Simulation and Statistical Techniques to Explore Lymphoid Tissue Organogenesis

Kieran James Alden

PhD

University of York

Biology

August 2012

Abstract

Secondary lymphoid organs have a key role in the initiation of adaptive immune responses to infection. Organogenesis occurs in foetal development, and the use of genetic tools, imaging technologies, and *ex vivo* culture systems has provided significant insights into the cellular components and associated signalling pathways that are involved. However such approaches tend to be reductionist and descriptive, focusing on the contribution of individual components, and cannot fully explain how lymphoid organs develop through interaction between biological components.

In this study, a set of simulation and statistical tools have been developed that provide further insights into the molecular and biophysical mechanisms of lymphoid tissue organogenesis. Specifically, the formation of Peyer's Patches, gut-associated secondary lymphoid organs, is examined. In collaboration with experimental immunologists, a structured process in the design and calibration of a computer simulation of the biological process has been conducted, leading to the development of a publicly accessible scientific tool where cell behaviour emerges that is statistically similar to that observed in *ex vivo* culture. Robust biological hypotheses can be generated through use of the tool to perform *in silico* experimentation that simulates different physiological conditions. A lack of available statistical tools to analyse *in silico* simulation results has been addressed through the development and release of the *spartan* toolkit, a set of techniques that can suggest the influence that pathways and components have on simulation behaviour, offering valuable biological insight into the system being explored.

An analysis of simulation results using *spartan* suggests the influence of biological pathways on tissue formation changes during development, in contrast to hypotheses in the literature that suggest the process is chemokine driven. Data presented suggests the development period is biphasic, with cell adhesion the key factor early in development, and chemokine expression influential at later point. Through novel application of the statistical tools in *spartan* to perform a time-lapse analysis of cell behaviour, it is suggested this change in phase occurs between hours 24 and 36. Novel *in silico* experimentation performed has suggested the key biological factors in causing cell aggregation, and suggested a role for LTin cells in limiting size and number of Peyer's Patches. A range of potential laboratory investigations have been suggested that could validate whether these simulation derived hypotheses are valid.

Contents

List of Figures

List of Tables

Acknowledgements

There are a large number of people who deserve to be acknowledged in this thesis, and to whom I owe my sincere thanks:

To Mark Coles and Jon Timmis for their excellent supervision and advice; Henrique Veiga-Fernandes for his biological insight and access to experimental data on which this work is based; Paul Andrews and Mark Read for their input into the design of the simulation and statistical analysis of the results; Susan Stepney and Jennifer Southgate for their advice and guidance as the project has progressed.

To Dhanajay Desai and Angela Privat Maldonado - rotation project students I have had the pleasure of working with who have contributed towards the development of this study.

To Roger Leigh, Patty Sachamitr, Bridget Glaysher, Steven Cuss, Amanda Barnes, Amy Sawtell, Rita Pinter, Matt Lakins, Richard Berks, and Daniel Peters - for their patience in helping a computer scientist get to grips with a lot of immunology, and for making the Coles lab a fantastic place to work.

To Heather Rowley, David Sharp, Kay Mitchell, Karin Diaconu, Michael A Kirkpatrick, and Duncan Lean - for their friendship and support over the last few years.

And last but certainly not least, to my parents Stephen and Pauline, for all they have done to make this possible.

Declaration

This thesis has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree other than Doctor of Philosophy of the University of York. This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by explicit references. I hereby give consent for my thesis, if accepted, to be made available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)

Date ....

Publications

Aspects of the work presented in this thesis have previously been presented in the following publications:

K. Alden, M. Read, J. Timmis, P.S. Andrews, H. Veiga-Fernandes, M.C. Coles (2012): **Spartan: A Comprehensive Tool for Understanding Uncertainty in Simulations of Biological Systems.** *PLoS Computational Biology: In Press*

A. Patel, N. Harker, L. Moreira-Santos, M. Ferreira, **K. Alden**, J. Timmis, K. Foster, A. Garefalaki, P. Pachnis, P.S. Andrews, H. Enomoto, J. Milbrandt, V. Pachnis, M.C. Coles, D. Kioussis, H. Veiga-Fernandes (2012): **Differential RET responses orchestrate lymphoid and nervous enteric system development.** *Science Signalling.* **Volume 5, Number 235**.

K. Alden, J. Timmis, P.S. Andrews, H. Veiga-Fernandes, M.C. Coles (2012): **Pairing experimentation and computational modelling to understand the role of tissue inducer cells in the development of lymphoid organs**. *Frontiers in Immunology.* **Volume 3, Number 172**.

