outlined and the need for its modelling established.
2. Appropriate modelling methods are identified and previous attempts reviewed.
3. A new modelling approach is proposed and developed to offer alternative benefits to existing work.
4. The new model is verified at the level of chemical interactions and pathway operation.
5. The model is used to investigate the pathway in detail, confirming assumptions of the pathway and raising a new question about its operation.
6. The model provides a quantitative prediction about the pathway, which is confirmed through biological experiment.
7. The model is used to investigate physical effects on the pathway.
8. The model is reviewed and conclusions about modelling the pathway are drawn.
9. Appropriate directions for future work are suggested.

1.6 Outline of Thesis

The thesis is organised as follows:

Chapter 2: a literature review of the NF-κB pathway. The main details of the operation of the NF-κB pathway are outlined. Different levels of organisation are considered, from the construction of proteins to the entire pathway. Details of protein and enzyme interactions are described; NF-κB and IκB proteins are considered in further detail; signal transduction through surface receptors is noted; nuclear transport is described; the cytoskeleton and physical effects are discussed; and gene transcription and translation are considered. Biological experimental procedures are summarised and significant recent findings reported.

Chapter 3: a literature review of suitable modelling methods, an analysis of major existing models, and an outline for the direction of modelling work in the thesis. The

main modelling approaches are described, including reaction kinetics, stochastic differential equations, petri nets, process algebra and cellular automata. Major existing models are reviewed, including that of Hoffmann et al. [HLSB02]. Agent-based modelling is presented in detail, describing its background, finite state machines, X-machines, previous agent-based modelling, and how agents can be applied to model the NF-κB signalling pathway.

Chapter 4: agent-based modelling of chemical interactions. A method for modelling chemical interactions is presented and tested against reaction kinetics in appropriate situations. The model is then investigated in detail, establishing both its scope and limitations.

Chapter 5: agent-based modelling of the NF-κB signalling pathway. The chemical model is used as a basis for modelling the pathway, with additional details described. The model is successfully verified against biological results for pathway dynamics and equilibrium. Effects of changing the levels of surface receptors and signalling intermediates are investigated in the model, which are in agreement with existing understanding about the pathway. NF-κB:IκB ratios in the pathway are investigated with the model, which raises a question about experimental observations; this provides direction for further investigation of the model.

Chapter 6: modelling physical effects of the pathway. The question about NF-κB:IκB ratios in the pathway is investigated further, and a proposed mechanism based on existing knowledge is tested in the model. This provides a quantitative prediction about the pathway which is confirmed experimentally. The model is then used to investigate the effect of different cell shapes on the pathway.

Chapter 7: conclusions and further work. Conclusions about the pathway and how the pathway is modelled are presented. Further work based on these conclusions is suggested.

Appendix: program code.

Chapter 2

Biology of the NF-κB Signalling Pathway

2.1 Motivation

Modelling the NF-κB signalling pathway depends strongly upon existing biological knowledge. If the model is to be successful, it must be informed by biology as far as possible. However, it is also important to consider the necessity and relevance of available biological information in relation to the level at which the system is to be modelled, the aims of the model and its underlying assumptions.

The model must not simply include all available information but make sense of it as straightforwardly as possible. Furthermore, the model must be sufficiently computationally inexpensive to permit reasonable execution times. The biology of the signalling pathway must therefore be investigated at different levels of detail in order to determine how and what the model should describe.

2.2 Proteins

2.2.1 Protein Structure

NF-κB and most other major molecules in the intracellular NF-κB signalling pathway are proteins. Protein molecules are constructed from a group of twenty different smaller molecules called amino acids [AJL+02, Nel04, LBZ+00]. Every amino acid has the same basic backbone, with a 'carboxyl' group at one end and an 'amino' group at the other,

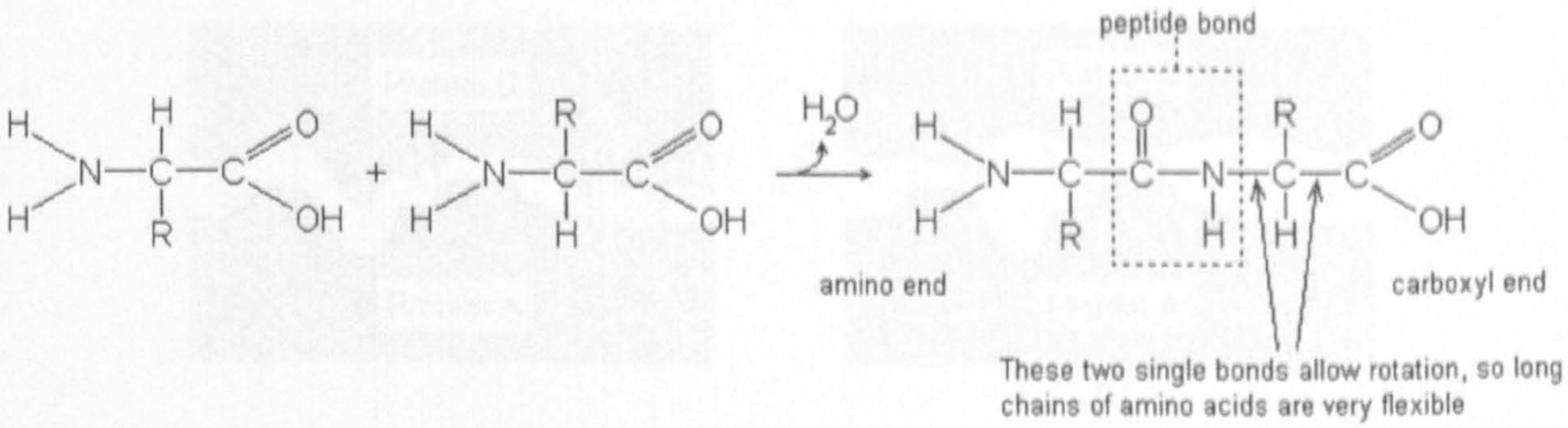

Figure 2.1: **Molecular structure diagram of the bonding of two amino acids (peptide bonding).** The side group (generically denoted by R) determines the identity of the amino acid. Proteins are made from long chains of up to thousands of amino acids. Figure adapted from Alberts et al. [AJL+02] and Lodish et al. [LBZ+00].

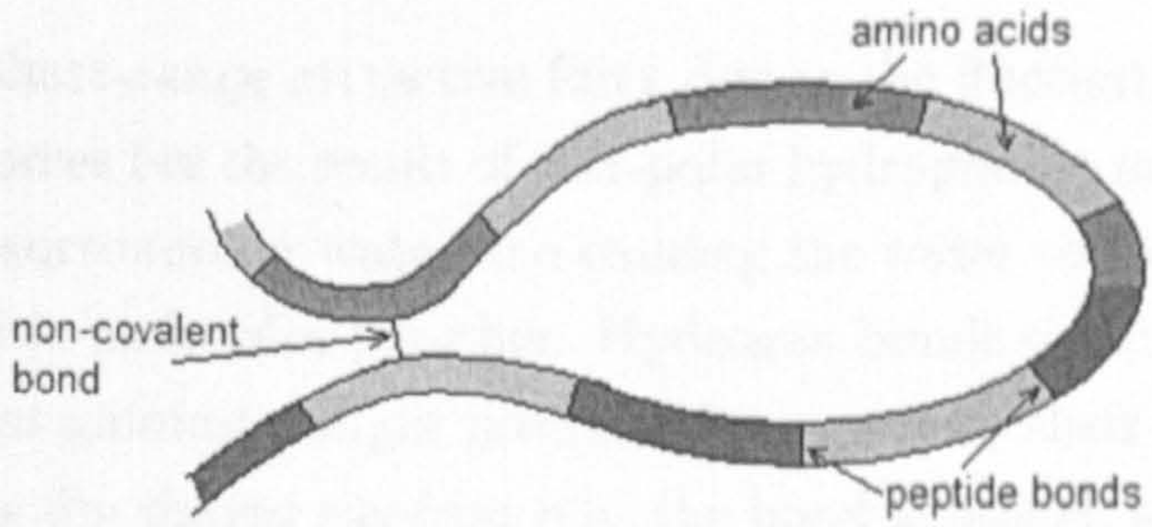

Figure 2.2: **Representation of an interaction giving rise to protein conformation.** A section of a protein molecule is shown, composed of several amino acids that fold and form a weak bond that contributes to the overall shape of the protein. Figure adapted from Alberts et al. [AJL+02].

which act as a plug and socket to bond, as shown in Figure 2.1. Amino acids form long chains through covalent peptide bonds to form proteins, which are generally between fifty and two thousand amino acids long.

Each protein has its own particular three-dimensional structure, which is determined by the sequence of its amino acids; because of the flexibility of an amino acid chain, the protein is able to fold and minimise its free energy, interacting with itself through non-covalent bonds, as in Figure 2.2. Each of these interactions is very weak, but the net effect of several simultaneous non-covalent bonds means the protein is firmly held in a very specific shape.

The principal types of non-covalent bonds that hold such macromolecules in shape are Van der Waals attraction, hydrophobic forces, hydrogen bonds, and ionic bonds. Van der

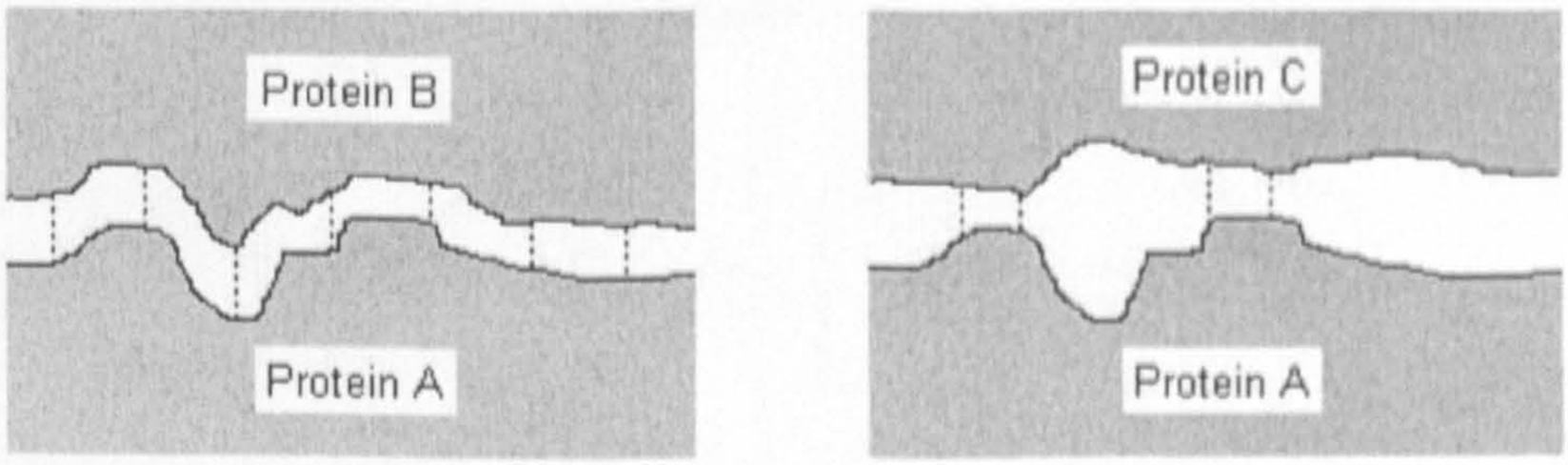

Figure 2.3: **Multiple weak bonds forming a stable complex.** The surface of Protein A fits closely with the surface of Protein B, with many weak bonds forming a stable complex. In contrast, the surface of Protein A does not fit well with the surface of Protein C, with few weak bonds making binding unlikely. Figure adapted from Alberts et al. [AJL+02] and Lodish et al. [LBZ+00].

Waals attraction is a short-range attractive force due to the fluctuating dipole moments of nuclei. Hydrophobic forces are the result of non-polar hydrophobic molecules distorting the usual structure of the surrounding water and causing the water to counteract this effect by pushing the hydrophobic molecules together. Hydrogen bonds occur because of covalently bonded hydrogen atoms gaining a slight positive charge due to their bound partner having a greater attraction for the shared electron (i.e. the bond is polar); a hydrogen bond forms when a similarly slightly negatively charged atom is electrically attracted to the slight positive charge of the hydrogen. Lastly, ionic bonds are the attraction of two oppositely charged ions.

2.2.2 Protein Interactions

The same non-covalent bonds that give proteins their three-dimensional structure (or 'conformation') also allow proteins to bind to each other and to other molecules [AJL+02, Nel04, LBZ+00]. Any region of a protein surface that can interact with another molecule through non-covalent bonds is called a binding site. The shape of two interacting molecules' binding sites must be a very close fit to form a stable complex, as displayed in Figure 2.3.

Proteins are thus able to interact, and in conjunction with other molecules can act as a means of communication inside the cell. Key to protein communications are transcription factors—proteins that bind to DNA, activating the transcription of specific genes to messenger RNA (see Section 2.5).

Molecules in the cell encounter each other very frequently due to their continual random

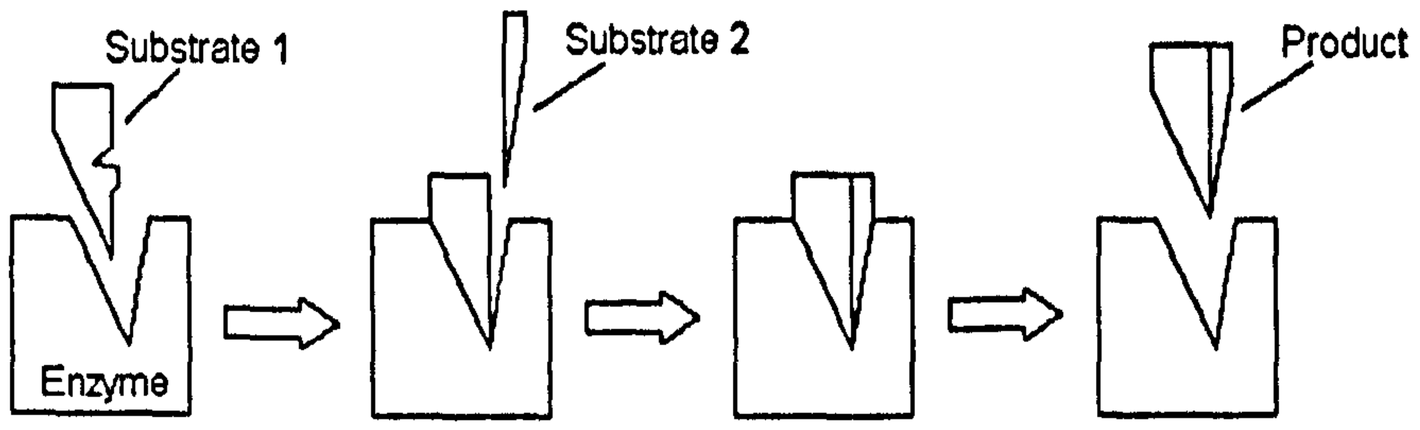

Figure 2.4: **Representation of enzyme action.** An enzyme mediates the combination of two substrates. The substrate binds to the enzyme, thereby undergoing a conformational change. The second substrate then binds to the enzyme, and the resulting product detaches from the enzyme. Figure adapted from Nelson [Nel04].

thermal movements. When two colliding molecules have poorly matching surfaces, few non-covalent bonds are formed between them, and the two dissociate almost immediately. Conversely, if two colliding molecules fit well and form many bonds, they can stay together for a very long time, remaining bound until they are separated by either thermal motion or a molecular signal. Dimers are molecules formed from the non-covalent combination of two similar or identical molecules (forming hetero- and homodimers respectively), which exist together indefinitely.

2.2.3 Enzymes

Large proteins generally consist of several distinct domains—structural subunits whose conformations are essentially independent of each other [AJL+02, Nel04, LBZ+00]. Large proteins with many domains are able to perform elaborate functions.

Enzymes are large proteins that are of vital significance to cellular activity. Enzymes catalyse chemical reactions in the cell by holding molecules (substrates) in such a way that it lowers the energy required for them to interact, as in Figure 2.4. Enzymes are not used up or altered in the process, and work by causing conformational changes in the substrates.

Enzymes can be categorised according to the effect they have on their particular substrates. Kinases and phosphatases are two types of enzyme that are crucial to regulating the activity of many cellular proteins, including those in the NF-κB pathway.

Kinases catalyse the transfer a phosphate group (a compound of phosphorous and oxygen) from an ATP (adenosine triphosphate) molecule to a particular region of a protein [Wil88]. The process is called phosphorylation, and to occur, the target protein must possess a specific binding site called a phosphorylation site. The addition of a phosphate

group to the protein can cause major conformational changes in the protein due to the double negative charge of the phosphate group. This affects the binding sites on the protein, and can also form part of a new binding site, hence changing the behaviour of the protein.

Phosphate groups are attached to one another by high-energy bonds, which in the case of ATP are formed by the metabolism. The release of one of the three phosphate groups in ATP is an energetically favourable reaction; ATP thus acts as an energy store that allows the execution of otherwise energetically unfavourable or impossible reactions in an organism.

Phosphatases perform the opposite function to kinases: they catalyse the removal of a phosphate group from a protein, returning the protein to its pre-phosphorylated state. Kinases and phosphatases can therefore act as 'on' and 'off' switches for actions of certain proteins.

2.2.4 Nuclear Factor κB and its Inhibitor

NF-κB transcription factors are hetero- or homodimers formed from a family of proteins composed of two subfamilies: those containing DNA activation domains, of which there are three proteins, and those lacking, of which there are two proteins [CCDQ99]. Each dimer combination activates its own characteristic set of genes. The prototypical NF-κB is a heterodimer of p65-RELA (containing an activation domain) and p50 (lacking an activation domain). Most NF-κB dimers contain DNA activation domains, although there exist those without.

In the inactive state of the signalling pathway, NF-κB is bound by its inhibitor molecule IκB, preventing its association with DNA and causing its localisation to the cytoplasm by masking the relevant binding sites on its surface. There are at least seven IκB isoforms (isoforms are molecules with the same or similar functions but the product of different genes), though the activity of most NF-κB in mammals is largely controlled by just three, namely IκBα, -β, and -ϵ, of which IκBα is the most critical [LPB+04].

NF-κB is so named because the family of proteins were first identified through their interaction with enhancers (regulators of DNA transcription) of the κ light chain (a particular sequence of amino acids) in B cells (a type of lymphocyte, or white blood cell, that makes antibodies) [SB86].

2.3 Signal Transduction

Signal transduction is the conversion of a signal from one physical form to another. In the cell, this is the conversion of an extracellular signal into an intracellular signal without physical flow through the plasma membrane, and it results in specific cellular responses [AJL+02, LBZ+00, HP96, Han97].

2.3.1 Cell Surface Receptors

There are three main signal transduction cell surface receptors in the NF-κB pathway: toll-like receptors (in particular toll interleukin-1 receptors, or TIRs), protein tyrosine kinase receptors, and adhesion receptors, each of which transmits different forms of extracellular signal into the cell, as shown in Figure 2.5.

TIRs are activated by cytokine signalling molecules interleukin-1 (IL-1), which are produced by cells of the innate immune system. Upon these soluble agonists binding to the extracellular domain of a TIR, conformational changes are induced in the intracellular domain of the receptor, thus relaying the signal into the cell by making it possible for certain receptor complex component proteins to bind. The receptors continually move around the plasma membrane.

2.3.2 Cytoplasmic Pathway

IL-1 activation is a main channel of the NF-κB pathway [YRQ03, YCQ01, CKTC+01, CDQ01]. Following activation of an IL-1 receptor, scaffold proteins MyD88 and TRAF6 bind to the inner surface of the receptor. These adaptor proteins then bind IRAK 1/2/3/4 kinases and phosphatases together in a functional complex. The resulting intermediate interactions lead to the phosphorylation and activation of IκB kinase (IKK).

Active IKK in turn phosphorylates IκB, causing small ubiquitin molecules to covalently bind to the IκB. This results in the destruction of IκB by a proteasome—a large cylindrical protein that degrades ubiquitin-tagged proteins in an ATP-dependent process. The newly released NF-κB is then free to translocate to the nucleus and initiate gene transcription.

Although structurally unrelated, protein tyrosine kinase receptors operate in much the same way as TIRs in response to their relevant cytokines, although the resulting cytoplasmic protein interactions that lead to the activation of NF-κB are somewhat different and not discussed here. Adhesion receptors activate in response to cell adhesion and mechanical

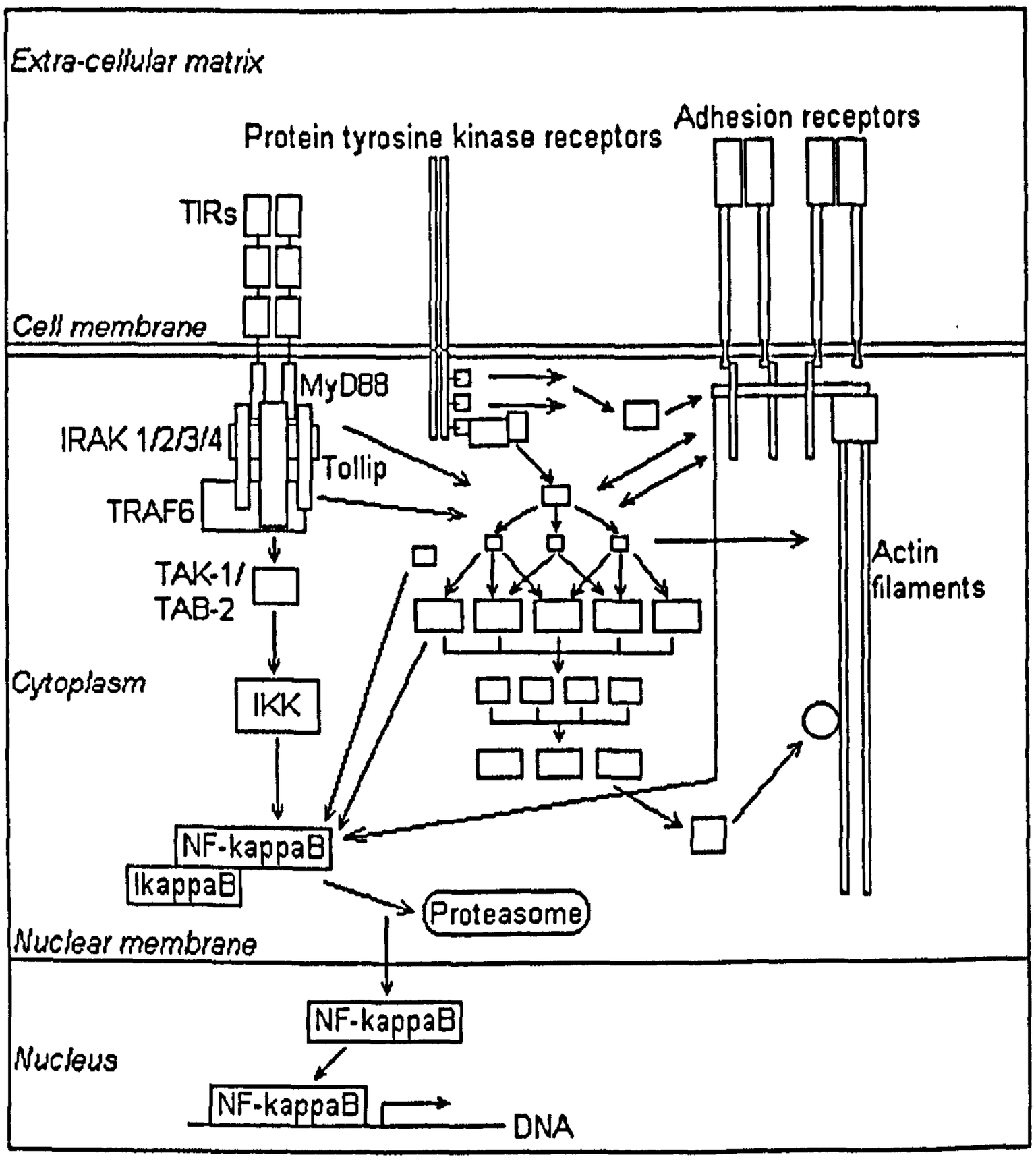

Figure 2.5: **Signal transduction in the NF-κB pathway.** The pathway is triggered by either cytokine activation of the TIRs or tyrosine kinase receptors, or by cell adhesion or mechanical stimulation of the adhesion receptors. Activation of TIRs results in the linear pathway on the left of the diagram. The other receptors trigger complex cross-signalling networks involving many proteins, some of which are included in the diagram, though most are unlabelled for simplicity. The target of the pathway is the activation of NF-κB, which translocates to the nucleus and activates appropriate genes. Figure adapted from Qwarnstrom [Qwa03].

stress; the resulting pathway is less well understood, although nonetheless vital to the system. The cytoskeleton (a system of protein filaments inside the cytoplasm giving the cell its shape and ability to move) is able to bind proteins, and changes in its conformation may affect protein binding; this may be vital to the activation of the pathway by mechanical stress (see Section 2.4).

2.3.3 Nuclear Transport

NF-κB is transported into the nucleus by nuclear import receptors [AJL+02, LBZ+00]. While very small molecules can freely diffuse through the nuclear pore complexes on the nuclear membrane, and medium-sized molecules can enter directly through active nuclear pores, large molecules like NF-κB require nuclear import receptor molecules to bind them as cargo to transfer them through nuclear pores (see Figure 2.6). Nuclear import receptors exist in the cytoplasm awaiting molecules to transport. When NF-κB is separated from IκB, a particular binding site called its nuclear localisation signal is exposed; this is recognised and bound by an appropriate nuclear import receptor. The nuclear import receptor then binds to the nucleoporins that make up the nuclear pore complex, and a GTP-dependent process (GTP is similar to ATP) 'walks' the NF-κB into the nucleus.

Free IκB molecules can be imported into the nucleus in the same manner (their nuclear localisation signal is mutually masked by NF-κB). Nuclear export of the molecules works similarly, but requires nuclear exporting receptors (which exist in the nucleus) instead of nuclear importing receptors. Although nuclear pores allow passage through in both directions, it is not properly understood how congestion and collisions are avoided. The energy for the GTP-dependent import and export process comes from Ran, a molecule that exists in two different states according to whether GTP or GDP is bound, thus coordinating import and export (see Figure 2.7).

2.3.4 Pathways, Networks and Modules

The NF-κB signalling pathway is sometimes also referred to as the NF-κB signalling module [HLSB02], since it can be conceptualised as a modular entity in a larger network of interactions [YYGK02]; the NF-κB signalling pathway can interact with many other pathways such as the MAP kinase pathway [Moy05]. However, usage of the terms 'network' and 'module' is often contextually dependent. The term 'pathway' appropriately suggests a definite input, a chain of interactions, and a definite output, and is a widely understood

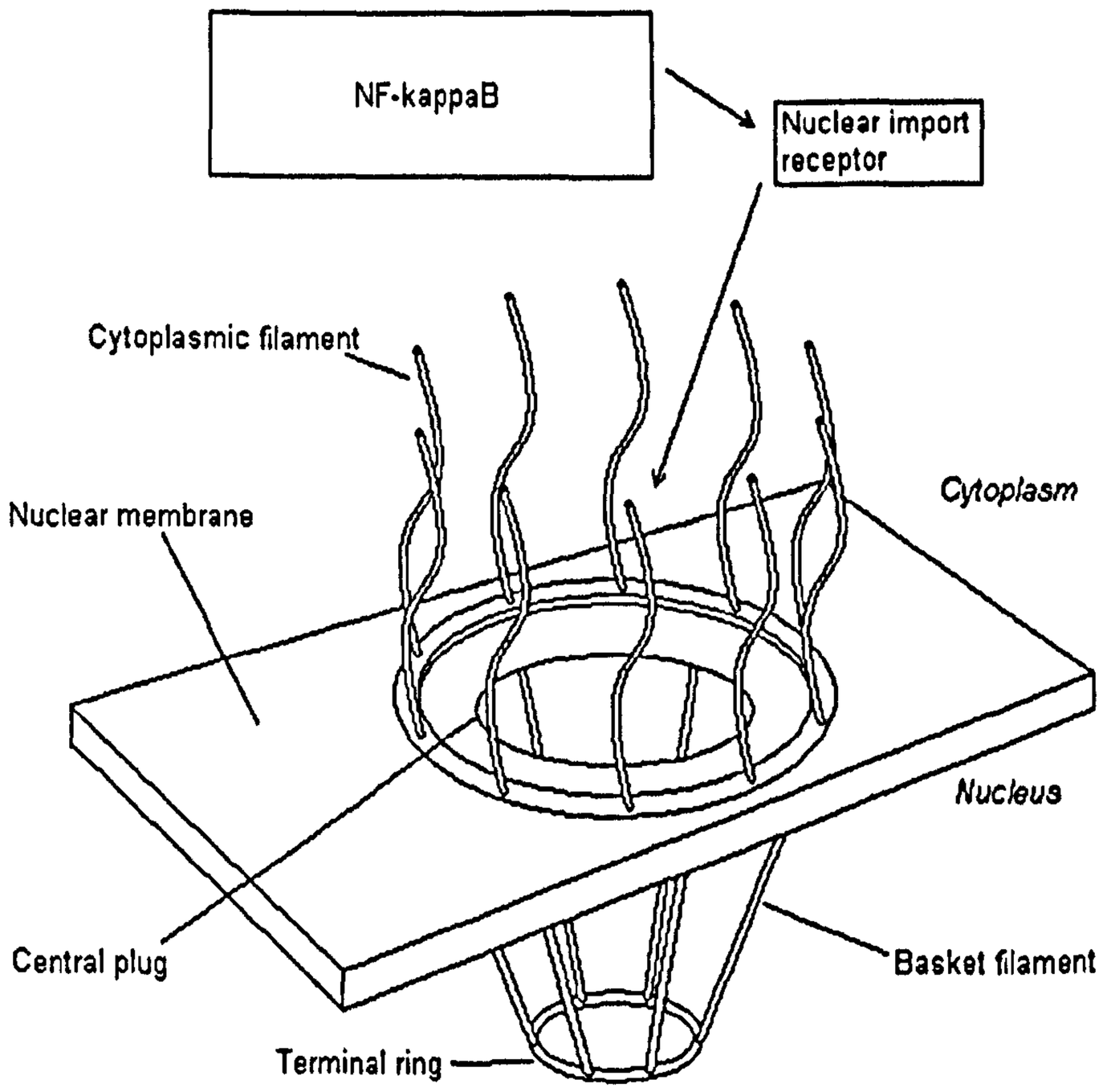

Figure 2.6: **NF-κB import through a nuclear pore complex.** While small molecules can move freely through the nuclear pore, and medium sized molecules require the complex to actively transport them through, large molecules such as NF-κB require a nuclear receptor molecule to move them through the complex in a GTP-dependent process. Figure adapted from Alberts et al. [AJL+02] and Lodish et al. [LBZ+00].

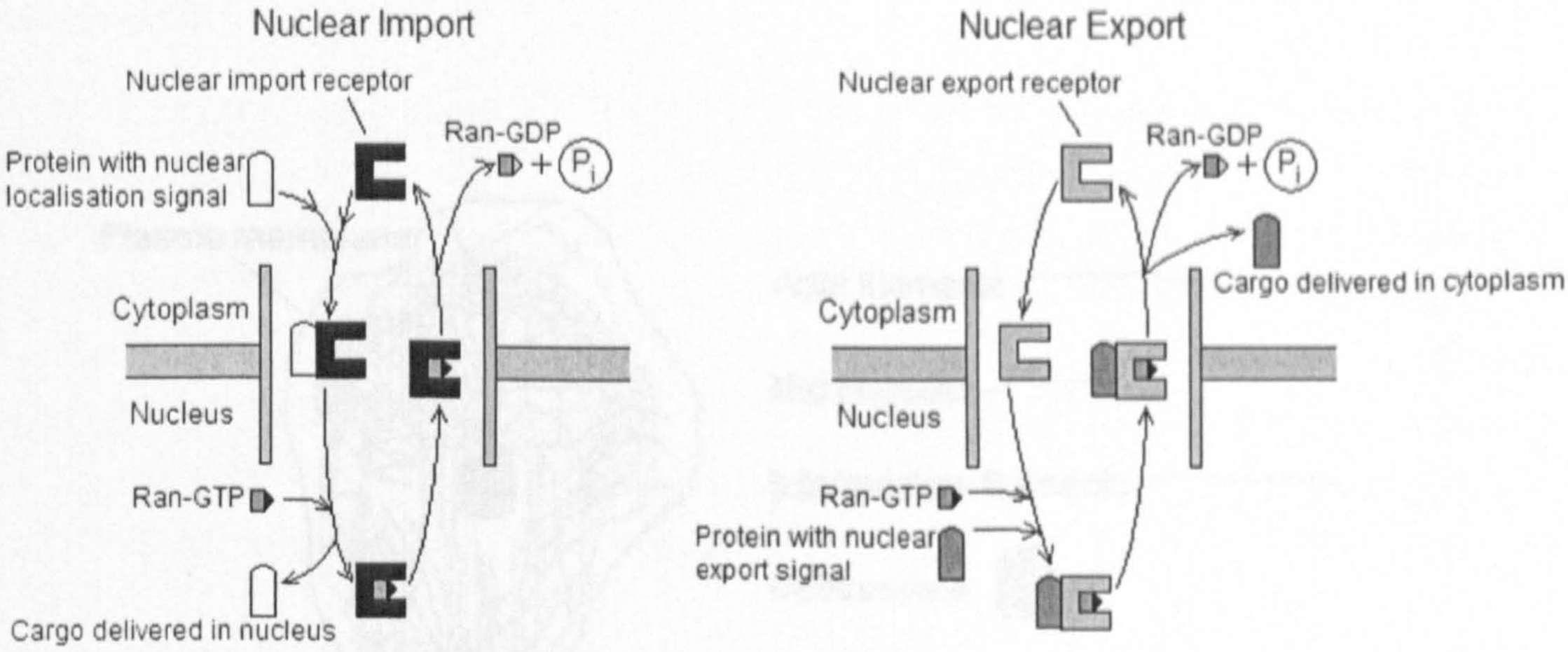

Figure 2.7: **Representation of the involvement of Ran in nuclear transport.** Energy for the transport of the protein cargo comes from the release of a phosphate group (P_i) from GTP, which is mediated by a further reaction, which is not shown. Figure adapted from Alberts et al. [AJL+02].

description of the system.

2.4 The Cytoskeleton

The cytoskeleton is a complex structure with many functions which exists inside the cell [AJL+02]. It provides spatial organisation within cells and a means of mechanical interaction with their environments; it allows cells to change shape and move, and facilitates the rearrangement of internal components of the cell during growth, division and changing circumstances. Although not all aspects of the cytoskeleton have a direct relation to the operation of the NF-κB pathway, certain features of the cytoskeleton have a crucial impact on particular mechanisms in the pathway, and must be understood in order to explain specific behaviours of the pathway. The varied functions of the cytoskeleton depend on the behaviour of its different components, the main types of which are discussed below.

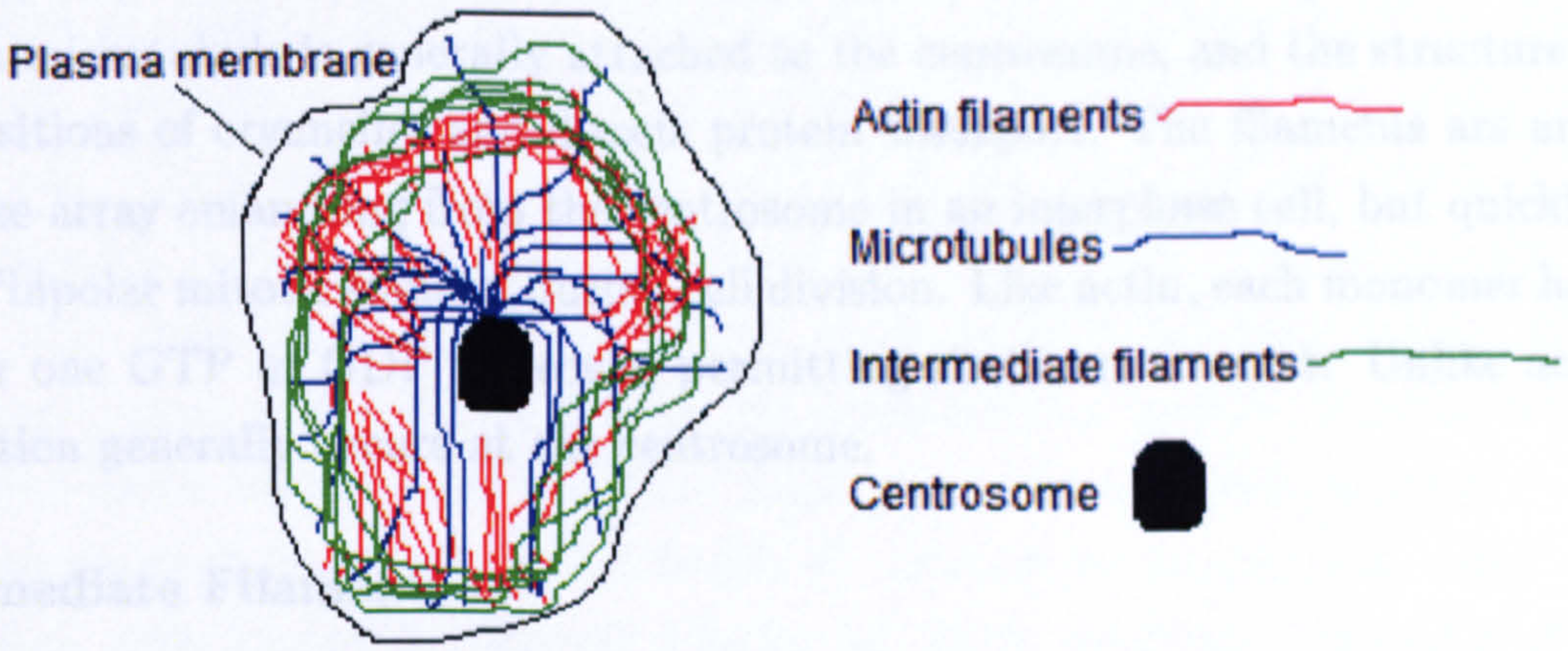

Figure 2.8: **Diagram of main components of the cytoskeleton and their general distribution in a typical cell.** The cell is an interphase cell, i.e. it is not undergoing mitosis (cell division). Figure adapted from Alberts et al. [AJL+02].

2.4.1 Major Types of Protein Filaments that form the Cytoskeleton

The cytoskeleton is composed of three main types of protein filaments—actin filaments, microtubules and intermediate filaments—which are in part organised by the centrosome, as shown in Figure 2.8.

Actin Filaments

Actin filaments are flexible protein polymers with a diameter of 5-9nm [AJL+02]. They exist throughout the cytoplasm, especially near the plasma membrane. The filaments determine cell shape, and are necessary for cell motility; the filaments can form projections from the cell surface to move the cell around. Actin filaments also line the inner face of the nuclear envelope, providing protection for DNA. Actin nucleation (where sufficient numbers of monomers are bound such that polymer formation becomes more rapid) often occurs at the cell cortex (the region near the plasma membrane), and is often regulated by external signals. Each monomer (or 'G-actin' (globular); 'F-actin' is the polymer (filamentous)) has a binding site for one ATP or ADP molecule, permitting a form of movement of the filament, the description of which is unnecessary here.

Although actin can attach to proteins embedded in the plasma membrane, most attachments are through other proteins (such as talin), which attach to proteins embedded in the membrane.

Microtubules

Microtubules are long hollow cylinders with an outer diameter of 25nm made from tubulin proteins α-tubulin and β-tubulin. The filaments are much more rigid than actin. One end of each microtubule is generally attached to the centrosome, and the structure determines the positions of organelles and directs protein transport. The filaments are arranged in a star-like array emanating from the centrosome in an interphase cell, but quickly rearrange into a 'bipolar mitotic spindle' during cell division. Like actin, each monomer has a binding site for one GTP or GDP molecule, permitting similar movement. Unlike actin, tubulin nucleation generally occurs at the centrosome.

Intermediate Filaments

Intermediate filaments are rope-like fibres, with a diameter of about 10nm, which attach to different parts of the cytoskeleton. The are made from a large family of different proteins, and provide mechanical strength.

2.4.2 Cytoskeletal Organisation, Operation and Dynamics

The centrosome exists at the centre of the cell (next to the nucleus), and 'organises' the dynamics of microtubules [AJL+02]; the details of its operation are not required here.

Accessory proteins link filaments to each other and to other cell components. They are essential for the controlled assembly of cytoskeletal filaments in particular locations; the accessory proteins bind to filaments to determine the sites of assembly of new filaments. Motor proteins are a class of accessory protein that move organelles along filaments or even move filaments themselves.

Large-scale cytoskeletal structures can alter depending on requirements, with time scales of change ranging from under a minute to the life of the cell. Individual macromolecules of the structure constantly change; their relatively small size allows rapid diffusion. All filaments are formed from helical assemblies of monomers, and their non-covalent bonding allows rapid assembly and disassembly; large number of bonds are therefore required for strength. Linear single-stranded protofilaments combine laterally.

2.4.3 Cell Communication

Adherens junctions are where actin binds to the cell membrane [AJL+02, LBZ+00]. Actin filaments close to the membrane are either linked to other cells or the extra cellular matrix

(a 'focal adhesion') through integral membrane proteins (namely 'cadherins' or 'integrins' respectively). Adherens have a complicated structure, with multiple different types, depending on the cell. Their complicated structure permits dynamic control of mechanical properties of different regions of the cell, which affects cellular function. The receptors also relay information about external mechanical and chemical properties to the inside of the cell. Focal adhesions bind to proteins of the extra cellular matrix (ECM), such as fibronectin, laminin and collagen. Some proteins in focal adhesions have a signalling rather than mechanical or structural role (such as focal adhesion kinase (FAK)). Focal adhesions receive multiple inputs from the extra cellular matrix—mechanical tension and phosphorylation signals—and integrate them, producing multiple outputs.

Integrins are a large and diverse family of transmembrane proteins, which bind as heterodimers, each having a different affinity for ECM proteins. The inner surface of an integrin binds with a complex cluster of cytoskeletal and accessory proteins; some of these proteins (including α-actinin, viculin and tensin) can also associate with cadherins, while others (such as talin) only associate with integrins. Actin filaments insert into a protein plaque [Bra01]. The binding of integrins to ECM proteins promotes their aggregation in the plasma membrane and their interaction with the cytoskeleton and signalling molecules [WMG99]. This leads to the assembly of focal adhesions and activation of signalling pathways which regulate gene expression.

Rho is a small GTP-containing protein that is essential to form focal adhesions. It has a central role in controlling many actin-membrane associations. Rho belongs to a family of over 40 GTP-binding proteins (G proteins) called the Ras superfamily, which relay intracellular signals and regulate many processes in cells. Rho is 'active' when bound to GTP, and 'inactive' when bound to GDP, with the addition and loss of the phosphate group catalysed by other proteins.

2.5 Gene Transcription and Translation

Once NF-κB is free in the nucleus, it is able to activate the transcription of appropriate genes, resulting in particular cell responses [AJL+02, LBZ+00]. Every gene has a region called a promoter site, which is a sequence in the DNA indicating where transcription of that gene should begin. The genes that NF-κB activates are those whose promoter sites can interact with the relevant binding sites of the particular NF-κB dimer.

Following the binding of NF-κB to a promoter site, many other proteins bind around it,

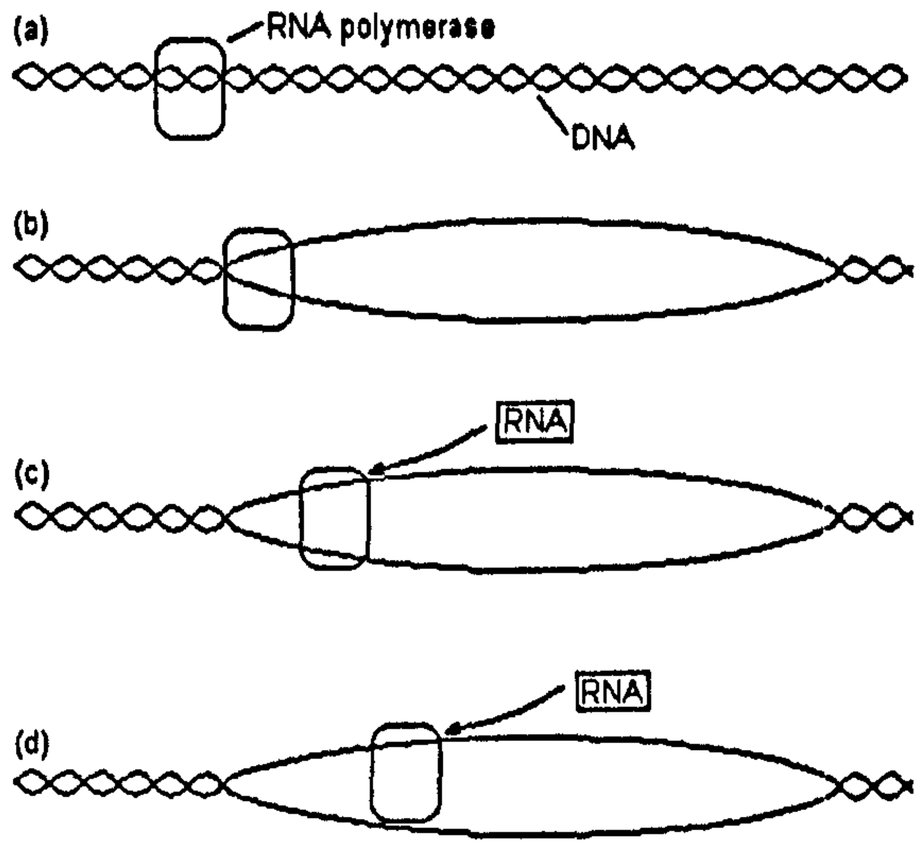

Figure 2.9: **Catalysis of transcription of DNA into RNA by RNA polymerase.** (a) Following gene activation, RNA polymerase binds to DNA. (b) DNA separates its strands. (c) The first RNA monomer unit is paired to the starting base of the DNA (only one DNA strand need be copied). (d) RNA polymerase moves along the DNA strand, continuing the process until the transcription is complete. Figure adapted from Alberts et al. [AJL+02].

reinforcing the connection with the DNA and encouraging the binding of RNA polymerase. RNA polymerase is an enzyme that catalyses the synthesis of RNA on the DNA template, as shown in Figure 2.9. Once RNA polymerase has bound, the section of DNA uncoils its double helix and the RNA polymerase moves along its length, causing the transcription of its information into RNA. The movement is jerky (or 'stepwise') and several proteins are involved to decrease the likelihood that the RNA polymerase dissociates before it reaches the end of the gene.

NF-κB activates the transcription of thousands of genes, although the genes of most interest to modelling the negative feedback of the system are the IκB-encoding genes, of which there is one gene for each IκB isoform.

Having transcribed a gene, the newly formed RNA molecule leaves the DNA and exits the nucleus into the cytoplasm. Here it associates with a ribosome, which catalyses the synthesis of the encoded protein. The process, known as translation, is complex and requires the interaction of many amino acids.