K. Alden, P.S. Andrews, J. Timmis, H. Veiga-Fernandes, M Coles (2011): **Towards Argument-Driven Validation of an *in silico* Model of Immune Tissue Organogenesis**. *Proceedings of the 10th International Conference on Artificial Immune Systems (ICARIS).* **LNCS 6011**.

Abbreviations

ABV   Argument-Based Validation

ARTN   Artemin

CoSMoS   Complex Systems Modelling and Simulation Infrastructure

E14.5   Embryonic Day 14.5

EAE   Experimental Autoimmune Encephalomyelitis

eFAST   Extended Fourier Amplitude Sampling Test

GSN   Goal Structuring Notation

ICAM   Intercellular Adhesion Molecule

IL-7Rα   Interleukin 7 receptor α

LTi   Lymphoid Tissue Inducer Cells

LTin   Lymphoid Tissue Initiator Cells

LTo   Lymphoid Tissue Organiser Cells

MAdCAM   Mucosal Addressin Cellular Adhesion Molecule

ODE   Ordinary Differential Equations

PP   Peyer's Patches

PRCC   Partial Rank Correlation Coefficient

Si   eFAST First-Order Sensitivity Index

SCi   eFAST Complementary Parameters Sensitivity Index

STi   eFAST Total-Order Sensitivity Index

spartan   Simulation Parameter Analysis R Toolkit Application

UML   Unified Modelling Language

VCAM   Vascular Cell Adhesion Molecule

Chapter 1

Introduction

1.1 Overview of the Immune System

The immune system is comprised of white blood cells or leukocytes that protect the organism from disease-causing pathogens. Protective immune responses are dependent on innate immune cells, including macrophages and dendritic cells, that present antigens from pathogens to the adaptive immune system, generating protective cellular and humoral responses. Unlike the innate immune system, adaptive immune cells, or lymphocytes, can differentiate into memory cells that provide long term protection from pathogens, ensuring a more rapid and stronger immune response upon future encounters with the same antigen.

T and B lymphocytes develop in the bone marrow, and undergo a process of maturation in the thymus and spleen respectively. Upon maturation lymphocytes enter circulation and migrate to secondary lymphoid tissue including lymph nodes, Peyer's Patches, tonsils and the spleen. These organs are strategically located at drainage points in lymphatic vessels to initiate protective immune responses to antigens from peripheral tissues (Randall *et al.*, 2008). Although each secondary lymphoid organ differs in their architecture, there are some common features, including distinct B and T cell zones (Fu and Chaplin, 1999). The structural organisation of these tissues permits efficient interactions between antigen-presenting cells and lymphocytes, subsequently aiding rapid removal of the pathogen at the site of infection (Goodnow, 1997). This process is initiated in part by antigen-bearing dendritic cells that transport antigens from surrounding tissues into the lymphoid organ, either through lymphatic vessels (in the case of lymph nodes), or through the epithelium (in the case of Peyer's Patches) (Cyster, 1999). Mature lymphocytes continuously circulate through these organs via the blood stream, constantly surveying each organ for the presence of their specific antigen. If that lymphocyte fails to interact with antigen, the lymphocyte can remain within the B or T cell region for up to 4 days before returning to circulation (Fu and Chaplin, 1999; Goodnow, 1997). T cells are activated if their specific antigen is encountered, and specialised T helper cells recruited to B cell areas by chemokines,

bringing together antigen specific B and T cells to provide the signals required for an efficient adaptive immune response. This response has two distinct phases: the clonal expansion of antigen specific B cells to secrete an antibody specific for the pathogen, followed by the development of immune memory that permits an accelerated, higher affinity immune response in the case of re-infection (Goodnow, 1997).

Secondary lymphoid organogenesis occurs in foetal development, but formation can also be caused by chronic infection, cancer and autoimmune disease (Randall *et al.*, 2008). A range of experimental studies has provided insight into the key factors required for lymph node, spleen and Peyers Patch development (Mebius, 2003; van de Pavert and Mebius, 2010; Randall *et al.*, 2008; Veiga-Fernandes *et al.*, 2007). There may be key differences between the mechanisms driving secondary lymphoid organogenesis in different tissues, yet many similarities. Thus developing new insights into the molecular and biophysical mechanisms that contribute to the formation of one set of secondary lymphoid organs, Peyer's Patches, will provide a platform to understand how lymphoid tissues develop.