2.6 Experimental Procedures

The NF-κB pathway has been subject to extensive biological experimentation in recent years. Much work has involved experimenting on several thousand cells at a time, which must be killed in order for results to be obtained. The results are for the entire group of cells, effectively averaging the behaviour of all cells. Due to the large differences in cell behaviours, this averaging does not necessarily provide a reasonable description of what is actually occurring in any cell. More recently, single-cell analysis of live cells has been possible using confocal microscopy. This allows observation of the varied behaviours of individual live cells [YRQ03, YCQ01, CKTC+01, CDQ01].

Confocal microscopy is able to observe a thin layer of a cell at a time, which means that the nucleus and cytoplasm can be clearly distinguished. To be observed, molecules must possess fluorescent tags (which do not affect the behaviour of the molecule), so molecules occurring naturally in the cell cannot actually be observed. Instead, artificial DNA, which is encoded to produce (or 'express') particular molecules with fluorescent tags, is placed (or 'transfected') inside the cytoplasm, thus introducing observable molecules into the cell.

The cell can then be experimented on, and the movement of the newly expressed molecules can be followed. Individual molecules cannot be identified; rather it is net movement of molecules that is observed. To observe the molecules, focused laser beams induce the fluorescence of molecules in a slice of the cell approximately one micron thick. Human fibroblast cells (found in connective tissue) are commonly used due to their resilience to the experimental conditions and procedures, though similar results should be obtained with many other cell types. An example of the images obtained is shown in Figure 2.10.

Large numbers of fluorescent molecules are necessary to be observable by confocal microscopy—far greater than endogenous levels—though if too many are introduced to the cell, the behaviour of the cell is altered [CCDQ99]. Care must therefore be taken to ensure optimal transfection levels. Another problem is that not many different colours of fluorescent tags exist, so only a few types of molecule can be observed at the same time. Furthermore, because only one slice of the cell is viewed, the behaviour of the cell must be assumed to be uniform in order to draw conclusions about its overall behaviour. It is difficult to observe several slices of the cell to overcome this, though the issue can be minimised by repeated tests and comparisons. To give a true representation of events, the entire cell should be observed layer-by-layer, rather than relying on a single layer as being representative of the whole.

NF-κB is generally observed by tagging p65-RELA, which is the active part of most

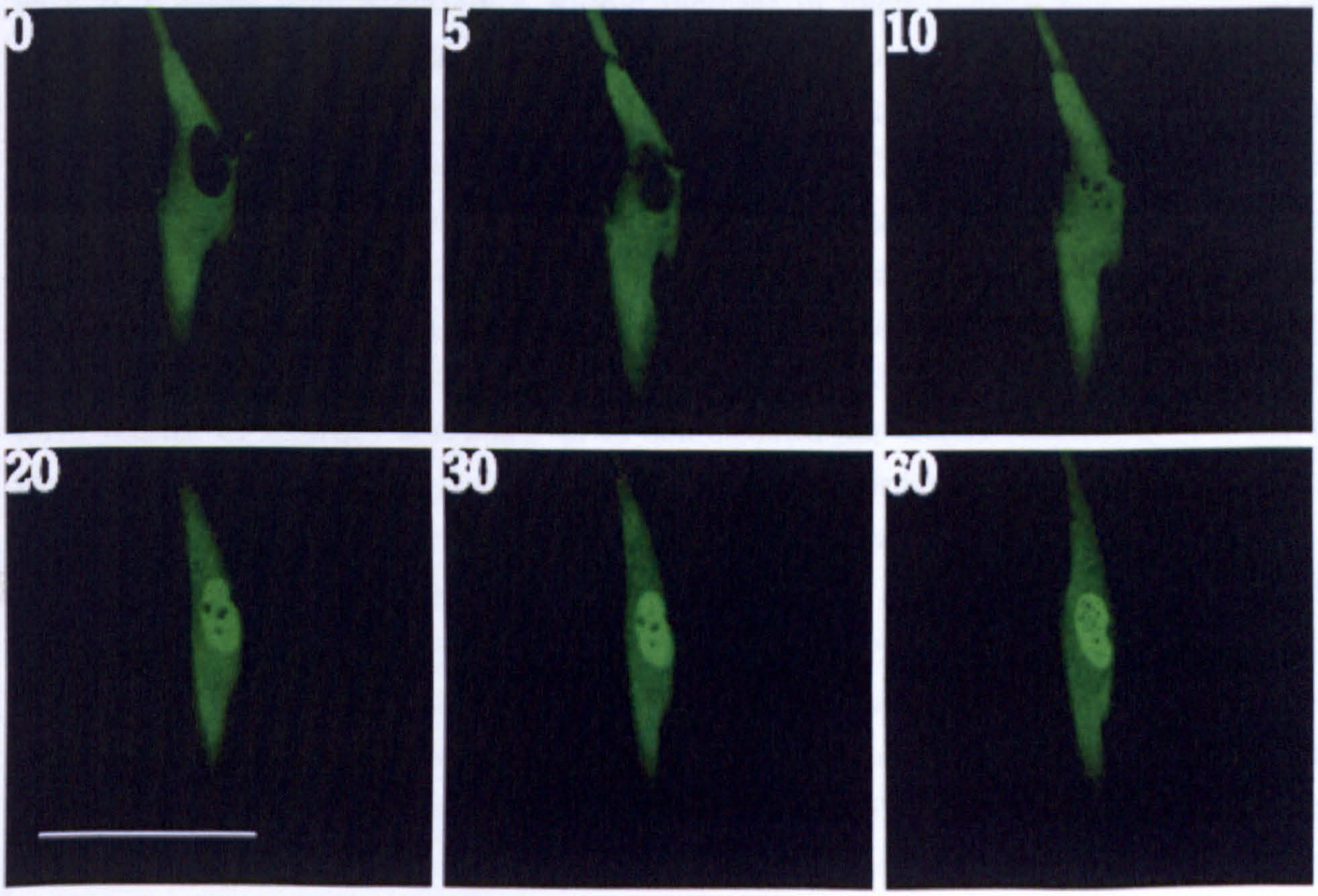

Figure 2.10: **Confocal microscope images of a cell's NF-κB molecules following IL-1 stimulation.** The number in the corner of each image shows the time in minutes following IL-1 stimulation. The dark patch in the centre of the cell at zero minutes is the nucleus, which is visibly free of NF-κB. As time progresses, the concentration of NF-κB in the nucleus increases markedly. The intensity of the fluorescence can be used to plot the concentration of NF-κB with time in each compartment (see Figure 2.11). Scale bar 10μm. Figure from Carlotti et al. [CDQ01].

NF-κB dimers; while this is not exactly the same as tagging every NF-κB molecule, it is a good guide. In this section of the thesis, the term NF-κB is used to refer to the tagged RELA subunit. Similarly, IκB is used to refer to the IκBα isoform tagged in many experiments.

2.7 Experimental Observations

Studying the dynamics of nuclear transport of NF-κB following stimulation of the cell with IL-1 shows approximately a 30-fold increase in nuclear levels and 20% decrease in cytoplasmic levels [CCDQ99]. The nuclear to cytoplasmic ratio of NF-κB concentration is around 1:10 in the unstimulated cell and around 4:1 in its maximally stimulated state.

The kinetics of nuclear transport depends on the IL-1 dose and the transfection levels of NF-κB (the latter being a consequence of the experimental method). The rate of nuclear uptake of NF-κB is saturable; this occurs at less than saturating doses of IL-1, suggesting the rate-limiting step lies in the signal transduction cascade.

The estimated maximum rate of flow of NF-κB into the nucleus, which is calculated to have volume of 100μm^3, is around 40 to 60molecules per second [CCDQ99]. Rather than plateau at this maximum when the NF-κB concentration is high, the nuclear import rate peaks and subsequently declines with increasing fluorescent NF-κB expression levels; the cell's response to IL-1 is biphasic, with a decline in the nuclear import rate of NF-κB at expression levels above 3 to 4 times endogenous (naturally occurring) levels. This correlates with the anti-apoptotic function of NF-κB, which is more prominent at low expression levels, and provides successively less protection at higher levels. In cells overexpressing NF-κB 30-fold, the minimum nuclear concentration of NF-κB is comparable to endogenous levels of nuclear NF-κB in a stimulated cell, which could explain the impairment of nuclear localisation in overly transfected cells.

Nuclear import is a continuous process, and does not display rapid rate changes. Following the nuclear concentration of NF-κB peaking, it takes long time for the cell to return to its basal state (more than 90 minutes), so the signal transduction pathway can remain activated for prolonged periods.

The anti-apoptotic function of NF-κB depends on its induction of protective genes. Since low levels of NF-κB have an anti-apoptotic effect, anti-apoptosis might be determined by free IκB levels. However, the loss of anti-apoptotic function at high expression levels is probably a consequence of the activation of different a set of genes (possibly those with

weaker NF-κB binding sites) as opposed to a drop in free IκB levels, since IκB is efficiently induced.

Using the nuclear export inhibitor leptomycin B (LMB), it is shown that shuttling between the nucleus and cytoplasm of NF-κB and IκB occurs even in unstimulated cells [CDQ01]. The shuttling is suggested to be a consequence of the cytoplasmic dissociation of the NF-κB-IκB complex rather than its direct nuclear import or degradation and resynthesis of IκB. Determination of the kinetics of nuclear entry shows that import of IκB occurs more rapidly than import of NF-κB. The rate of LMB-induced nuclear accumulation of NF-κB is lower than the rate of import in cells containing physiological levels of NF-κB following activation by IL-1, suggesting that the shuttling is not simply a consequence of high levels of NF-κB in the experiment.

It is deduced that around 17% of NF-κB is not bound to IκBα in a resting cell. It is suggested that non-saturated interaction of NF-κB with the inhibitor may enhance the specificity of action of IκB proteins on different NF-κB dimers and allow additional modes of regulation of IκB function. Levels of both NF-κB and IκB proteins are specific to cell type and developmental stage.

Whereas LMB-induced nuclear localisation of NF-κB is almost linear during the first 60 minutes of treatment and displays minimal time lag, the IL-1-induced response has a distinct lag (as expected for the pathway to react) and reaches plateau. Cells treated with both IL-1 and LMB also plateau, but at a level six-fold above the initial cytoplasmic concentration, which is more than that found by adding the effect for LMB and IL-1 alone. This level of nuclear translocation depletes approximately 70% of cytoplasmic NF-κB. The time lag in NF-κB nuclear uptake following IL-1 stimulation increases at higher transfection levels.

The maximum absolute nuclear translocation rate in response to LMB is higher than that in IL-1-stimulated cells, again indicating that the limiting factor for nuclear localisation in the latter is not the import mechanism. Since the nuclear import effects of IL-1 plus LMB are greater than additive, it would appear that increased nuclear to cytoplasmic transport of NF-κB occurs following IL-1 stimulation than in a resting cell. However, blocking the nuclear export pathway does not significantly reduce the saturability of IL-1-induced nuclear translocation of NF-κB at high expression levels, suggesting the restriction lies in the pathway.

Single-cell analysis shows the diversity of responses of cells. For example, some cells are fully stimulated at the lowest IL-1 dose, whilst others are only partially stimulated

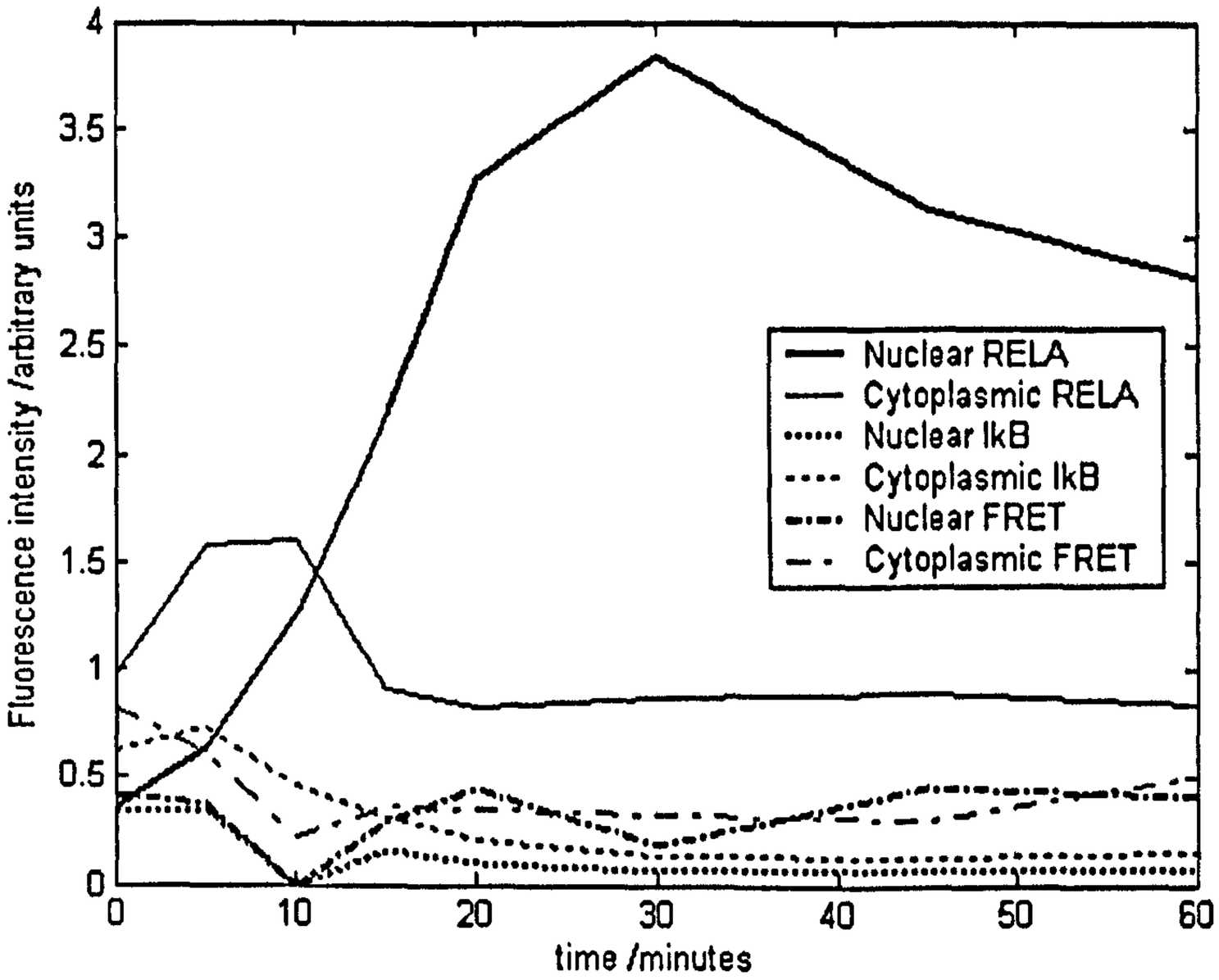

Figure 2.11: Graph of fluorescent intensity of molecules against time for a single cell following stimulation by IL-1. The fluorescent intensity is proportional to the concentration of the molecules. FRET refers to bound RELA and IκB. The fluorescent tags of the molecules are chosen such that the frequency of light emitted by RELA stimulates any nearby IκB to fluoresce. RELA is then excited by a laser of a particular frequency, and any IκB fluorescence detected is recorded as FRET, although the number of FRET recorded will be slightly higher than the number of bound RELA and IκB since the required proximity of an IκB molecule to a RELA molecule to exhibit FRET is slightly greater than the distance between bound molecules. The freeing of cytoplasmic RELA and its subsequent nuclear translocation is clear in the graph. Data from Yang [Yan03].

by the highest dose. Experimental results vary greatly between different cells, but general patterns of behaviour are often similar. An example of the typical results for an experiment on a single cell following stimulation by IL-1 is shown in Figure 2.11.

Despite the level of understanding obtained from experiment, the limitations of experimental procedures show the need for a model to allow any number of types of molecule to be observed under any conditions. The model must be verified against the large amounts of existing biological data on the pathway.

Chapter 3

Modelling the NF-κB Signalling Pathway

3.1 Key Features to Model

The purpose of the model is to determine and understand the key elements of the pathway—not to replicate the system in its entirety and thereby create something equally incomprehensible—so choosing the appropriate level of complexity for the model is essential if it is to be successful.

Many details of the signalling pathway are of fundamental importance to its behaviour but need not be modelled directly. Indeed, if the model is to succeed in reducing the system to its crucial elements, it is necessary that many such details are not modelled directly but instead accounted for at a higher level. For example, the peptide bonding of amino acids to form proteins underpins the activity of the entire pathway, though it is the actual behaviour of the proteins that is of interest to the model, and not the details of their formation. Similarly, while NF-κB are dimers, they act as single entities whose behaviour can be explained by their components, and the model need generally only consider the behaviour of the whole NF-κB.

The model must of course deal with the interaction of molecules in the pathway. While details of binding sites, non-covalent forces and molecular orientation are beyond the scope of the model, they must be accounted for (or 'abstracted away') by the rules and calculations of interactions (see Chapter 4). Details may also have to be abstracted away due to computational limitations.

A more obvious issue for the model is to decide what molecules are important to the

behaviour of the system. For example, although there are at least seven different IκB types, it may not be necessary to model them all, at least initially. While biological knowledge provides clues as to which are most important, it is part of the model's purpose to help determine this, as well as what the role of each molecule is.

There is a great deal of biological knowledge of certain aspects of the system, though much of this is too detailed for the model. Conversely, many details of the system are not understood or perhaps not even known, and abstraction of these details is imposed on the model rather than chosen. This issue is central to most models, though it is particularly apparent here due to the interdependence of many aspects of the system and the lack of quantitative knowledge of key processes.

It is impossible to fully model such a complex and wide-ranging system as the NF-κB pathway. It is important, however, to account for the limitations of the model; while it may not be possible to explicitly model certain aspects of the pathway, their effects on the system often cannot simply be ignored. It can be difficult to determine the significance of many components of the system, and features that appear not to be directly related to the concerns of the model may in fact be vital elements. There may, of course, also be important aspects of the pathway that are not even known. However, these issues must be overcome or else the model would have to include all the activities of the cell (or even the entire organism) in an effort to deal with the complexities and coordinated response of the system.

3.2 Modelling Methods

Since reaction kinetics is by far the most widely used and accepted approach to model the dynamics of the NF-κB signalling pathway (as well as biochemical dynamics in general), this is considered first. Various methods besides reaction kinetics have also been proposed to model biological signalling pathways [BCT+04a, HS98]. Some provide a quantitative description of pathway dynamics similar to reaction kinetics, while others offer a qualitative description of the topological structure and steady-state properties of pathways. A qualitative model can produce useful results, and is less hindered by the lack of reliable data for kinetic interactions, though such a model would be inappropriate for the aims of understanding the activation and dynamics of the NF-κB pathway. Other modelling methods are evaluated with respect to pathway modelling.

3.2.1 Reaction Kinetics

Reaction kinetics is concerned with the rate at which chemical reactions occur [DS92, PS95]. It is able to mathematically describe reaction rates by combining experimental data with theoretical equations. It is a well-understood and accurate model of chemical reactions that involve sufficiently large numbers of molecules that the fluctuations of individual stochastic chance molecular interactions may be assumed to be unnoticeable in the overall process; hence the overall chemical reactions are effectively deterministic.

The 'order' of a reaction refers to the number of molecules engaged in the reaction. A first order reaction is a spontaneous change of a molecule A into another molecule B without having to interact with any other molecules. Or, in chemical notation, with 'rate constant' k:

$$A \xrightarrow{k} B \tag{3.1}$$

Experimental observation shows that the reaction proceeds at a rate proportional to the concentration of the original molecule, with k the proportionality constant. Expressing Equation 3.1 mathematically, using lower-case letters to denote concentration:

$$\frac{da(t)}{dt} = -ka(t) \tag{3.2}$$

$$\frac{db(t)}{dt} = ka(t) = -\frac{da(t)}{dt} \tag{3.3}$$

It is thus simple to describe the concentration of the chemicals with time by integrating Equation 3.2. An analytical solution is straightforward here, but in more complicated cases a numerical solution is often required.

In reality, there is always a reverse reaction, even if negligibly small, since one-way reactions are not thermodynamically possible. Consequently, with l the reverse rate constant, Equations 3.1, 3.2 and 3.3 become:

$$A \underset{l}{\overset{k}{\rightleftarrows}} B \tag{3.4}$$

$$\frac{da(t)}{dt} = -ka(t) + lb(t) \tag{3.5}$$

$$\frac{db(t)}{dt} = ka(t) - lb(t) = -\frac{da(t)}{dt} \tag{3.6}$$

Of greater interest in many circumstances is the reaction of two chemicals to form a third compound substance in a second order reaction:

$$A + B \underset{l}{\overset{k}{\rightleftarrows}} C \tag{3.7}$$

The rate at which the reaction proceeds is proportional to both a and b in the forward direction, and to c in the reverse (since the reverse reaction is first order):

$$\frac{da(t)}{dt} = -ka(t)b(t) + lc(t) \tag{3.8}$$

$$\frac{db(t)}{dt} = -ka(t)b(t) + lc(t) = \frac{da(t)}{dt} \tag{3.9}$$

$$\frac{dc(t)}{dt} = ka(t)b(t) - lc(t) = -\frac{da(t)}{dt} \tag{3.10}$$

When modelling chemical systems, if a chemical can react with multiple chemicals then the appropriate terms for each possible reaction can simply be added to the chemical's rate equation, just as forward and reverse reaction terms are added together. Due to the interdependence of all the concentrations, however, large systems can quickly become intractable, and care must be taken when obtaining solutions.

It is possible to model the dynamics of many chemical systems by using these equations in combination with the experimentally determined rate constants of each reaction. However, biological systems usually also depend upon enzyme activity. Enzyme-mediated reactions are of the form:

$$S + E \underset{l}{\overset{k}{\rightleftarrows}} C \underset{n}{\overset{m}{\rightleftarrows}} P + E \tag{3.11}$$

where S is the substrate, E the enzyme, C the complex, and P the product.

Due to the nature of enzyme interactions, n will be negligible (see Section 2.2.3). In Equation 3.11 it is assumed that only one substrate is present, yet this assumption may seem unrealistic since most enzyme interactions involve two substrates. However, in most cases one of the substrates will be present in excess, so will not influence the reaction rate of the process. In the unusual case where both substrates are present in similar quantities, Equation 3.11 does not hold.

The dynamics of the process under the appropriate conditions, with E and C present in very much lower concentration than S, are:

$$\frac{ds(t)}{dt} = -ke(t)s(t) + lc(t) \tag{3.12}$$

$$\frac{de(t)}{dt} = -ke(t)s(t) + lc(t) + mc(t) - np(t)e(t) \tag{3.13}$$

$$\frac{dc(t)}{dt} = ke(t)s(t) - lc(t) - mc(t) + np(t)e(t) = -\frac{de(t)}{dt} \tag{3.14}$$

$$\frac{dp(t)}{dt} = mc(t) - np(t)e(t) \tag{3.15}$$

By combining these basic reaction rate equations, many biological pathways have been modelled, including the intracellular NF-κB pathway. Due to the large number of chemicals involved, it is likely for there to be several interdependent equations to be solved. Consequently, small changes in parameters can completely upset the entire model. Although procedures exist to enable variations of parameters in systems involving large numbers of parameters, it is a problem inherent to the model [MK98].

Because of the different volumes of the nucleus and cytoplasm, movement of chemicals between them affects their concentration, since it is molecules that move, not the concentration. Care must be taken to account for this, requiring affected terms to be multiplied by the relative volumes of the compartments.

Nuclear membrane transport of a chemical is generally treated as a first order reaction of a molecule simply 'changing' from being in one compartment to the other; due to the nature of the model, there is no actual spatial concept of a compartment. This simple diffusion clearly does not closely resemble the complicated biology of the situation. The transport of molecules across the nuclear membrane is a crucial step in the signalling pathway, and more detailed modelling of the process would be desirable.

Owing to the largely unknown nature of the underlying regulatory processes involved in transcription and translation, rate equations to describe them are difficult to formulate, and are largely based on guesswork. While this would be true for any modelling method used to describe transcription and translation, the strong interdependence of equations in a reaction kinetics model means the unavoidable inaccuracy is of critical significance to the entire model. Moreover, because of the low numbers of molecules involved in certain parts of the processes, such as having just one gene for the transcription of each protein, reaction kinetics is not necessarily an appropriate model to apply to the situation anyway. The difficulty of incorporating the time delays of the processes into differential equations

is another problem with the method.

While reaction kinetics can provide a very accurate description of many chemical processes—including some parts of the intracellular NF-κB signalling pathway—its use to describe the entire intracellular NF-κB signalling pathway is not always appropriate, dependent on the purpose of the model. A different modelling method is certainly required if mechanical stimulation of the pathway is to be given proper consideration. A more suitable model may include reaction kinetics for some parts of the pathway, but it would be used in conjunction with a different approach for key sections.

3.2.2 Stochastic Simulation

Stochastic differential equations are able to describe systems where the number of particles involved is so small that microscopic fluctuations can produce macroscopic effects, as with genetic and enzymatic reactions in living cells [BT03, BBJ+03]. Simulation of the equations requires a variable time step in which only one reaction occurs. Recent work has developed a method that allows more than one reaction to occur at each time step, known as the τ-leap method. This eases computation, but introduces some errors (though these are controllable). The method clearly overcomes the issue of treating discrete molecules as a continuum in standard reaction kinetics equations, but is still unable to explicitly account for spatial issues. The technique partially overcomes problems of ODE descriptions but does not provide a new approach to modelling.

3.2.3 Petri Nets

Petri nets are a mathematical formalism developed in computer science [GP98]. Originally created to model concurrent systems, they are well suited to deal with biological processes, and have recently been used for such purposes. They have a standard graphical representation that is simple to interpret and to use to define models. Models can be implemented by specific software rather than a lower-level programming language, making them relatively easy to create and modify.

Different forms of Petri net exist, though stochastic Petri nets are of particular interest to modelling intracellular signalling pathways. The mathematics of stochastic Petri nets can be used to describe the details of molecular interactions, specifically: the individual molecule, the molecular species, the number of molecules, the reaction, the reactant, the chemical product, the rate of reaction, the possibility of a reaction, and the occurrence

of a reaction. This provides a clear description of signalling events, allowing quantitative modelling of the dynamics of biochemical pathways.

As suggested by their name, stochastic Petri nets can account for variations caused by stochastic interactions of small numbers of molecules, and are also able to model large numbers of different types of molecule. They have been used to quantitatively model various intracellular signalling pathways. Other forms of Petri net have also been used to model biochemical pathways, including the steady-state behaviour of pathways [HKV01].

3.2.4 Process Algebra

Process algebra is the use of algebraic constructions to describe parallel systems in computer science [Bae03]. Many different types exist; π-calculus is discussed here as a representative example.

π-Calculus

π-calculus is a formal language for specifying concurrent computational systems, and has a solid and widely investigated theoretical basis [Mil99]. Biochemical systems such as transcription circuits, metabolic pathways and signal transduction networks can be modelled by π-calculus by treating molecules and their individual domains as computational processes, with potential interactions corresponding to communication channels [RSS01]. Stochastic π-calculus is an extension of standard π-calculus process algebra, and it accounts for probabilistic distributions that govern pathways, allowing quantitative descriptions [CDPT03].

π-calculus is able to provide a similar description to Petri nets, but with a linguistic rather than graphical basis. While both can incorporate spatial information in modelling, the description is topological and thus somewhat unsuitable for the desired simulation of the NF-κB pathway. The ability to 'model check' (a rigorous computational analysis of the model [CGP99]) process algebra models is often used to argue for their use in describing signalling pathways. However, due to the computational intensity of model checking, it is not possible to model check anywhere near the level of complexity present in signalling pathways, thus severely reducing the benefits of the technique.

3.2.5 Cellular Automata

Cellular automata are dynamical systems in which space and time are discrete [Gut91, TM87]. Space is represented by a uniform grid, with each lattice point (or 'cell', though

this can be confused with a biological cell) containing a few bits of information. The lattice points are updated synchronously, with their behaviour completely specified by local interaction rules. They are thus able to describe many types of concurrent dynamical system.

Cellular automata rules have previously been developed to model biochemical molecular interactions. Each lattice point (representing a unit of volume) may be empty, or occupied by up to a specified maximum number of molecules being modelled, with the maximum number being dependent on the type of molecule [Wei02b, Wei02a].

Despite their theoretical scope, cellular automata are often limited by their computational requirements. To model systems of significant complexity requires large numbers of lattice points and large numbers of iterations, and can result in impossibly long computation time (a state space explosion). Moreover, although space is explicitly modelled, its uniformity prevents the incorporation of structure, and the simplicity of interaction rules places restrictions on what can be modelled. These concerns would severly restrict the ability of cellular automata to properly model signalling pathways.

3.2.6 Different Approaches

As mentioned earlier, representation and reasoning about the topology of signalling networks has also been investigated [BCT+04a]. Rather than focussing on the simulation aspect of modelling (as with the techniques generally considered so far), such models treat signalling networks as a knowledge base from which queries can be answered. The ability of some of these methods to handle incomplete or partial information, as well as the capability to update models with new information ('elaboration tolerance'), makes them well suited to modelling biochemical networks, albeit with restricted scope.

Pathway logic is an algebraic formalism used to model and analyse signalling pathways at an abstract level higher than simulation. Biological structures are represented by algebraic expressions, and biological processes correspond to rules that transform expressions from one to another. Pathway logic can predict and explain certain pathway behaviours.

Other modelling techniques exist that provide a similar qualitative description, such as Boolean networks and recently BioSigNet-RR ('biological signal networks representation and reasoning'), which was developed to allow reasoning about networks with incomplete information [GCC+03]. Knowledge Representation in the area of Artificial Intelligence has also been used to describe biochemical pathways, including the NF-κB pathway, again allowing reasoning with incomplete information. Cellulat is an agent-based approach to

investigate pathway interactions [BCT+04b].

All the above quantitative and qualitative computational models treat biochemical networks as communication systems, focussing to a greater or lesser degree on the movement of information as opposed to the physics of interactions. While this is appropriate in many circumstances—indeed, even deliberately 'unrealistic' models can provide useful insights[Bed99]—a more lifelike description of the cell (rather than an abstract pathway) is necessary in the case of the NF-κB pathway, where cell structure and mechanical stimulation are particularly important. Although several models overcome some of the problems associated with reaction kinetics models, none consider intracellular pathways as anything more than just the chemicals involved. For many aspects of the NF-κB pathway, this must be rectified.

3.3 Previous Models

3.3.1 Hoffmann et al.

The first widely accepted model of the intracellular NF-κB signalling pathway was that by Hoffmann et al. in 2002 [HLSB02]. In the model, reaction kinetics equations describe the concentrations with time of cytoplasmic and nuclear NF-κB, cytoplasmic and nuclear IκBα, -β, -ϵ, cytoplasmic IKK, and complexes thereof. Association, dissociation, degradation and catalysis of molecules are considered, and transcription and translation are accounted for, as shown in Figure 3.1.

The rate constants used in the model are obtained from published results, previous experimental work, and model fitting to new experimental data. When implementing the model, initial NF-κB concentration is 0.1μM, with all other concentrations being zero. The model is allowed to run for 2000 minutes while IκB is produced and binds to NF-κB, causing its cytoplasmic localisation. Following this, IKK concentration is raised as a step function to 0.1μM, corresponding to pathway activation (where IKK refers to 'active' IKK). Activation of the pathway is described as being by TNF binding to tyrosine kinase receptors, though so much of the pathway is abstracted away that the model could be of any stimulation resulting in IKK activation. Removal of the activation signal is modelled by a large increase in the IKK degradation coefficient.

Transcription of IκBβ and IκBϵ is modelled as a continual zeroth order process (i.e. independent of any chemical concentration), which is completely unaffected by the pathway. Transcription of IκBα is modelled in the same way, but with additional transcription caused

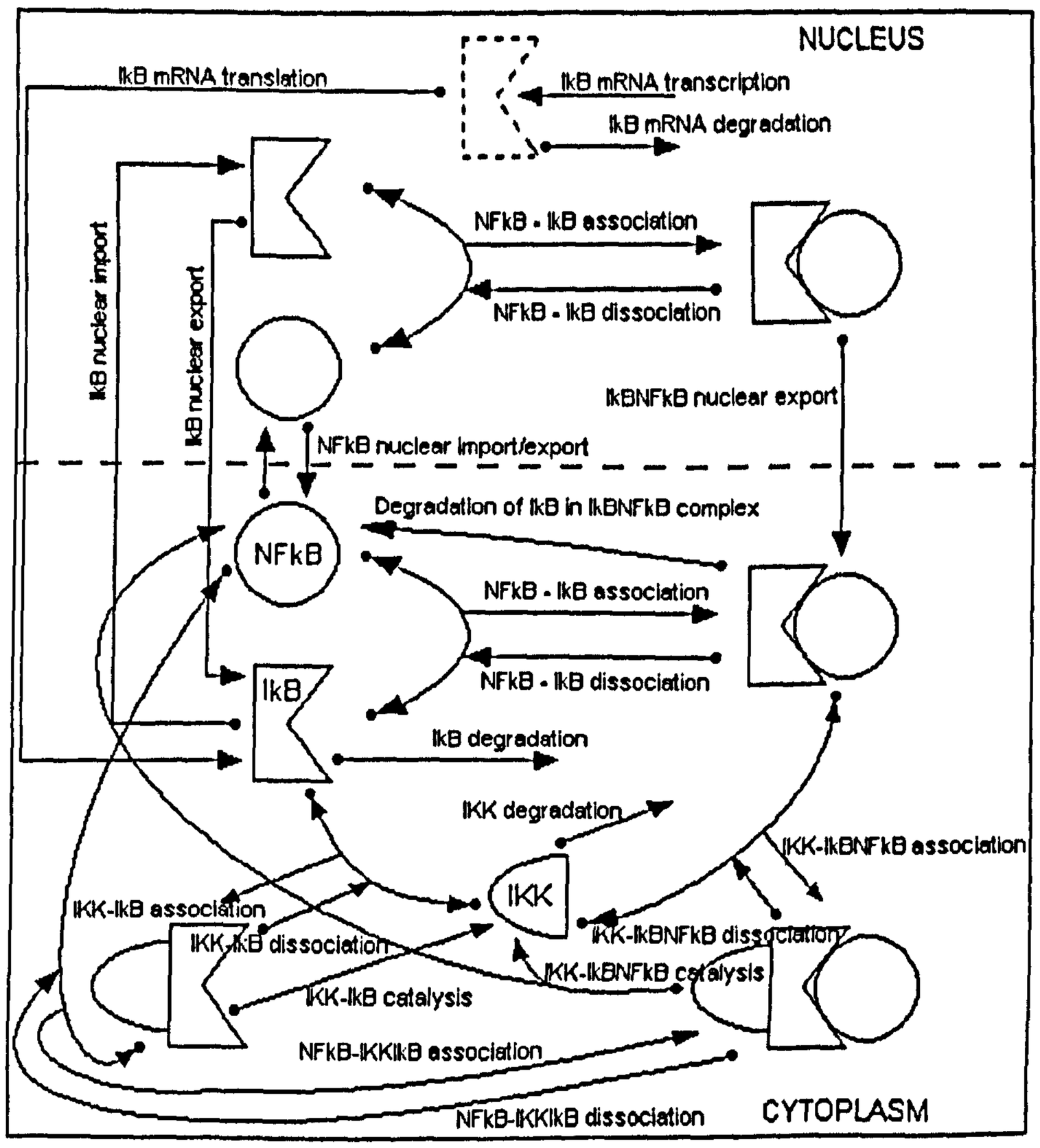

Figure 3.1: **Schematic diagram of the processes modelled by Hoffmann et al.** Arrows denote possible molecular processes, with molecular complexes shown by the grouping together of the constituents. The diagram shows all possible states of the molecules. Nuclear transport is modelled by changing a chemical into a 'different' chemical rather than actual movement; the nuclear and cytoplasmic compartments in the diagram represent this change. IκB-bound NF-κB is assumed not to enter the nucleus at all. IκBα, -β, -ϵ all behave in the same way, and are shown as one molecule type in the diagram, though they each posses different rate constants for the reactions. IκB transcription is modelled by a constant production of mRNA that is independent of any events in the cell; IκBα transcription in the model is additionally induced by the temporary binding of two nuclear NF-κB molecules (not shown in the diagram as the event does not affect the NF-κB). This is reportedly because NF-κB binds DNA in a non-linear cooperative manner, although DNA itself is not modelled. IκB translation is modelled as a first order conversion of mRNA into cytoplasmic IκB.

by a second order process proportional to the square of nuclear NF-κB concentration. This corresponds to the binding of two nuclear NF-κB molecules, and is supposed to capture the fact that NF-κB is a dimer, and so can apparently be presumed to bind DNA in a non-linear, cooperative fashion. The non-linearity is reportedly not essential for the qualitative behaviour of the model, though it is not clear why it is included since the reason provided is physically unreasonable. There is no evidence to suggest that two NF-κB molecules are required to cooperatively initiate transcription (there is only one binding site), and the fact that NF-κB is a dimer is irrelevant since in the model it is treated as a single molecule.

The model was used to investigate the roles of different IκB isoforms in signal responses. Results for the model with each of the three isoforms included on its own were compared with biological results for mice cells with the relevant IκB-encoding genes deleted. Experiments used the average of multiple cells rather than single-cell analysis. Agreement between the model and experiment was good, and IκBα was shown to be responsible for strong negative feedback of the pathway, providing a fast turn-off of the NF-κB response. IκBβ and IκBϵ were shown to reduce oscillations in the system and stabilise responses during longer stimulations. The different isoforms thus apparently enable the system to provide both rapid short-term reactions to stimulation as well as stable long-term responses.

Despite the successes of the model, it also has many flaws. The transcription terms in particular are debatable, with only IκBα transcription being affected in any way by NF-κB, and, as mentioned, even the way in which this occurs is questionable. So much detail of the transcription process is abstracted away that it is difficult to determine whether it is a good approximation to reality or whether the terms and constants used simply help replicate results rather than describe and explain the pathway.

It is reported that the concentrations of nuclear IκBβ-NF-κB complexes and nuclear NF-κB are summed when plotting graphs of nuclear NF-κB since both have a similar transcriptional capacity, though this is not a true reflection of the actual model, where only nuclear NF-κB has any effect on transcription.

The different volumes of the nucleus and cytoplasm are not considered with respect to their effect on the concentration of chemicals moving between the two compartments. This is a major issue, affecting to some degree the results of almost every chemical process modelled, including translation since the mRNA is formed in the nucleus but is translated in the cytoplasm.

Time delays in the processes of transcription and translation are not considered, though they are known to be of key significance to the pathway, being of the order of several

minutes. It is possible to include time delays in differential equations, though solutions can be difficult. Furthermore, the discrete nature of transcription (with just one gene for each IκB isoform) is also neglected. Although the rate constants employed for these processes might be argued to account to some degree for both these issues, it is a rather unsatisfactory and inaccurate description of the situation.

The biological results used to test the model were from multiple-cell averaging, thus being subject to the concerns discussed in Section 2.6.

Much of the success of the model lies in its detailed and comprehensive rate constants and data fitting. However, possibly because the model is so finely tuned, small changes in some of these values can completely upset its behaviour. Since not all the values used can actually be completely accurate, and because biology itself is not so fragile, it suggests that the model is not necessarily a good description of the pathway but rather a successful replication of particular results. This severely limits it as a predictive tool.

The model was originally implemented using Gepasi (version 3.1), which is a software package intended for simulating the kinetics of systems of chemical and biochemical reactions. Differential equations are created and numerically solved by the package, with the user supplying information about the possible reactions in the pathway, the rate constants of each reaction, and initial concentration of all chemical species. The E-Cell software package is able to perform a similar role, but with specific emphasis on intracellular reactions rather than general chemical systems. Other differential equation solvers such as those in Cellerator or Matlab could also be used.

3.3.2 Lipniacki et al.

The model by Hoffmann et al. was developed by Lipniacki et al. in 2004 [LPB+04]. A key amendment was that the different volumes of the nucleus and cytoplasm were given due consideration with respect to nuclear transport. In addition, the model was fitted with better data for the number of free IκB, and transcription and translation coefficients were re-estimated by using data on a molecular level. Some processes such as free nuclear NF-κB export and dissociation of IκB and NF-κB were disregarded as having no significant impact on the model.

Rather than investigating the effects of different IκB isoforms, only IκBα was modelled in order to examine other pathway features. IκBα was chosen as it is the most active and abundant isoform, and its deletion, unlike that of the others, is lethal. The effect of changing the relative volumes of the nucleus and cytoplasm was investigated, which

suggested that the ratio has a considerable effect on the overall dynamics and oscillation of the system. It is acknowledged that more detailed analysis of this is required however, as other parameters that were not modelled would undoubtedly also be affected by different ratios of nuclear and cytoplasmic volumes.

The role of A20, the inhibitor molecule of IKK, was also explored. Like IκB, A20 transcription is induced by NF-κB, and it also helps to deactivate the pathway. As opposed to inhibiting NF-κB directly, A20 stops the phosphorylation of IκB by IKK, thus causing the accumulation of IκB in the cytoplasm, which inhibits NF-κB and deactivates the pathway. If A20 is removed from the model, IκB is continually degraded by IKK, and so cannot deactivate the pathway.

The model is a successful progression of the model by Hoffmann et al., rectifying some of the original problems, and investigating new areas of the pathway, although time delays and stochastic effects are still not accounted for. Its consideration of the different nuclear and cytoplasmic volumes, as well as its use of more accurate data, shows some sections of the model by Hoffmann et al. to be unrealistic. However, the basic framework of the model is the same, and with some key alterations it is still successful in its description of the pathway, suggesting the fundamental approach does have scope. It models NF-κB-induced transcription of IκBα as a first order rather than second order process, which is more realistic, and is still able to exhibit the required behaviour. The model was run using Matlab equation solvers.

3.3.3 Developments with ODE Models

In 2004, Nelson et al. reported a model exhibiting oscillatory behaviour in agreement with single cell data [NIE+04] rather than population data used by Hoffmann et al. They predict a critical role for the period of oscillation in determining the response of the cell and to permit the cell to distinguish between different pathways. While this is very plausible, no suggestion is given as to how this temporal signalling is made sense of. Furthermore, a more realistic consideration of oscillations would include time delays, but this would alter the quantitative behaviour of the model [Mon03].

The work of Nelson et al. prompted a debate with Barken et al. (based on the work of Hoffmann et al. [HLSB02]), raising concerns about the experimental data used in their respective models [NHS+05, BWK+05].

In 2005, Werner et al. developed the work of Hoffmann et al. by modelling interactions leading to IKK activation in further detail, investigating the effects of different antagonists

on pathway activation [WBH05].

In 2006, Lipniacki et al. developed their ODE model by introducing stochastic differential equations to describe the transcription of genes: due to the relatively large numbers of proteins and mRNA molecules in the cell, gene regulation is considered the main source of stochasticity [LPB+06]. The model demonstrates a significant variation between cells in different simulations; however, TNF stimulation causes synchronisation of cells, particularly in the first 90 minutes following activation, progressively digressing thereafter. This suggests that population-level modelling of cells is valid within reason, while individual-level modelling of cells is also justified.

3.3.4 Other Models

Other pathway models include signal transduction models based on process algebra [CVGO, CGH]. An aim of these is to use model checking [CGP99] to investigate in detail the behaviour of the model; this is extremely computationally intensive, so the models themselves must be relatively straightforward to be of use. Agent-based models of other biochemical systems have also been proposed, as discussed in Section 3.5.4.

3.3.5 Review of Previous Models

Although reaction kinetics models have been shown to be a useful method to investigate the NF-κB pathway, they are restricted in what they can model. They are unable to provide further insight into key transcription and translation processes due to their macroscopic view of all processes, and cannot deal with spatial considerations that can be of great significance to the activation of the pathway. Although time delays can theoretically be incorporated into the equations, little work has been done on this, and their solutions are especially difficult for large systems of equations.

The existing reaction kinetics models should be referred to when developing a new model. A new model should build on their successful treatment of many of the pathway interactions, and utilise much of their comprehensive rate constants, but should also consider their assumptions and limitations. How previous models have dealt with unclear sections of the pathway should also help inform a new model, as should their results.

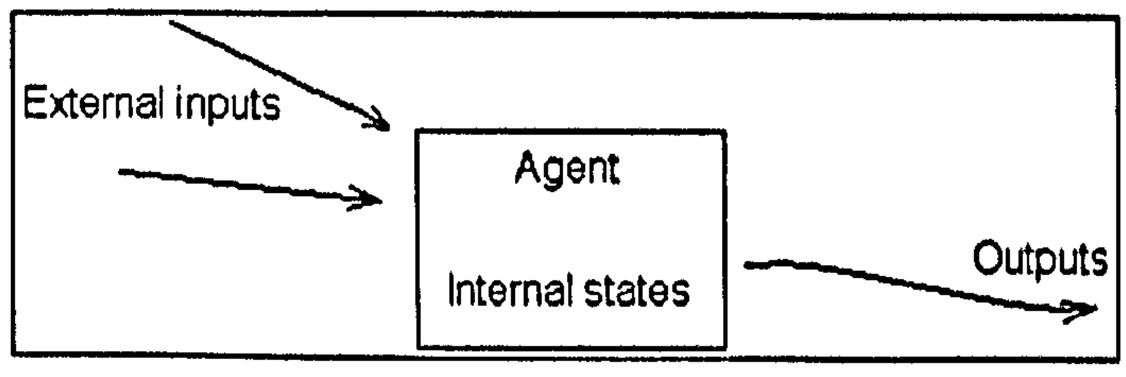

Figure 3.2: **Representation of an agent interacting with its environment.** External inputs are received by the agent and responded to by the rules of the agent according to its internal state. The resulting 'outputs' could be a change in its state, communication with its environment, or to continue its current behaviour.

3.4 Agent-Based Modelling

Agent-based modelling originated in the 1940s with McCulloch and Pitts with finite state machines and von Neumann with cellular automata (as described in Section 3.2.5), though has since developed in generality and scope [Gut91, TM87, Ele03]. Agents can model any system of discrete units. Applications range from modelling the effects of workers of different personality types in businesses, to assisting with air-traffic management systems in airports [VdVP96].

An agent is an autonomous entity whose behaviour is governed by the response of its own internal rules to interactions with its immediate environment. In agent-based (also known as individual-based) modelling, a system is described by a number of such agents, each of which behaves entirely according to its own internal set of rules triggered by the individual conditions that it experiences at any moment in time, as in Figure 3.2. Responses may be based on a number of key factors, namely the current internal state of the agent, its history or 'memory state', and the influences that it is experiencing from either adjacent agents or the environment. Such influences are treated as communications of specific information.

Agents can be represented by various different computational models. Representations are based on state machines, which are systems where the current input in combination with all past inputs determines the output. The effect of all previous inputs is represented by a state.