1.2   Intestinal Immune Responses

The intestinal mucosa is the largest area that is in contact with the exterior environment, and is constantly exposed to bacteria and other pathogens (Reis and Mucida, 2012). The potential for infection caused by bacteria, viruses, parasites, and fungi is thus significant. Although 100 times the area of the skin, which is protected by a physical barrier of several layers, the intestine is comprised of a single layer of absorptive epithelial cells that create a barrier between the interior lamina propria and the external lumen (Reis and Mucida, 2012). This may explain why up to 70% of the body's lymphocytes are found to reside within gut-associated lymphoid tissue (GALT) (Jung *et al.*, 2010).

GALT consists of Peyer's Patches (PP), isolated lymphoid follicles, intraepithelial lymphocytes, lamina propria leukocytes, and mesenteric lymph nodes that together have an essential role in the generation of protective antibody responses to infection within the gastrointestinal tract. PP are secondary lymphoid organs that were initially described by Severino in 1645, but later named following Peyer's investigations in 1677 (Jung *et al.*, 2010). PP are domed structures that consist of 1-5 B cell follicles containing follicular dendritic cells (FDC), a T cell zone and associated fibroblastic reticular cell (FRC) network (Jung *et al.*, 2010). A schematic representation of a PP can be seen in Figure 1.1. Lymphocytes migrate into the PP via high endothelial venules, and continue their circulation via efferent lymphatic vessels that connect to mesenteric lymph nodes (Jung *et al.*, 2010). Unlike lymph nodes, there is no lymphatic input to PP; instead antigen uptake occurs via specialised epithelial cells, Microfold or M-cells, in the follicle-associated epithelium (red in Figure 1.1). This antigen is

transferred to local dendritic cells in the FRC (orange area in Figure 1.1), (Owen and Bhalla, 1974) and these dendritic cells then migrate into the T-cell zone for antigen presentation to lymphocytes. Pre-natal studies show that an average of sixty PP develop in the human foetal gut (Cornes, 1965), and seven to eleven in the mouse (Figure 1.2), distributed along the length of the small intestine, with a large variation in the location, number, and size of PP between different genetically identical mice.
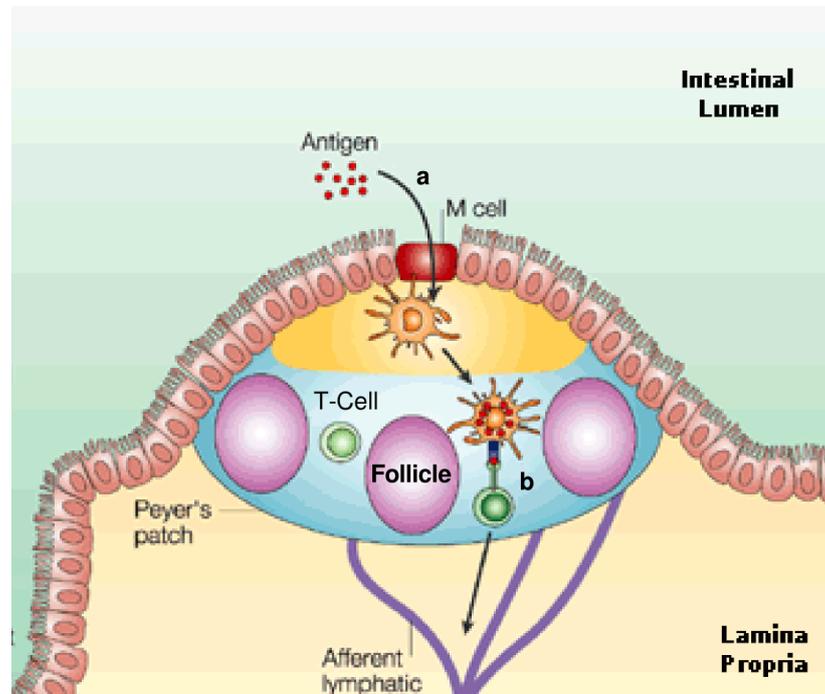
Figure 1.1: Schematic representation of Peyer's Patch in the small intestine. (a) Antigen enters the Peyer's Patch through the Microfold, or M-Cells in the follicle-associated epithelium. The antigen is transferred to local dendritic cells in the subepithelial dome (the orange area); (b) The antigen is then presented by the dendritic cell to T-Cells (green cells) in the T-Cell zone (blue), triggering an adaptive immune response. Alternatively, the antigen-loaded dendritic cell may migrate to mesenteric lymph nodes via draining lymph, and a response triggered in the lymph node. Image adapted from Figure 3 in Mowat (2003).