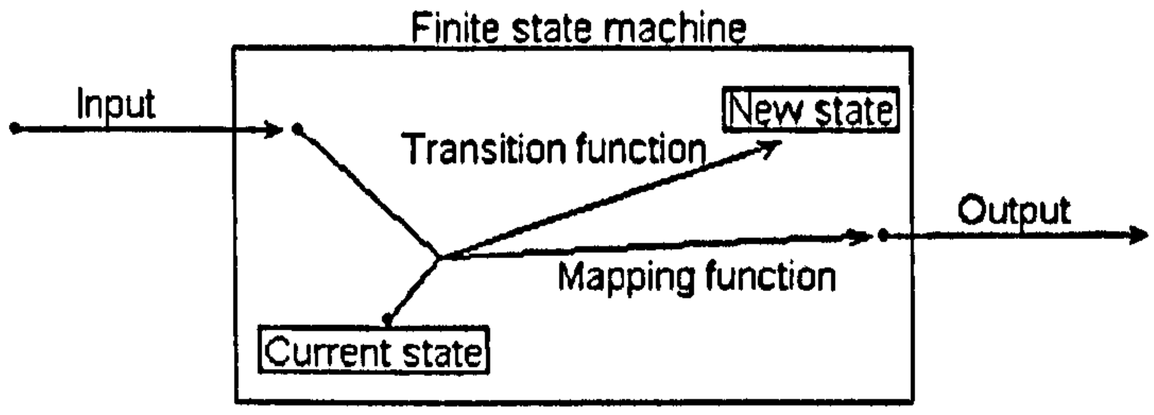

Figure 3.3: **Representation of a finite state machine.** Arrows represent processes and functions, as labelled.

3.4.1 Finite State Machines

A finite state machine is an abstract mathematical model of computation that represents computation as transitions through a finite set of states, as shown in Figure 3.3 [Elc03].

A stochastic finite state machine is defined as a tuple (N, O, S, F, A, I), where:

- N is a finite set representing all possible inputs
- O is a finite set representing all possible outputs
- S is the finite set of states
- F is the transition function operating on an input and state in N and S that causes changes in state
- A is the mapping function operating on an input and state in N and S that generates output
- I is the set of initial states (a subset of S)

(a tuple is simply an array of data that is not necessarily all of the same type).

The finite state machine is a valuable means to describe the dynamical behaviour of systems, providing an intuitive and easily comprehensible description. However, there are no communication channels between machines, meaning it cannot be used for agent modelling, and just one machine can be used to describe the entire system. A new state is therefore required to describe every possible state of the system, which can result in huge complication for even relatively simple systems. For example, to model the NF-κB pathway including spatial dimensions would require a different state for each possible combination of positions of molecules even before interactions are considered. There is also a lack of

modularity in the model, meaning the system is quite inflexible and requires an overall change to account for new or different requirements.

It is possible for communication channels between machines to be defined to allow agent-based modelling, though the number of states required for each agent can still be unreasonably large. This is a similar situation to cellular automata.

These problems make finite state machines inappropriate for modelling large systems; hence variations on finite state machines have been proposed for such purposes. Petri nets, as previously described (see Section 3.2.3), are a restricted version of finite state machines, providing greater analytical power. Extended versions of finite state machines have also been proposed to improve modelling capabilities, most notably of which is the X-machine.

3.4.2 X-machines

The X-machine is a formalism introduced by Eilenberg in the 1970s as a mathematical model of computation, and further developed by Holcombe and others from the 1980s onwards [Ele03]. It provides an intuitive method of modelling the functional behaviour of systems.

The X-machine is similar to a finite state machine except that transitions are not simply labelled by inputs, but include functions that operate on inputs and a memory state, which allow the machine to be more expressive and flexible than a finite state machine. Stream X-machines (often referred to as simply X-machines) are defined as X-machines but with input and output restricted as being processed in a straightforward manner (effectively a string of symbols), as shown in Figure 3.4.

A stream X-machine is defined as a tuple (N, O, S, M, P, F, s, m), where:

- N is the set of input symbols
- O is the set of output symbols
- S is the set of states
- M is the set of memory states
- P is the type of the machine: a set of functions p that map an input and memory in N and M to an output and new memory in O and M
- F is the function that gives the next state in S from the current state and function p in S and P

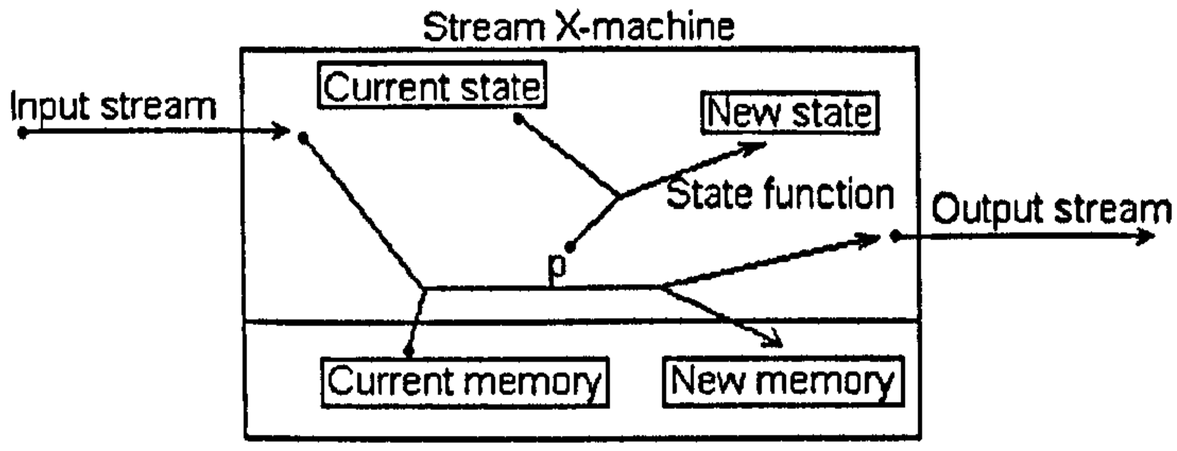

Figure 3.4: **Representation of a stream X-machine.** Arrows represent processes and functions, as labelled. p is a function in P, as defined in the above text.

- s is the initial state, a subset of S
- m is the initial memory state, a subset of M

Such an X-machine can be used to model various situations [Bar93]. For example, the input stream could be information supplied from a computer keyboard, and the output stream information to be passed to a monitor, though the model can be applied to many situations aside from those relating to computational processes. However, like finite state machines, basic stream X-machines cannot be used to formulate agent models; to achieve this, communication channels must be defined between machines. This creates the communicating stream X-machine, which is extremely well suited to modelling complex concurrent systems due to the flexibility and modularity of the machines [BHC+99]. Communication occurs by a 'communication matrix', which contains a space for reading and writing information between every X-machine, allowing each machine to generate and receive influences from other machines.

Communicating X-machines can therefore be used to define a system of agents where each agent interacts with other agents and changes state depending on both its environment and internal state. This is a natural and powerful method to model many systems. Each agent behaves in a structured way as follows: At any moment in time, the agent exists in a given internal state and a given internal memory state. Upon receiving an external influence, such as a communication from its neighbours or its environment, the agent will act according to one of a number of possibilities depending on its state. Following this it will change internal and/or memory state, generate a communication, or continue its current behaviour.

An important property of such a model is that it is relatively straightforward to develop by adding new agents to the system, provided the state behaviour of the new agents and

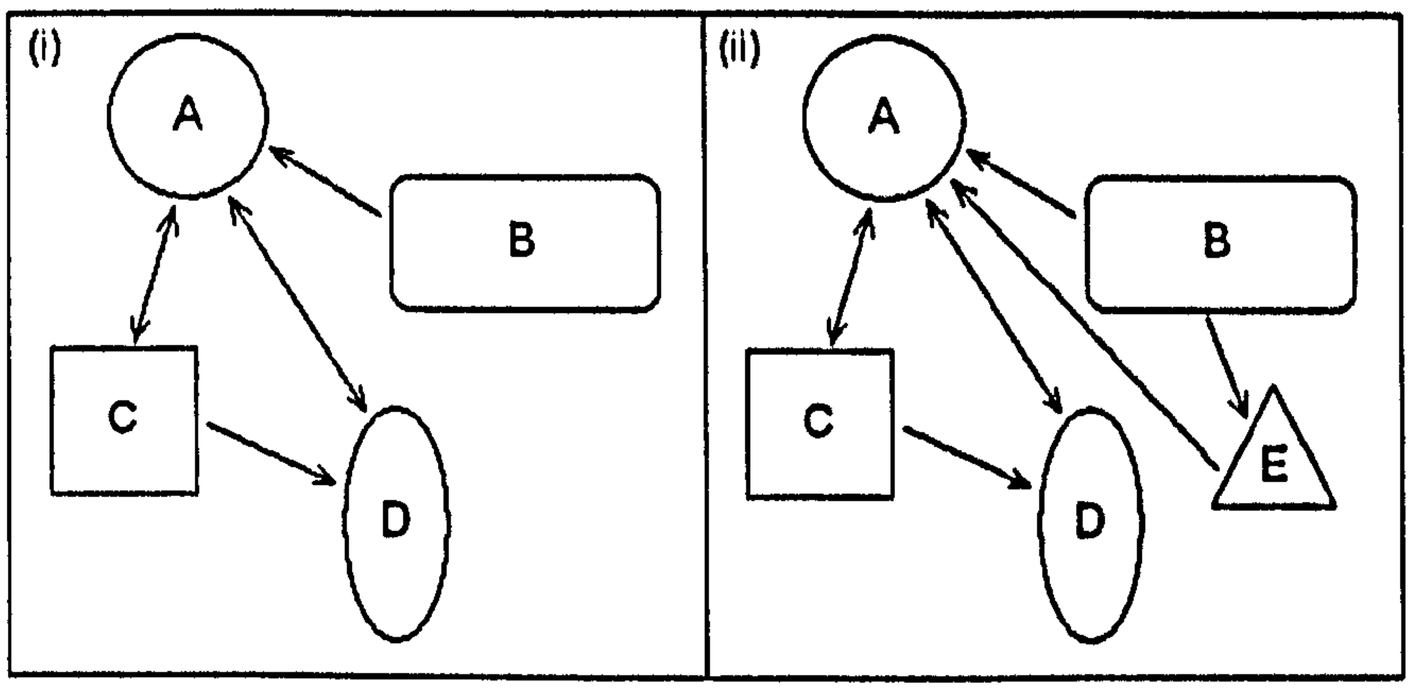

Figure 3.5: **Example of expanding a system.** System (ii) is created simply by adding agent E to system (i) and defining the new communication channels.

their communication channels with other agents in the system are precisely defined, as in Figure 3.5. The modelling process is thus extensible—an essential feature when modelling a complex system such as the NF-κB pathway.

It is possible to generate computer simulations of such systems and experiment with the agent types and communication mechanisms; this provides a very powerful approach to help understand complex systems.

3.5 Application of Agents to Pathway Modelling

3.5.1 Previous Agent-Based Modelling

Similar approaches have previously been used to successfully model a range of biological systems with a social element to their behaviour, including insect communities and epithelial tissue [SHW04, WSH+04, JHR04b, RJHR05, JHR04a, WHW+04]. Many of the principles of previous agent-based models can be applied to an agent-based pathway model, such as communication methods and discrete time steps. Agent-based models depend on identifying coherent local entities that can be modelled individually, with a general method of communication that allows these agents to interact and adapt to circumstances. The choice of agent should be biologically natural and computationally meaningful; in the case of the NF-κB signalling pathway it is not so straightforward, hence a variety of types of agent will be required to decompose the system into meaningful sections. It is essential to ensure that the agents are both biologically plausible as entities and that their behaviour

is based on experimental measurements.

3.5.2 Modelling the Pathway

An agent-based model of the NF-κB signalling pathway must clearly be based around an agent-based method of describing chemical interactions. Since this is crucial to the model, an appropriate chemical modelling method must be investigated.

3.5.3 Agent-Based Chemical Interactions

A successful model of the intracellular NF-κB signalling pathway depends on an accurate model of chemical interactions. While an agent-based approach offers greater modelling scope than many other methods, it also requires greater definition of details due to its lack of implicit assumptions, such as molecular movement and how interactions occur. With the aim of using agent-based chemical interactions to model the NF-κB pathway, the chemical model must not only accurately describe chemical interactions, but should be sufficiently computationally efficient to permit simulation of such a complex system as the NF-κB pathway in reasonable time. The chemical model must strike a balance between the level of spatial resolution achievable with agents (precision) and the need for acceptable simulation times when applied to a large system (achievability).

The most rigorous treatment of the movement of molecular agents is through the use differential equations that describe Brownian dynamics [Ris96]. With typical reactant concentrations in biochemical networks, conventional Brownian dynamics simulations require time steps smaller than 10^{-10}–10^{-9}s; the use of Green's functions can increase time steps to 10^{-6}–10^{-4}s, which in combination with the increase in CPU time required per time step results in 2–5 times greater computational efficiency [vZRtW05]. While this provides a significantly more viable method to model individual particles than conventional Brownian dynamics, it is still prohibitively slow if thousands of molecules are to be modelled over several minutes and even hours.

Andrews and Bray [AB04] formulated an agent-based chemical interaction model in 2004, implemented in a computer program called 'Smoldyn'. The method is spatially accurate to almost the size of individual molecules. Each molecular agent is a point-like particle diffusing freely in three dimensions. The collision of reactive molecules results in the simulated reactants being replaced by a product or products. Using the rate constant, diffusion coefficient and simulation time step, the 'binding radius' (the proximity

required to permit molecules to react) and 'unbinding radius' (the separation required for molecules to move apart upon separation to prevent immediate re-binding) can be calculated. Molecules move in a close approximation to Brownian motion at each time step. A detailed derivation of how to calculate the binding and unbinding radii for different types of reaction is presented, which permits time steps much larger than in Brownian dynamics. A method that holds for very large time steps is also mentioned, but dismissed as being unable to provide sufficient spatial detail. The same large time step method is used in the ChemCell model [PS05]. MCell is a similar program to Smoldyn but currently only permits reactions to occur at surfaces [SB01], which is rather restrictive.

3.5.4 Previous Agent-Based Chemical Models

The Smoldyn program has been used by Lipkow et al. to model signal transduction in the bacterium Escherichia coli [LAB05]. The large time step method is used to calculate interaction radii. A rectangular container is used to represent the cell. The model demonstrates the feasibility of modelling thousands of individual molecular agents; it also highlights the importance of physical effects by changing the dimensions of the container.

3.6 Modelling Summary

Many modelling approaches and examples have been considered. Agent-based modelling has been demonstrated to be an appropriate and feasible approach to model the dynamics and physical effects of the NF-κB signalling pathway.

Chapter 4

Agent-Based Chemical Interaction Modelling

4.1 Formulating the Agent-Based Chemical Model

As discussed in Section 3.5.3, the agent-based chemical model must not only be accurate but also sufficiently efficient, with a large enough time step, to be of use in modelling the NF-κB pathway; a balance must be struck between the benefits of agent-based modelling and the ability to run the model in reasonable time (no more than of the order of days). Since the primary concern is the investigation of the entire pathway over several minutes and hours, for which data exist, rather than individual molecular detail over short time scales, the large time step method of interaction as mentioned by Andrews and Bray [AB04] and Plimpton and Slepoy [PS05] is the most appropriate. A detailed description of this follows.

4.1.1 Agent Interactions

Reaction kinetics treats reacting chemicals as being uniformly well mixed and implicitly assumes that the number of molecules involved is large [PSQH06]. In contrast, an agent-based approach can treat each chemical molecule as a single identifiable entity, as shown in Figure 4.1. Thus any distribution and number of molecules (within computational limitations) is permissible.

The agent-based model must of course agree with the corresponding reaction kinetics model in the circumstance where reaction kinetics can reasonably be applied (i.e. with

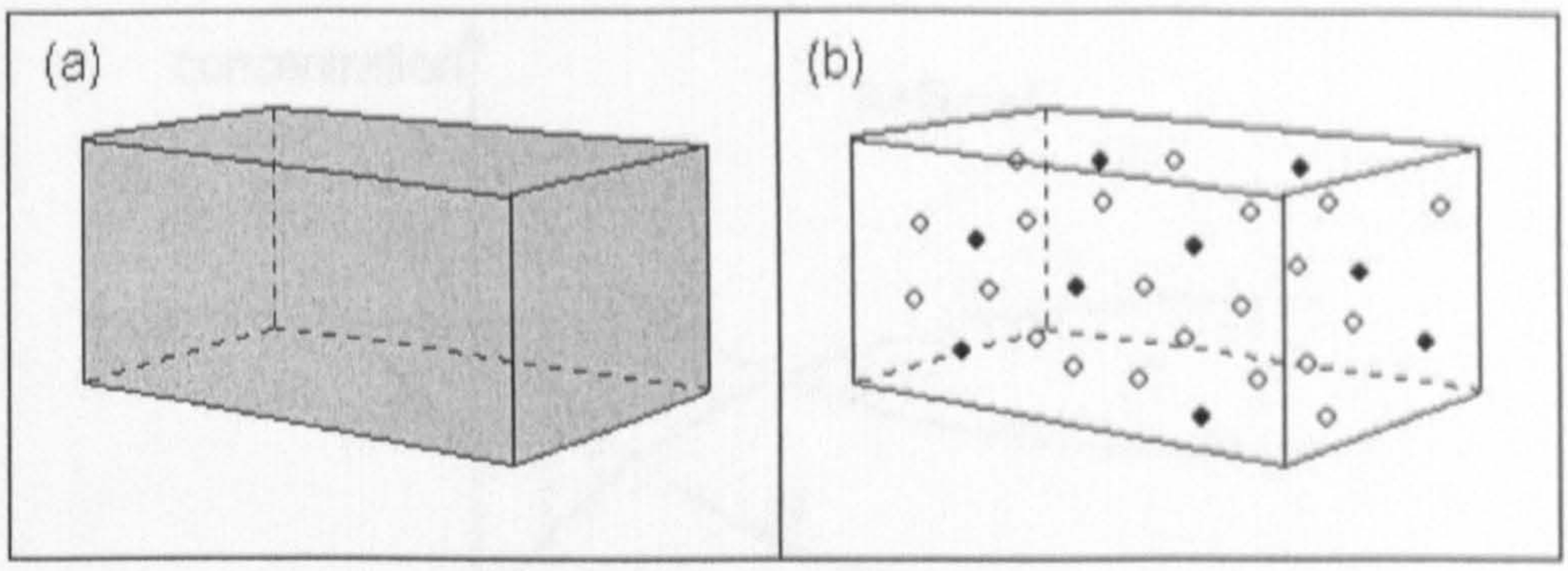

Figure 4.1: **Chemical reactions.** Reaction kinetics treats reacting chemicals as well mixed and uniform, as in (a), whereas an agent-based approach models each individual molecule, as in (b).

large numbers of molecules of well-mixed chemicals). Since information about reacting chemicals is invariably given for such a situation, and because little information exists about individual molecular interactions, it is important that the necessary data for the agent-based model can be inferred from reaction kinetics.

Rates of Interaction

Take two reacting chemicals A and B combining to form C, with rate constant k (neglecting the reverse reaction for now):

$$A + B \xrightarrow{k} C \tag{4.1}$$

If the concentrations of A, B and C at time t are $a(t)$, $b(t)$ and $c(t)$ respectively, then from reaction kinetics:

$$\frac{da(t)}{dt} = -ka(t)b(t) \tag{4.2}$$

Now if the magnitude of the change in concentration of A during a time step of Δt is Δa, as shown in Figure 4.2, then:

$$a(t + \Delta t) = a(t) - \Delta a \tag{4.3}$$

The iterative process of Equation 4.3 acts as the basis for the agent-based interactions, with Δt a constant time step.

Assuming Δt is sufficiently small, then:

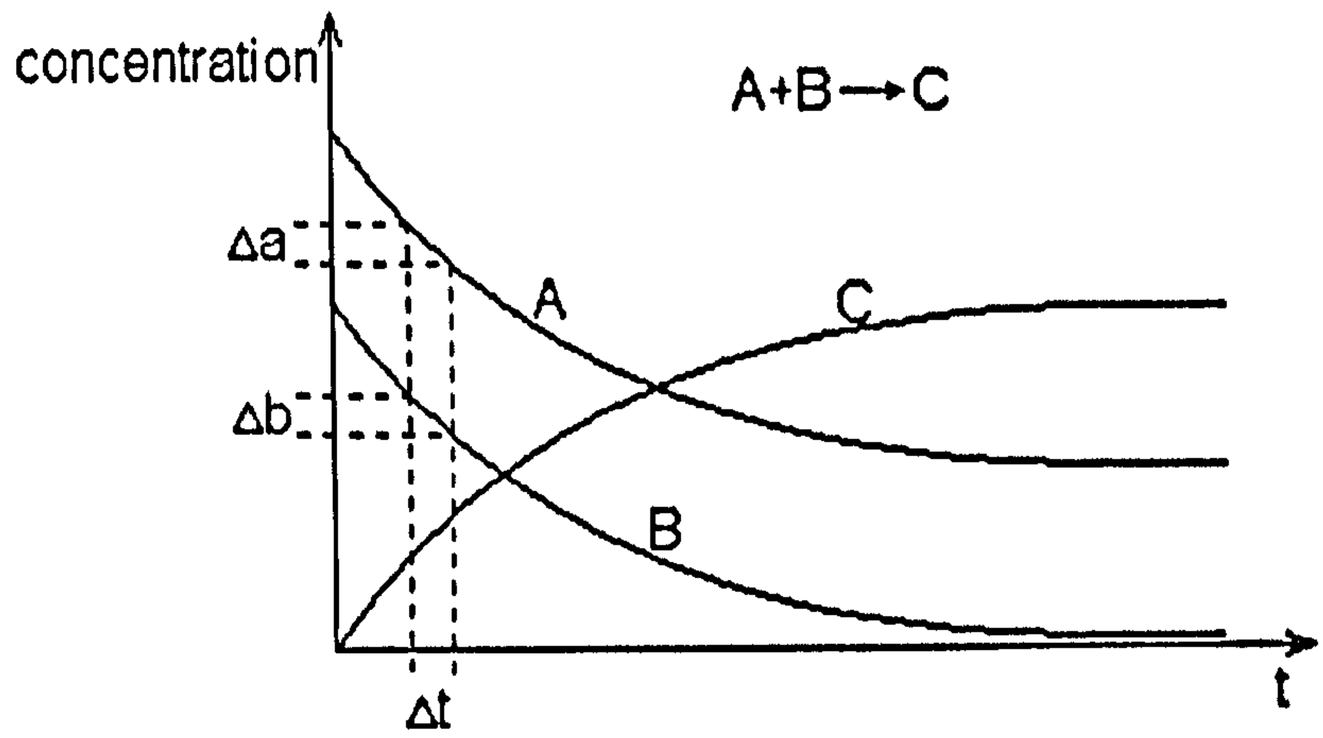

Figure 4.2: **Generalised graph of concentration against time t for two chemicals A and B interacting to form C.** With a time step of Δt, the change in A concentration is Δa, and the change in B concentration Δb. Since one A molecule interacts with one B molecule to form one C molecule, it follows that $\Delta a = \Delta b$

$$\Delta a \approx -\Delta t \frac{da(t)}{dt} \tag{4.4}$$

Substituting Equation 4.2 into Equation 4.4 gives:

$$\Delta a = ka(t)b(t)\Delta t \tag{4.5}$$

Since one A molecule interacts with one B molecule to form one C molecule, it follows that $\Delta a = \Delta b$, where Δb is the magnitude of change in concentration of B in time step Δt. Therefore the proportion of B molecules that interact with A at each time step is:

$$\frac{\Delta b}{b(t)} = \frac{\Delta a}{b(t)} = \frac{ka(t)b(t)\Delta t}{b(t)} = ka(t)\Delta t \tag{4.6}$$

If V is the volume of the container, and A and B are randomly distributed, then in order for the proportion of B molecules that interact with A to equal that obtained in Equation 4.6, it must hold that the total volume inside which A interacts with B (denoted V_i, as shown in Figure 4.3) is:

$$V_i = ka(t)\Delta t V \tag{4.7}$$

The number of A molecules n_a at any time is given by:

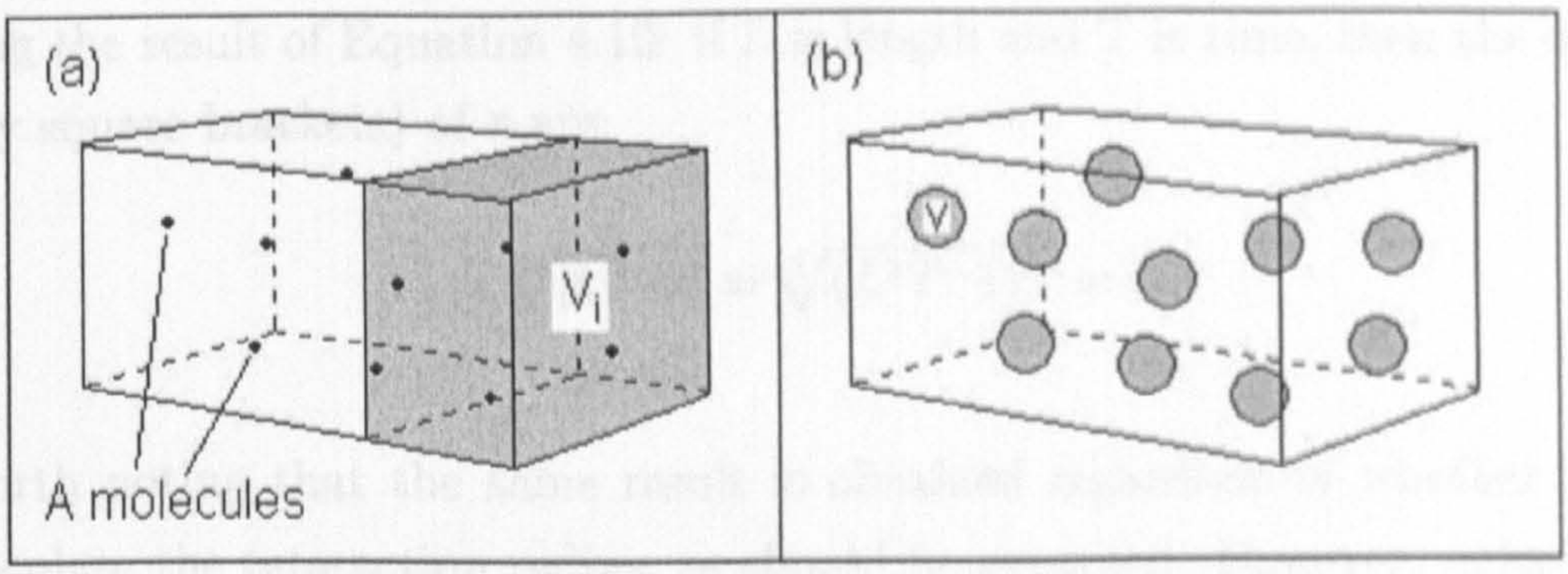

Figure 4.3: **Interaction volumes.** In (a), the shaded volume shows the size of the total 'interaction volume' V_i; (b) shows this volume 'redistributed' over all the A molecules, giving each its own interaction volume v. B molecules are not shown for simplicity.

$$n_a(t) = 10^3 a(t) LV \tag{4.8}$$

where L is Avogadro's constant, and the factor of 10^3 is required if the concentration $a(t)$ is measured in molars, with the rate constant defined accordingly (as is generally the case), and V measured in cubic metres.

So each A molecule can interact with a B molecule if the B molecule is within an 'interaction volume' v surrounding A of size:

$$v = \frac{V_i}{n_a(t)} = \frac{ka(t)\Delta tV}{10^3 a(t) LV} = \frac{k\Delta t}{10^3 L} \tag{4.9}$$

Assuming each individual interaction volume is spherical, then the maximum possible separation, or 'interaction radius' r, of two interacting A and B molecules is:

$$r = \sqrt[3]{\frac{3k\Delta t}{4\pi 10^3 L}} \tag{4.10}$$

The rate constant can now therefore be used to deduce information on local interactions. The distance r will be larger than the combined molecular radii of A and B due to the length of the time step, since molecules move around in this time. The time step used must of course be sufficiently small for this approach to hold (as assumed in Equation 4.4), though it must also be large enough for the average movement of each agent to exceed the binding radius so there is sufficient mixing of agents. Lengthening the time step will still yield accurate results over large time periods, but clearly limits both spatial and temporal resolution.

Checking the result of Equation 4.10: if L is length and T is time, then the dimensions (denoted by square brackets) of r are:

$$[r] = \sqrt[3]{[k][\Delta t]} = \sqrt[3]{(L^3T^{-1})T} = L \tag{4.11}$$

as required.

It is worth noting that the same result is obtained regardless of whether A or B is used to calculate the interaction radius, as should be expected. However, unless A and B are both of the same concentration, it seems reasonable to question this statement: if the interaction radius is the same no matter how it was calculated, then the total interaction volume V_i will be different depending on whether A or B are 'assigned' this radius, when it might be expected that it should be the same. On the other hand, if the total interaction volume were to remain the same, then this would require the interaction radius to change, which certainly should not be the case.

It is actually correct that the total interaction volume changes (recall that Equation 4.7 does indeed depend on the choice of molecule). Say there are more A molecules than B: if A molecules are each assigned the interaction radius, then V_i will be larger than if B molecules were assigned the radius, but the number of molecules with which they seek to interact is smaller. Likewise, if B molecules are assigned the interaction radius, then V_i is smaller, but the number of molecules they are seeking is larger. The number of A and B molecules in each other's interaction volume is consequently the same regardless. This is apparent in the derivation of r, though it may not be immediately obvious.

This local interaction method can be used not only for large numbers of randomly distributed molecules, but also for any molecular distributions since the interaction radius obtained should hold whatever the circumstance. Such local interactions are key to the model, and allow its expansion to deal with any interactions—not just straightforward large-scale chemical reactions.

Binding Decisions

Only the possibility of interaction has been considered so far, and not the actual process of interaction. For instance, if more than one B molecule lies inside an A molecule's interaction volume, it must have some method to 'choose' the most suitable B molecule with which to bind. This could be done at random, and with infrequent decisions to be made this would be of little consequence, but it is worth considering possible 'decision' processes.

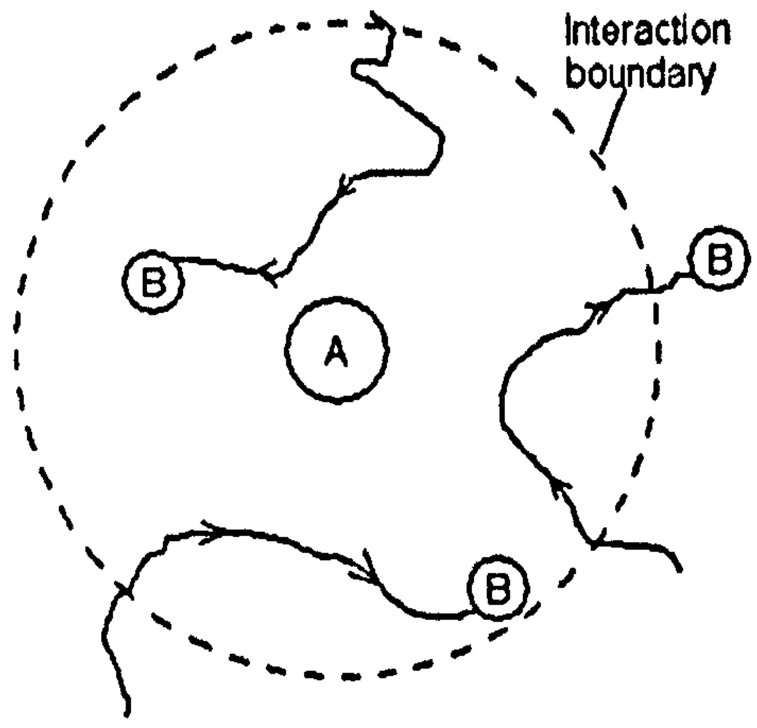

Figure 4.4: **Possible movement of B molecules during a time step Δt in the centre-of-mass frame of an A molecule.** Since the movement and orientation of molecules is not known during the time step, a random element is required in the interaction decision.

A spherical interaction volume, uniform in all directions, is of course an abstraction of the process of interaction, since the orientation of molecules is vital for interaction. Clearly knowledge of orientation of molecules is beyond the scope of the model, especially since interactions occur during a time step where the orientations will have changed and the movement of the molecules is not exactly known, as represented in Figure 4.4. So although a spherical interaction volume is used, its shape is fairly arbitrary, and a randomness in the decision of interaction could be included to account for these issues—perhaps with an increasing probability of interaction for closer molecules, rather than just randomly choosing any molecule within the interaction radius. It is possible that even if a molecule is within the required distance to interact, interaction may not occur. This could be modelled by including a 'threshold' over which the decision metric must be for molecules to interact, similar to the probability of interaction considered by Plimpton and Slepoy [PS05], remembering that this would affect the reaction rate since slightly fewer molecules would interact, thus requiring the interaction radius to be adjusted. In addition to this concern, it is clear in Figure 4.4 that it is possible for a molecule to have been able to interact during the time step, but due to the time at which calculations are performed it would not be considered. In fact, these concerns may be disregarded as having no observable effect on results, being beyond the resolution of the technique.

So far, only association of molecules has been considered, though dissociation is of course also possible. Dissociation can easily be accounted for by making bound products separate randomly at a rate specified by the dissociation rate constant (using a uniform

random distribution initially, though this could be modified as appropriate). Although this is straightforward, it may be an unnecessary consideration since the rate of dissociation is often negligible, and its inclusion may simply unnecessarily overcomplicate computation. However, some features of the NF-κB pathway may depend on dissociation, such as the shuttling of NF-κB and IκB between the cytoplasm and nucleus in unstimulated cells [CDQ01], so the model must be able to deal with it if necessary. Unbinding molecules must separate sufficiently so as not to re-bind immediately [AB04]; this can be achieved either through an enforced separation, or an adequately large time step (or molecular speed) that they will separate far enough anyway (see Section 4.1.2).

First-order reactions can be modelled by using a similar random change as would be used for two-particle dissociation. Higher-order reactions are modelled by the relevant number of two-particle local interactions (where the two 'particles' become products of earlier interactions), as are enzyme interactions.

Just as terms are added to reaction kinetics differential equations to deal with several chemical reactions, each molecular agent can seek interaction with several relevant types of molecule, with the appropriately sized interaction radius for each type. Instead of a list of linked differential equations to be solved, the agent model executes rules for each agent at each time step in accord with the state of the agent and its surroundings. In reality, molecular interactions are not tested for sequentially but are simply the result of continual attractions of all molecules involved. Without having a computer processor to simulate each agent (an impossibility for large systems), this clearly cannot be replicated in a model; instead, the order in which different agent interactions are computed must be randomised at each time step in order not to produce any artificial systematic preferences.

The method used to calculate interaction is not perfect, but it potentially allows pathways to be modelled in greater detail than with differential equations. The discontinuity of attraction at the 'interaction radius' is not quite a true reflection of reality, but it is sufficiently accurate for the purposes of many models, and it allows rate constant data to be used to infer information for local interactions.

4.1.2 Agent Movements

At every time step in the simulation, each molecular agent must move according to predefined rules. The above consideration of molecular interactions simply assumes good mixing of molecules, so some straightforward random motion of agents should be sufficient for the model to produce accurate results. Although detailed consideration of Brownian

motion [Cra75] would be the most rigorous treatment of this, the level of detail would be unnecessarily high, since the model does not seek to provide single-molecule detail. Furthermore, the movement of molecules in a cell is not always simple diffusion, but often active transport along scaffolds and membranes. Since the model will sometimes have to abstract such movement as random, it seems misleadingly and unnecessarily precise to have very detailed movement rules. More desirable is a simple but realistic rule for random movement that can easily be changed as necessary if different kinds of movement are to be modelled.

The size of the molecules is assumed to be sufficiently small that collisions between them can be neglected for the purposes of their movement (molecules are treated as weightless particles). Molecules must collide with the sides of their container in order to stay within their confines, with a reasonable assumption being an elastic collision, which is simply a reversal of the component of their velocity which is perpendicular to the container at the point of collision.

A simple rule for random movement of molecular agents is that each agent is initially randomly assigned a direction vector in 3 dimensions (assuming no net flow of molecules, just random movement), with a fixed average speed. At each time step, each agent moves a distance according to its speed, which is a random speed bounded around the average speed. The agent deviates from its direction by a random angle from the direction vector; this new direction replaces the previous direction from which to deviate at the next time step, as shown in Figure 4.5. The average speed of agents must be defined to reflect true molecular speeds, and the range of speed and direction deviation must be reasonable (see Section 4.4.4).

These movement rules are reviewed in Section 6.4, and concerns with their results highlighted.

Although not necessary for the chemical model, the movement of membrane-bound agents must also be accounted for to model the cell; this will be considered when formulating the pathway model (see Section 5.1.3).

Due to the nature of agent movements, the overall energy of the system fluctuates slightly at each time step, but on average remains constant. A proper treatment of the energy of the system would be too complicated for the purposes of the model, especially since energy is provided for many processes by the metabolism.

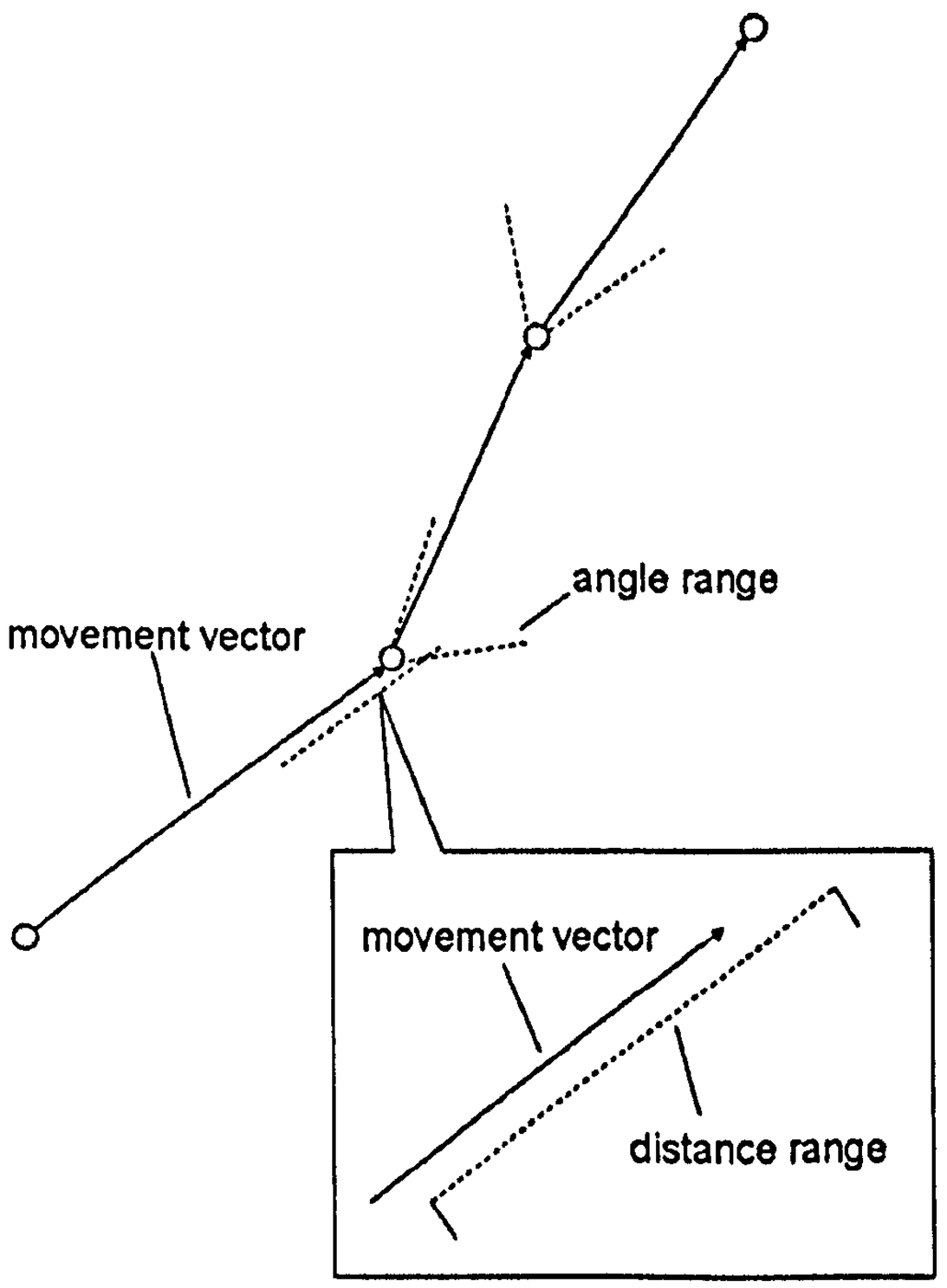

Figure 4.5: **Example of the movement of an agent at three successive time steps.** Agent moves a random distance at each time step, with deviation from the average distance limited by the distance range (shown in inset box for the first time step), and deviation from previous direction limited by the angle range (which is 3-dimensional, i.e. conical).

4.2 Model Simulation

Each agent is represented by an X-machine for the purposes of computation (see Section 3.4.2). The state of each agent's X-machine records to what (if anything) the agent is bound, while the memory of each agent's X-machine contains its physical location, meaning the number of states required to model the system is manageably small, since a different state of the agent is not required for each position it occupies.

The physics of a molecule is modelled in the rules of the agent: namely, with what molecules it can interact, and how close it must be for this to be able to occur (as described in Section 4.1.1). If two molecules are able to interact (i.e. their rules state that it is possible), to do so they must both be in a state such that interaction at that time is allowed, as well as satisfying the criteria on proximity. If interaction occurs, then the state of each agent changes to a 'bound' state; a new agent is not created, but each agent involved changes to a state such that the product behaves as a single entity, and can randomly separate. Agents could instead be destroyed and created for binding and dissociation, but this would unnecessarily remove information from the system.

From a computational perspective, when agents interact, they directly communicate this to all other involved agents. This could instead be achieved by sending and receiving messages, which would give the same results and slightly complicate the simulation, but is important with regard to possible parallelisation of the model [CSH06].

4.3 Chemical Reaction Results

The chemical model described above is firstly tested against some arbitrary but generalisable reaction kinetics cases before being investigated in further detail (see Section 4.4). Specific cases of two-particle association, two-particle association and dissociation, and competing association are presented. Matlab is used throughout to run the model using a Pentium computer (see Appendix A for program code and Appendix C for a generic description of the algorithm used).

4.3.1 Two-Particle Association

With arbitrary concentrations of chemicals A and B combining to form C, the agent-based chemical model is initially as shown in Figure 4.6.

Running the simulation with an arbitrary rate constant (neglecting dissociation) gives

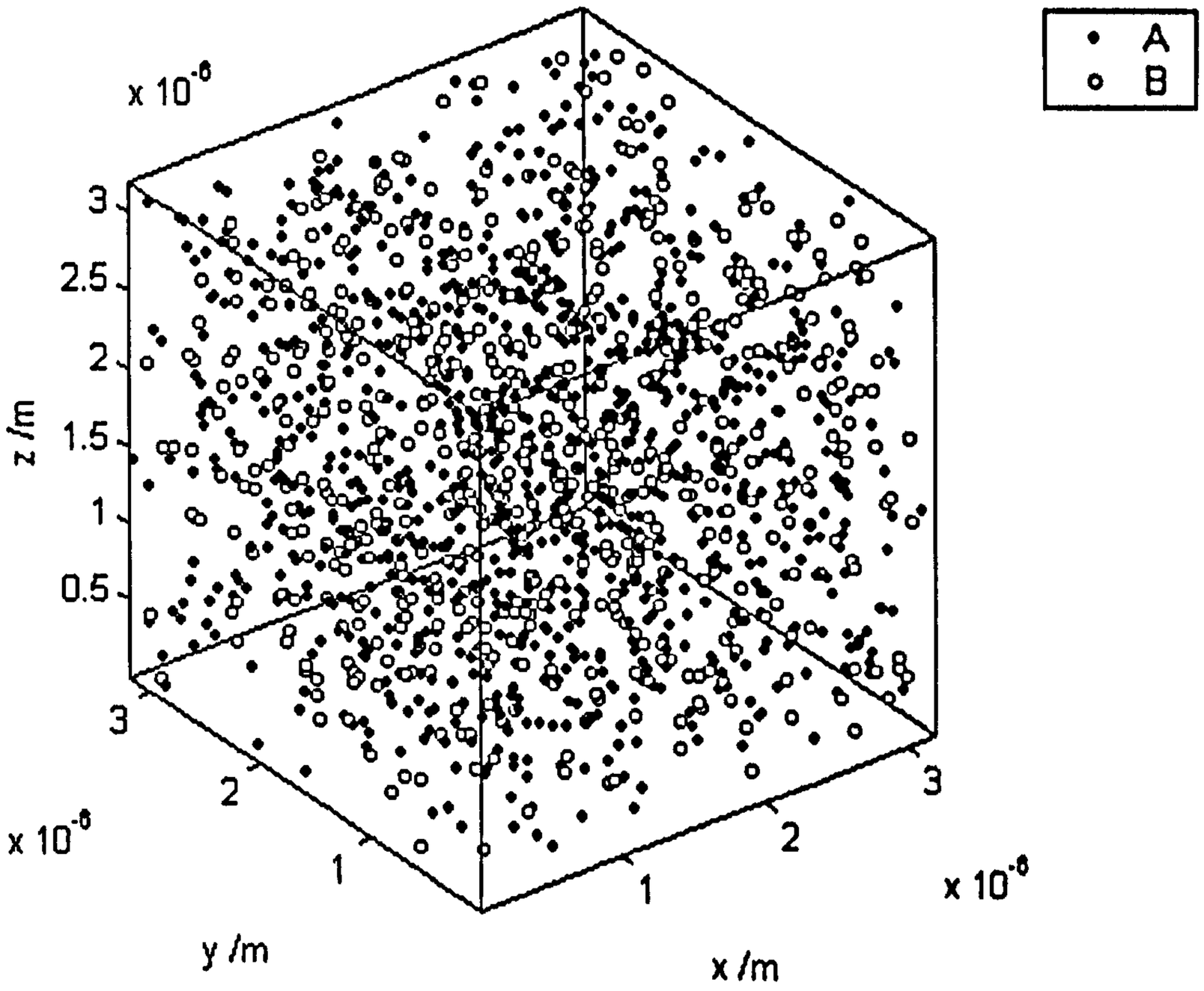

Figure 4.6: **Initial positions of A and B molecules interacting in a box.** Concentrations of A and B are arbitrary but reasonable with respect to biochemical pathways (50nM and 30nM respectively), as is the size of the box (a cube of side length 3.2μm). No C molecules exist at the start of the simulation. Positions of the molecules are random.