1.3 Peyer's Patch Development

1.3.1 Basic Model of Development in Mice

The use of genetic tools, imaging technologies and *ex vivo* culture systems has provided significant insights into the cellular components and associated signalling pathways that are involved in the formation of gut-associated secondary lymphoid tissue in mice (Mebius, 2003; van de Pavert and Mebius, 2010; Randall *et al.*, 2008; Veiga-Fernandes *et al.*, 2007). One example of such an investigation is demonstrated in Figure 1.3, showing the use of imaging of both adult and foetal mice intestines to determine the location of PP in the small intestine and the behaviour of cells that lead to organ
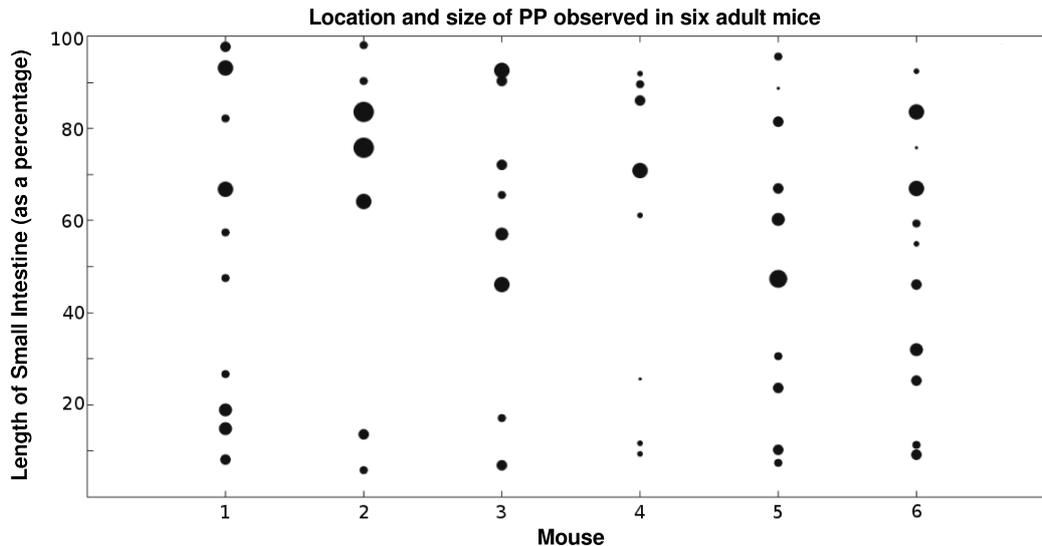
Figure 1.2: Peyer's Patch size, location, and number in six adult human-CD2-GFP mice. Position on the length of the small intestine has been normalised to a percentage to counter variance in small intestine length between mice. These data indicate that the development process is highly stochastic. *Graph taken from Alden et al* (2012b).

development. The combination of a number of experimental studies has led to the generation of an accepted model of pre-natal secondary lymphoid tissue development. The understanding of the mechanisms involved in human PP development is limited, mainly due to the difficulty of performing research using human foetal organs, and thus studies attempt to align observations made in humans with data from murine studies (Hoorweg and Cupedo, 2008). This section details the current understanding gleamed from explorations using mouse models, summarised diagrammatically in Figure 1.4.

In the mouse, the migration of hematopoietic cells from the foetal liver into the small intestine has been detected from Embryonic Day 14.5 (E14.5) (Mebius *et al.*, 2001). These hematopoietic cells can be divided into two populations, $CD4^-CD3^-IL-7R\alpha^-$c-kit$^+$CD11b$^+$CD11c$^+$ cells termed Lymphoid Tissue Initiator Cells (LTin), and $CD4^+CD3^-$IL-7Rα^+c-kit$^+$ termed Lymphoid Tissue Inducer Cells (LTi) (Fukuyama and Kiyono, 2007; Veiga-Fernandes *et al.*, 2007). Aggregations of these hematopoietic cells in the small intestine can be observed by E17.5, with the completion of the first processes in compartmentalising the organ and formation of follicles detectable via whole-mount immunostaining at E18.5 (Hashi *et al.*, 2001). It is thought that the process of organ formation between E14.5 and E18.5 occurs in three distinct phases (Adachi *et al.*, 1997). The first is the appearance on the epithelium of VCAM-1$^+$ stromal cells, termed Lymphoid Tissue Organiser Cells (LTo) (Adachi *et al.*, 1997; Fukuyama and Kiyono, 2007). This is followed by the identification of clusters of hematopoietic cells around VCAM-1$^+$ expressing stromal cells from E14.5. As previous imaging investigations have shown that hematopoietic cells (LTin and LTi) are evenly distributed across the gut by E15.5 (Randall *et al.*, 2008), it is assumed that this cell aggregation phase must occur after this point, yet before E18.5, deemed to be the

final phase of development where lymphocytes are recruited and the follicle structure developed.