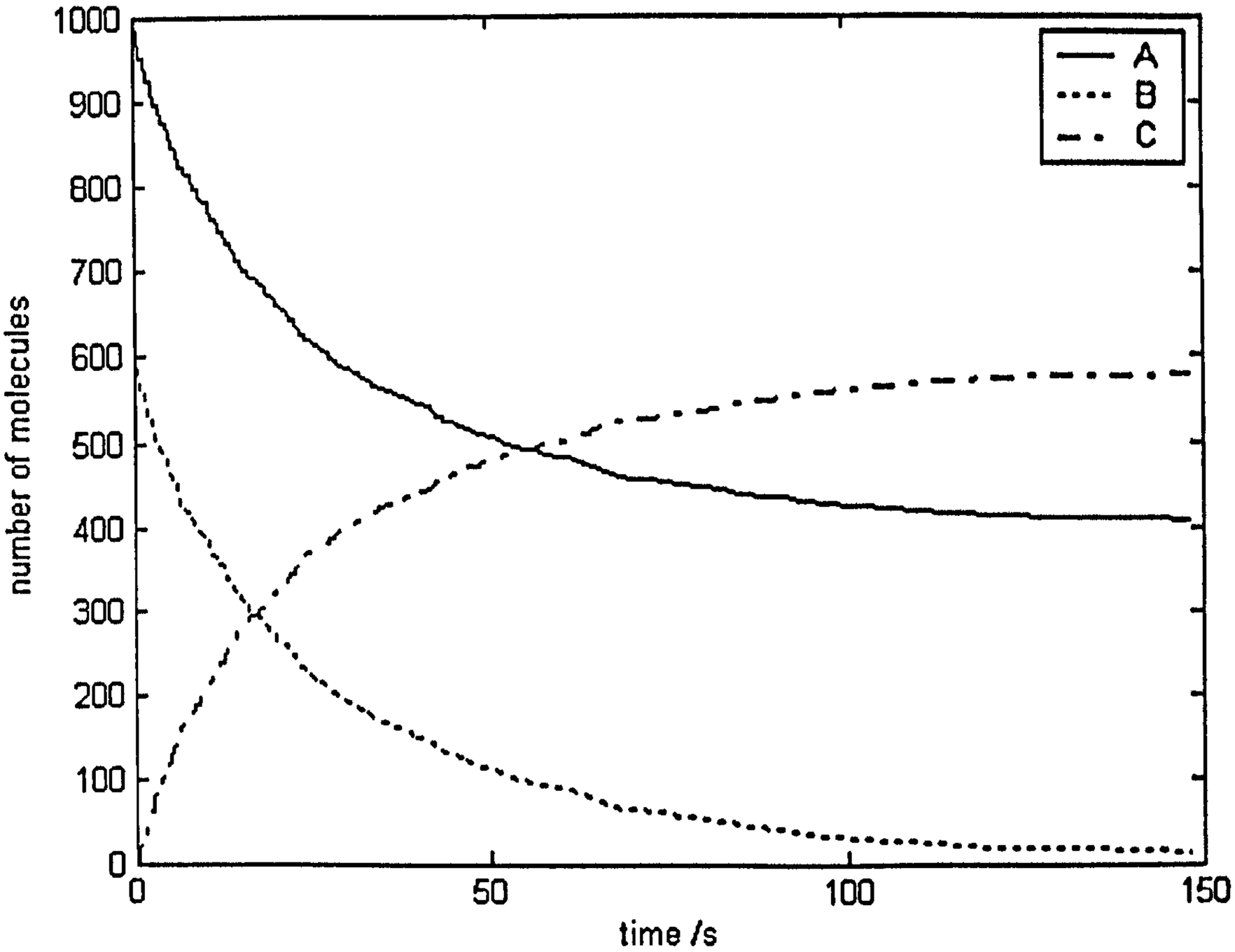

Figure 4.7: **Agent model graph of number of molecules against time for chemicals A and B combining to form C.** The rate constant is arbitrary but reasonable ($10^6 M^{-1}s^{-1}$), and is similar to the rate constant of many reactions in the NF-κB pathway. Initial concentrations are as in Figure 4.6 (with no C initially), but are expressed as the number of molecules to reflect the nature of the model. A time step of 0.5s is used here and in subsequent simulations.

the results of Figure 4.7. A time step of 0.5s is used here and in subsequent simulations; further consideration of this is given in Section 4.4. The same rate constant and initial chemical concentrations are used in a numerical solution of the corresponding reaction kinetics differential equations to give the results of Figure 4.8.

The agreement of the agent model and reaction kinetics model is clearly visible in Figure 4.7 and Figure 4.8. It is worth noting that the same results of Figure 4.7 are achieved regardless of whether A or B is assigned the interaction radius, as stated in Section 4.1.1. The random nature of the interactions can be seen in the fluctuations of the graph in Figure 4.7; these fluctuations decrease as the number of molecules involved increases, and become more pronounced as the number of molecules decreases, as would be expected (results not displayed).

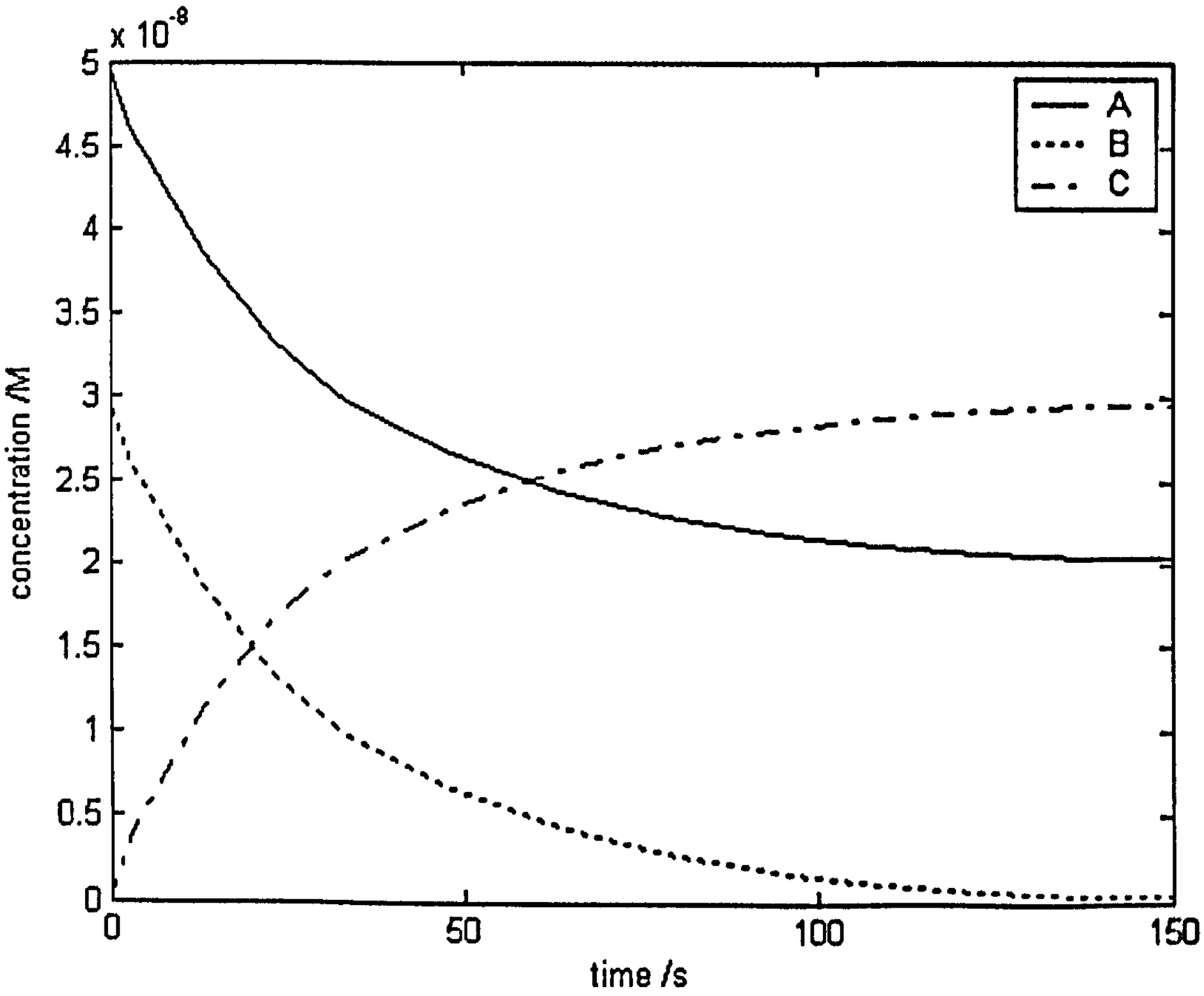

Figure 4.8: **Reaction kinetics graph of chemical concentration against time for chemicals A and B combining to form C.** The rate constant and initial concentrations are the same as used in Figure 4.7.

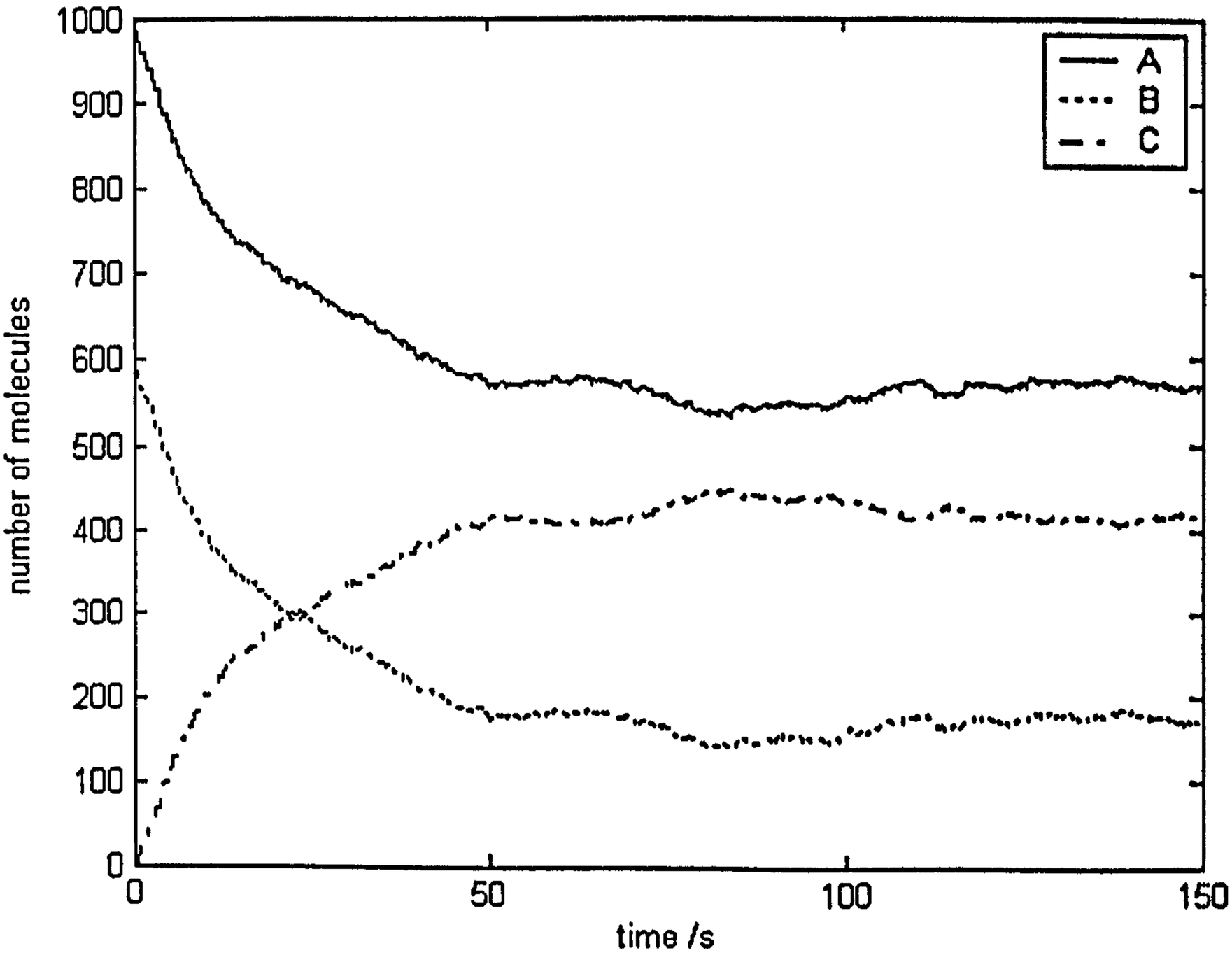

Figure 4.9: Agent model graph of number of molecules against time for chemicals A and B combining to form C, including the reverse reaction. The forward reaction rate constant and initial concentrations are the same as previously. The reverse rate constant is arbitrary but reasonable ($0.01s^{-1}$), although to make its effect more obvious it is about an order of magnitude larger than the dissociation rate constant of many of the reactions in the NF-κB pathway.

4.3.2 Two-Particle Association and Dissociation

Although it may not always be necessary to include in the pathway model, it is worth testing that agents can also be used to accurately model dissociation. A similar process would also be used to model a simple first order chemical reaction. The results for a reaction including dissociation are shown in Figure 4.9, with the differential equation solution in Figure 4.10.

The agreement between the agent model and reaction kinetics model in the circumstance where standard reaction kinetics may reasonably be applied is again clear in Figure 4.9 and Figure 4.10. The increase in fluctuations in Figure 4.9 over Figure 4.7 is due to the randomness of dissociation increasing the overall randomness of the process.

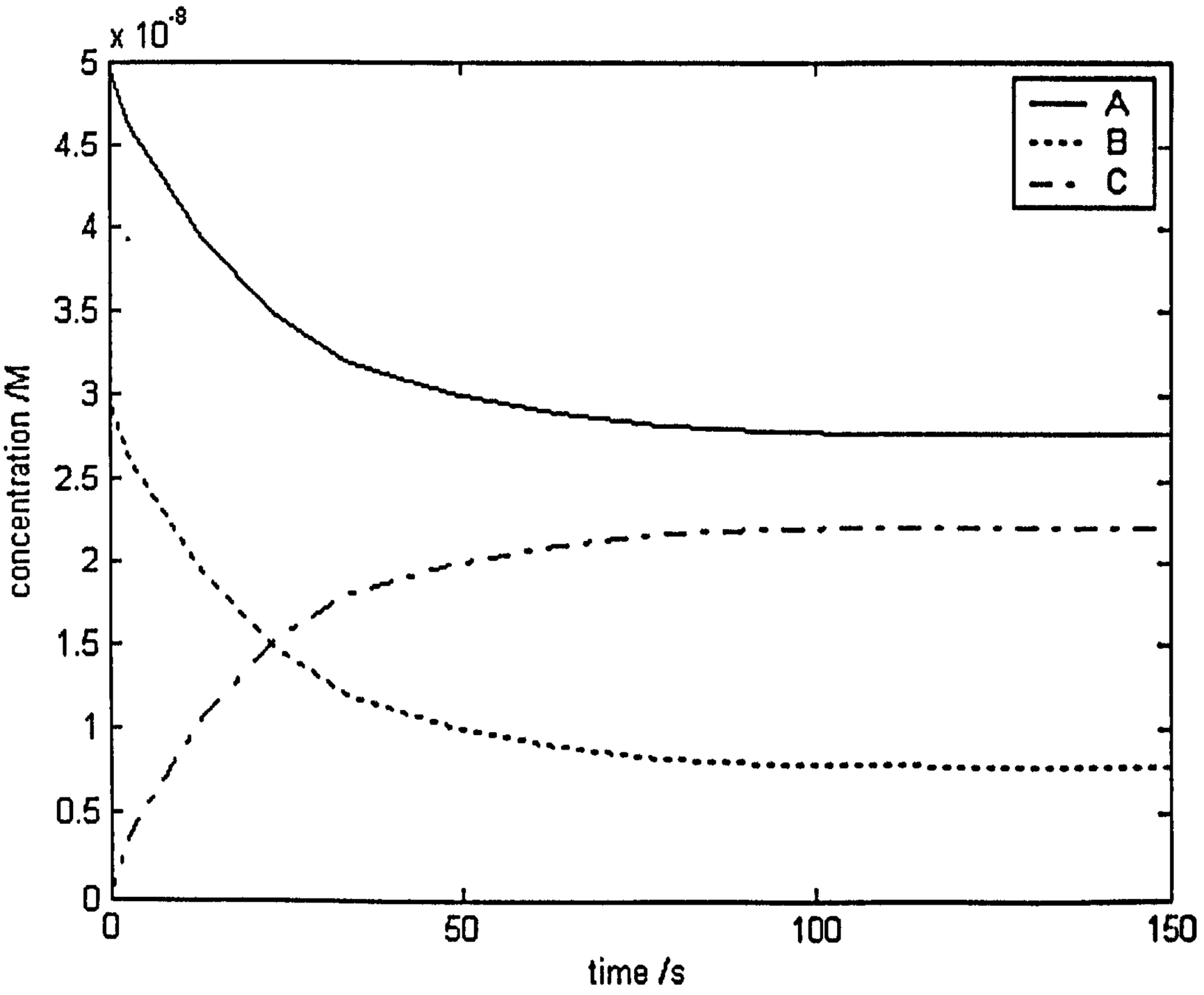

Figure 4.10: **Reaction kinetics graph of chemical concentration against time for chemicals A and B combining to form C, including the reverse reaction.** The values used are the same as those in Figure 4.9.

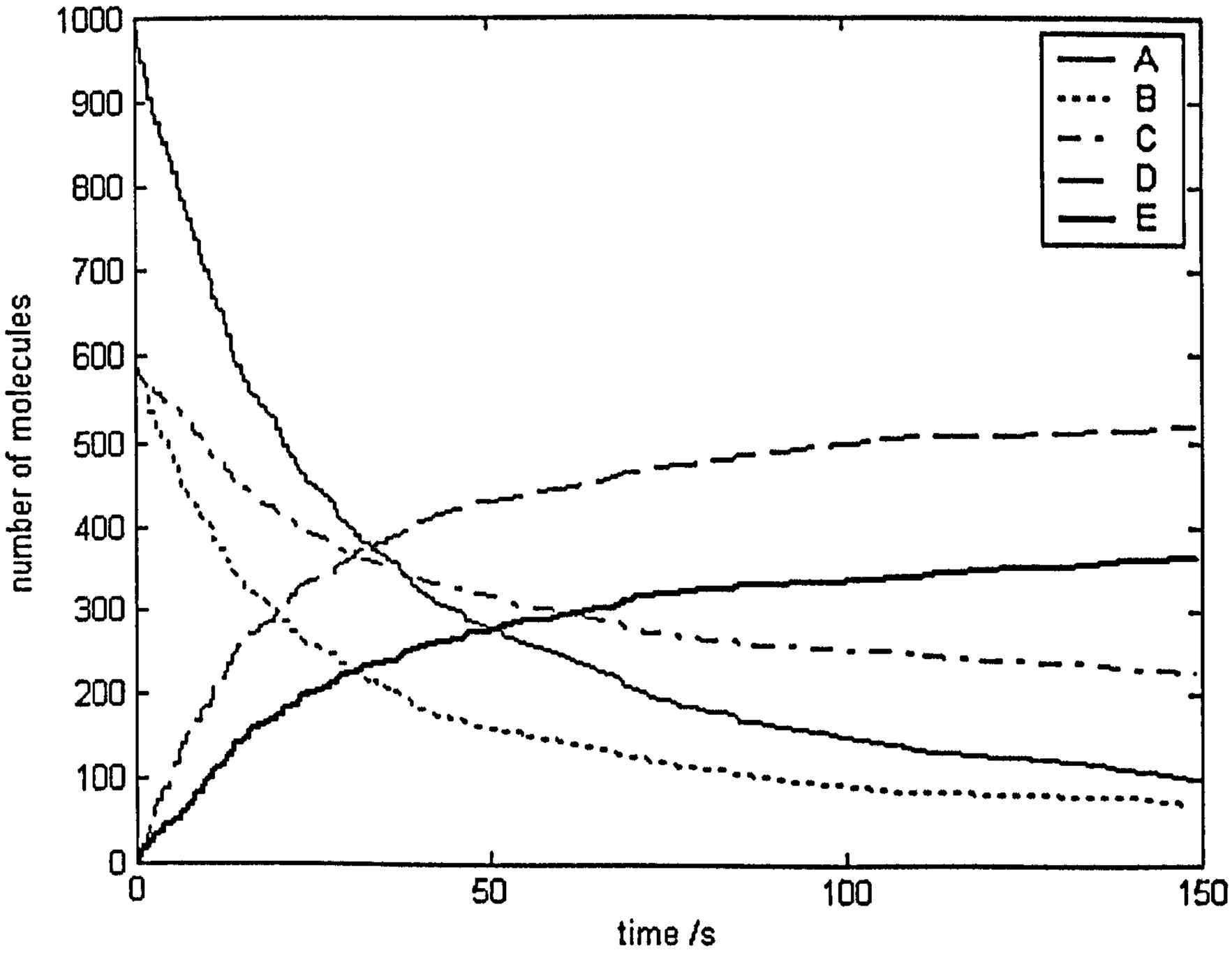

Figure 4.11: **Agent model graph of number of molecules against time for chemicals A and B combining to form D combined with A and C combining to form E.** The rate constant for A and B association is $10^6 M^{-1}s^{-1}$, and for A and C association is $5 \times 10^5 M^{-1}s^{-1}$. The initial concentration of A is 50nM, and initial concentrations of B and C are both 30nM. There are no D or E initially.

4.3.3 Competing Association

As molecules are often able to interact with more than one other type of molecule, this must also be tested in the agent-based chemical model. Neglecting dissociation, Figure 4.11 shows the agent model result of the reaction A + B → D combined with A + C → E. The equation result for the same reaction is shown in Figure 4.12.

Agreement of the agent model with reaction kinetics equations is again very good. The agent model appears to successfully recreate the basic reactions necessary to describe the NF-κB pathway, and indeed many other biochemical pathways. However, a more detailed investigation of the model—beyond simple comparisons with specific reactions—should be carried out to ensure its accuracy and understand its limitations. A consideration of this follows.

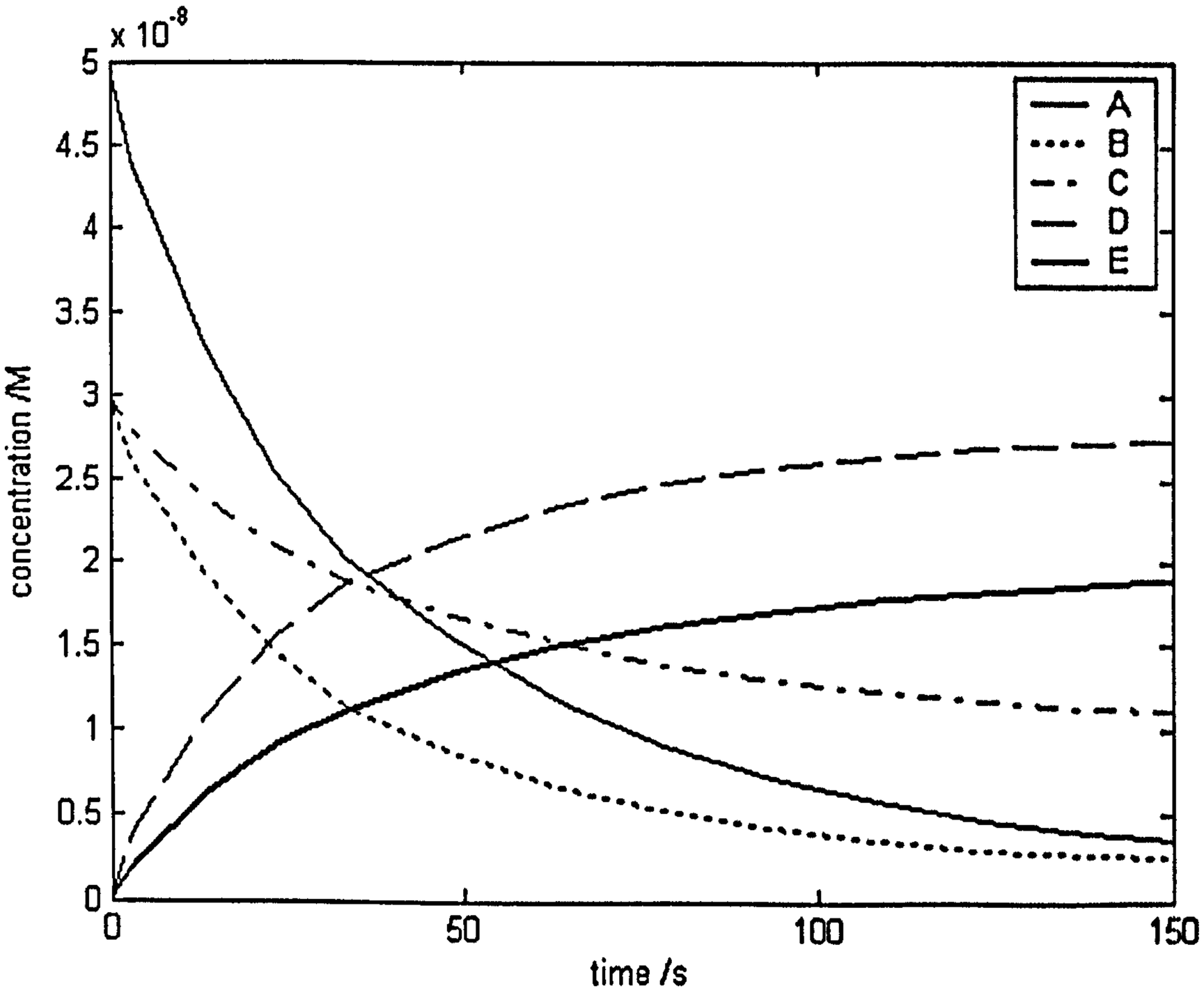

Figure 4.12: **Reaction kinetics graph of chemical concentration against time for chemicals A and B combining to form D combined with A and C combining to form E.** The values used are the same as those in Figure 4.11.

4.4 Performance of Model

4.4.1 Features to be Investigated

Although time course results from the agent-based chemical model show good agreement with ODE solutions for specific chemical systems, the model must be investigated in greater detail in order to understand better the accuracy and performance of the model, specifically as regards the effects of:

- the number of agents on the accuracy and time taken to run the model.
- the length of the time step on the accuracy and time taken to run the model.
- the movement of agents on the accuracy and time taken to run the model.

To evaluate these effects, the model is executed for a particular chemical system and each parameter is changed in turn. Accuracy is defined by how close the model comes to replicating results obtained by ODEs over the same time period; the accuracy of results in situations where the assumptions implicit in ODEs no longer hold (with low numbers of molecules) is also considered. Accuracy will be tested at two different time points, one near saturation and one away from saturation; this will indicate how well the model performs throughout a simulation. The effects of changing initial concentrations and rate constants are also investigated; these should have no effect but must be verified.

4.4.2 Effect of the Number of Agents on Accuracy and Simulation Time

The effect of different numbers of agents on the accuracy of results and simulation time is shown in Figure 4.13. The model is run for the reaction A + B $\rightarrow$ C with initial concentrations of A and B as 50nM and 30nM respectively, and association rate constant $10^6 M^{-1}s^{-1}$ (as in Section 4.3.1); the container size is scaled appropriately to maintain the correct molecular concentrations with different numbers of agents. Results from the simulation are compared with those obtained from ODEs after 50s and 150s, where the system is close to and far from saturation respectively; concentration of C molecules is used to measure accuracy.

Figure 4.13 shows that at both 50s and 150s, with more than 1,000 agents model accuracy is very good and is invariant, which is as should be expected. Simulation times

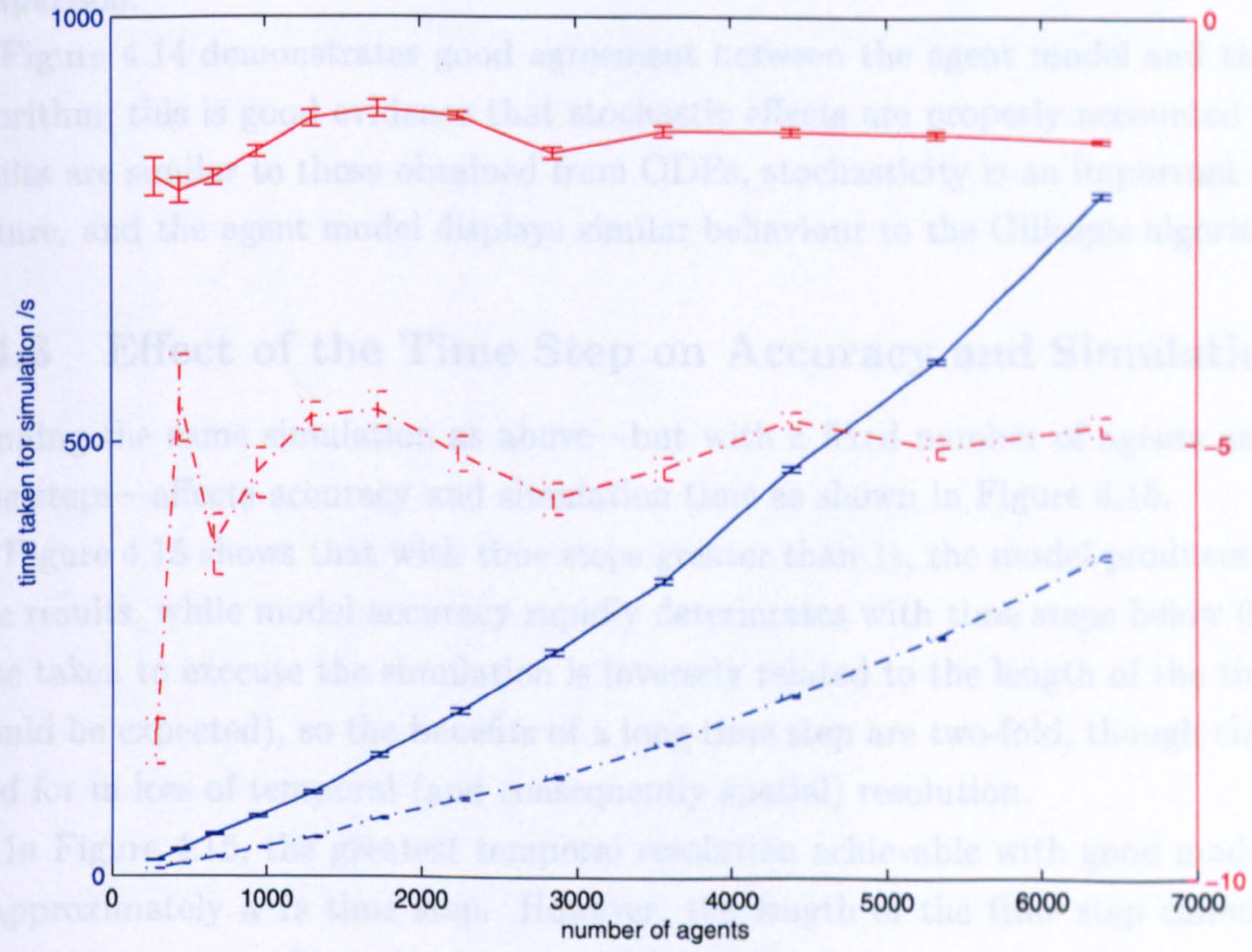

Figure 4.13: **Dependence of model accuracy on number of agents.** Effect of changing number of agents on simulation time and accuracy (as compared with ODEs). Time step fixed at 1s. Average speed fixed at 1.4×10^{-7}ms^{-1}. Accuracy of results tested after 50s (away from saturation, dashed line) and after 150s (close to saturation, solid line). Error bars show one standard error of the mean from 10 simulations.

increase with the numbers of agents, as expected. Accuracy close to saturation is consistently better than away from saturation due to the decrease in the number of degrees of freedom for the system near saturation. Accuracy as compared with ODEs deteriorates below around 1,000 agents; this requires further investigation as the assumptions implicit in ODEs no longer hold with low numbers of molecules, so a different measure of model accuracy is required.

To test accuracy with low numbers of agents, the agent model is compared with results from the Gillespie algorithm for the same system, as shown in Figure 4.14 [Gil77, Sca06]. Results are normalised around ODE results at the appropriate time for straightforward comparison.

Figure 4.14 demonstrates good agreement between the agent model and the Gillespie algorithm; this is good evidence that stochastic effects are properly accounted for. While results are similar to those obtained from ODEs, stochasticity is an important conceptual feature, and the agent model displays similar behaviour to the Gillespie algorithm.

4.4.3 Effect of the Time Step on Accuracy and Simulation Time

Running the same simulation as above—but with a fixed number of agents and different time steps—affects accuracy and simulation time as shown in Figure 4.15.

Figure 4.15 shows that with time steps greater than 1s, the model produces very accurate results, while model accuracy rapidly deteriorates with time steps below 0.5–1s. The time taken to execute the simulation is inversely related to the length of the time step (as should be expected), so the benefits of a long time step are two-fold, though clearly this is paid for in loss of temporal (and consequently spatial) resolution.

In Figure 4.15, the greatest temporal resolution achievable with good model accuracy is approximately a 1s time step. However, the length of the time step cannot be taken in isolation for its effect on accuracy—the length of the time step affects accuracy due to the change in the mixing of agents at each iteration, so it is the combination of the length of the time step with the average speed of agents that is important (i.e. average distance moved). So in the above simulation, the best spatial resolution attainable while still achieving accurate results is 1.4×10^{-7}m (a time step of 1s). Although this is orders of magnitude larger than molecular sizes, it offers an acceptable level of precision.

In the simulation in question, the interaction radius of agents is 7.3×10^{-8}m. From the above empirical results, it appears that agents must move approximately twice as far as this per iteration to achieve sufficient mixing, which is intuitively reasonable—it seems the

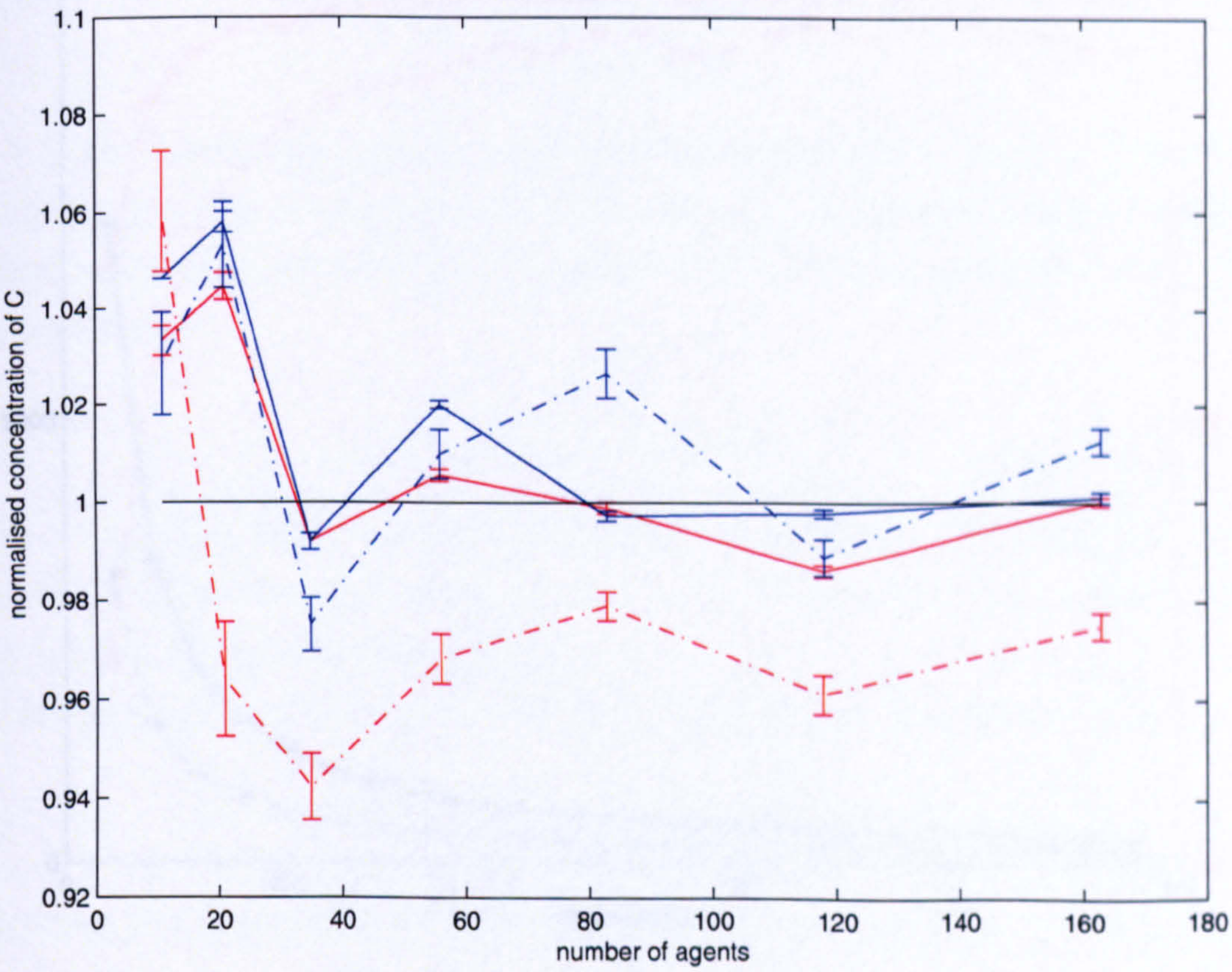

Figure 4.14: **Dependence of model accuracy on number of agents with low numbers of agents.** Plot of concentrations normalised around the ODE concentration at respective simulation times. Time step fixed at 1s. Average speed fixed at $2.2\times10^{-7}\mathrm{ms}^{-1}$ (see Section 4.4.4). Black line shows normalised ODE results. Agent results shown after 50s (away from saturation, dashed red line) and after 150s (close to saturation, solid red line). Gillespie results shown as blue lines for comparison. Error bars show one standard error of the mean from 20 simulations.

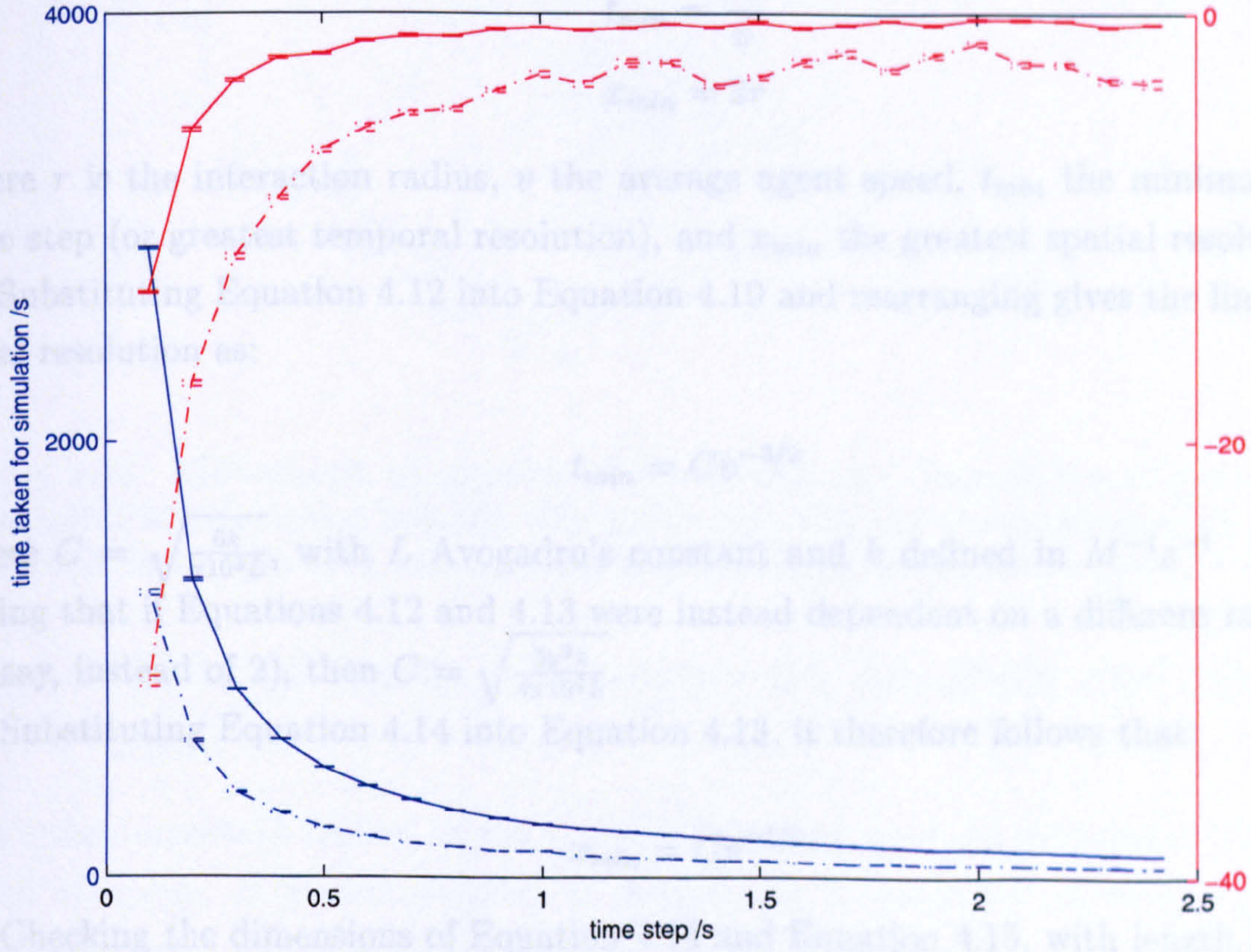

Figure 4.15: **Dependence of model accuracy on length of time step.** Effect of changing time step on simulation time and accuracy (as compared with ODEs). Total number of agents fixed at 3,568. Average speed fixed at $1.4 \times 10^{-7}\text{ms}^{-1}$. Accuracy of results tested after 50s (away from saturation, dashed line) and after 150s (close to saturation, solid line). Error bars show one standard error of the mean from 10 simulations. Change in accuracy is due to change in distance moved per time step (see Figure 4.17)

interaction volume of an agent should not on average overlap with its interaction volume at the previous time step, since this area has already been tested for interaction, and despite the fact that both chemical species move, the content of the volume in question is more likely than not to be the same. Hence temporal and spatial resolution are dictated by the rate constant and agent speed, with greatest possible temporal resolution limited to the time required for an agent to move twice the length of the interaction radius (which of course itself depends on the time step), and spatial resolution limited to twice the interaction radius, or:

$$t_{min} = \frac{2r}{v} \tag{4.12}$$

$$x_{min} = 2r \tag{4.13}$$

where r is the interaction radius, v the average agent speed, t_{min} the minimum possible time step (or greatest temporal resolution), and x_{min} the greatest spatial resolution.

Substituting Equation 4.12 into Equation 4.10 and rearranging gives the limit for temporal resolution as:

$$t_{min} = Cv^{-3/2} \tag{4.14}$$

where $C = \sqrt{\frac{6k}{\pi 10^3 L}}$, with L Avogadro's constant and k defined in $M^{-1}s^{-1}$. It is worth noting that if Equations 4.12 and 4.13 were instead dependent on a different multiple of r (y, say, instead of 2), then $C = \sqrt{\frac{3y^3k}{4\pi 10^3 L}}$.

Substituting Equation 4.14 into Equation 4.13, it therefore follows that:

$$x_{min} = Cv^{-1/2} \tag{4.15}$$

Checking the dimensions of Equation 4.14 and Equation 4.15, with length L and time T:

$$[t_{min}] = [C][v^{-3/2}] = (L^{3/2}T^{-1/2})(L^{-3/2}T^{3/2}) = T \tag{4.16}$$

$$[x_{min}] = [C][v^{-1/2}] = (L^{3/2}T^{-1/2})(L^{-1/2}T^{1/2}) = L \tag{4.17}$$

as required.

These expressions for temporal and spatial resolution will be tested below. Investigation of the effects of agent movements on model accuracy follows.

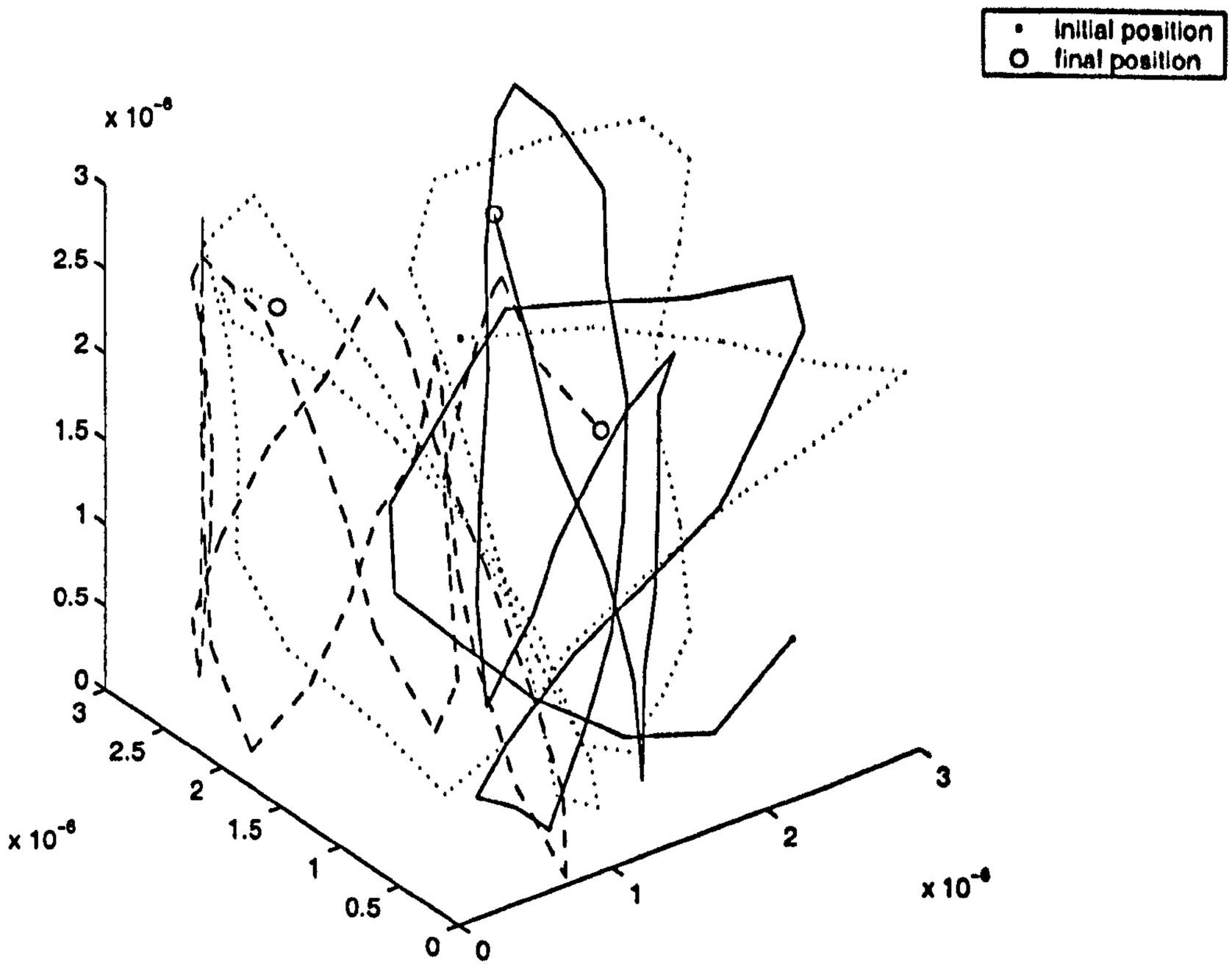

Figure 4.16: **Movement of agents in a simulation.** Visualisation of the movement of three agents in a simulation over 50s, with a time step of 1s and average speed of $5.9 \times 10^{-7} \mathrm{ms}^{-1}$. Initial and final positions of each agent are shown. Marked lengths are in metres.

4.4.4 Effect of Agent Movements on Accuracy and Simulation Time

The most important feature of the model's movement rules to investigate is the average speed of agents. It is also necessary to explore the 'randomness' of movement of agents, both the change in direction of agents and the change in distance moved at each time step. Before investigating these effects, it is worth seeing how agents move in the model; Figure 4.16 shows the movement of agents from a simulation.

Figure 4.16 shows 'reasonable'-looking movements of agents, with some randomness of movement between collisions with the container walls. Further investigation of this will begin with the average speed of agents. Clearly the average speed of agents in the model should be a 'fixed' parameter that is dependent on data about the system in question. However, the effects of different speeds of agents on model results must be investigated; this will evidently be closely linked with the effects of different time steps, as discussed

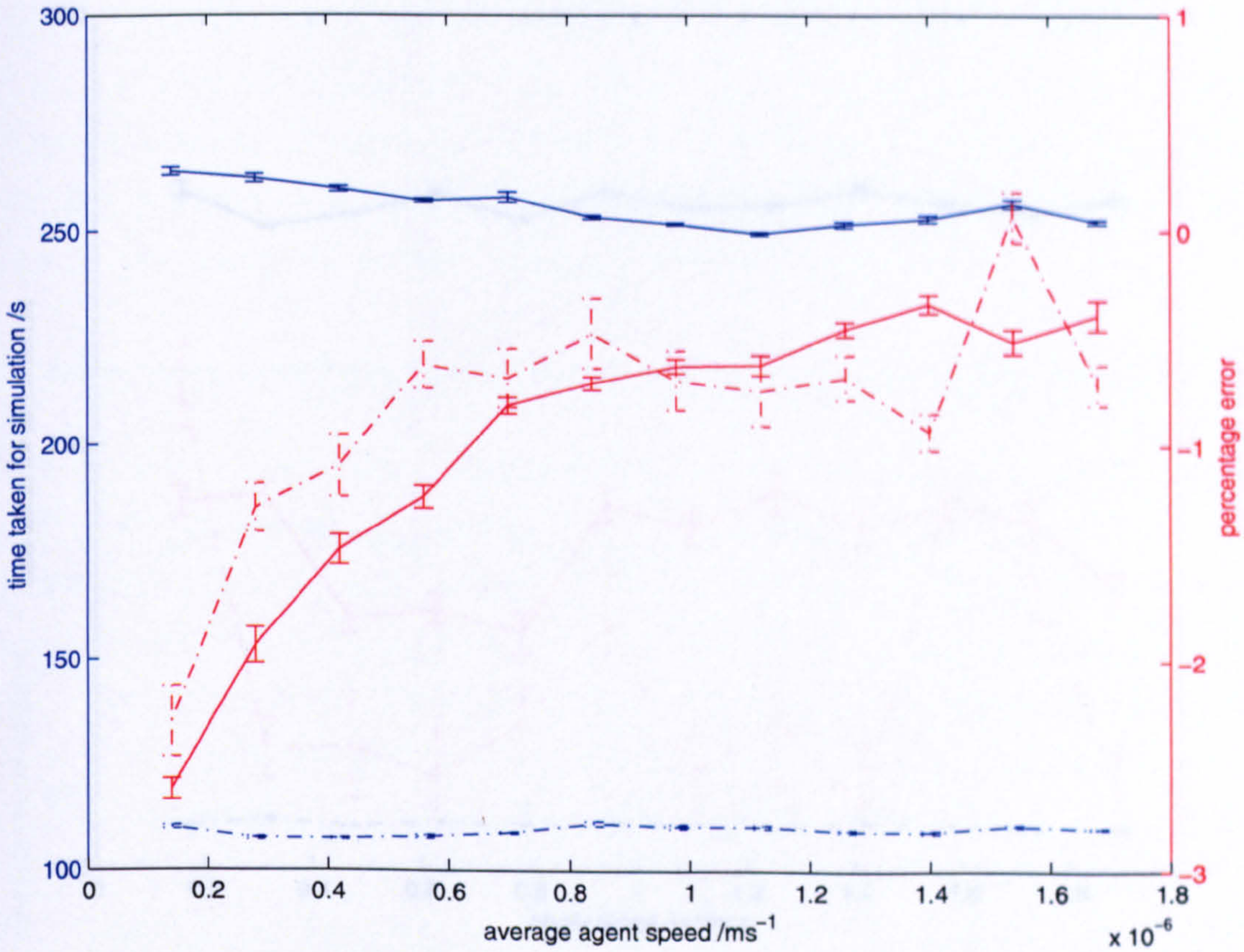

Figure 4.17: **Dependence of model accuracy on speed of agents.** Effect of changing agent speeds on simulation time and accuracy (as compared with ODEs). Total number of agents fixed at 3,568. Time step fixed at 1s. Accuracy of results tested after 50s (away from saturation, dashed line) and after 150s (close to saturation, solid line). Error bars show one standard error of the mean from 10 simulations.

above. Altering the average speed of agents, with other parameters fixed, produces the results of Figure 4.17.

As with changing time steps, model accuracy increases up to a plateau with increasing average speeds of agents. Simulation time is unaffected by different average speeds, as expected.

Using either Equation 4.14 or Equation 4.15, with the rate constant and time step used in Figure 4.17, an average speed of at least $5.9 \times 10^{-7}\text{ms}^{-1}$ is necessary for maximal accuracy of the model; this is evident in Figure 4.17 where a plateau is reached at this average speed (more markedly in the results away from saturation, which are less influenced by other factors). This strongly supports the validity of Equation 4.14 and Equation 4.15. For the same reaction as used in Figure 4.17 but with agents with an average velocity lower than $5.9 \times 10^{-7}\text{ms}^{-1}$, a larger time step would be required to give accurate results.

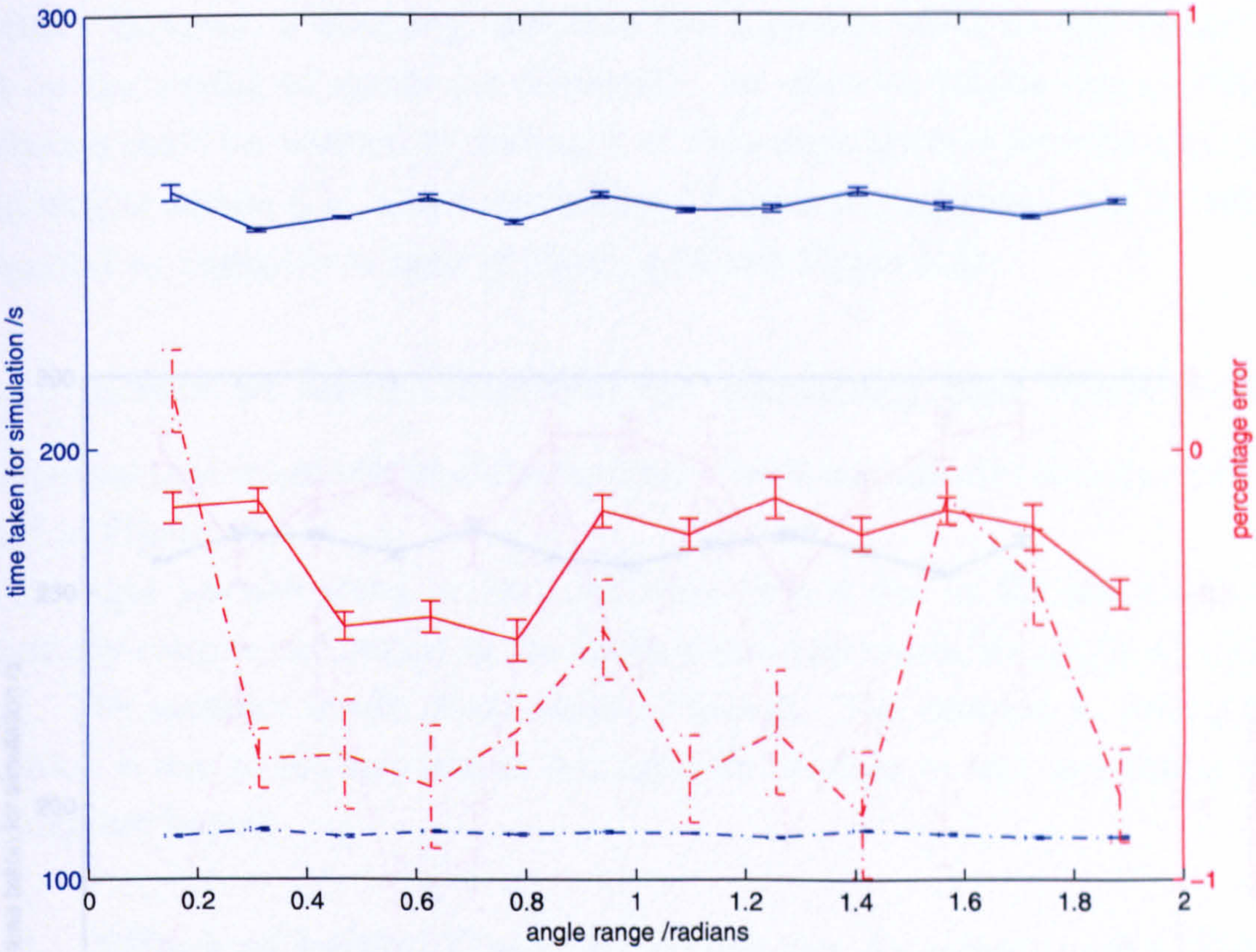

Figure 4.18: **Dependence of model accuracy on angle through which agents can move at each time step.** Effect of changing agent movements (magnitude of change in direction) on simulation time and accuracy (as compared with ODEs). Total number of agents fixed at 3,568. Time step fixed at 1s. Average speed fixed at $5.9 \times 10^{-7} \mathrm{ms}^{-1}$. Accuracy of results tested after 50s (away from saturation, dashed line) and after 150s (close to saturation, solid line). Error bars show one standard error of the mean from 10 simulations.

Although a minor feature of movement rules, it is worth checking that changing the angle range within which agents can deviate from their present direction, and also the distance range (as shown in Figure 4.5), does not have a significant effect on results, as displayed in Figure 4.18 and Figure 4.19.

Figure 4.18 and Figure 4.19 both show no effect on model accuracy when altering the parameters for the 'randomness' of movement, showing they are of minor importance for the model, which is a desirable feature of the model. This is expected due to the limits of spatial resolution in the model (stated in Equation 4.15), meaning that alterations in movement which are smaller than this should not affect results. However, this does not mean that randomness of movement is unnecessary since some randomness is required to minimise systematic bias towards any particular behaviour and over-dependence on initial

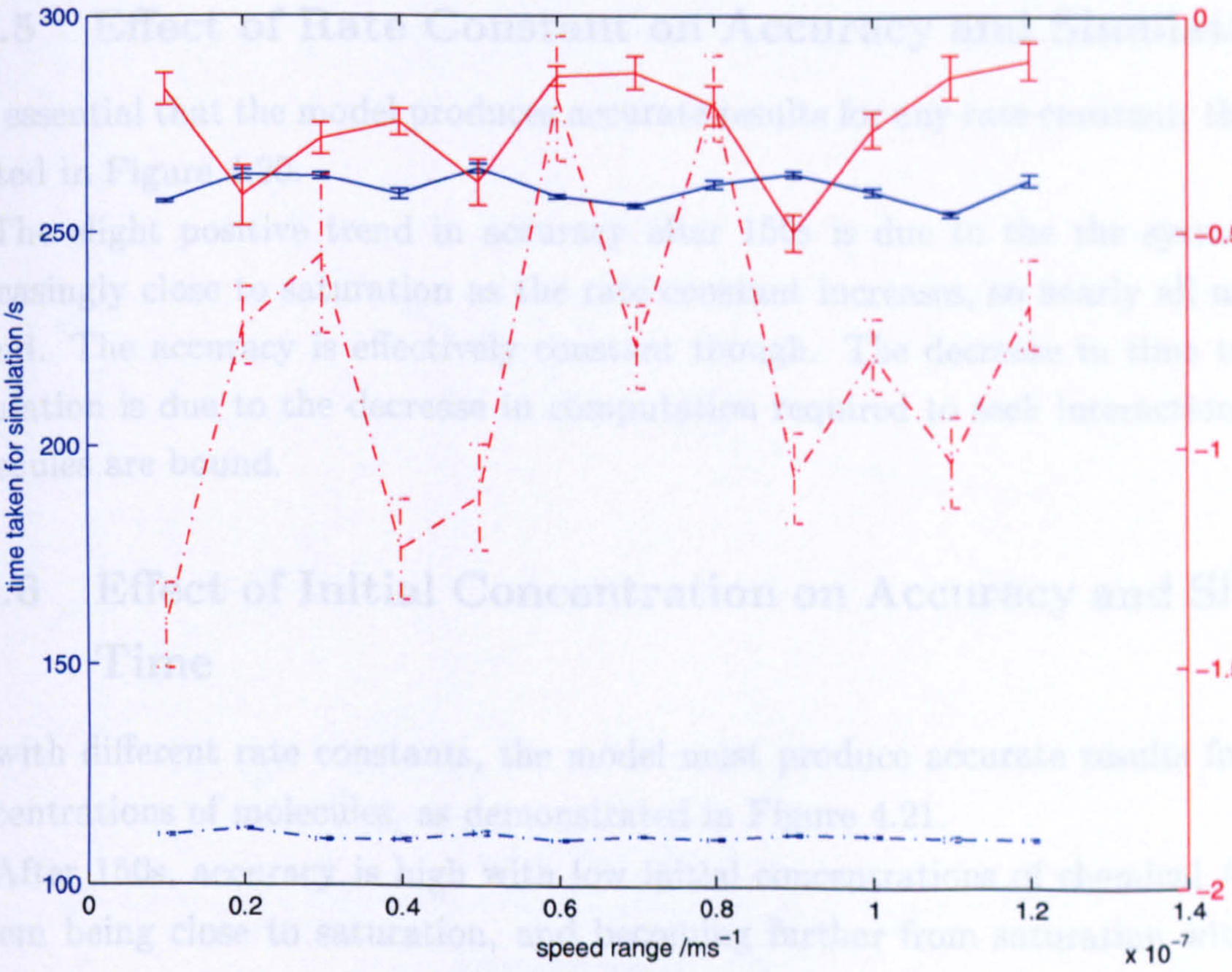

Figure 4.19: **Dependence of model accuracy on change in speed of agents at each time step.** Effect of changing agent movements (magnitude of possible variation of distance moved) on simulation time and accuracy (as compared with ODEs). Total number of agents fixed at 3,568. Time step fixed at 1s. Average speed fixed at 5.9×10^{-7}ms^{-1}. Accuracy of results tested after 50s (away from saturation, dashed line) and after 150s (close to saturation, solid line). Error bars show one standard error of the mean from 10 simulations.

conditions.

Because of the nature of the movement rules, changing the time step will change overall agent movements slightly because it changes the frequency with which they can change direction. However, a changing time step has a greater effect on the model due to its effect on the mixing of agents (as discussed). Its effect on results due to altering agent movements could be assessed by testing it at time steps above where the time step affects the mixing of agents (i.e. above the minimal temporal resolution), but its effect can be disregarded as negligible in light of Figure 4.18 and Figure 4.19.

4.4.5 Effect of Rate Constant on Accuracy and Simulation Time

It is essential that the model produces accurate results for any rate constant; this is demonstrated in Figure 4.20.

The slight positive trend in accuracy after 150s is due to the the system becoming increasingly close to saturation as the rate constant increases, so nearly all molecules are bound. The accuracy is effectively constant though. The decrease in time taken for the simulation is due to the decrease in computation required to seek interaction when more molecules are bound.

4.4.6 Effect of Initial Concentration on Accuracy and Simulation Time

As with different rate constants, the model must produce accurate results for any initial concentrations of molecules, as demonstrated in Figure 4.21.

After 150s, accuracy is high with low initial concentrations of chemical A due to the system being close to saturation, and becoming further from saturation with increasing concentrations of chemical A. This very slight decrease in accuracy is 'over-ridden' by the larger number of agents involved as the concentration increases (as in Figure 4.13). Away from saturation there is simply a slight increase in accuracy with increasing numbers of agents. Accuracy is effectively constant though. The increases in simulation times are due to increasing numbers of agents.

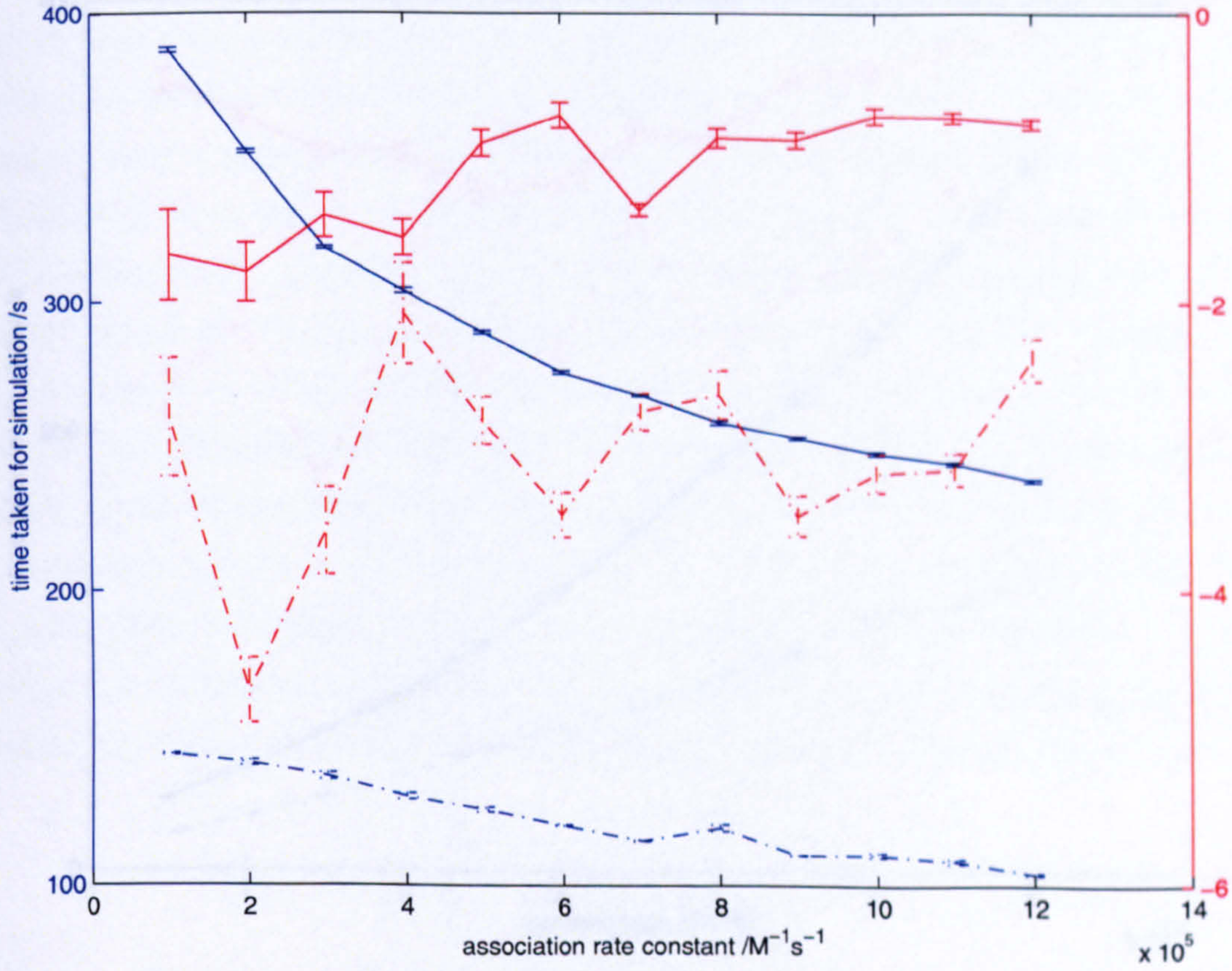

Figure 4.20: **Dependence of model accuracy on rate constant.** Effect of changing association rate constant on simulation time and accuracy (as compared with ODEs). Total number of agents fixed at 3,568. Time step fixed at 1s. Average speed fixed at 1.4×10^{-7}ms^{-1}. Accuracy of results tested after 50s (away from saturation, dashed line) and after 150s (close to saturation, solid line). Error bars show one standard error of the mean from 10 simulations.

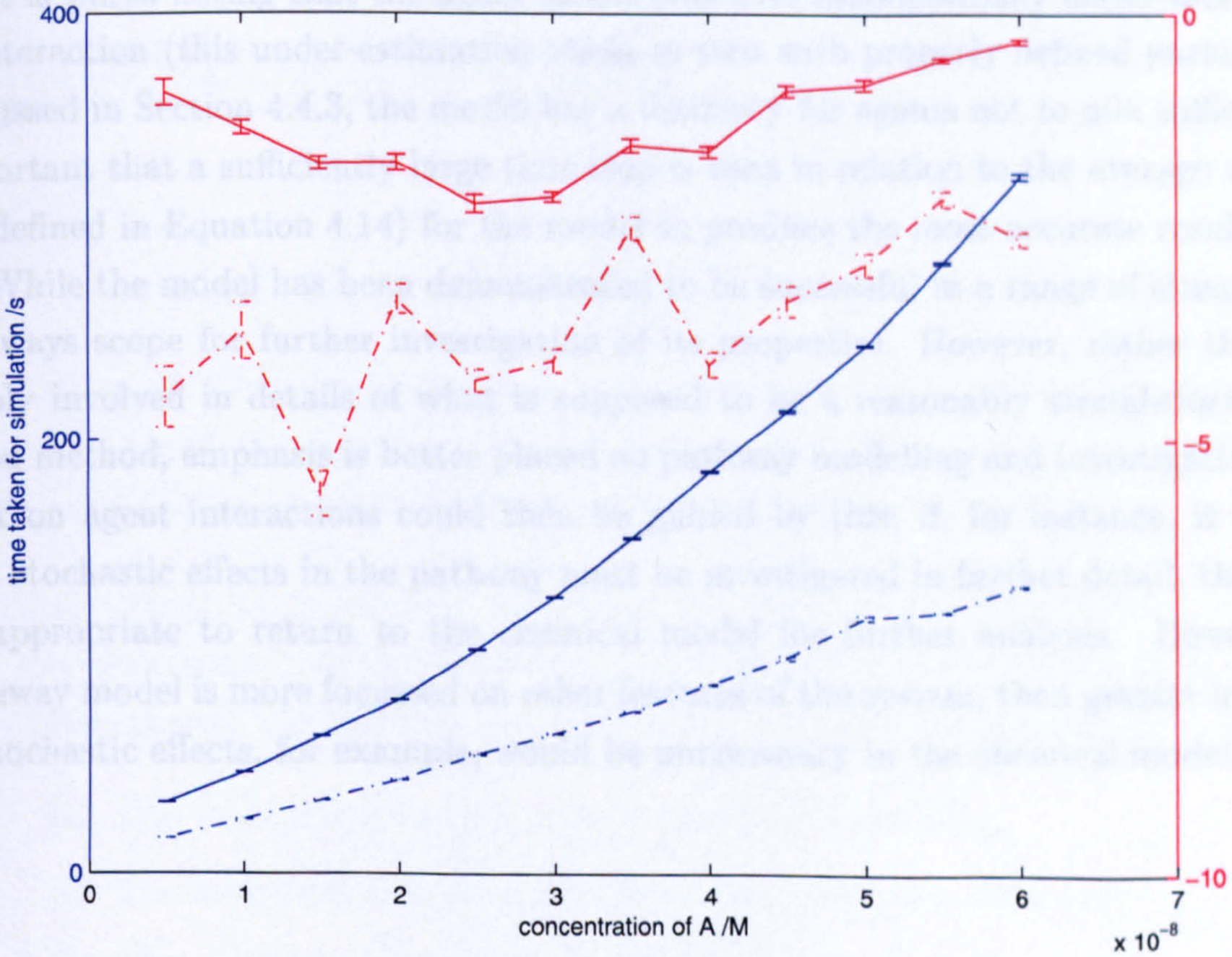

Figure 4.21: **Dependence of model accuracy on initial concentration.** Effect of changing initial concentration of A on simulation time and accuracy (as compared with ODEs). Total number of agents ranges from 1,561–4,014 depending on initial A concentration. Time step fixed at 1s. Average speed fixed at $1.4 \times 10^{-7}\text{ms}^{-1}$. Accuracy of results tested after 50s (away from saturation, dashed line) and after 150s (close to saturation, solid line). Error bars show one standard error of the mean from 10 simulations.

4.5 Summary of Chemical Model

An agent-based model of chemical interactions has been formulated for use in systems where spatial detail is required, but also where reasonable simulation times are more important than single-molecule resolution. This means the model is very suitable for modelling intracellular pathways, fulfilling the objectives stated in Section 4.1.

The model has been investigated in some detail to ensure it behaves as necessary for modelling chemical pathways; its spatial and temporal resolution have been defined, and its results compare well with both ODE and stochastic results in appropriate situations.

It is worth noting that the agent model only ever systematically under-estimates rates of interaction (this under-estimation tends to zero with properly defined parameters). As discussed in Section 4.4.3, the model has a tendency for agents not to mix sufficiently; it is important that a sufficiently large time step is used in relation to the average agent speed (as defined in Equation 4.14) for the model to produce the most accurate results possible.

While the model has been demonstrated to be successful in a range of situations, there is always scope for further investigation of its properties. However, rather than become deeply involved in details of what is supposed to be a reasonably straightforward agent-based method, emphasis is better placed on pathway modelling and investigation. Further work on agent interactions could then be guided by this; if, for instance, it were found that stochastic effects in the pathway must be investigated in further detail, then it would be appropriate to return to the chemical model for further analysis. However, if the pathway model is more focussed on other features of the system, then greater investigation of stochastic effects, for example, would be unnecessary in the chemical model.

Chapter 5

Agent-Based Modelling of the NF-κB Signalling Pathway

5.1 Formulation of Model

Formulation of the pathway model requires the consideration of what the model seeks to achieve, how and why it will do this, and how its suitability will be measured. Successful agent-based modelling of the pathway depends on the following:

1. Specifying the 'population' level to be modelled—what areas of the pathway should be investigated; where the start and end are defined; and what external influences interact with the pathway, including other pathways. This should be directed by:
   (a) The questions about the system that need answering.
   (b) The biological data that exist to inform and verify the model.
   (c) The direction future work may take.
2. Identifying the individual components of the system to be modelled as agents. This should be directed by the above, as well as:
   (a) The levels at which constituents of the system may be treated as a single entity.
   (b) The areas of the system that are most important for its operation.
   (c) The areas of the system that are most central to the questions the model seeks to address.

(d) The level of detail required, both to form a realistic model and for results to be useful.

(c) The amount of computing power and time available.

3. Verifying the model—both at the individual and system level—against existing data.

4. Forming testable predictions from the model, which should be directed by the same considerations as used to specify what should be modelled at the population level, as well as:

(a) The level of detail the model provides and the limits in which it operates.

5. Validating predictions of the model with experimental results.

Since the pathway is a set of molecular interactions, the individual molecule is a natural choice of agent for the model; a 'lower-level' agent would be not only unnecessary and very difficult to include, but an actual hindrance to the model both in terms of computation and tractability; and a 'higher-level' agent would be too much of an abstraction for much of the system, providing little benefit over other modelling approaches.

To a certain extent, the pathway model can be formulated simply by defining the necessary chemical reactions in the agent-based chemical model (see Chapter 4), which has already been largely verified. However, further model details must also be defined, including cellular compartments and shape, more abstracted peripheral agents, and pathway activation.

5.1.1 Main Features to Model

Due to its importance to the NF-κB signalling pathway, the available biological data, and its relative simplicity, TIR pathway activation is the starting point for modelling.

Each NF-κB, IκB and IκB-kinase (IKK) molecule is an individual agent, as are the TIRs, as depicted in Figure 5.1. Importing and exporting nuclear receptors are also modelled as individual agents, which is an abstraction of the transport mechanism described in Section 2.3.3 (the nuclear pores are not directly modelled, but nuclear receptors are restricted to the nuclear membrane and assumed to each have free access to a pore).

In addition to the model's key agents, there also exist temporary agents and other devices to deal with the complexities and computational limitations of the system. IL-1 cytokines are not modelled as agents but are instead treated as a whole chemical entity

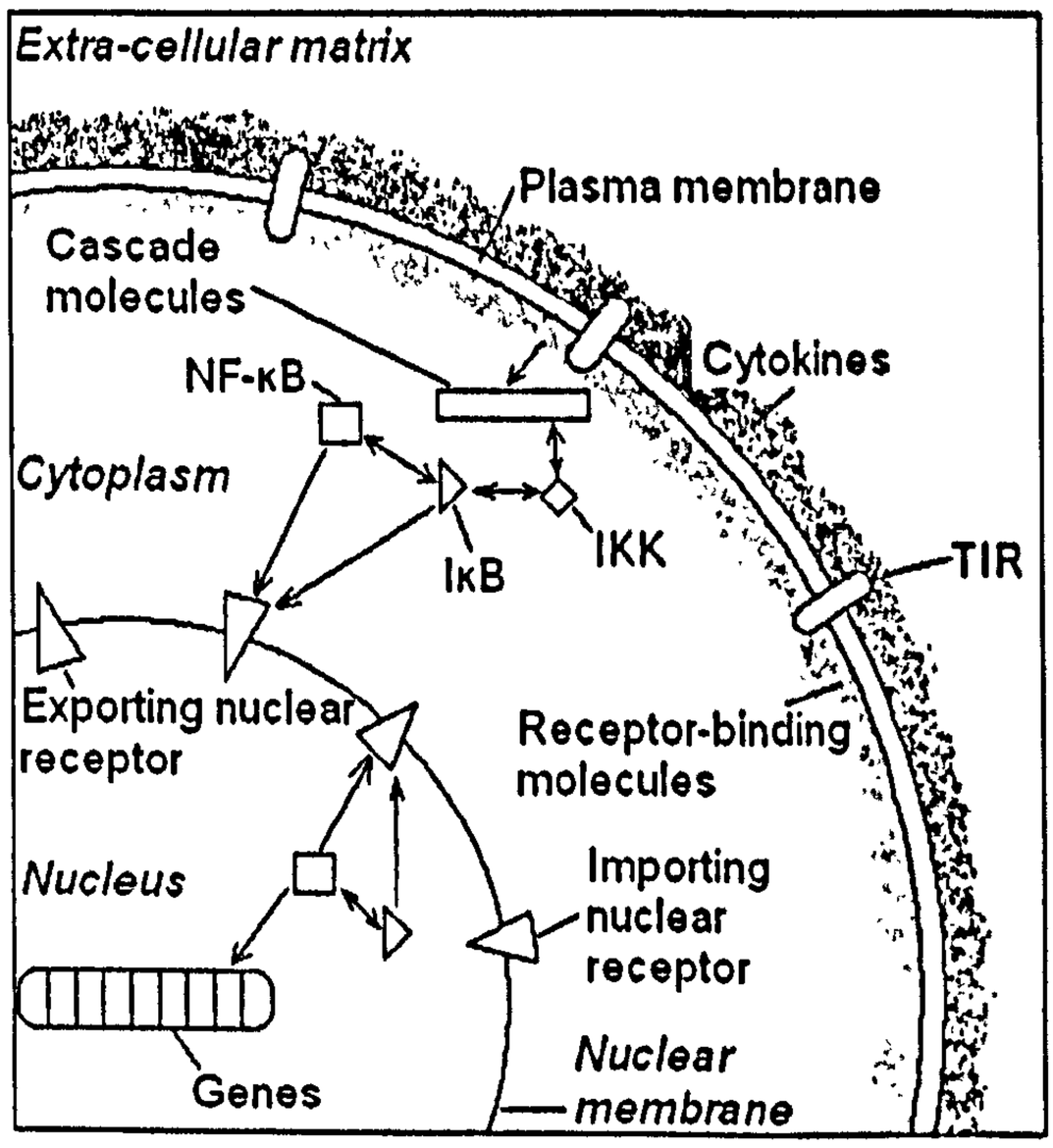

Figure 5.1: **Simplified diagram of principal pathway agents.** Interactions between two agents may only occur if both are within a defined proximity and their internal states permit. Defined internal states of the receptors and genes are shown; other agents have more possible states and are not marked on the diagram. Soluble agonists and receptor-binding molecules are not modelled as agents, as described in the text.

whose fluctuating local concentrations on the cell surface are required to rise above a certain level in order to initiate signal transduction in nearby dormant TIR agents.

A similar method is used at the intracellular domain of the TIR's signal transduction, where local concentrations of certain molecules in the cytoplasm must be above a certain level in the vicinity of an active TIR to complete the process. Following this, a temporary agent with an internal time delay is created to account for the cascade that triggers the upstream kinases, including IRAKs. This allows for the complex molecular interactions and cross signalling with other trigger mechanisms prior to IKK activation. Clearly this can be developed as required by decomposing the cascade into as many agents as necessary. At the other end of the pathway, a similar temporary agent method is used to account for transcription and translation, resulting in the creation of new IκB agents. The inhibitor of primary interest to the model is IκBα—the most critical and abundant inhibitor (see Section 2.2.4)—but other isoforms can also be included as necessary.

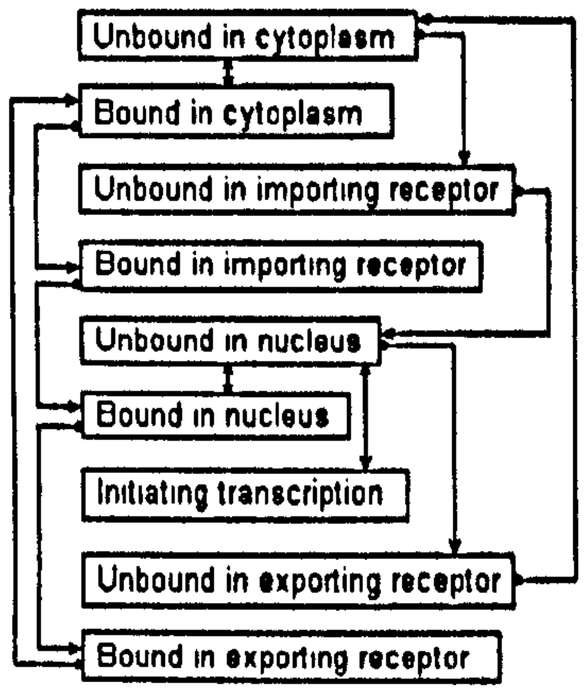

Figure 5.2: **States and transitions of NF-κB.** The possible state transitions shown depend on an agent's immediate environment. For example, bound to unbound transitions in the cytoplasm may be due to random thermal dissociation, or may involve IKK; in the cytoplasm the same transition can only occur by random thermal dissociation since IKK does not exist in the nucleus.

5.1.2 States of Agents

Each agent exists in one of a set number of possible states at any time. For example, an NF-κB molecule may be free in the cytoplasm, free in the nucleus, bound to IκB in the cytoplasm, bound to IκB in the nucleus, being transported in an importing or exporting nuclear receptor while being either bound or unbound to IκB, or initiating gene transcription, as shown in Figure 5.2. At every time step, each agent observes its surrounding environment, seeking interaction in accord with its present state, and if successful, consequently changing state. For example, if NF-κB is unbound in the nucleus, it may bind to nuclear IκB, move through the nuclear membrane with an exporting nuclear receptor, initiate gene transcription, or remain unbound in the nucleus, depending on the proximity and availability of the appropriate agents, with a weighting for the likelihood of each, and an appropriate time delay where necessary. If the same agent were instead unbound in the cytoplasm, for example, it may bind to cytoplasmic rather than nuclear IκB, move through the nuclear membrane with an importing as opposed to exporting nuclear receptor, and would not be able to initiate gene transcription.

5.1.3 Movement of Agents

Membrane-Bound Movements

Non-membrane-bound agent movements are as described in Section 4.1.2. However, the movement of membrane-bound agents must also be considered. This follows the same principles as other agent movements, except the movement must be constrained to two dimensions. The movement rules therefore rely on the geometry of the surface; as a first approximation to the shape of cells, this is a spherical cell boundary with a concentric spherical nucleus.

Each membrane-bound agent is assigned a random initial position on the membrane, an initial direction of movement, an average speed (converted to an angular speed) on the membrane, a possible range of deviation from this speed at each time step, and a range of angles through which its movement can deviate at each time step. Each agent is also assigned a direction of movement in terms of theta and phi values in spherical polar coordinates (with the origin at the centre of the cell, as opposed to the origin being at the position of the agent for non-membrane-bound movements). At every time step each agent deviates from its average speed by a random amount and from its previous direction by a random amount, both within defined limits, as though Figure 4.5 were curved over the membrane (see Appendix B for movement code).

Boundary Collisions

Clearly, membrane-bound agents do not undergo boundary collisions. However, the movement of all other agents requires boundaries to be considered. Since the boundaries are spherical, elastic collisions are simply a reflection of the radial component of the movement of the agent, as shown in Figure 5.3. It is worth considering the length of the movement at each time step in relation to the size of the container, since this could result in the spurious results also shown in Figure 5.3, though with reasonably sized time steps this is not a problem.

5.1.4 Model Simulation

When running the model, the latent system is allowed to reach a steady state prior to activation. The soluble agonist is then introduced to the cell surface for a fixed period. At each time step, the order in which each set of interactions is computed, such as NF-κB association with IκB, is randomised to ensure that no interaction is given priority in the

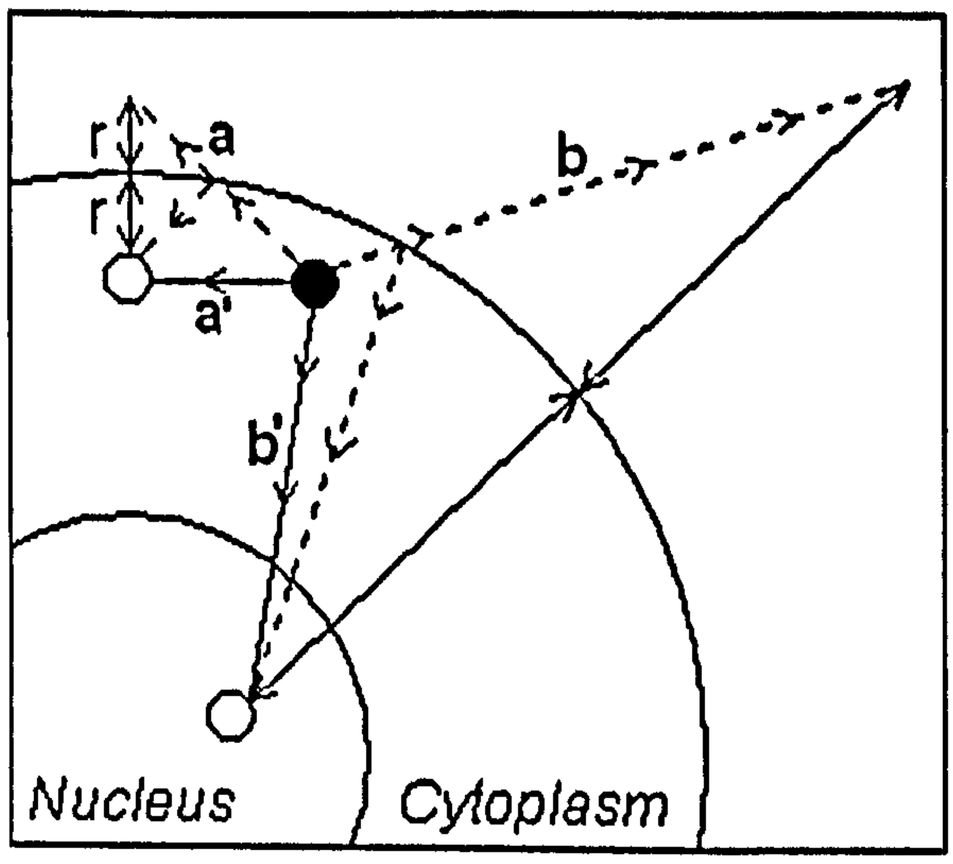

Figure 5.3: **Example of an agent undergoing an elastic 'collision' with a membrane.** The random movement vector a of a cytoplasmic agent has moved it outside the cell. Reversing the radial component r of its displacement outside the cell puts it back into the cytoplasm, as though an elastic collision with the membrane has occurred, with resultant vector a'. Note that if the speed of the agent is too great, as with vector b, it could move so far out of its compartment that this method would not return it to the cytoplasm.

model as a whole, as discussed in Section 4.1.1. The generic algorithm is displayed in Appendix C and the parameter values used are shown in Appendix D.

An image of agent positions in the model is shown in Figure 5.4. The cell is scaled down to reduce computation; this is justified in light of the results of Figure 4.13, and is accepted practice in comparable stochastic differential equation modelling [BT03]. Scaling will affect noise in the system, which should be accounted for if stochastic effects are to be considered in detail.

5.2 Basic Pathway Operation

5.2.1 Biological Results for Basic Pathway Operation

Biological results for a relatively simple area of pathway operation are necessary to verify the model. Data for IL-1 stimulation of the pathway are used, which were obtained from monitoring the pathway in live cells using GFP-tagged NF-κB and IκBα molecules by confocal microscopy, as described in Section 2.6 [YRQ03, CDQ01, CCDQ99]. Simultaneous observation of the NF-κB subunit relA and IκBα were carried out using cyan (ECFP) and

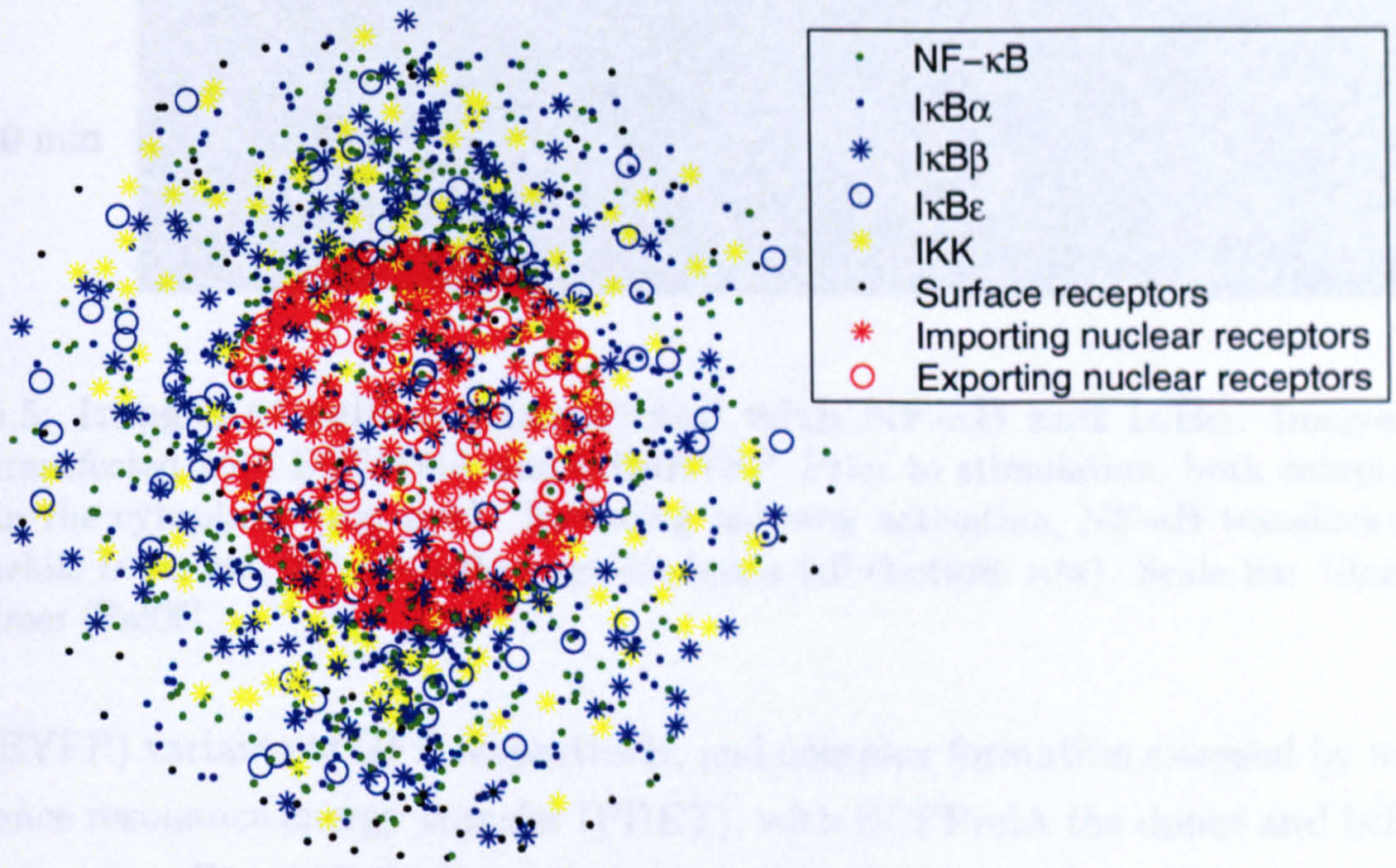

Figure 5.4: **Three-dimensional visualisation of the positions of agents in the model at a moment in time.** Positions are random within each agent's confines. The cell is spherical, with a cell membrane, nuclear membrane, cytoplasm and nucleus. The genes that NF-κB can activate are placed randomly along a line at the centre of the nucleus, though are not shown. The cell is scaled down from an average diameter of around 10μm to a diameter of 2μm to reduce computation, containing of the order of 1,000 agents. Concentrations and rate constants of molecules are based on biological data.

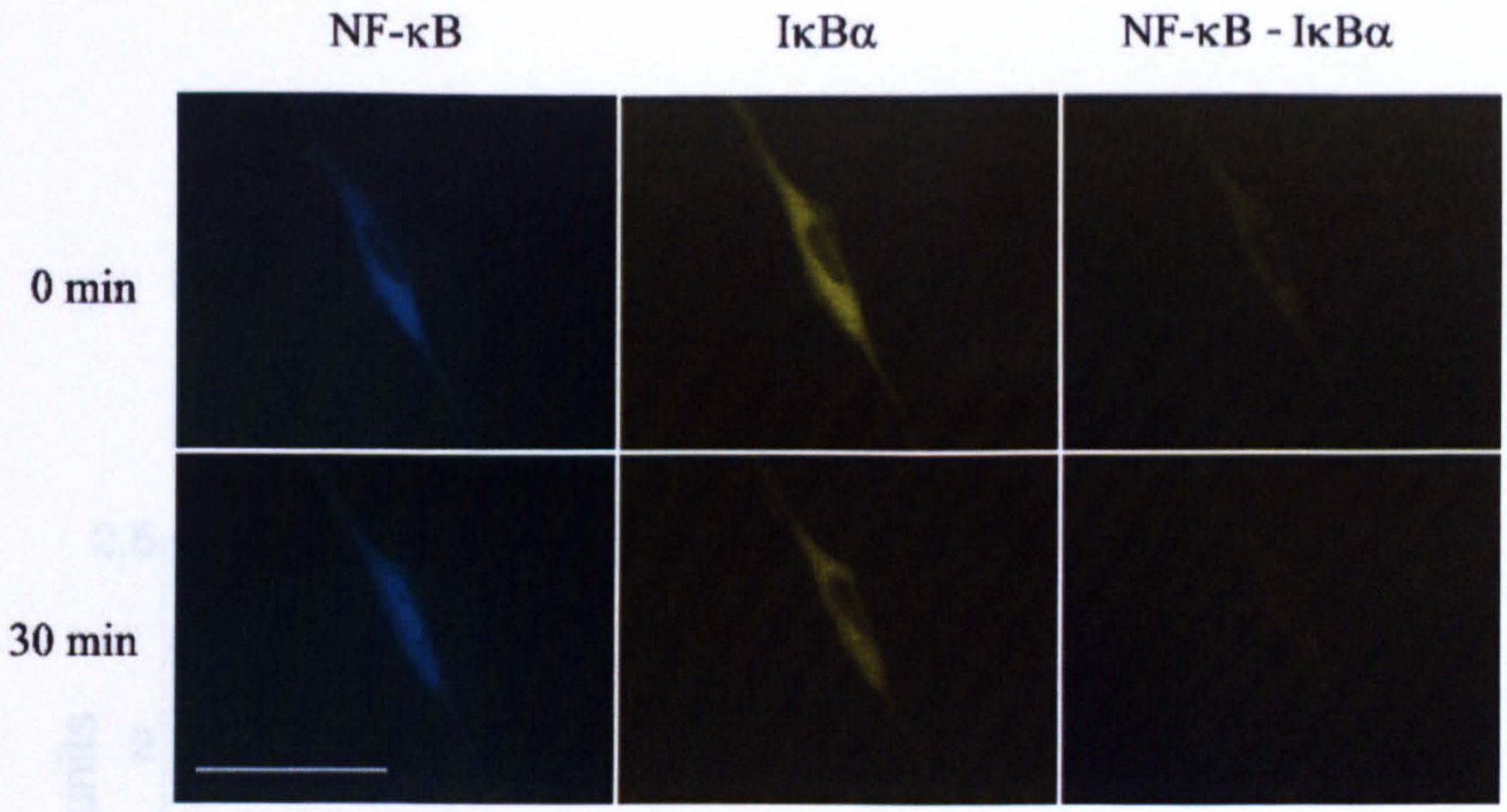

Figure 5.5: **Images of cells co-transfected with NF-κB and IκBα.** Images of single cells co-transfected with ECFPrelA and IκBαEYFP. Prior to stimulation, both components are located in the cytoplasm (top row). Following pathway activation, NF-κB translocates to the nucleus while IκBα and NF-κB-IκBα complex levels fall (bottom row). Scale bar 10μm. Figure from Palmer [Pal06].

yellow (EYFP) variants of GFP respectively, and complex formation assessed by measuring fluorescence resonance energy transfer (FRET), with ECFPrelA the donor and IκBαEYFP the acceptor (see Figure 5.5). Experimental results for one such experiment are graphed in Figure 5.6.

5.2.2 Comparison of Model with Biological Results

Running the model for the same situation as shown in Figure 5.6 gives the results in Figure 5.7. The model shows close similarity to single-cell results, suggesting the model works well. In particular, nuclear translocation of NF-κB following stimulation and the reduction of cytoplasmic molecular levels are in good agreement with experimental observations over the same time period.

Constants used in the agent-based model give reasonable results over a range of parameter values. Figure 5.8 shows model results for the same situation as in Figure 5.7, but with rate constants changing between half and double the values used previously.

Figure 5.8 demonstrates that—as would be hoped—rate constants have a clear effect on results (i.e. the model does not produce the same outcomes regardless of inputs), but

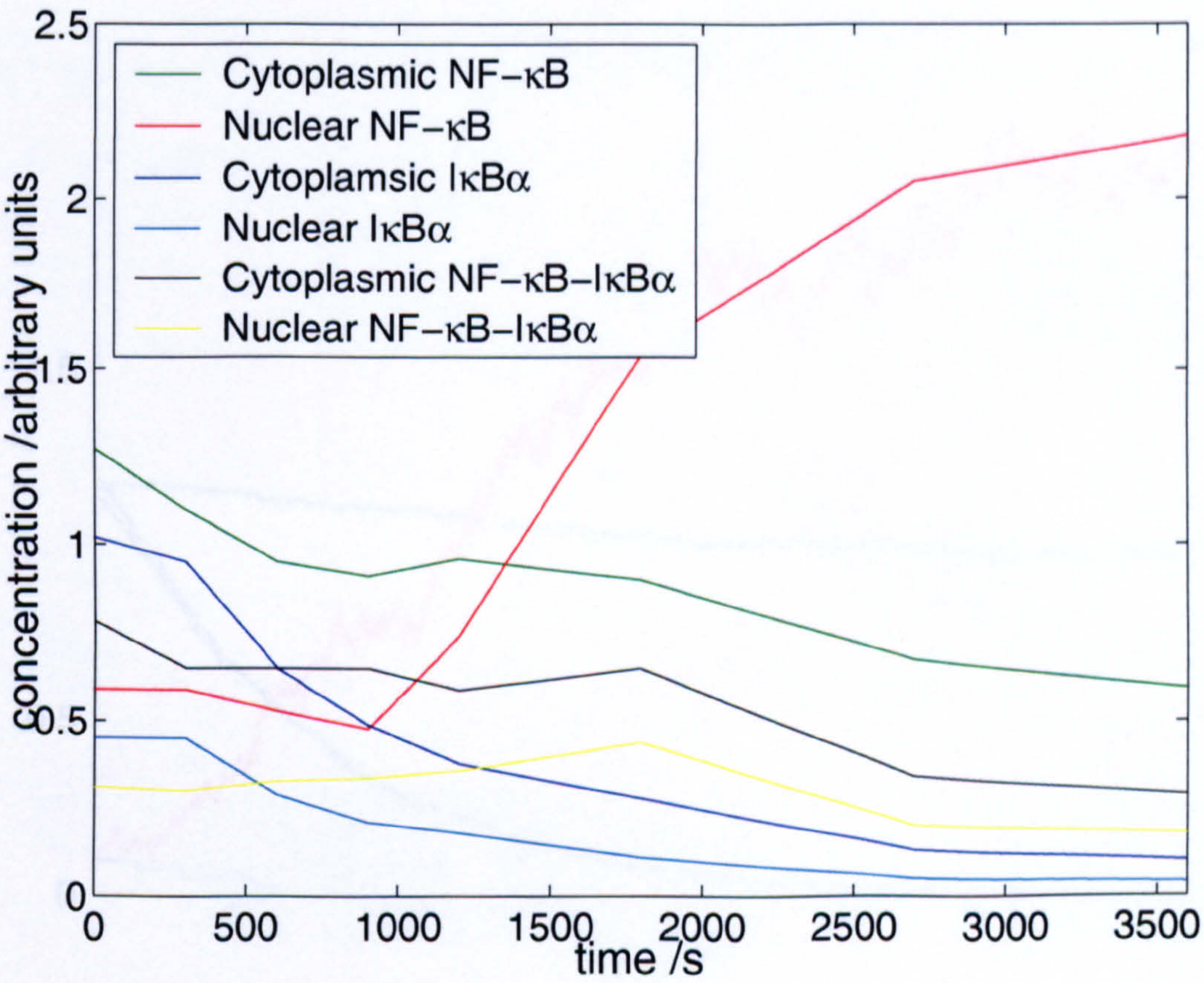

Figure 5.6: **Single-cell experimental results for 1 hour following IL-1 stimulation.** Quantitation of data as shown in Figure 5.5. Data from Yang [Yan03].

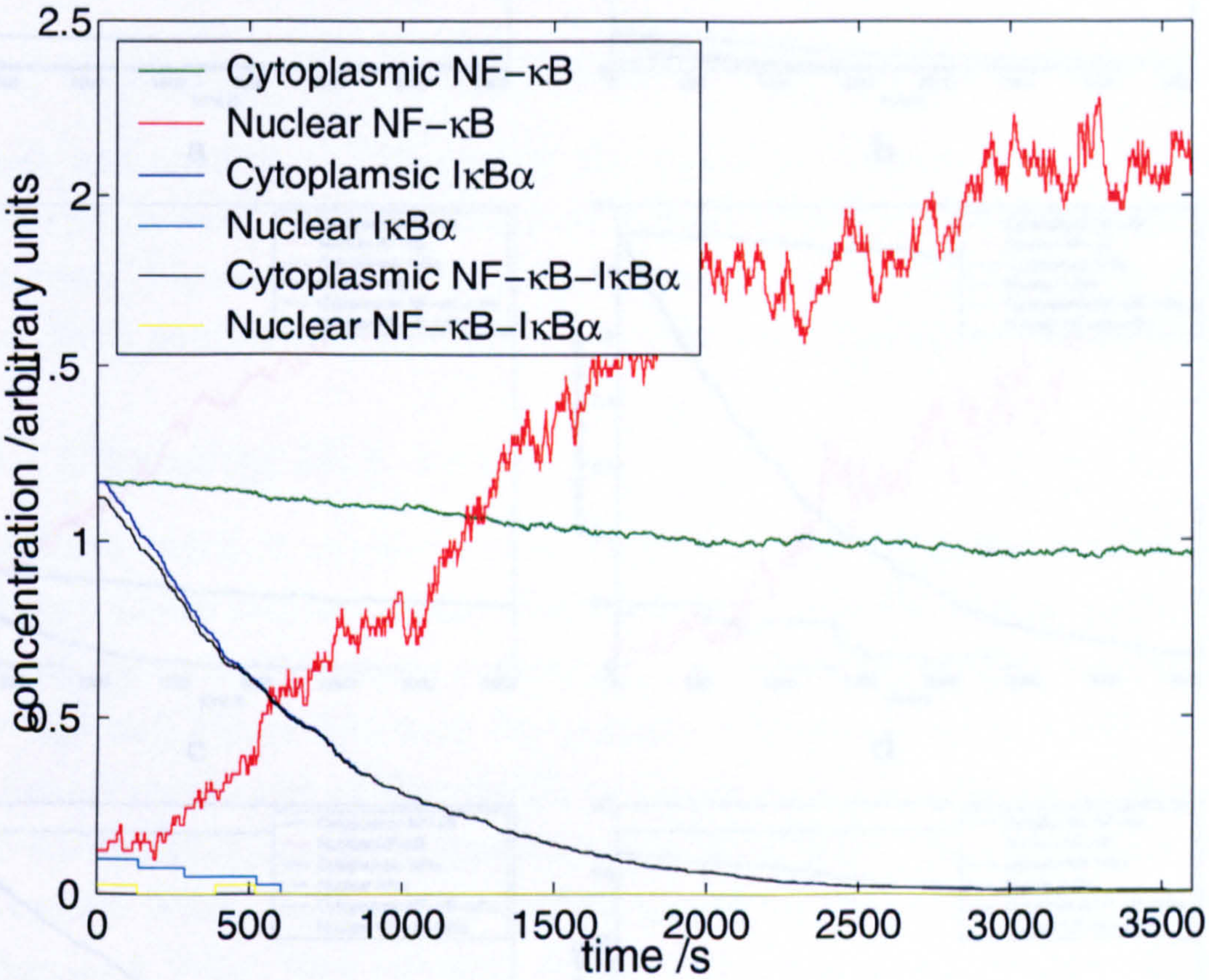

Figure 5.7: **IL-1 stimulation in model.** Model results for the same stimulation as Figure 5.6. Results shown are for a single cell. Standard error 5.4% from 12 simulations for final NF-κB nuclear concentration.

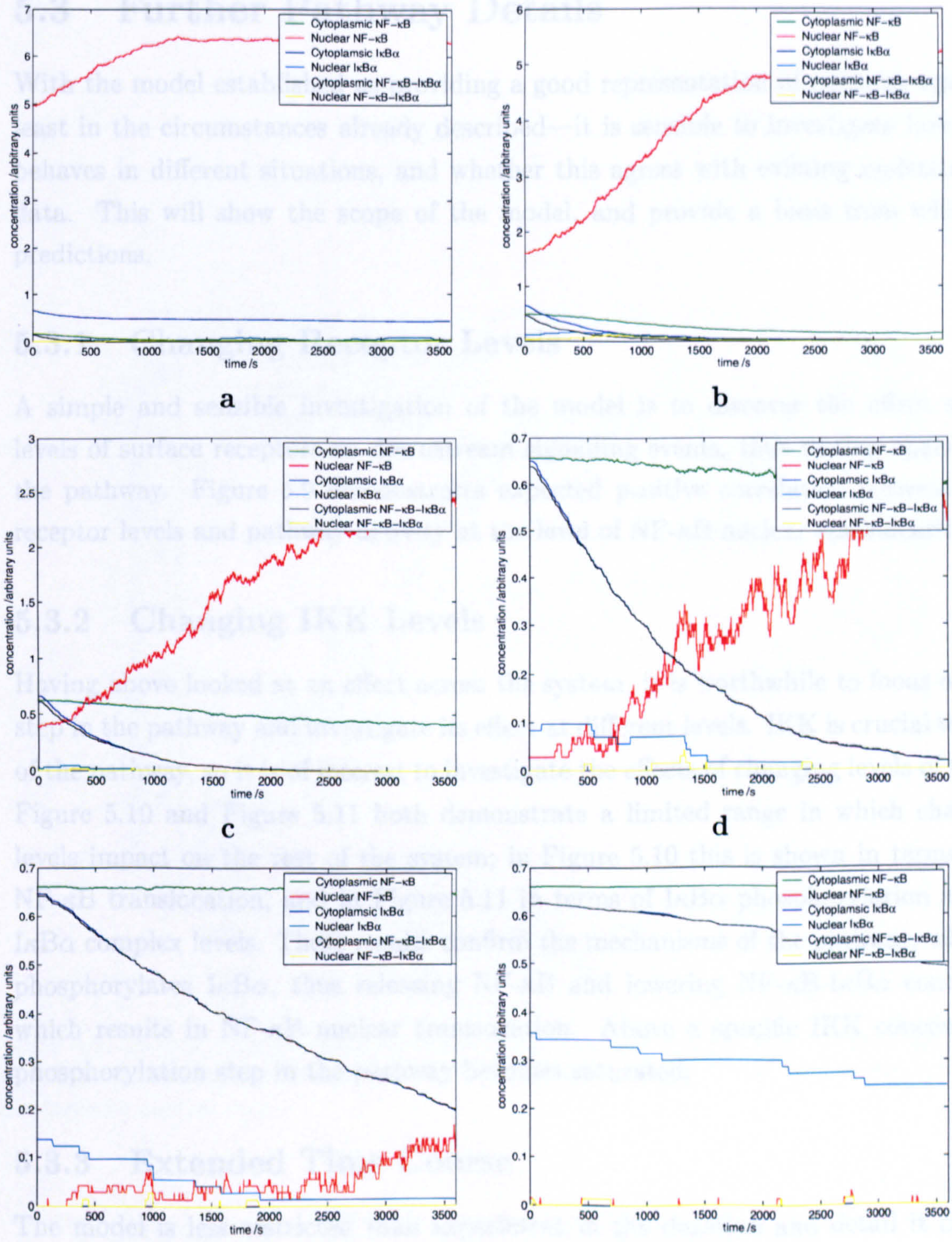

Figure 5.8: **Varying rate constants in the simulation.** All the rate constants for chemical reactions are varied in equal steps between half and double those used for Figure 5.7. Half the rate constants increase from a factor of 0.5 to 2 in panels **a** to **c**, while the rest decrease from 2 to 0.5, to remove any particular bias in results, since the variation is supposed to be quite arbitrary.

equally the model is not over-dependent on data, with similar behaviour being produced within reasonable ranges of rate constants.

5.3 Further Pathway Details

With the model established as providing a good representation of pathway operation—at least in the circumstances already described—it is sensible to investigate how the model behaves in different situations, and whether this agrees with existing understanding and data. This will show the scope of the model, and provide a basis from which to form predictions.

5.3.1 Changing Receptor Levels

A simple and sensible investigation of the model is to discover the effect of changing levels of surface receptors on downstream signalling events, thus testing operation across the pathway. Figure 5.9 demonstrates expected positive correlation between increasing receptor levels and pathway activity at the level of NF-κB nuclear translocation.

5.3.2 Changing IKK Levels

Having above looked at an effect across the system, it is worthwhile to focus on a specific step in the pathway and investigate its effect at different levels. IKK is crucial to activation of the pathway, so it is of interest to investigate the effects of changing levels of the protein. Figure 5.10 and Figure 5.11 both demonstrate a limited range in which changing IKK levels impact on the rest of the system; in Figure 5.10 this is shown in terms of nuclear NF-κB translocation, and in Figure 5.11 in terms of IκBα phosphorylation and NF-κB-IκBα complex levels. These results confirm the mechanisms of the pathway, whereby IKK phosphorylates IκBα, thus releasing NF-κB and lowering NF-κB-IκBα complex levels, which results in NF-κB nuclear translocation. Above a specific IKK concentration the phosphorylation step in the pathway becomes saturated.

5.3.3 Extended Time Course

The model is less restricted than experiment in the duration and detail it can observe. Figure 5.12 shows behaviour of the system over 10 hours following activation; gene tran-

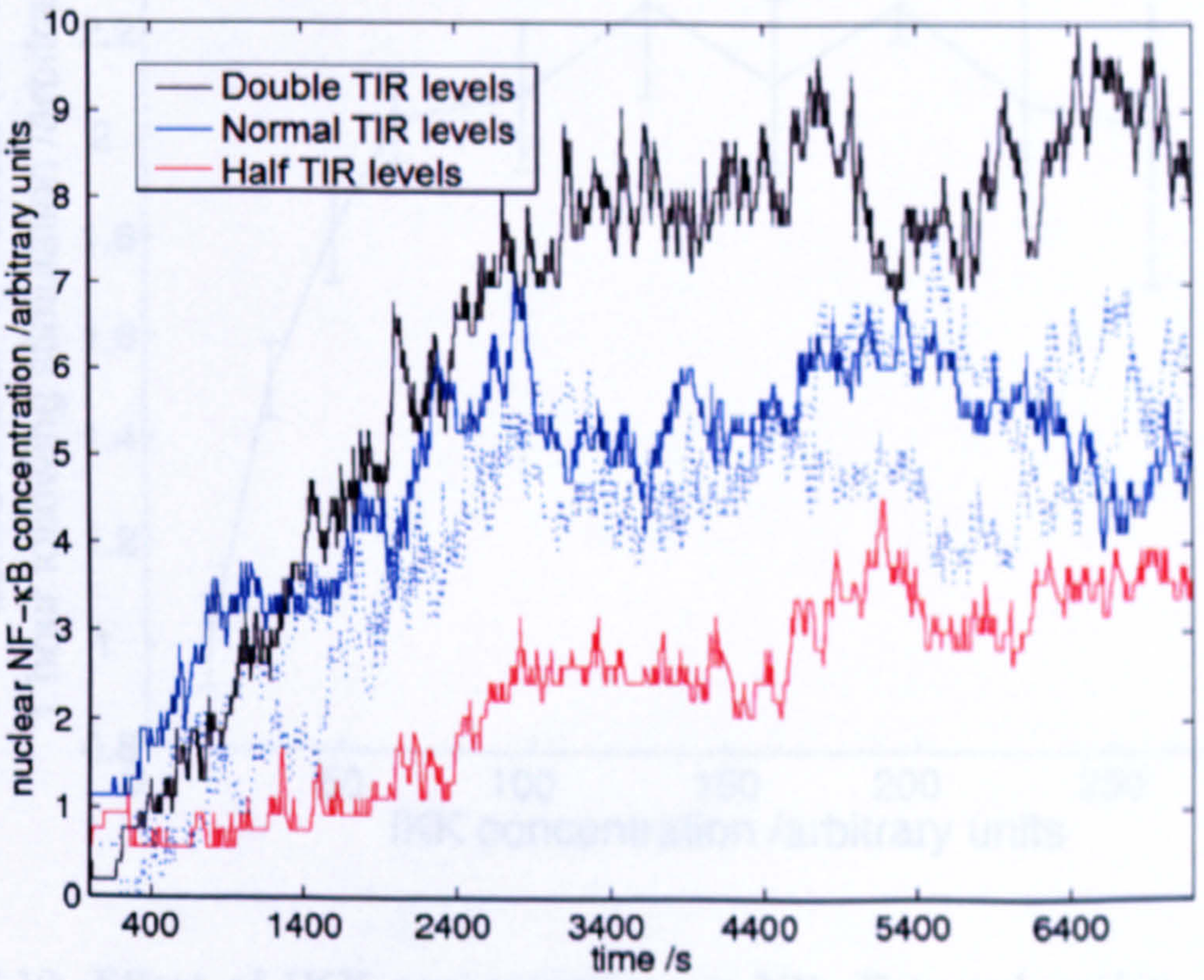

Figure 5.9: **Effect of changing levels of cell surface receptors.** The model is run with three different numbers of receptors (TIRs), and nuclear NF-κB levels plotted for each. Dashed lines indicate control runs.

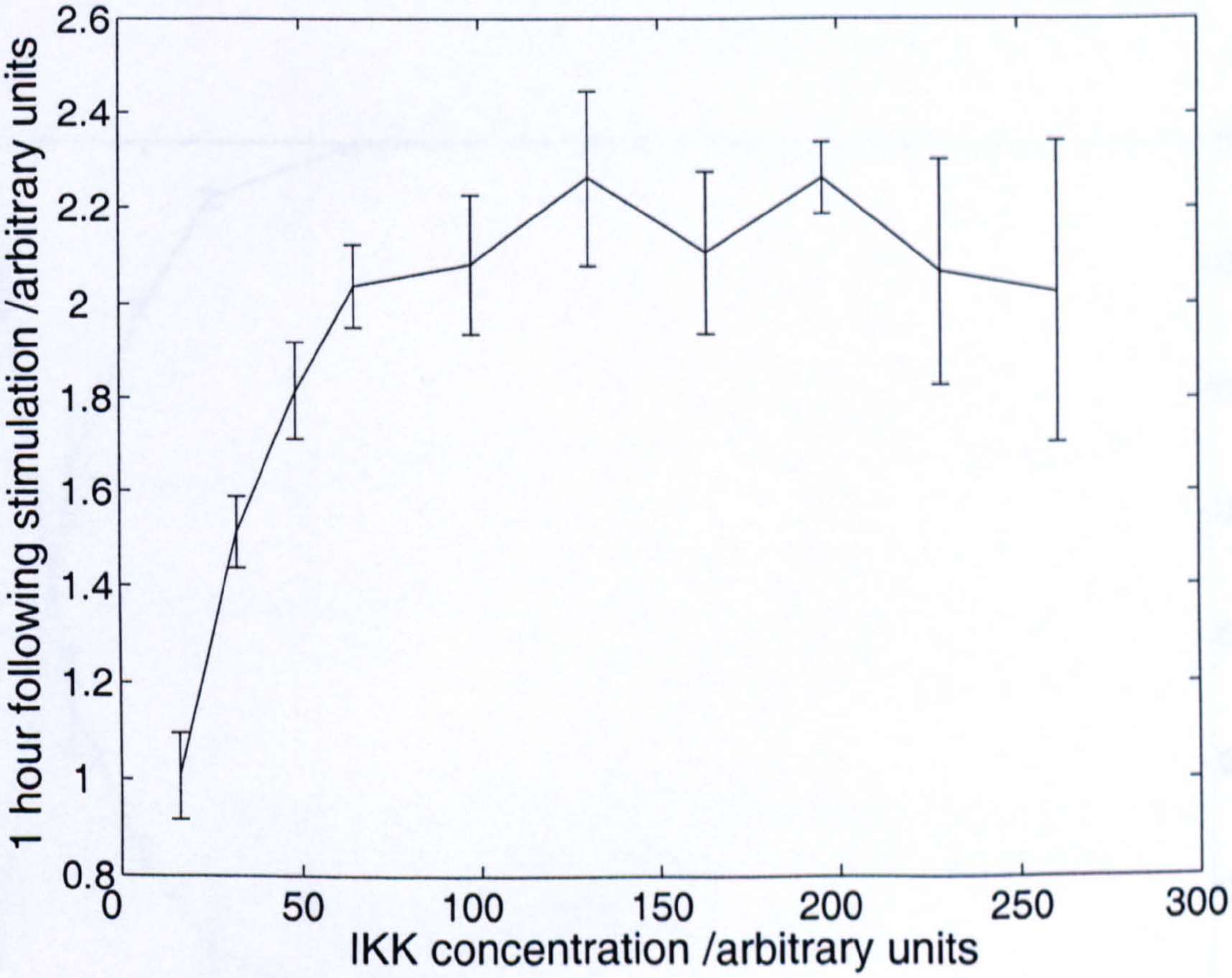

Figure 5.10: **Effect of IKK concentration on NF-κB translocation.** Different IKK levels have a downstream effect on NF-κB translocation over a limited range.

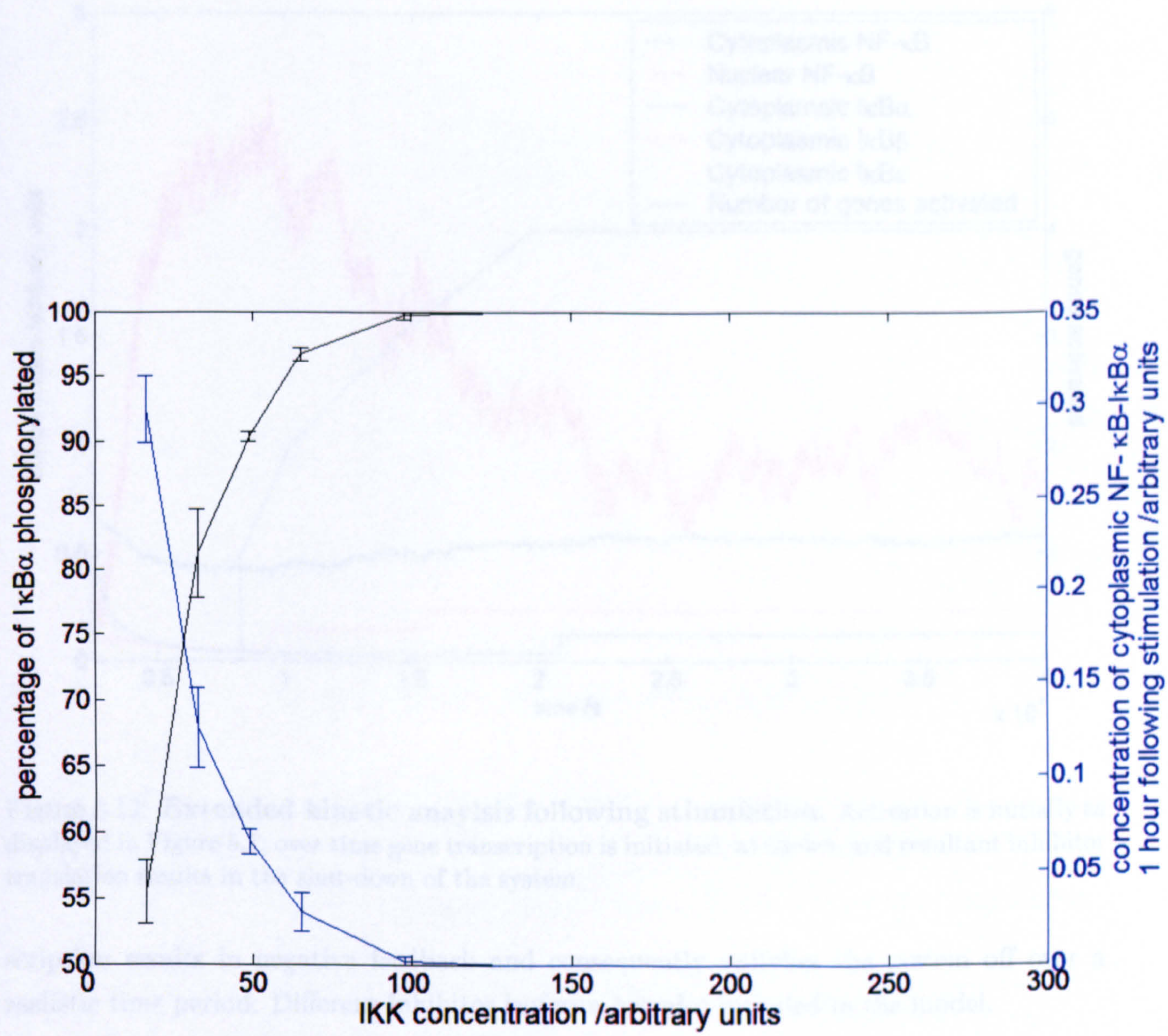

Figure 5.11: **Effect of IKK concentration on IκBα phosphorylation and NF-κB-IκBα complex levels.** Different IKK levels affect IκBα phosphorylation and NF-κB-IκBα complex levels in accord with Figure 5.10.

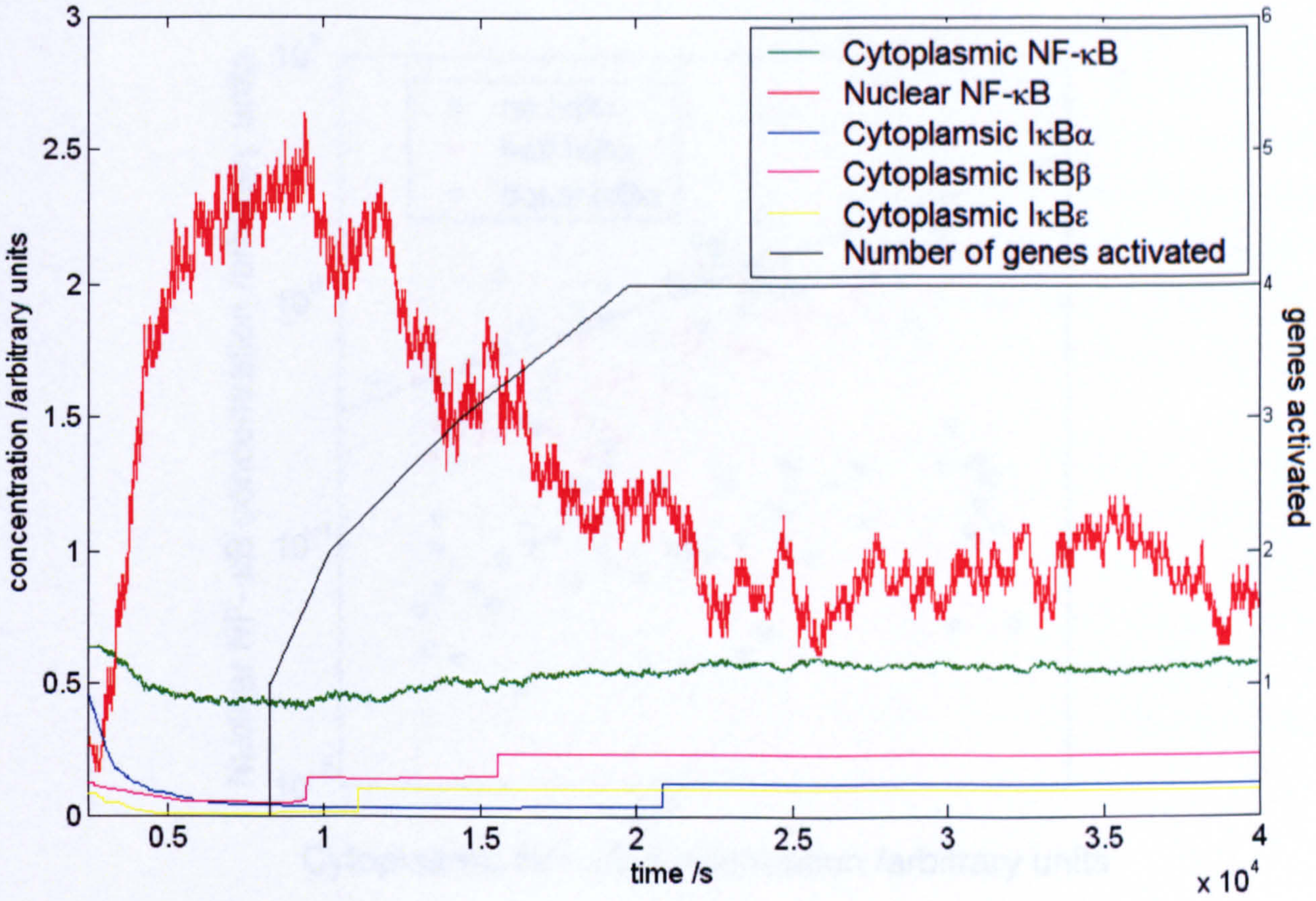

Figure 5.12: **Extended kinetic anaylsis following stimulation.** Activation is initially as displayed in Figure 5.7; over time gene transcription is initiated, as shown, and resultant inhibitor translation results in the shut-down of the system.

scription results in negative feedback and consequently switches the system off over a realistic time period. Different inhibitor isoforms are also included in the model.

5.3.4 Pathway Equilibrium

Not only are the dynamic properties of the system following stimulation of interest, but steady-state pathway behaviour is also important. Nuclear and cytoplasmic concentrations of NF-κB at equilibrium with changing relative levels of inhibitor are shown in Figure 5.13. This demonstrates that complex formation of NF-κB with IκBα causes its cytoplasmic retention, in close agreement with existing biological data (Figure 5.14). The results again verify the suitability of the model, matching well with experimental findings.

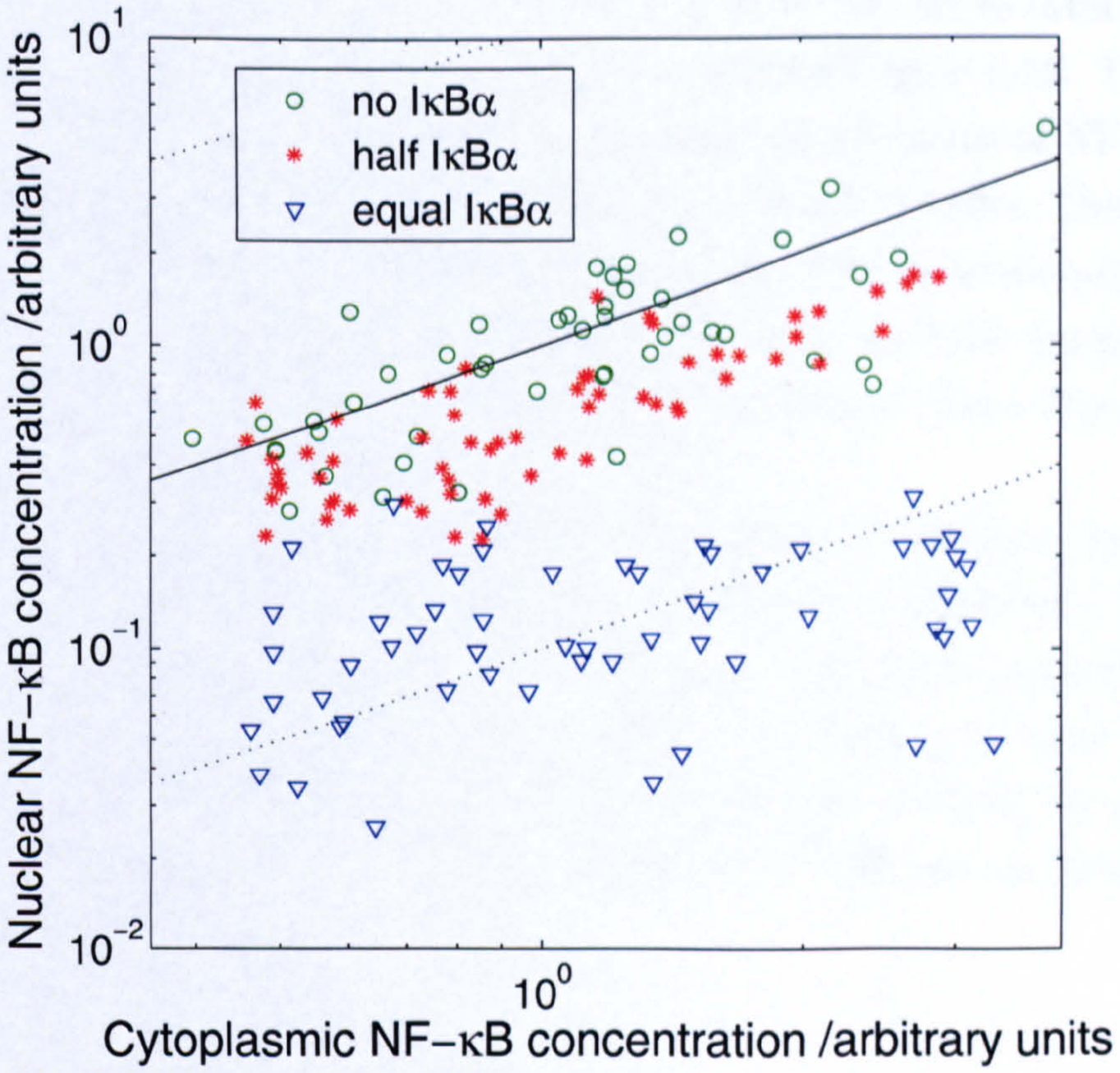

Figure 5.13: **Relative nuclear and cytoplasmic concentrations of NF-κB with varying IκBα levels at equilibrium.** Uninhibited, the cytoplasmic:nuclear concentration of NF-κB is approximately 1:1 (solid diagonal line); increasing inhibitor levels to equal those of NF-κB results in a 10:1 cytoplasmic:nuclear concentration of NF-κB (lower dashed diagonal line).

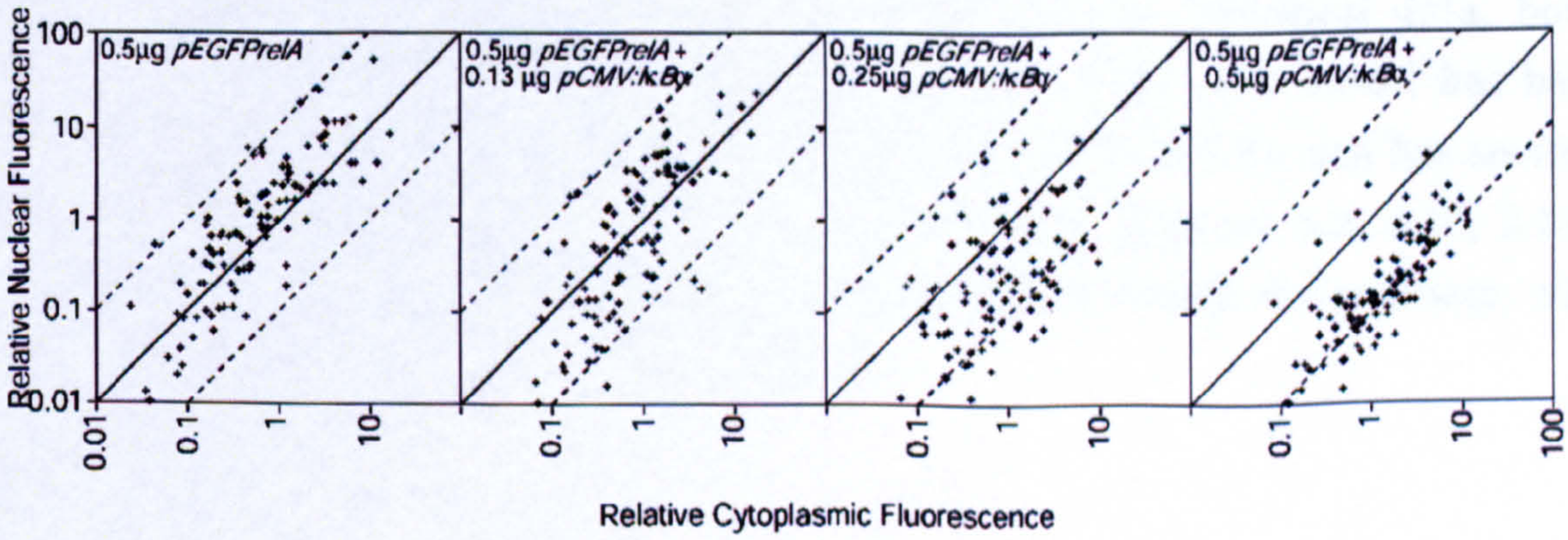

Figure 5.14: **Localisation of NF-κB at equilibrium.** The same quantitative results are displayed as in Figure 5.13. In the left panel there is no IκBα present (green circles in Figure 5.13); this increases through to the right panel where equal amounts of NF-κB and IκBα are present (blue triangles in Figure 5.13). Figure taken from Carlotti et al. [CCDQ99].

5.3.5 Optimum Pathway Operation with Different Inhibitor Ratios

Investigating further the effects of different relative NF-κB to inhibitor ratios, but instead with regard to pathway dynamics, gives the results of Figure 5.15. The 'effectiveness' of the pathway (here defined as the amount of nuclear translocation of NF-κB one hour following stimulation) is plotted against relative levels of NF-κB to IκBα. This demonstrates optimal pathway operation with a 1:1 NF-κB:IκBα ratio. This is intuitively reasonable, since too much inhibitor would simply hinder activation, and too little would mean that more NF-κB would already be in the nucleus prior to activation (from Figure 5.13), meaning less would be available to translocate.

However, this finding is at odds with quantities of inhibitor found in live cell experimentation, where three times the level of inhibitor is observed [CCDQ99]. Reasonably assuming that the pathway has evolved to be as effective as possible, this excess inhibitor must either serve some other more vital role, or somehow be held inactive in the resting state so as not to interfere with pathway activation, at least by cytokines. This could involve the cytoskeleton [CKL+97, QMPD91, QOT+94], and is investigated further in the following chapter.

5.4 Review of Model

5.4.1 Summary

A working model of the IL-1-activated section of the intracellular NF-κB signalling pathway has been formulated, which has been developed from the chemical model of the previous chapter. The model has been verified against exisiting biological data, both following activation (Figure 5.7) and at equilibrium (Figure 5.13). The model has been found to have an appropriate dependence on its parameters (Figure 5.8), and has successfully been used to investigate different features of the pathway (Figures 5.9, 5.10, 5.11, 5.12, and 5.15). Results from the model have raised questions about the pathway, which will be considered further in the following chapter.

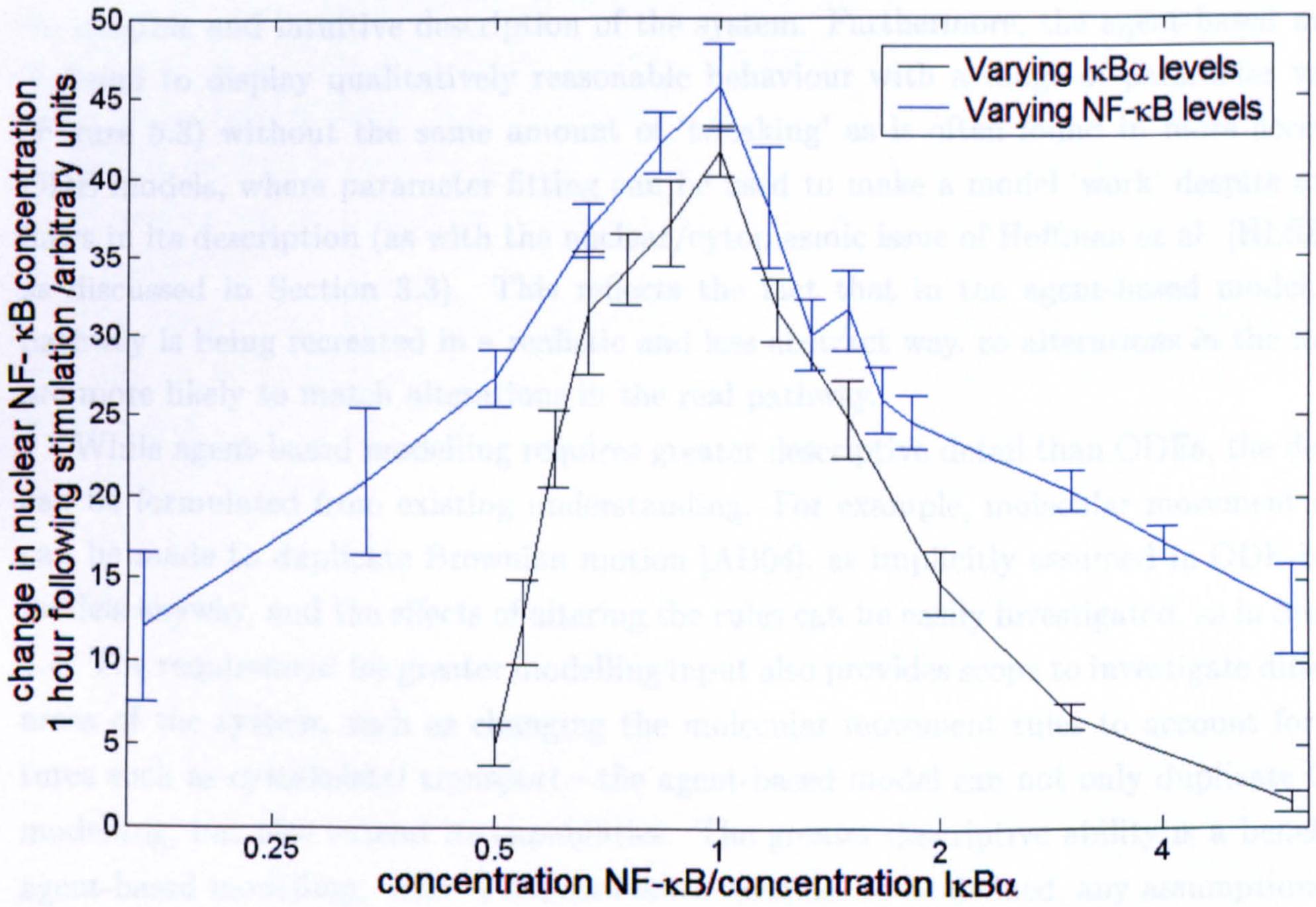

Figure 5.15: **Effect of changing NF-κB:IκBα ratios on pathway response to stimulation.** Maximal response corresponds to a 1:1 NF-κB:IκBα ratio.

5.4.2 Modelling Concerns

Although the model has been successful in its aims so far, issues and limitations in its capabilities must of course be considered.

A major concern in all modelling is that reproduction of experimental data can often be achieved simply by fixing parameters to reflect the desired outcome. Since the NF-κB pathway is a complex multi-component system, this is especially true.

This issue exists for all modelling methods, but it is perhaps less problematic with agent-based modelling. As it requires greater explicit input of detail than many modelling approaches, its operation is more tractable, and potentially incorrect details in the model are obvious, due both to the lack of implicit assumptions in the model and also because of its tangible and intuitive description of the system. Furthermore, the agent-based model is found to display qualitatively reasonable behaviour with a range of parameter values (Figure 5.8) without the same amount of 'tweaking' as is often found in more accepted ODE models, where parameter fitting can be used to make a model 'work' despite major flaws in its description (as with the nuclear/cytoplasmic issue of Hoffman et al. [HLSB02], as discussed in Section 3.3). This reflects the fact that in the agent-based model, the pathway is being recreated in a realistic and less abstract way, so alterations in the model are more likely to match alterations in the real pathway.

While agent-based modelling requires greater descriptive detail than ODEs, the details can be formulated from existing understanding. For example, molecular movement rules can be made to duplicate Brownian motion [AB04], as implicitly assumed in ODE-based models anyway, and the effects of altering the rules can be easily investigated, as in Section 4.4. The requirement for greater modelling input also provides scope to investigate different areas of the system, such as changing the molecular movement rules to account for features such as cytoskeletal transport—the agent-based model can not only duplicate ODE modelling, but also extend its capabilities. The greater descriptive ability is a benefit of agent-based modelling; while it requires more variables to be defined, any assumptions are explicit and can be fully investigated, which ultimately allows more details of the system to be explored. It is worth noting that the agent-based model requires only the same rate constant data as an ODE model, and other details such as movement can simply be made to match the implicit assumptions of ODEs if necessary, so the model is not restricted by lack of data any more than any other kinetic model.

A drawback of agent-based modelling over many other modelling approaches is that it often requires greater computing time. In some circumstances this is not a problem,

but if repeated simulations of large systems are required then measures must be taken to ensure that the model does not become prohibitively time-consuming. The modelling thus far has not been hindered by this, but it is worth considering the efficiency of the computer code and possible parallelisation of the simulation [CSH06], which will permit greater complexity in the model than might otherwise be possible.

The agent-based model has not been used to recreate existing ODE models of the NF-κB pathway because this would be a rather time-consuming process—both in its formulation and simulation—with little benefit. There would be little to be gained scientifically from the exercise, and the worth of the model can be proved in other ways; in its mathematical formulation, comparison with existing biological results, and most importantly as a predictive tool. However, from the results of the previous chapter it would be entirely possible to recreate ODE models. The model has not been compared with existing ODE models because they model different areas of the pathway over different time periods; an appropriate modelling method is used for the models' respective purposes. Even comparisons between ODE models are often hampered due to different models being fitted with different sets of data [LPB+06].

The agent-based model does not aim to replace ODE or other modelling methods, but provides an alternative in instances where explicit spatial resolution is required, and time periods and numbers of molecules are sufficiently small for the simulation to be computed in a reasonable time. For the model to incorporate and be incorporated into other models, it would be useful to write it in a common language such as CellML or the Systems Biology Markup Language (SBML). However, this is not a major aim of the model; these languages are unable to represent individual-based models [Wal06], and modelling should certainly not be restricted by this.

It might be questioned why it is worth using a different modelling method to previous work. It is in fact important to try a different approach for many reasons; simply because other methods have been successful does not mean that they are the only or best way, nor the most appropriate for all parts of the system. Different approaches offer different benefits, and at least offer a different way of thinking about the system in question, requiring different details and raising different questions at the very least. The aim of modelling is to assist with understanding; this can be achieved not only through results from simulations, but also through the formulation of the model itself. An added benefit of using agents in particular is that they are intuitive and easily understood, which is useful for discussing the model with biologists, and also for potential future use of the model by biologists.

5.5 Conclusions and Further Work from Agent-Based Pathway Modelling

Agent-based pathway modelling has been shown to be an appropriate and successful representation of the intracellular NF-κB signalling pathway. Results suggest that further investigation of relative levels of NF-κB and IκB at equilibrium is necessary, specifically why there are excess inhibitors in live cells. This is the direction that modelling will take in the next chapter, which is appropriate both in terms of the strengths of the model, such as its explicit spatial resolution, and in terms of specific questions about the pathway that can be explored using such a model. Indeed, the focus of modelling should be directed by what the model can achieve, what needs explaining biologically, and what can be realistically verified experimentally. For example, further study of transcription would not be sensible within the scope of the pathway model. Instead, investigation of spatial phenomena, such as binding of IκB to the cytoskeleton, is both biologically relevant and can be easily undertaken with the agent-based model.

Chapter 6

Physical Effects on Pathway

6.1 Features to Investigate

The previous chapter established a successful model of the NF-κB signalling pathway that incorporates spatial detail. Having verified its basic operation, the model can now be used with greater confidence to investigate the pathway further, in particular focussing on spatial details of its operation. Following on from the previous chapter, the model should:

1. From the data of Figure 5.15, investigate whether excess inhibitor in the system is held inactive in the resting cell, and if so, how.

2. Using the spatial resolution of the model, investigate a simple but verifiable physical effect on the system, namely that of the shape of the cell.

The aim of this work is to form biologically testable predictions about the pathway in order to gain further insight into its operation.

6.2 Actin Involvement in Pathway

6.2.1 Background to Investigating Actin Involvement

Results from the previous chapter suggest that in live cells there exist three times the number of inhibitor molecules than are required for optimal operation of the TIR-activated pathway. As mentioned, it seems that the excess inhibitor must either serve some other more vital role, or else must be held inactive. The latter option is both the more likely and the more sensible first line of investigation with the model.

The cytoskeleton is important to the pathway (see Section 2.4), and it is known that actin filaments of the cytoskeleton possess a binding site for IκB [CKL+97]. Since pathway response is different following mechanical stimulation than with cytokines, it seems realistic that the cytoskeleton could be involved in sequestering excess IκB, which could be released mechanically and contribute to alternative pathway behaviour. A sensible investigation of how excess IκB molecules are sequestered is therefore to introduce actin filaments to the model to which IκB can bind.

6.2.2 Introducing Actin Filaments to the Model

Actin filaments are introduced to the model by defining radial lines of random direction that extend from the centre of the nucleus to the cell membrane, to which IκB can bind in the cytoplasm (the 'actin' in the nucleus is disregarded, simply being an artefact of the method), as depicted in Figure 6.1. This is a reasonable approximation to the biology described in Section 2.4.

The model assumes an interphase cell: large-scale cytoskeletal structures are treated as static, hence there is no need to consider individual molecules of the structure (sufficient numbers of binding sites are assumed). Intermediate filaments are ignored as the physics of mechanical stress is not explicitly modelled. Accessory proteins are assumed to be present in large numbers and readily available, so need not be modelled explicitly.

IκB can bind to actin in the same way as it binds to other agents: it must be within a defined distance of any point on the filament to be able to bind, as described in Section 4.1.1—the filament is in effect an agent that is a line rather than a particle. Once bound to actin, the IκB is effectively removed from the system for the purposes of investigating equilibrium; reversible binding is ignored for the purposes of this investigation.

6.2.3 Actin Binding Results

NF-κB and IκBα are allowed to reach equilibrium in the cell, both in the presence and absence of actin. When actin is present, IκBα can not only bind to NF-κB, but also to the actin filaments (see Figure 6.2). The number of actin filaments and the interaction radius for binding are chosen to be reasonable, as recorded in Appendix D.

Figure 6.2 shows that the expected 1:1 NF-κB:IκBα binding ratio is achieved with approximately a 1:3 ratio of NF-κB to total IκBα in the presence of actin. This suggests that actin could indeed sequester excess inhibitor in the resting cell; in the absence of actin,

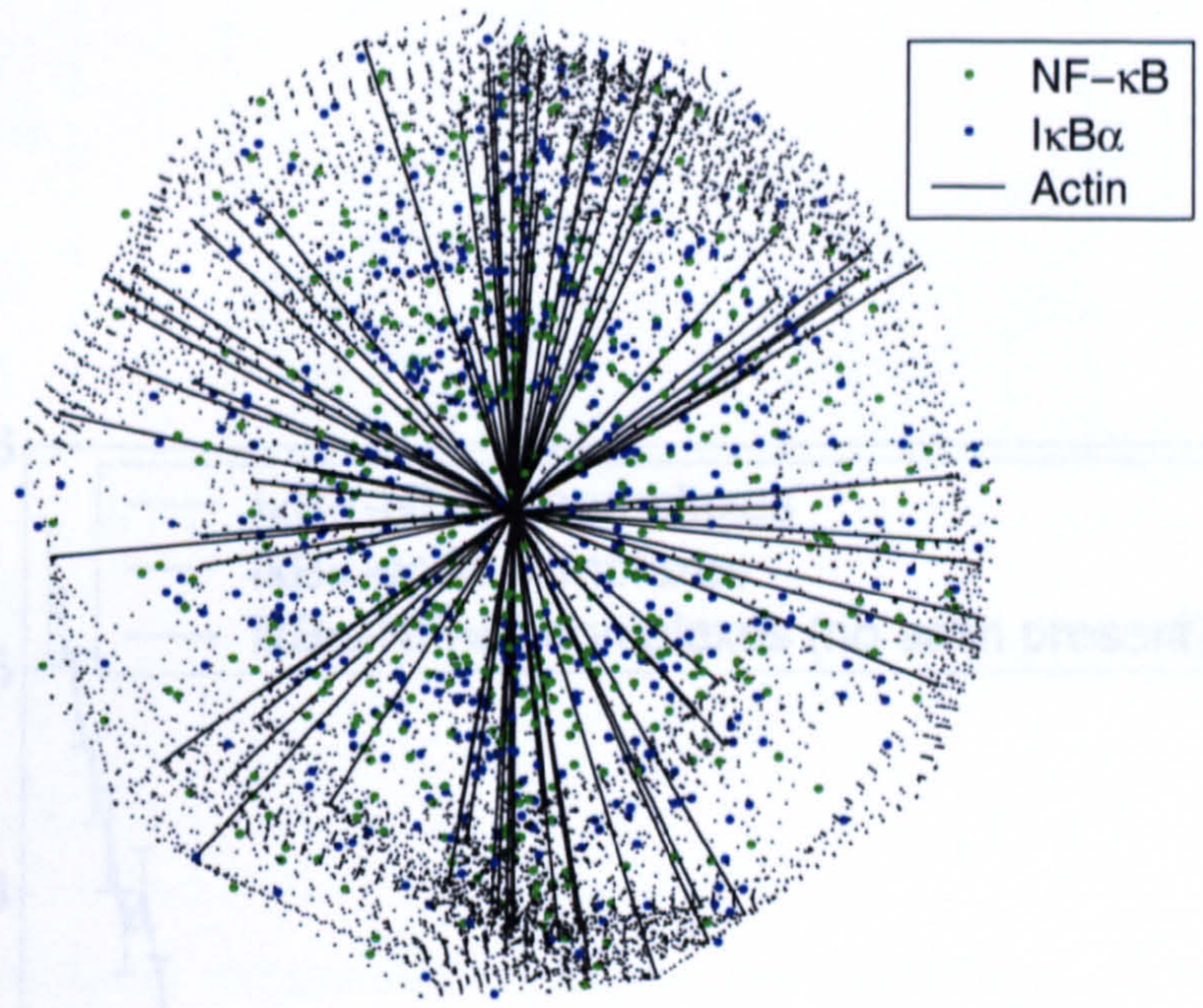

Figure 6.1: **Three-dimensional visualisation of agent positions in the model including actin filaments.** Visualisation of the cell membrane is also shown.

the same maximum is reached at around 1:1 NF-κB to total IκBα.

This result is a clear and testable hypothesis from the model: namely, approximately two-thirds of all IκBα is bound to actin in the resting cell. Although other explanations exist for the earlier observations from the model (as discussed in Section 5.3.5), the combination of existing understanding of the pathway and the model results make this a sensible result to validate biologically.

6.2.4 Biological Validation of Model Results

Experimental work was carried out by Anderson [PHS+06, And06]. Assessment of IκB-actin binding was performed by immunoprecipitation and Western blotting, as described below.

In immunoprecipitation, an 'antibody' is used to bring a corresponding protein out of solution [AJL+02]. The particular antibody is chosen to bind to a specific protein of interest. The antibody is introduced to cell extracts and incubated. A centrifuge is then used to draw out the specific antibody and associated protein, as shown in Figure 6.3. The antibody-bound proteins form a pellet at the bottom of the container.

To assess complex formation, an antibody is chosen to bind to one of the protein

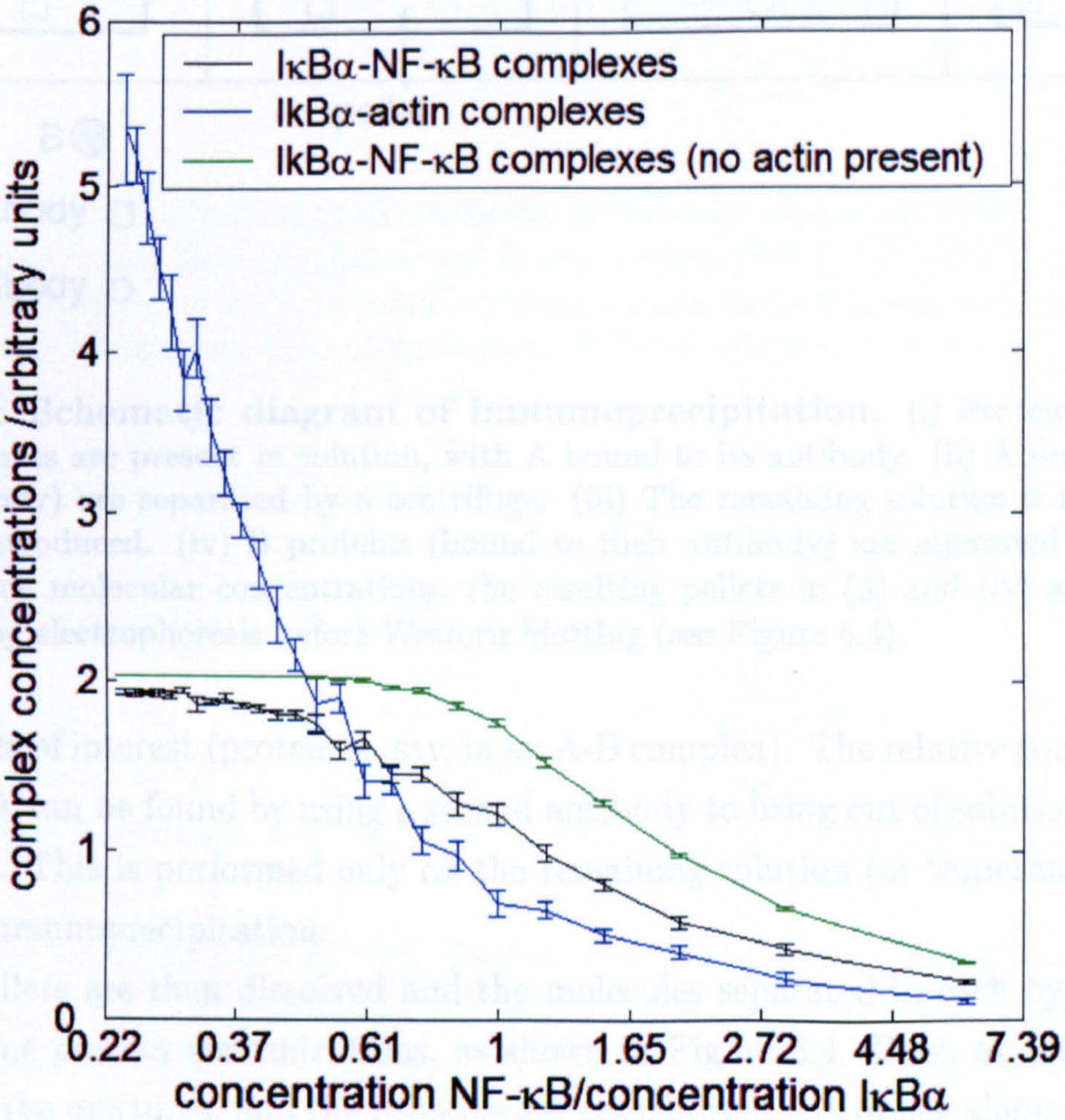

Figure 6.2: **NF-κB-IκBα and actin-IκBα levels at equilibrium with changing NF-κB:IκBα ratios (NF-κB concentration constant).** With actin present, the maximum NF-κB-IκBα complex level is reached at around 1:3 NF-κB:IκBα concentration (black line), with the excess IκBα bound to actin (blue line). The same maximum is reached at around 1:1 concentration in the absence of actin (green line).

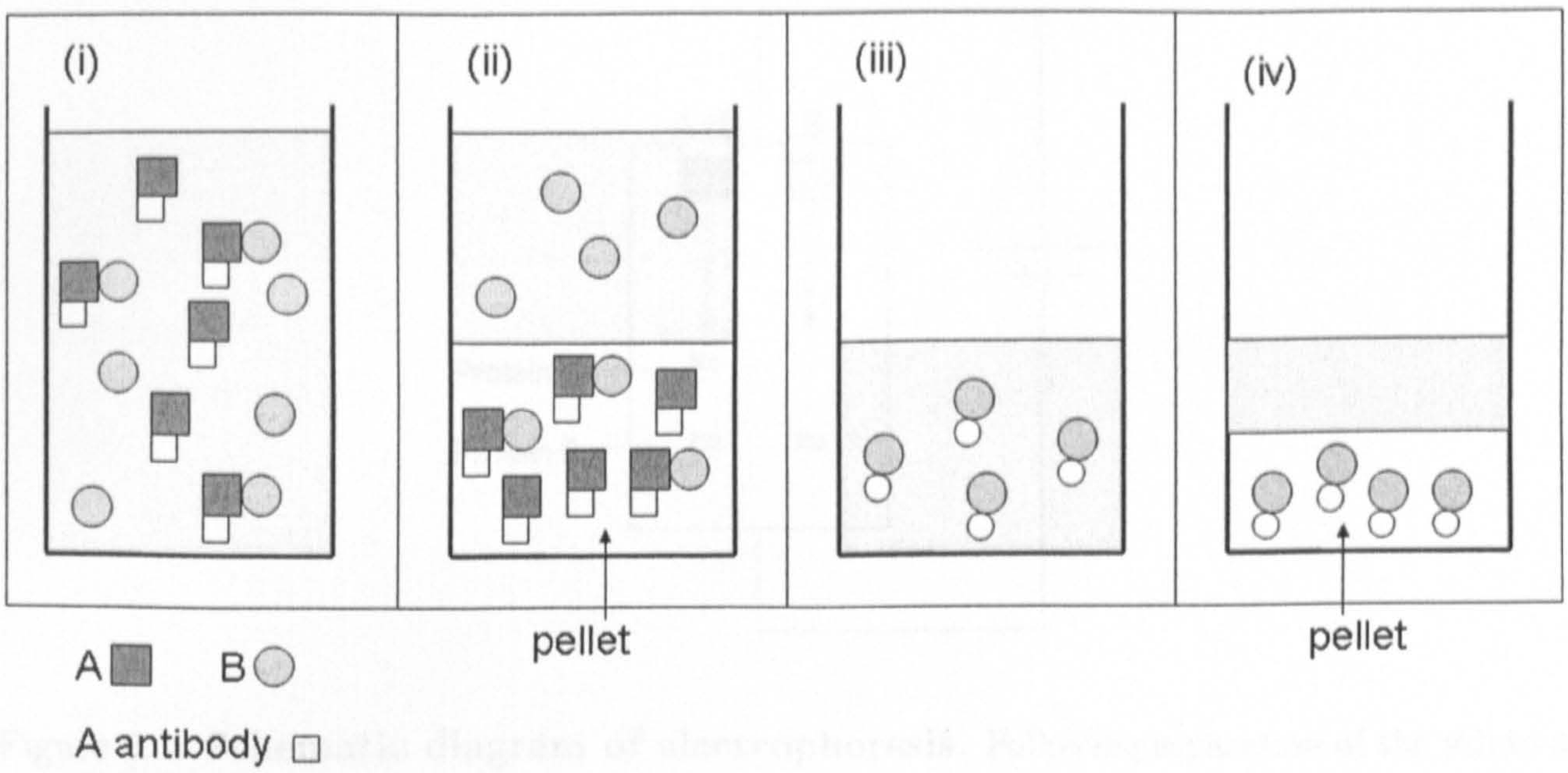

Figure 6.3: **Schematic diagram of immunoprecipitation.** (i) Proteins A and B, and A-B complexes are present in solution, with A bound to its antibody. (ii) A proteins (bound to their antibody) are separated by a centrifuge. (iii) The remaining solution is removed, with B antibody introduced. (iv) B proteins (bound to their antibody) are separated by a centrifuge. To determine molecular concentrations, the resulting pellets in (ii) and (iv) are dissolved and separated by electrophoresis before Western blotting (see Figure 6.4).

components of interest (protein A, say, in an A-B complex). The relative complex formation of B with A can be found by using a second antibody to bring out of solution the remaining B proteins. This is performed only on the remaining solution (or 'supernatant') following the first immunoprecipitation.

The pellets are then dissolved and the molecules separated in each by electrophoresis to determine protein concentrations, as shown in Figure 6.4. Here, an electric current is applied to the mixtures, and the proteins are transported a distance along a gel dependent on their mass and charge. The relative levels of complexed and uncomplexed proteins can be determined by the relative stain intensities, as observed by a process called Western blotting (results shown below).

Results

The above procedure was performed with actin as 'protein A' and IκBα as 'protein B'. The results of a Western blot is shown in Figure 6.5. Only IκBα is stained for. Biological

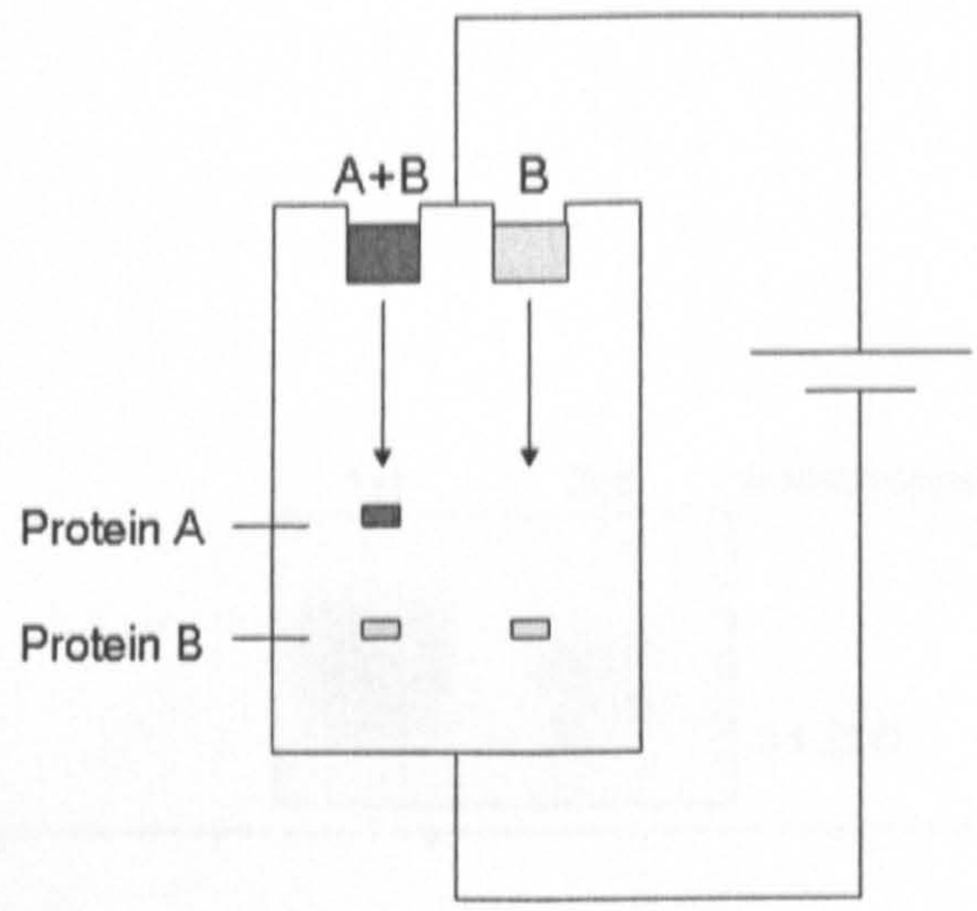

Figure 6.4: **Schematic diagram of electrophoresis.** Following separation of the solution in Figure 6.3, A and A-B complexes are in one column (labelled A+B), and B proteins in a second column. Relative levels of complexed and uncomplexed B proteins can be determined by the relative stain intensities in the two columns. A Western blot is used to observe this (procedure not described).

results agree with the quantitative predictions of the model.

6.2.5 Conclusions of Actin Binding

Based on preliminary modelling results and existing knowledge of the pathway, a mechanism for sequestering the apparent excess inhibitor in the inactive pathway has been successfully hypothesised with the model; subsequent experimental work has supported this.

Although the same investigation of actin binding could be performed using ODEs, with actin binding a first-order reaction, this would be less realistic and more restricted in what could be observed; it is also unlikely that the question of excess inhibitor would have even been raised using ODEs. ODE models tend to initialise the system by allowing NF-κB to cause transcription of IκB until equilibrium is reached; results in this chapter demonstrate that this does not produce physiological levels of inhibitor unless actin is included in the model, either explicitly or implicitly. Although it does not invalidate ODE models—the excess inhibitor is treated as inactive anyway—it is something that would most likely not be predicted without a new modelling approach.

Only IκBα was included in the model since it is the most critical isoform, it is representative of other isoforms, and is experimentally verifiable.

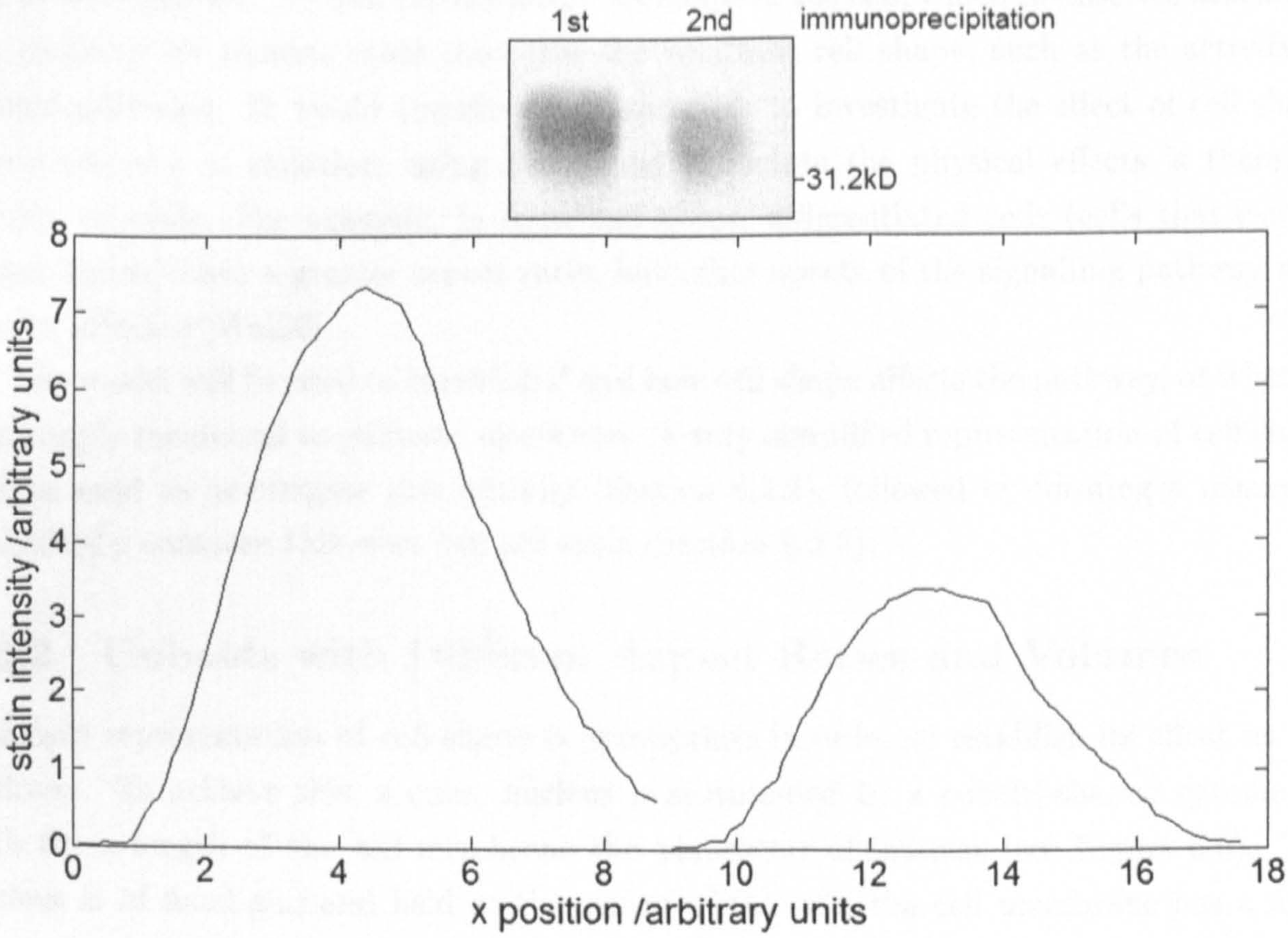

Figure 6.5: **Immunoprecipitation of actin (left column) and secondary IκBα immunoprecipitation (right column).** Cell extracts were subjected to a first immunoprecipitation using an actin antibody (left column), followed by a second immunoprecipitation using an IκBα antibody (right column). Detection is by a Western blot; IκBα is observed in each column, with the left column showing the amount of actin-bound IκBα, and the right column the amount of remaining IκBα. Graphing the stains shows an approximately 2:1 ratio of actin-bound IκBα to free IκBα, corresponding to a 1:1 NF-κB:free IκBα ratio given a 1:3 NF-κB:total IκBα ratio, as predicted by the model. One of three experiments is shown.

6.3 Effect of Cell Shape on Pathway

6.3.1 Aims of Cell Shape Investigation

The effect of the shape of the cell on pathway operation is a relatively simple but significant line of investigation. Several factors affect the shape of the cell, which themselves also affect the pathway for reasons other than just the resultant cell shape, such as the activity of related pathways. It would therefore be impossible to investigate the effect of cell shape experimentally in isolation; using the model to isolate the physical effects is therefore doubly valuable. For example, in some cell types, differentiated cells (cells that can no longer divide) have a greater aspect ratio, but other aspects of the signalling pathway may also be affected [Wal06].

The model will be used to establish if and how cell shape affects the pathway, or whether it is simply incidental to pathway operation. A very simplified representation of cell shape will be used to investigate this initially (Section 6.3.2), followed by forming a means to realistically compare this with live cell data (Section 6.3.3).

6.3.2 Cuboids with Different Aspect Ratios and Volumes

A simple representation of cell shape is appropriate in order to establish its effect on the pathway. To achieve this, a cubic nucleus is surrounded by a cuboid-shaped cytoplasm, with the x-length of the cell membrane the parameter of interest (see Figure 6.6). The nucleus is of fixed size and held at the centre of the cell; the cell membrane has a fixed square cross-section in the y-z plane, and varies from equal length (cubic) to ten times the length in the x-direction. Cytoplasmic concentrations of molecules are held constant with different cell shapes, hence more molecules are present in larger cells.

The pathway is activated by IL-1 as previously, and the time taken for nuclear import of NF-κB to begin following stimulation is observed with different x-lengths of the cell (Figure 6.7). An apparent linear dependence of this delay on the distance of the receptors to the nucleus is observed, which is reasonable since the signal has further to travel. This demonstrates a simple observable effect of the shape of the cell on the pathway. Although a fairly elementary observation, it helps establish the importance to the pathway of cell shape, and is not necessarily immediately obvious since although the signal has further to travel, the concentration of molecules is held constant with larger cells in the simulations, which could have countered the increase in distance to travel.

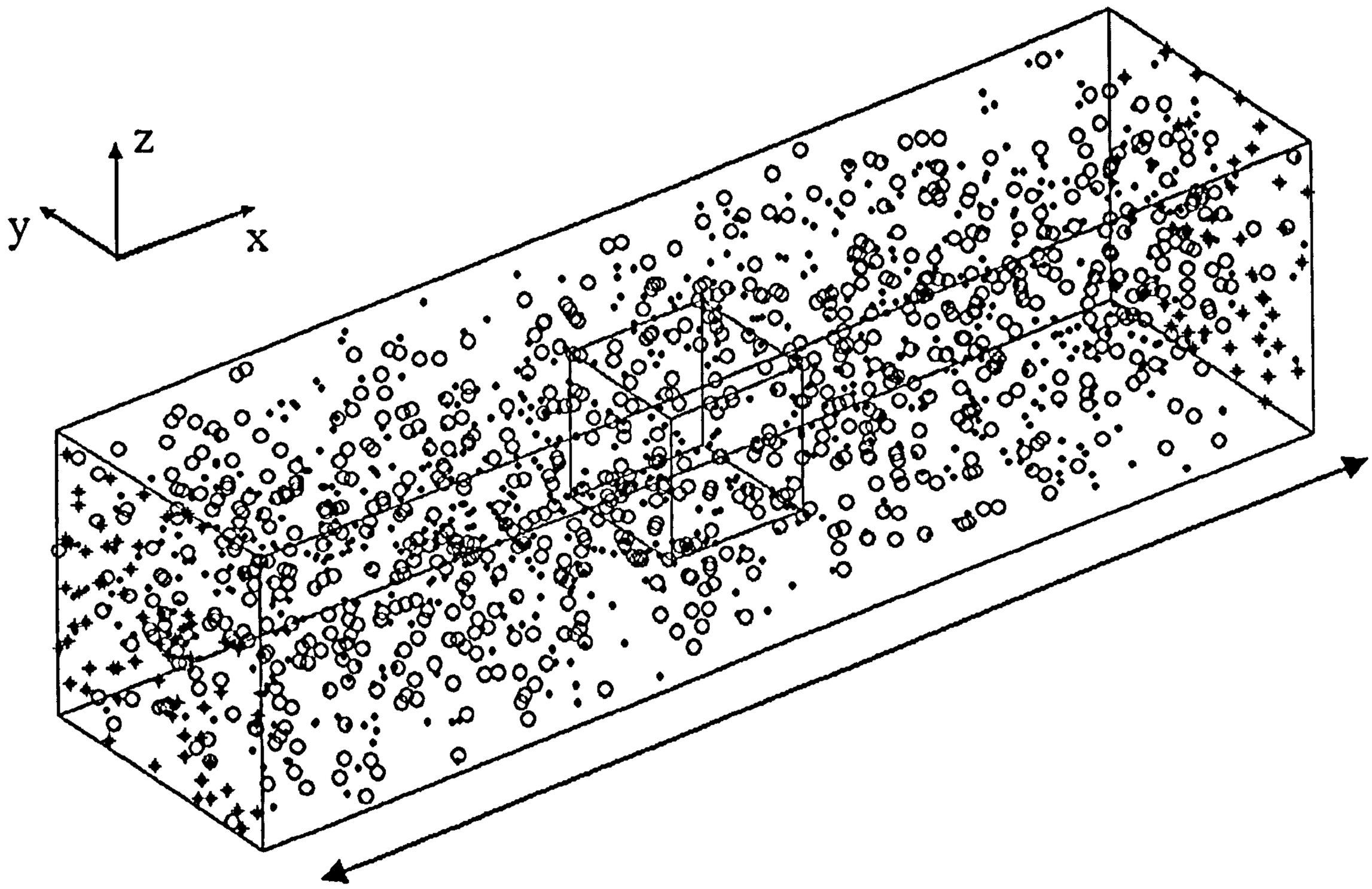

Figure 6.6: **Image of cuboid-shaped cell.** Cuboid-shaped cell with NF-κB and IκB shown. Different aspect ratios of the cell are achieved by keeping two dimensions fixed and varying the third dimension (the x-axis, say, indicated by arrows in the diagram), with the nuclear dimensions fixed and maintained at the centre of the cell, as shown. TIRs (shown as asterisks) are present only at the 2 varying ends of the cell to isolate the effect of changing the aspect ratio.

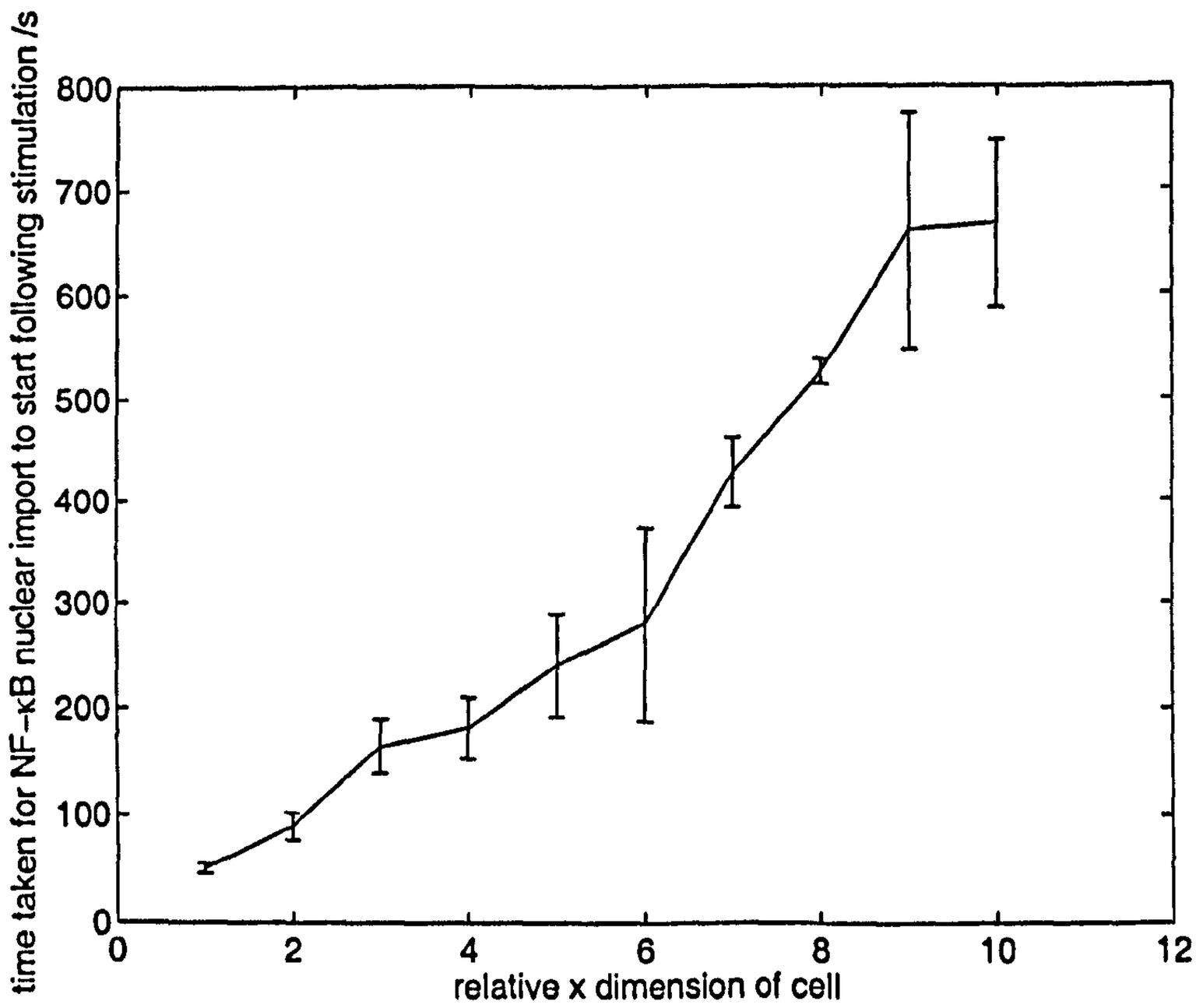

Figure 6.7: **Delay following stimulation for NF-κB nuclear transport to begin with differently lengthed cells.** With TIRs closer to the nucleus, the time delay between stimulation and nuclear transport is small, and increases fairly linearly with increasing distances.

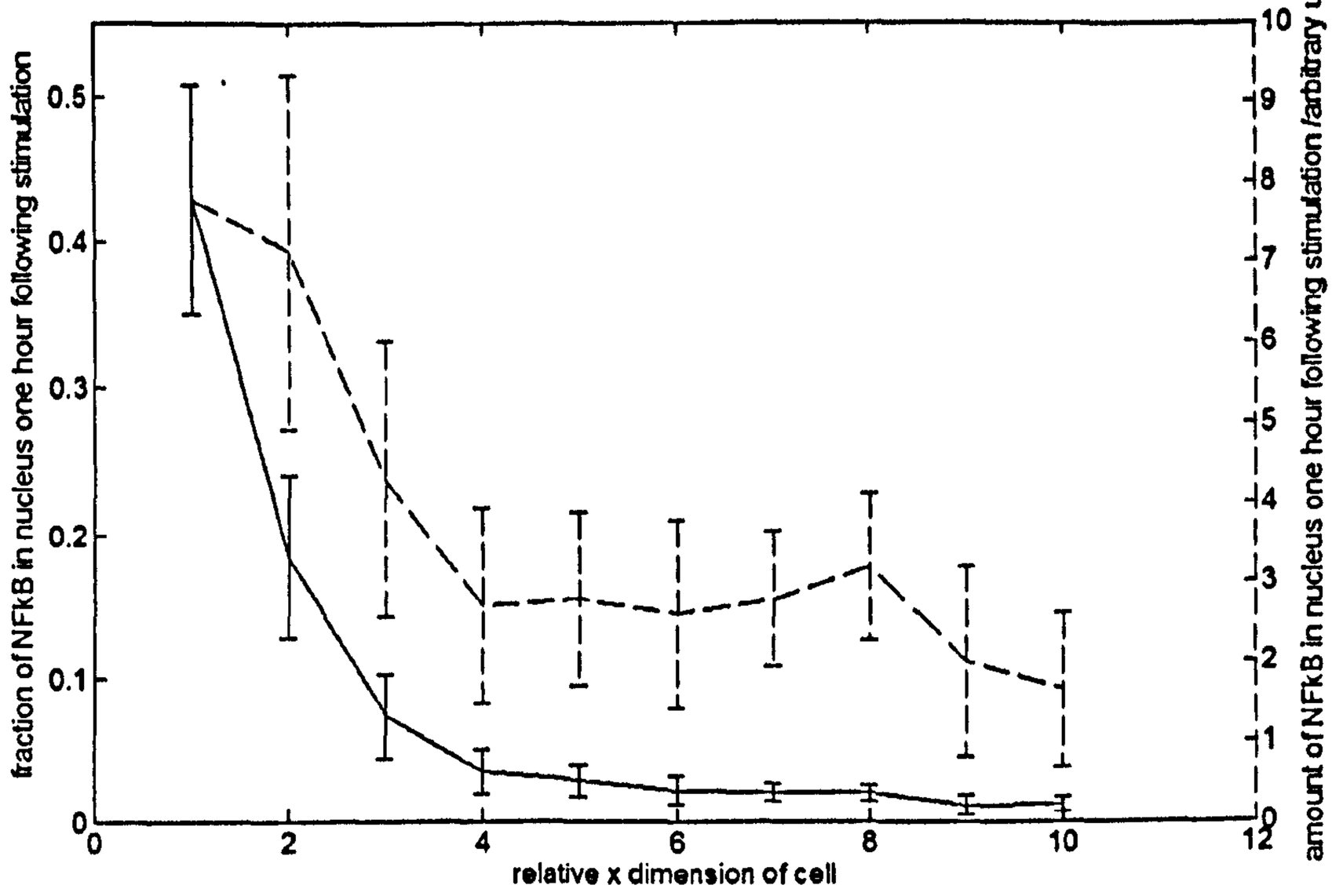

Figure 6.8: **Amount of nuclear translocation of NF-κB one hour following stimulation of cells with varying aspect ratios.** With TIRs closer to the nucleus, nearly half the NF-κB translocates within one hour, decreasing almost exponentially with increasing distances (solid line). The same effect is observed with total NF-κB translocation (dashed line), demonstrating that this is not a consequence of nuclear receptors limiting translocation (see text).

In the same virtual experiment as above, the fraction of NF-κB located in the nucleus one hour following stimulation is observed (Figure 6.8, solid line). This shows a near-exponential decrease in nuclear localisation with increasing x-lengths of the cell. The total amount of NF-κB localised in the nucleus following one hour is also plotted (Figure 6.8, dashed line) to confirm that this is an effect of the change in cell shape rather than an artefact of increased numbers of molecules in larger cells. If this were the case, the total level of NF-κB in the nucleus would remain constant due to nuclear import restricting import; since this is not so, the effect can be reasonably attributed to the shape of the cell. As opposed to the linear relation seen previously, the exponential dependence here is likely due to the accumulation of the increased delay for nuclear translocation over the hour—not just for translocation to begin.

It has been established that cell shape has an observable effect on the pathway in the simulation, not only on the time taken for activation to affect nuclear translocation of NF-κB, but perhaps more significantly also on the level of nuclear translocation of NF-

κB. These results clearly require validation, and although biological experiment cannot replicate the method in the stimulation, some kind of comparison should be possible.

Notable is the dependence of the pathway not only on the shape of the cell but also on the distribution of surface receptors; in the simulation, receptors are present only at the 'far' ends of the cell to emphasise the effect of the shape of the cell, so the modelled effects may be an over-estimation. It should also be noted that the model cuboid cell is a particularly crude representation that should be improved upon, particularly with regards to biological validation, as performed in the following section.

6.3.3 Cell Morphing

To enable closer comparison with biological results, a realistic representation of cell shape in the simulation is desirable. Although it is not possible to eliminate the effects of the various other parameters that affect the pathway, having a close match to the actual shape of a cell in the simulation would permit a much fairer comparison, and would be a suitable direction for future work to take.

A toolkit exists for Matlab that can morph any geometric shape onto any three-dimensional object (the object being a collection of images, which are a 'z-stack' of slices through the real object) [Woo05]. The software was originally designed for exploring the biomechanics of the knee, mapping a known topology onto specific knee scans. The same principle can be used in modelling intracellular pathways, mapping the model cell from a known geometry (which permits straightforward boundary collisions) to a three-dimensional representation of a real cell. Thus real cell shapes can be modelled, at each iteration performing a reverse transformation on the 'real'-shaped cell to check for boundary collisions of molecules (Figure 6.9). This provides a basis for closer comparison of the model with biological experiment in the future.

6.3.4 Conclusions of Cell Shape Effects

Simulation has shown that cell shape can directly affect pathway operation, both in terms of the time taken for nuclear translocation of NF-κB to begin, and in the quantity of NF-κB translocation. Important to these observations is not only the shape of the cell but also the distribution of surface receptors; this adds to the difficulty of biologically validating the results of the model. However, such a validation should be possible to a certain extent; to assist with this the model has been developed to replicate the shapes of

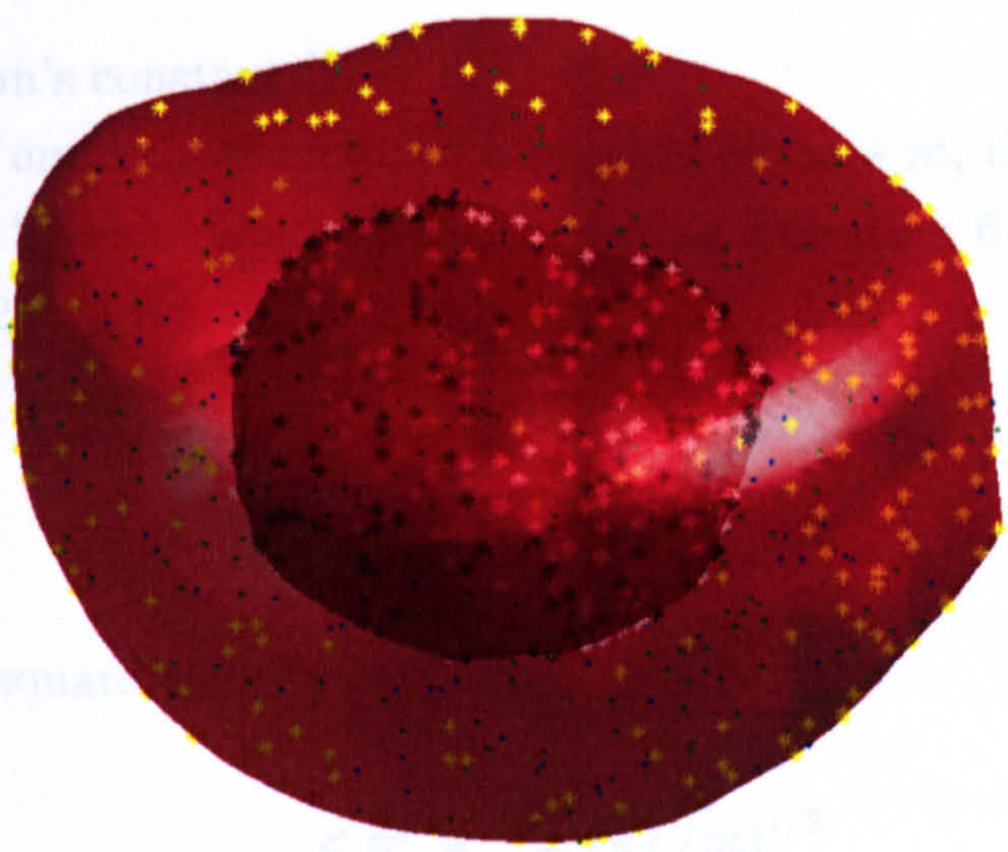

Figure 6.9: **Three-dimensional image of model cell with membranes defined by confocal images of real cell.** Morphing program [Woo05] can define cell shape from a 'z-stack' of confocal images, and permits consideration of boundary collisions of molecules in a simulation. Confocal images used with kind permission from Fang Zhao, Advanced Microscopy Unit of the Haartman Institute, University of Helsinki [Zha05].

real cells, permitting a more direct comparison between simulation and experiment.

6.4 Diffusion Concerns

Investigation of cell shape has produced predictions about the translocation times of molecules, though these do not necessarily agree with diffusion equations. Although previously assumed not to be of vital importance to the model in Section 4.1.2, in light of the above modelling results this should be reviewed.

Molecules move around the cell due to continual random thermal motions known as diffusion [Ber93, BR93, JD83]. The average kinetic energy E of the motion in each axis of a particle at temperature T is given by:

$$E = kT/2 \tag{6.1}$$

where k is Boltzmann's constant [Ber93].

Considering just one axis, then given a particle of mass m, the average velocity $< v >$ along that axis can be calculated by incorporating Equation 6.1 into the kinetic energy equation $E = mv^2/2$ and rearranging to give:

$$< v^2 >= kT/m \tag{6.2}$$

giving a root-mean-square velocity of:

$$< v^2 >^{1/2}= (kT/m)^{1/2} \tag{6.3}$$

Thus, from Equation 6.3, the instantaneous velocity of a particle can be estimated from its mass and temperature. Given that NF-κB is found to have a molecular mass of approximately 50kilodaltons [RBW+90], then at body temperature of 310K, it would travel at around $7ms^{-1}$—several-fold larger than assumed in the model. However, the migration of molecules depends on their random motion, not just their kinetic energy; this aspect of the movement rules must also be considered.

Constant collisions mean that each molecule is forever changing direction. Given the randomness of changes in direction, then if a group of molecules starts at length $x = 0$ at time $t = 0$, their root-mean-square position at time t is given by:

$$< x^2 >^{1/2} = (2Dt)^{1/2} \tag{6.4}$$

where D is known as the coefficient of diffusion, and is a measure of the migration of a particular type of molecule in a particular medium [Ber93].

It is evident from Equation 6.4 that the displacement of molecules is proportional to the square-root of time, in contrast to the linear results of Figure 6.7.

For a small molecule in water at room temperature, D $\approx 10^{-9}m^2s^{-1}$. Thus to traverse the length of a cell of diameter 10^{-5} would take approximately 0.05s—many times faster than in the model. This estimation of the movement of a typical molecule inside a cell has again highlighted problems in the movement rules of the model; both the speed and motion of molecules in the model have been shown to be inaccurate and the consequences of this must be noted.

Although many results from the model do not rely on spatial detail, and the results of Section 4.4 still largely hold, the problems raised here are clearly critical flaws for physical modelling, with molecules in reality traversing the diameter of the cell several times over in the time steps used for simulation.

Given the much larger speed of molecules than previously assumed, spatial resolution of the model could be drastically improved—recall Equation 4.15—though this would require the use of a far smaller time step, in accord with Equation 4.14. Thus, although larger time steps can still be used in the model to give macroscopically correct results, at the time steps used in the simulations which yield reasonable execution times, spatial resolution would be far more limited than assumed until now.

While the limitations of the movement rules in the model do not invalidate all modelling results, they must be re-considered to draw further conclusions on spatial detail.

6.5 Conclusions of Spatial Modelling

The first consideration of spatial detail in the NF-κB pathway has produced a quantitatively validated prediction about the involvement of actin in the inactive cell, as well as qualitative predictions about the effect of cell shape on nuclear translocation of NF-κB. The movement rules of molecules would have to be re-considered to permit further investigation of spatial detail.

Chapter 7

Conclusions and Future Work

7.1 Summary of Achievements

The thesis has achieved the following:

1. The biology of the NF-κB signalling pathway has been outlined, and different modelling approaches have been identified. Based on their various advantages and disadvantages, the most appropriate approach has been selected for the purpose of gaining new insight into the pathway, incorporating spatial and stochastic detail.

2. The model has been successfully verified at a chemical and pathway level.

3. The model has been used to investigate pathway operation, with predictions formulated and tested.

Further consideration of the modelling method follows.

7.2 Review of Modelling

An agent-based modelling approach has been chosen to model the NF-κB pathway due both to its suitability in representing the pathway and its ability to investigate areas of the pathway that are often abstracted away in existing models, such as spatial details.

The model has succeeded in its functional requirements: it is extensible, sufficiently flexible to incorporate abstractions where necessary, provides a clear visualisation of the pathway, is based on mathematical structures that can be investigated in detail, and has scope to be manipulated by biologists. Agent-based modelling has not only provided novel

results due to its different approach to modelling the NF-κB signalling pathway, but also due to the new questions it raises in its formulation.

No model can ever provide a complete description of the system it seeks to recreate (indeed, the aim of modelling is not to do this, but simply to understand better the system as straightforwardly as possible), but it is desirable that details can be added or removed from the model as necessary to model different features of the system. This is very true of the agent-based model, where, for instance, molecular movement rules or binding details can be changed as required without altering the rest of the model, permitting as much or as little detail as is needed.

The agent-based modelling approach not only has its own benefits, but also overcomes many of the problems of widely-accepted reaction kinetics models. While many of these issues are not inherent to reaction kinetics modelling, but rather its execution, they must still be addressed. Spatial considerations are essential to properly model the NF-κB signalling pathway due to its susceptibility to mechanical stimulation [KOTS04]; this cannot be understood by simply using reaction kinetics. Another spatial issue is that of transport of molecules between the cytoplasm and nucleus, with the differing volumes of each compartment sometimes being neglected with concern to their effects on concentration. Time delay in transcription is a further complication that is often not given proper consideration, but is easily incorporated with agents. Agent-based modelling can provide a more tangible view of what is occurring in the pathway, incorporating spatial considerations and allowing detailed observation of specific mechanisms in the pathway. Although agent-based modelling has its own drawbacks, its advantages make it a useful complement to existing models.

7.3 Review of Results

Results from chemical interaction modelling in Chapter 4 demonstrate a sound basis for pathway modelling. Initial results from pathway modelling in Chapter 5 establish the model as a good representation of the NF-κB signalling pathway. Further investigation raises an issue about excess inhibitor molecules in the pathway. Results in Chapter 6 provide a prediction about this issue, which is experimentally validated. Further modelling work demonstrates an observable effect of the shape of the cell on the pathway, though this also reveals limitations of the model as it stands.

7.4 Future Work

Obvious further work to the model would be to continue investigating the effect of cell shape with the real cell images, and comparing this with experimental results. This would require further collaboration with experimental cell biologists, as well as a review of the movement rules. A progression of this in the model would be to begin modelling mechanical stress and mechanical stimulation, although this would be a major undertaking.

Much of the modelling of the NF-κB pathway has used approximations and estimates of chemical rate constants; indeed, this was part of the motivation to use agents rather than an ODE model, since rate constant data is often rather unreliable. The model could be used in the future to investigate this further, for example to check or suggest rate constant data by comparing expected behaviour with model results using different rate constants.

Unlike many other models of the NF-κB pathway, the thesis has not concentrated much on transcription and translation of molecules. This is because it was found that there was much to investigate aside from transcription which was more appropriate and achievable in the model. Future work could investigate transcription and translation in greater detail, making use of the stochasticity and time delays in the model, although work may be hampered by lack of biological data. Interactions of the NF-κB pathway with other pathways and with other cells would also be a constructive line of investigation.

Although agents have been used to model the NF-κB pathway, their potential has not necessarily been fulfilled, though this would be near impossible, and the most has been made of available information in terms of molecular data and biological results against which to verify the model. Agents have prompted original findings about the pathway and a different approach to biochemical modelling. Future work could clearly expand upon this, whether at the small scale with increased spatial and temporal detail, or in terms of increased biological complexity, both in pathway component detail and effects on and from other pathways.

Appendix A

Two-Particle Chemical Interaction Matlab Code

```
% Chemicals A and B inside a cubic box of side length L
% A interacts with B
% Have matrix for each molecule of:

% time of event
% unbound
% bound

clear statematrix_A
clear statematrix_B

% constants
dt = 1;                          % time step
t = dt;                          % current time
t_stop = 150;                    % length of simulation
speed = 5.9e-7                   % average speed of particle (ms^-1)
speed_range = 1e-8;              % maximum possible deviation from the average speed (ms^-1)
angle_range = pi/10;             % maximum possible deviation from current direction (radians)
concentration_A = 50e-9;         % concentration of A in molars
concentration_B = 30e-9;         % concentration of B in molars
L = 3e-6;                        % box side length L metres
total_A = round(6.02e26*concentration_A*L^3);
```

```
    % initial number of A molecules (it is 10^26 not 10^23 to account for use of m^3 not dm^3
total_B = round(6.02e26*concentration_B*L^3); % initial number of B molecules
kab = 1e6;                                    % rate constant of a and b association
threshold_A_B = ((3*kab*dt)/(4e3*pi*6.02e23))^(1/3);    % interation radius

% initial positions
A = zeros(3*total_A, 1);
    % This is all the A molecule matrices (there are 3 rows for each A molecule matrix)
A(3*(0:total_A-1)+2) = 1; % putting free cytoplasmic label to each A molecule
x_A(1:total_A, 1) = L*rand(total_A, 1); % x position
y_A(1:total_A, 1) = L*rand(total_A, 1); % y position
z_A(1:total_A, 1) = L*rand(total_A, 1); % z position

B = zeros(3*total_B, 1);
B(3*(0:total_B-1)+2) = 1;
x_B(1:total_B, 1) = L*rand(total_B, 1);
y_B(1:total_B, 1) = L*rand(total_B, 1);
z_B(1:total_B, 1) = L*rand(total_B, 1);

a = ones(total_A, 1);                 % counter for each A molecule
b = ones(total_B, 1);                 % counter for each B molecule
boundindexA = zeros(total_A, 1);      % index of the B molecule to which each A is bound
boundindexB = zeros(total_B, 1);      % index of the A molecule to which each B is bound

% Simulation

while t <= t_stop


    for n = 1:total_A
        if A(3*(n-1)+2, a(n)) == 1    % A molecule is free in cytoplam
            for p = 1:total_B
                busy_B(p) = B(3*(p-1)+2, b(p)) - 1; % used to see if B is busy
            end
            % Looking to bind (from perspective of each A molecule)
        position_matrix_with_B = [x_A(n) - x_B, y_A(n) - y_B, z_A(n) - z_B];
            % difference in x,y,z coordinates between 'n'th 'A' agent and each 'B' agent
```

```
distance_matrix_with_B =
    [(position_matrix_with_B(1:total_B, 1)).^2 +
    (position_matrix_with_B(1:total_B, 2)).^2
    + (position_matrix_with_B(1:total_B, 3)).^2];
        % squared distance between 'n'th A and each B
        %(squared distance is used to determine interaction)
temp = find(distance_matrix_with_B == 0);
    % preventing division by zero
    % (temp is used as a temporary matrix for various purposes)
distance_matrix_with_B(temp) = 0.000001;    % preventing division by zero
decision_matrix_with_B = rand(total_B, 1)./distance_matrix_with_B;
    % inroducing randomness to the decision of interaction,
    % decreasing with the square of the distance
temp = find(distance_matrix_with_B > threshold_A_B^2);
    % finding the indices of 'B' agents
    % too distant to interact with the 'n'th A agent
decision_matrix_with_B(temp) = 0;
    % marking 'B's too distant to interact.
    % Entry not deleted as this ruins indices.
temp = find(busy_B ~= 0);
    % agent 'B' is unavailable if the 'busy' matrix entry isn't zero
decision_matrix_with_B(temp) = 0;   % marking the 'B' agents that are unavailabl
% choose the best
index_best_B = find(decision_matrix_with_B == max(decision_matrix_with_B));
    % the index of the 'best' 'B' molecule
index_best_B = index_best_B(1);
    % the (1) means that if there are multiple equal candidates
    % (likely if they're all marked as unavailable) then the method still works
best_B = decision_matrix_with_B(index_best_B)/threshold_A_B;
decision = [best_B index_best_B];
    % outputting the 'decision number' and 'B' index
j = decision(2);
if decision(1) > 0
    a(n) = a(n) + 1;
    b(j) = b(j) + 1;
    A(3*(n-1)+1, a(n)) = t;
    B(3*(j-1)+1, b(j)) = t;
```

```
            A(3*(n-1)+2, a(n)) = -1;    % no longer free
            B(3*(j-1)+2, b(j)) = -1;
            A(3*(n-1)+3, a(n)) = 1;     % bound
            B(3*(j-1)+3, b(j)) = 1;
            boundindexA(n) = j;
                % inputting matrix index of molecule with which it is bound
            boundindexB(j) = n;
        end
    end
end

% movement
for n = 1:total_A
    theta_vec_A = -pi + 2*pi*rand(total_A, 1);  % random angle
    phi_vec_A = -pi + 2*pi*rand(total_A, 1);    % random angle
    r_vec_A = speed + (-speed_range+2*speed_range*rand(total_A, 1));
        % random speed within speed range of average speed
    [x_vec_A, y_vec_A, z_vec_A] = sph2cart(theta_vec_A, phi_vec_A, r_vec_A);
        % converting velocity vector to Cartesian
    x_A(n) = x_A(n) + dt*x_vec_A(n);
        % assigning the new position in Cartesian coordinates
    y_A(n) = y_A(n) + dt*y_vec_A(n);
    z_A(n) = z_A(n) + dt*z_vec_A(n);
    if x_A(n) > L   % i.e. molecule is outside box
        x_A(n) = 2*L - x_A(n);
        x_vec_A(n) = -x_vec_A(n);
            % reversing the stored direction to account for the collision
        [theta_vec_A(n), phi_vec_A(n), r_vec_A(n)]
            = cart2sph(x_vec_A(n), y_vec_A(n), z_vec_A(n));
    end
    if x_A(n) < 0   % i.e. molecule is outside box
        x_A(n) = - x_A(n);
        x_vec_A(n) = -x_vec_A(n);
        [theta_vec_A(n), phi_vec_A(n), r_vec_A(n)]
            = cart2sph(x_vec_A(n), y_vec_A(n), z_vec_A(n));
    end
    if y_A(n) > L   % i.e. molecule is outside box
```

```
        y_A(n) = 2*L - y_A(n);
        y_vec_A(n) = -y_vec_A(n);
        [theta_vec_A(n), phi_vec_A(n), r_vec_A(n)]
            = cart2sph(x_vec_A(n), y_vec_A(n), z_vec_A(n));
end
 if y_A(n) < 0   % i.e. molecule is outside box
        y_A(n) = - y_A(n);
        y_vec_A(n) = -y_vec_A(n);
        [theta_vec_A(n), phi_vec_A(n), r_vec_A(n)]
            = cart2sph(x_vec_A(n), y_vec_A(n), z_vec_A(n));
end
if z_A(n) > L   % i.e. molecule is outside box
    z_A(n) = 2*L - z_A(n);
    z_vec_A(n) = -z_vec_A(n);
    [theta_vec_A(n), phi_vec_A(n), r_vec_A(n)]
        = cart2sph(x_vec_A(n), y_vec_A(n), z_vec_A(n));
end
if z_A(n) < 0   % i.e. molecule is outside box
    z_A(n) = - z_A(n);
    z_vec_A(n) = -z_vec_A(n);
    [theta_vec_A(n), phi_vec_A(n), r_vec_A(n)]
        = cart2sph(x_vec_A(n), y_vec_A(n), z_vec_A(n));
    end
end

for n = 1:total_B
    if B(3*(n-1)+3, b(n)) == 1
        x_B(n) = x_A(boundindexB(n));
        y_B(n) = y_A(boundindexB(n));
        z_B(n) = z_A(boundindexB(n));
    else
        theta_vec_B = -pi + 2*pi*rand(total_B, 1);
        phi_vec_B = -pi + 2*pi*rand(total_B, 1);
        r_vec_B = speed + (-speed_range+2*speed_range*rand(total_B, 1));
[x_vec_B, y_vec_B, z_vec_B] = sph2cart(theta_vec_B, phi_vec_B, r_vec_B);
x_B(n) = x_B(n) + dt*x_vec_B(n);
    % assigning the new position in Cartesian coordinates
```

```
y_B(n) = y_B(n) + dt*y_vec_B(n);
z_B(n) = z_B(n) + dt*z_vec_B(n);
if x_B(n) > L   % i.e. molecule is outside box
    x_B(n) = 2*L - x_B(n);
    x_vec_B(n) = -x_vec_B(n);
        % reversing the stored direction to account for the collision
    [theta_vec_B(n), phi_vec_B(n), r_vec_B(n)]
        = cart2sph(x_vec_B(n), y_vec_B(n), z_vec_B(n));
end
if x_B(n) < 0   % i.e. molecule is outside box
    x_B(n) = - x_B(n);
    x_vec_B(n) = -x_vec_B(n);
    [theta_vec_B(n), phi_vec_B(n), r_vec_B(n)]
        = cart2sph(x_vec_B(n), y_vec_B(n), z_vec_B(n));
end
if y_B(n) > L   % i.e. molecule is outside box
    y_B(n) = 2*L - y_B(n);
    y_vec_B(n) = -y_vec_B(n);
    [theta_vec_B(n), phi_vec_B(n), r_vec_B(n)]
        = cart2sph(x_vec_B(n), y_vec_B(n), z_vec_B(n));
end
if y_B(n) < 0   % i.e. molecule is outside box
    y_B(n) = - y_B(n);
    y_vec_B(n) = -y_vec_B(n);
    [theta_vec_B(n), phi_vec_B(n), r_vec_B(n)]
        = cart2sph(x_vec_B(n), y_vec_B(n), z_vec_B(n));
end
if z_B(n) > L   % i.e. molecule is outside box
    z_B(n) = 2*L - z_B(n);
    z_vec_B(n) = -z_vec_B(n);
    [theta_vec_B(n), phi_vec_B(n), r_vec_B(n)]
        = cart2sph(x_vec_B(n), y_vec_B(n), z_vec_B(n));
end
if z_B(n) < 0   % i.e. molecule is outside box
    z_B(n) = - z_B(n);
    z_vec_B(n) = -z_vec_B(n);
    [theta_vec_B(n), phi_vec_B(n), r_vec_B(n)]
```

```
                = cart2sph(x_vec_B(n), y_vec_B(n), z_vec_B(n));
        end
    end
end

% visualisation
for n = 1:total_A
    if A(3*(n-1)+2, a(n)) == 1  % unbound
        plot3(x_A(n), y_A(n), z_A(n),'g.')
        hold on
    end
end
for n = 1:total_B
    if B(3*(n-1)+2, b(n)) == 1  % unbound
        plot3(x_B(n), y_B(n), z_B(n),'b.')
    else    % bound
        plot3(x_B(n), y_B(n), z_B(n),'bo')
    end
end
hold off
axis([0 L 0 L 0 L]) % makes all axes run from 0 to L
daspect([1 1 1])    % makes all axis the same scale
    drawnow

    t = t + dt;

end

% finding molecular levels with time
% A
temp = 0;
    % temp is used to put the 'arriving' matrix at the end of statematrix_A
for n = 1:total_A
    % putting matrices that are on top of one another (in matrix 'A') next to each other
    statematrix_A(1:3, (temp+1):(temp+a(n))) = A((3*n-2):3*n, 1:a(n));
    % this puts the matrices alongside each other without including the zero matrices
temp = length(statematrix_A);
```

```
        % temp is used to add the next matrix at the end of statematix_A
end
statematrix_A = (sortrows(statematrix_A'))'; % putting statematrix_A in chronological order
number_of_A = cumsum(statematrix_A(2,:));   % giving number of free A at each timestep
number_of_AB = cumsum(statematrix_A(3,:)); % giving number of bound A to B at each timestep

% B
temp = 0;
for n = 1:total_B
    statematrix_B(1:3, (temp+1):(temp+b(n))) = B((3*n-2):3*n, 1:b(n));
    temp = length(statematrix_B);
end
statematrix_B = (sortrows(statematrix_B'))';
number_of_B = cumsum(statematrix_B(2,:));
number_of_BA = cumsum(statematrix_B(3,:));
    % giving number of bound B at each timestep (this should equal the number of bound A)

% Graphing molecular levels against time
figure
plot(statematrix_A(1,:), number_of_A/(6.02e26*L^3), 'k:')
    % concentration of free A with time
hold on
plot(statematrix_B(1,:), number_of_B/(6.02e26*L^3), 'k-.')
    % concentration of free B with time
plot(statematrix_A(1,:), number_of_AB/(6.02e26*L^3), 'k-')
    % concentration of bound A to B with time
hold off
title('Concentration of Chemicals against Time')
xlabel('time /s')
ylabel('concentration')
legend('a', 'b', 'c')
end
```

Appendix B

Membrane Movement Matlab Code

```
% constants
total_A = 100;  % number of agents

dt = 0.5;       % time step
t_stop = 200;   % length of simulation

r_nuc = 4000;   % membrane radius
nuc_speed_ave = 20;   % average speed of agents
nuc_speed_range = 5;  % maximum possible deviation from average speed at each time step
nuc_ang_speed_ave = nuc_speed_ave/r_nuc;    % converting to angular average speed
nuc_ang_speed_dev = nuc_speed_range/r_nuc;  % converting to angular speed deviation
nuc_ang_dev = pi/10;
nuc_ang_speed = 500000/r_nuc;
memb_calc = (2.0*r_nuc*sin(1/(2*r_nuc)))^2;
    % square of |displacement| between two points that are 1 unit apart on the surface
alpha = 1.0/r_nuc;  % angle between 2 points on membrane that are 1 unit apart

% initial positions
theta = -pi + 2*pi*rand(1,total_A); % theta coordinates of agents
phi = -pi + 2*pi*rand(1,total_A);   % phi coordinates
r = r_nuc*ones(1,total_A);          % r coordinates
[x,y,z] = sph2cart(theta,phi,r);    % converting to Cartesian

% initial directions
for n = 1:total_A
```

```
length_test = memb_calc + 1;
    % arbitrary number larger than memb_calc for the while loop to work
while(length_test > memb_calc)  % while movement angle is too large
    theta_mov_temp = - 0.5*alpha + alpha*rand(1);
        % a random angle within permitted range
    theta_temp = theta(n) + theta_mov_temp;     % move the agent through this angle
    phi_mov_temp = - 0.5*alpha + alpha*rand(1); % a random angle within permitted range
    phi_temp = phi(n) + phi_mov_temp;           % move the agent through this angle
    [temp_x,temp_y,temp_z] = sph2cart(theta_temp,phi_temp,r(n));
        % converting new positions to Cartesian
        length_test = (temp_x - x(n))^2+(temp_y - y(n))^2+(temp_z - z(n))^2;
            % testing if the movement angles that have been generated are too large
            % (think of a circle inside a square of equal diameter; need movement to
            % be within the circle and not in the corners of the square)
    end
    theta_mov(n) = theta_mov_temp;  % storing acceptable angles
    phi_mov(n) = phi_mov_temp;
end

% simulation
for t = dt:dt:t_stop
    ang_change = 2.*rand(1,total_A)-nuc_ang_dev.*ones(1,total_A);
    theta_temp = theta + theta_mov + ang_change;
    phi_temp = phi + phi_mov - ang_change;
    mov_nucrec
        = nuc_ang_speed_ave - nuc_ang_speed_dev + 2.0*nuc_ang_speed_dev*rand(1,total_A);
    theta = theta + dt*mov_nucrec.*(theta_temp - theta);
    phi = phi + dt*mov_nucrec.*(phi_temp - phi);
    [x,y,z] = sph2cart(theta,phi,r);
    plot3(x,y,z,'.')
    axis equal
    axis off
    drawnow
end
```

Appendix C

Generic description of agent-based algorithm

This 'pseudo-code' supplements the descriptions in the thesis of the workings of the model, particularly movement rules (Sections 4.1.2 and 5.1.3), interaction of agents (Sections 4.1.1 and 5.1.2) and running the model (Section 5.1.4).

```
//define parameters
input time end
input time step

input cell radius
input nuclear radius

input quantities of each agent type
input rate constants for possible interactions (and thus calculate interaction radii)
input agent speeds

set each agent to unbound state

assign random positions of agents according to their confines (in cytoplasm
                                                              or appropriate membrane)
assign random direction vector for movement of each agent

//run simulation
for time = t_0, t_0+time_step, ..., time_end
```

```
//calculate possible interactions in randomised order
[for all interaction pairs, say species A and B, one type (A) is arbitrarily chosen
to 'seek' interaction with the other (B); interaction is not sought twice.  Call the
ones chosen to seek interaction as 'interacting_types']

for all interacting_types [in randomised order]
    for m = agent_1, agent_2, ..., agent_number_of_agent_type [e.g. all 'A' molecules]
        find distances between 'm'th agent and all agents with which it can interact
            [i.e. agents which are able to interact and are in an appropriate state]

        if distances are found that exist within the appropriate interaction radius
            [call this list of distances 'dis']

            dis_squared = square of each element of dis
            rand = vector of random numbers [length equal to length of dis_squared]
            candidates = rand/dis_squared
                [Note: this is physically unrealistic, and is used simply to decide
                which agent should interact if more than one interaction is possible.
                If this program were used in future work, this step should be replaced
                by a random decision.  It is, however, a relatively rare occurence,
                and its effects small.]

                l = find maximum(candidates)    [i.e. agent l is chosen to interact
                        (its agent type and index is tracked throughout calculations)]
                update states such that 'm'th agent is bound to 'l'th agent
            end
    end
end

//perform movement rules
for all agent types
    for m = agent_1, agent_2, ..., agent_number_of_agent_type
        if 'm'th agent is either free or designated to
                                    be the interacting_type in a complex
deviate within pre-defined boundaries from existing movement vector
update movement vector
```

```
                move m along vector
                if agent outside confines (such as cytoplasm)
                    reverse movement vector component along radial line of cell
                    reflect agent position radially over boundary
                end
                if 'm'th agent is bound to another agent (1, say)
                    copy 'm'th agent position as '1'th agent position
                end
            else leave new position to be calculated by agent it is bound to
        end
    end

    end

    //graph data
```

Appendix D

Parameter values used in simulations

Unless otherwise marked in the text, the numbers of agents are scaled down in simulations from the following:

number of NFkB = 60000
number of IkBa = 60000 (subject to discussion in text)
number of surface receptors = 10000
number of importing nuclear receptors = 6000
number of exporting nuclear receptors = 6000
number of actin filaments (where included) = 8000

Interaction radii are calculated from the rate constant values included in the supplementary material of Hoffman et al. [HLSB02], namely:

NFκB-IκBα association = $5 \times 10^5 M^{-1}s^{-1}$
NFκBIκBα-IKK association = $1.85 \times 10^5 M^{-1}s^{-1}$
IκBα-IKK association = $2.25 \times 10^4 M^{-1}s^{-1}$
IKKK-IKK association = $5 \times 10^5 M^{-1}s^{-1}$ (estimated—not treated by Hoffmann et al.)
IκBα-actin association = $5 \times 10^5 M^{-1}s^{-1}$ (estimated—not treated by Hoffmann et al.)
NFκB-importing nuclear receptor association = $9 \times 10^{-2} s^{-1}$
NFκB-exporting nuclear receptor association = $8 \times 10^{-5} s^{-1}$
IκBα-importing nuclear receptor association = $3 \times 10^{-4} s^{-1}$
IκBα-exporting nuclear receptor association = $2 \times 10^{-4} s^{-1}$

(import and export rates depend on the number of receptors in the agent model for all receptor rate constants, since they are expressed in terms of a total rate as opposed to a rate per unit concentration of the components involved).

Bibliography

[AB04] S. S. Andrews and D. Bray. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Physical Biology*, 1:137–151, 2004.

[AJL+02] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell.* Garland Science, 4th edition, 2002.

[And06] D. Anderson. Unpublished data, 2006.

[ARM98] A. Arkin, J. Ross, and H. H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected escherichia coli cells. *Genetics*, 149(4):1633–1648, 1998.

[Bae03] J. C. M. Baeten. Over 30 years of process algebra: Past, present and future. In L. Aceto, Z. Ésik, W. J. Fokkink, and A. Ingólfsdóttir, editors, *Process Algebra: Open Problems and Future Directions*, volume NS-03-3 of *BRICS Notes Series*, pages 7–12. 2003.

[Bar93] J. Barnards. X-machine specification, verification and testing techniques. Master's thesis, Unversity of Sheffield, Department of Computer Science, 1993.

[Bas00] J. B. Bassingthwaighte. Strategies for the physiome project. *Annals of Biomedical Engineering*, 28:1043–1058, 2000.

[BB96] P. A. Baeuerle and D. Baltimore. NF-κB: ten years after. *Cell*, 87:13–20, 1996.

[BBJ+03] K. Burrage, P. Burrage, S. Jeffrey, T. Pickett, R. Sidje, and T. Tian. A grid implementation of chemical kinetic simulation methods in genetic regulation. In *Proceedings of the APAC03 Conference on Advanced Computing, Grid Applications and eResearch*, 2003.

[BCT+04a] C. Baral, K. Chancellor, N. Tran, N. L. Tran, A. Joy, and M. Berens. A knowledge based approach for representing and reasoning about signaling networks. *Bioinformatics*, 20(Supplement 1):i15–i22, 2004.

[BCT+04b] C. Baral, K. Chancellor, N. Tran, N. L. Tran, A. Joy, and M. Berens. A knowledge based approach for representing and reasoning about signaling networks. *Bioinformatics*, 20:i15–i22, 2004.

[Bed99] M. A. Bedau. Can unrealistic computer models illuminate theoretical biology? In *1999 Genetic and Evolutionary Computation Conference Workshop Program*, pages 20–23, 1999.

[Ber93] H. C. Berg. *Random Walks in Biology.* Princeton University Press, 1993.

[Bha03a] U. S. Bhalla. Models of cell signaling pathways. *Current Opinions in Genetics and Development*, 14:375–381, 2003.

[Bha03b] U. S. Bhalla. Understanding complex signaling networks through models and metaphors. *Progress in Biophysics and Molecular Biology*, 81(1):45–65, 2003.

[BHC+99] T. Balanescu, M. Holcombe, A. J. Cowling, H. Gheorgescu, M. Georghe, and C. Vertan. Communicating stream X-machines systems are no more than X-machines. *Journal of Universal Computer Science*, 5:494–507, 1999.

[BR93] D. Brown and P Rothery. *Models in Biology: Mathematics, Statistics and Computing.* Wiley, 1993.

[Bra01] D. Bray. *Cell Movements: From Molecules to Motility.* Garland Publishing, 2nd edition, 2001.

[Bra03] D. Bray. Molecular networks: the top-down view. *Science*, 301:1864–1865, 2003.

[BT03] K. Burrage and T. Tian. Poisson runge-kutta methods for chemical reaction systems. In *Proceedings of the Hong Kong Conference on Scientific Computing*, 2003.

[BWK+05] D. Barken, C. J. Wang, J. Kearns, R. Cheong, A. Hoffmann, and A. Levchenko. Comment on 'oscillations in NF-κB signaling control the dynamics of gene expression'. *Science*, 308:52a–52a, 2005.

[CCDQ99] F. Carlotti, R. Chapman, S. K. Dower, and E. E. Qwarnstrom. Activation of Nuclear Factor κB in single living cells. *Journal of Biological Chemistry*, 274(53):37941–37949, 1999.

[CDPT03] M. Curti, P. Degano, C. Priami, and Baldari C. T. Causal pi-calculus for biochemical modelling. *CMSB'03*, 2602:21–33, 2003.

[CDQ01] F. Carlotti, S. K. Dower, and E. E. Qwarnstrom. Dynamic shuttling of Nuclear Factor κB between the nucleus and cytoplasm as a consequence of inhibitor dissociation. *Journal of Biological Chemistry*, 275(52):41028–41034, 2001.

[CGH] M. Calder, S. Gilmore, and J. Hillston. Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA.

[CGP99] E. Clark, O. Grumberg, and D. Peled. *Model Checking*. The MIT Press, 1999.

[CKL+97] P. Crepieux, H. Kwon, N. Leclerc, W. Spencer, S. Richard, and J. Hiscott. IκBα physically interacts with a cytoskeleton-associated protein through its signal response domain. *Journal of Physical Chemistry*, 17(12):7375–7385, 1997.

[CKTC+01] C. J. Caunt, E. Kiss-Toth, F. Carlotti, R. Chapman, and E. E. Qwarnstrom. Ras controls Tumor necrosis factor Receptor-Associated Factor (TRAF)6-dependent induction of Nuclear Factor-κB. *Journal of Biological Chemistry*, 276(9):6280–6288, 2001.

[CLB98] J. W. Christman, L. H. Lancaster, and T. S. Blackwell. Nuclear factor κB: a pivotal role in the systemic inflammatory response syndrome and new target for therapy. *Intensive Care Medicine*, 24(11):1131–1138, 1998.

[Cra75] J. Crank. *The mathematics of diffusion*. Oxford University Press, 2nd edition, 1975.

[CSH06] S. Coakley, R. Smallwood, and M. Holcombe. Using X-machines as a formal basis for describing agents in agent-based modelling. In *Agent Directed Simulation 2006*, 2006.

[CVGO] M. Calder, V. Vyshemirsky, D. Gilbert, and R. Orton. Analysis of signalling pathways using the PRISM model checker.

[CW03] K. H. Cho and O. Wolkenhauer. Analysis and modelling of signal transduction pathways in systems biology. *Biochemical Society Transactions*, 31:1503–1509, 2003.

[DS92] P. Doucet and P. B. Sloep. *Mathematical Modelling in the Life Sciences.* Ellis Norwood, 1992.

[EI04] N. J. Eungdamrong and R. Iyengar. Modeling cell signaling networks. *Biology of the Cell*, 96:355–362, 2004.

[EL00] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[Ele03] G. Eleftherakis. *Formal verification of X-machine models: towards formal development of computer-based systems.* PhD thesis, Unversity of Sheffield, Department of Computer Science, 2003.

[ELSS02] M. B Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.

[GCC+03] P. P. González, M. Cárdenas, D. Camacho, A. Franyuti, O. Rosas, and J. Lagúnez-Otero. Cellulat: an agent-based intracellular signalling model. *Journal of Cell Science*, 68:171–185, 2003.

[Gil77] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[GP98] P. J. E. Goss and J. Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets. *Proceedings of the National Academy of Sciences USA*, 95:6750–6755, 1998.

[Gut91] H. Gutowitz. *Cellular Automata: Theory and Experiment.* The MIT Press, 1991.

[Han97] J. T. Hancock. *Cell Signalling.* Pearson Education Limited, 1997.

[HB03] P. Hunter and T. K. Borg. Integration from proteins to organs: the physiome project. *Nature Reviews Molecular Cell Biology*, 4:237–243, 2003.

[HKV01] M. Heiner, I. Koch, and K. Voss. Analysis and simulation of steady states in metabolic pathways with petri nets. In *CPN 01 - Third Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools*, pages 15–34, 2001.

[HLSB02] A. Hoffmann, A. Levchenko, M. L. Scott, and D. Baltimore. The IκB-NF-κB signaling module: temporal control and selective gene activation. *Science*, 298(5596):1241–1245, 2002.

[HN63] A. E. Hall and D. Noble. Transient responses of purkinje fibres to non-uniform currents. *Nature*, 199:1294–1295, 1963.

[HP96] C. H. Heldin and M. Purton, editors. *Signal Transduction.* Chapman and Hall, 1996.

[HRN02] P. Hunter, P. Robbins, and D. Noble. The IUPS human physiome project. *European Journal of Physiology*, 445:1–9, 2002.

[HS98] R. Heinrich and S. Schuster. The modelling of metabolic systems: structure, control and optimality. *Biosystems*, 47:61–77, 1998.

[JD83] D. S. Jones and Sleeman B. D. *Differential Equations and Mathematical Biology.* George Allen and Unwin, 1983.

[JHR04a] D. Jackson, M. Holcombe, and F. Ratnieks. Coupled computational simulation and empirical research into the foraging system of Pharaoh's ant (monomorium pharaonis). *Biosystems*, 76:101–112, 2004.

[JHR04b] D. Jackson, M. Holcombe, and F. Ratnieks. Trail geometry gives polarity to ant foraging networks. *Nature*, 432:907–909, 2004.

[KG95] E. B. Kopp and S. Ghosh. NF-κB and rel proteins in innate immunity. *Advances in Immunology*, 58:1–27, 1995.

[Kit02a] H. Kitano. Computational systems biology. *Nature*, 420:206–210, 2002.

[Kit02b] H. Kitano. Systems biology: A brief overview. *Science*, 295:1662–1664, 2002.

[KOTS04] A. Katsumi, A. W. Orr, E. Tzima, and M. A. Schwarz. Integrins in mechanotransduction. *Journal of Biological Chemistry*, 279(13):12001–12004, 2004.

[LAB05] K. Lipkow, S. S. Andrews, and D. Bray. Simulated diffusion of phosphorylated CheY through the cytoplasm of Escherichia coli. *Journal of Bacteriology*, 187:45–53, 2005.

[LBZ+00] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell. *Molecular Cell Biology*. W. H. Freeman and Company, 4th edition, 2000.

[Liu05] E. T. Liu. Integrative biology and systems biology. *Molecular Systems Biology*, 1:E1, 2005.

[LPB+04] T. Lipniacki, P. Paszek, A. R. Brasier, B. Luxon, and M. Kimmel. Mathematical model of NF-κB regulatory module. *Journal of Theoretical Biology*, 228(2):195–295, 2004.

[LPB+06] T. Lipniacki, P. Paszek, A. R. Brasier, B. A. Luxon, and M. Kimmel. Stochastic regulation in early immune response. *Biophysical Journal*, 90:725–742, 2006.

[MA97] H. H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A*, 94(3):814–819, 1997.

[Mil99] R. Milner. *Communicating and Mobile Systems: The π-Calculus*. Cambridge University Press, 1999.

[MK98] P. Mendes and D. B. Kell. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14(10):869–883, 1998.

[MN04] M. N. Moore and D. Noble. Editorial: computational modelling of cell and tissue processes and function. *Journal of Molecular Histology*, 35:655–658, 2004.

[Mon03] N. A. M. Monk. Oscillatory expression of hes1, p53, and NF-κB driven by transcriptional time delays. *Current Biology*, 13(16):1409–1413, 2003.

[Moy05] P. N. Moynagh. The NF-κB pathway. *Journal of Cell Science*, 118:4389–4392, 2005.

[MSSL02] I. I. Moraru, J. C. Schaff, B. M. Slepchenko, and L. M. Loew. The virtual cell: an integrated modeling environment for experimental and computational cell biology. *Ann. N Y Acad. Sci.*, 971:595–596, 2002.

[Nel04] P. Nelson. *Biological Physics.* W. H. Freeman and Company, 2004.

[NHS+05] D. E. Nelson, C. A. Horton, V. See, J. R. Johnson, G. Nelson, D. G. Spiller, D. B. Kell, and M. R. H. White. Response to comment on 'oscillations in NF-κB signaling control the dynamics of gene expression'. *Science*, 308:52b–52b, 2005.

[NIE+04] D. E. Nelson, A. E. C. Ihekwaba, M. Elliott, J. R. Johnson, C. A. Gibney, B. E. Foreman, G. Nelson, V. See, C. A. Horton, D. G. Spiller, S. W. Edwards, H. P. McDowell, J. F. Unitt, E. Sullivan, R. Grimley, N. Benson, D. Broomhead, D. B. Kell, and M. R. H. White. Oscillations in NF-κB signalling control the dynamics of gene expression. *Science*, 306:704–708, 2004.

[Nob02] D. Noble. The rise of computational biology. *Nature Reviews Molecular Cell Biology*, 3:460–463, 2002.

[O'N05] L. A. J. O'Neill. Immunity's early-warning system. *Scientific American*, pages 38–45, January 2005.

[Pal06] I. Palmer. Unpublished data, 2006.

[PHPS05] J. A. Papin, T. Hunter, B. O. Palsson, and S. Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nature Reviews Molecular Cell Biology*, 6(2):99–111, 2005.

[PHS+06] M. Pogson, M. Holcombe, R. Smallwood, D. Anderson, and E. Qwarnstrom. Introducing spatial information into predictive nf-κb modelling—an agent-based approach. in preparation, 2006.

[PS95] M. J. Pilling and P. W. Seakins. *Reaction Kinetics.* Oxford University Press, 1995.

[PS05] S. J. Plimpton and A. Slepoy. Microbial cell modeling via reacting diffusive particles. *Journal of Physics: Conference Series*, 16:305–309, 2005.

[PSQH06] M. Pogson, R. Smallwood, E. Qwarnstrom, and M. Holcombe. Formal agent-based modelling of intracellular chemical interactions. *Biosystems*, 85:37–45, 2006.

[QMPD91] E. E. Qwarnstrom, S. A. MacFarlane, R. C. Page, and S. K. Dower. Interleukin 1β induces rapid phosphorylation and redistribution of talin: A possible mechanism for modulation of fibroblast focal adhesion. *Proceedings of the National Academy of Sciences of the United States of America*, 88(4):1232–1236, 1991.

[QOT+94] E. E. Qwarnstrom, C. O. Ostberg, G. L. Turk, C. A. Richardson, and K. Bomsztyk. Fibronectin attachment activates the nf-κb p50/p65 heterodimer in fibroblasts and smooth muscle cells. *Journal of Biological Chemistry*, 269(9):30765–30768, 1994.

[Qwa03] E. Qwarnstrom. Private communication, 2003.

[RBW+90] D. Ron, A. R. Brasier, K. A. Wright, J. E. Tate, and J. F. Habener. An inducible 50-kilodalton nf kappa b-like protein and a constitutive protein both bind the acute-phase response element of the angiotensinogen gene. *Molecular and Cellular Biology*, 10(3):1023–1032, 1990.

[Ris96] H. Risken. *The Fokker-Planck equation: methods of solutions and applications*. Springer, 1996.

[RJHR05] E. J. H. Robinson, D. E. Jackson, M. Holcombe, and F. L. W. Ratnieks. Insect communication: 'no entry' signal in ant foraging. *Nature*, 438:442, 2005.

[RSS01] A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the pi-calculus process algebra. In *Proceedings of the Pacific Symposium of Biocomputing 2001 (PSB2001)*, pages 459–470, 2001.

[SB86] R. Sen and D. Baltimore. Inducibility of κ immunoglobulin enhancer-binding protein NF-κB by a posttranslational mechanism. *Cell*, 47:921–928, 1986.

[SB01] J. R. Stiles and T. M. Bartol. Monte carlo methods for simulating realistic synaptic microphysiology using mcell. In E. De Schutter, editor, *Compu-*

tational neuroscience: realistic modeling for experimentalists, pages 87–127. 2001.

[Sca06] S. Scarle. Private communication, 2006.

[SFL+93] M. L. Scott, T. Fujita, H. C. Liou, G. P. Nolan, and D. Baltimore. The p65 subunit of NF-κB regulates IκB by two distinct mechanisms. *Genes and Development*, 7(7):1266–1276, 1993.

[SHW04] R. H. Smallwood, W. M. L. Holcombe, and D. C. Walker. Development and validation of computational models of cellular interaction. *Journal of Molecular Histology*, 35:659–665, 2004.

[TF01] P. P. Tak and G. S. Firestein. NF-κB: a key role in inflammatory diseases. *Journal of Clinical Investigation*, 107(1):7–11, 2001.

[TM87] T. Toffoli and N. Margolus. *Cellular Automata Machines: A New Environment for Modelling.* The MIT Press, 1987.

[TRRG02] R. Taussig, R. Ranganathan, E. M. Ross, and A. G. Gilman. Overview of the alliance for cellular signalling. *Nature*, 420:703–706, 2002.

[VAB04] S. J. Vayttaden, S. M. Ajay, and U. S. Bhalla. A spectrum of models of signaling pathways. *Chembiochem*, 5(10):1365–1374, 2004.

[VdVP96] W. Van de Velde and J. W. Perram, editors. *Agents Breaking Away.* Springer, 1996.

[vZRtW05] J. S. van Zon and P. Rein ten Wolde. Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. *Physical Review Letters*, 94(12):1281031–4, 2005.

[Wal06] D. C. Walker. Private communication, 2006.

[WBH05] S. L. Werner, D. Barken, and A. Hoffmann. Stimulus specificity of gene expression programs determined by temporal control of ikk activity. *Science*, 309:1857–1861, 2005.

[Wei02a] J. R. Weimar. Cellular automata approaches to enzymatic reaction networks. In S. Bandini, B. Chopard, and M. Tomassini, editors, *Cellular Automata*

(Fifth International Conference on Cellular Automata for Research and Industry ACRI), pages 294–303. Springer-Verlag, 2002.

[Wei02b] J. R. Weimar. Three-dimensional cellular automata for reaction-diffusion systems. *Fundamenta Informaticae*, 2:275–282, 2002.

[WHW+04] D. C. Walker, G. Hill, S. M. Wood, R. H. Smallwood, and J. Southgate. Agent-based computational modelling of wounded epithelial cell monolayers. *IEEE Transactions Nanobioscience*, 3:153–163, 2004.

[Wil88] J. Williams. *Biochemistry.* Gower Medical Publishing, 1988.

[WMG99] K. K. Wary, A. Mariotti, and F. G. Giancotti. Specificity of integrin signalling. In J. L. Guan, editor, *Signalling Through Cell Adhesion Molecules*, pages 101–113. CRC Press LLC, 1999.

[Woo05] S. Wood. Private communication, 2005.

[WSH+04] D. C. Walker, J. S. Southgate, G. Hill, M. Holcombe, D. R. Hose, S. M. Wood, S. MacNeil, and R. H. Smallwood. The Epitheliome: modelling the social behaviour of cells. *Biosystems*, 76:89–100, 2004.

[Yan03] L. Yang. Unpublished data, 2003.

[YCQ01] L. Yang, H. Chen, and E. Qwarnstrom. Degradation of IκBα is limited by a postphosphorylation/ubiquitination event. *Biochemical and Biophysical Research Communications*, 285(3):603–608, 2001.

[YG01] Y. Yamamoto and R. B. Gaynor. Therapeutic potential of inhibition of the NF-κB pathway in the treatment of inflammation and cancer. *Journal of Clinical Investigation*, 107(2):135–142, 2001.

[YRQ03] L. Yang, K. Ross, and E. E. Qwarnstrom. RelA control of IκBα phosphorylation: a positive feedback loop for high affinity NF-κB complexes. *Journal of Biological Chemistry*, 278(33):30881–30888, 2003.

[YYGK02] T. Yamanda, Y. Yamanishi, S. Goto, and M. Kanehisa. Extraction of modules from the metabolic pathways with phylogenetic profile. *Genome Informatics*, 13:353–354, 2002.

[Zha05] F. Zhao. Private communication, 2005.