The basic model of PP development in Figure 1.4 captures the current understanding of the second of Adachi et al's (1997) phases of PP development, the aggregation of hematopoietic cells. LTin cells express the tyrosine kinase receptor RET and initiate the process of PP induction upon surface contact with an LTo cell expressing Artemin (ARTN), a known ligand for RET, leading to LTo cell differentiation (Fukuyama and Kiyono, 2007; Patel *et al.*, 2012; Veiga-Fernandes *et al.*, 2007). This differentiation leads to the upregulation of adhesion factors Vascular Cell Adhesion Molecule (VCAM), Intercellular Adhesion Molecule (ICAM) and Mucosal Addressin Cellular Adhesion Molecule (MAdCAM). LTi cells interact with VCAM-positive LTo cells through the expression of $LT\alpha\beta$, that stimulates the $LT\beta$ receptor expressed on the LTo cell. This induces the production of IL-7, and chemokines CXCL13, CCL19, and CCL21 by the LTo cell (Adachi *et al.*, 1997; Honda *et al.*, 2001; Luther *et al.*, 2003). These in turn stimulate the IL-7R and chemokine receptors CXCR5 and CCR7 expressed by the LTi cell, thus a self-sustaining process is created where each cell type has the ability to mutually stimulate its corresponding component (Nishikawa *et al.*, 2003). This attracts LTi cells to a forming aggregation through chemotaxis, and adhesion factor expression retains them within the primordial patch. Cell aggregation continues through to E18.5 where, for reasons not currently understood, further aggregation of hematopoietic cells ceases to occur (Randall *et al.*, 2008).
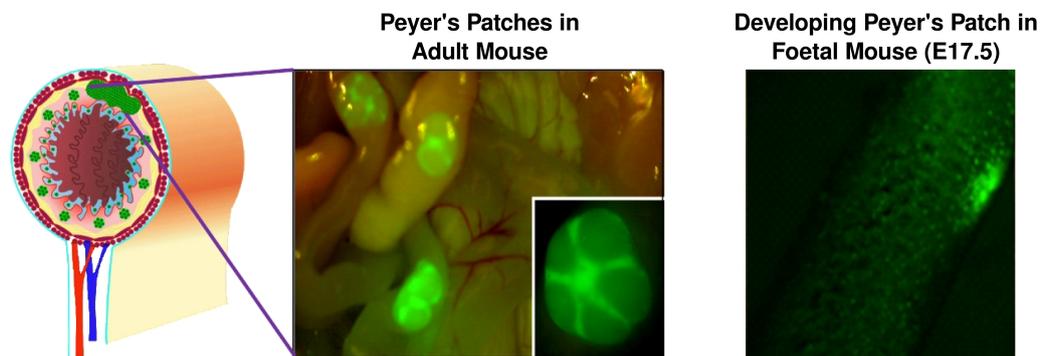
Figure 1.3: Peyer's Patches in the mouse small intestine. Left: Diagram representing a cross section of the intestine tract, large green patch represents where Peyer's Patches can be found in the small intestine. Source: Mark Coles, University of York, unpublished; Centre: GFP Stain of Peyer's Patches from a human-CD2-GFP adult mouse. Source: Mark Coles, University of York, unpublished; Right: *in vivo* confocal image of LTi cells at E17.5 in a human-CD2-GFP mouse foetal intestine. Source: Veiga-Fernandes et al (2007).

E14.5

RET

1

Gut

LTi   LTin   LTαβ

E15.5

ARTN - Ligand
for RET

Stromal Organiser
Cell (LTo)

LTβR

2

3

4

↑ CXCL13, CCL21 and CCL19
↑ VCAM1, ICAM1 and MADCAM1
↑ IL-7

E18.5

Aggregations of
hematopoietic cells

5

Lymphoid Tissue Initiator (LTin) Cell
(CD4⁻CD3⁻IL-7Rα⁻c-kit⁺CD11b⁺CD11c⁺)

Lymphoid Tissue Inducer (LTi) Cell
(CD4⁺CD3⁻IL-7Rα⁺c-kit⁺)
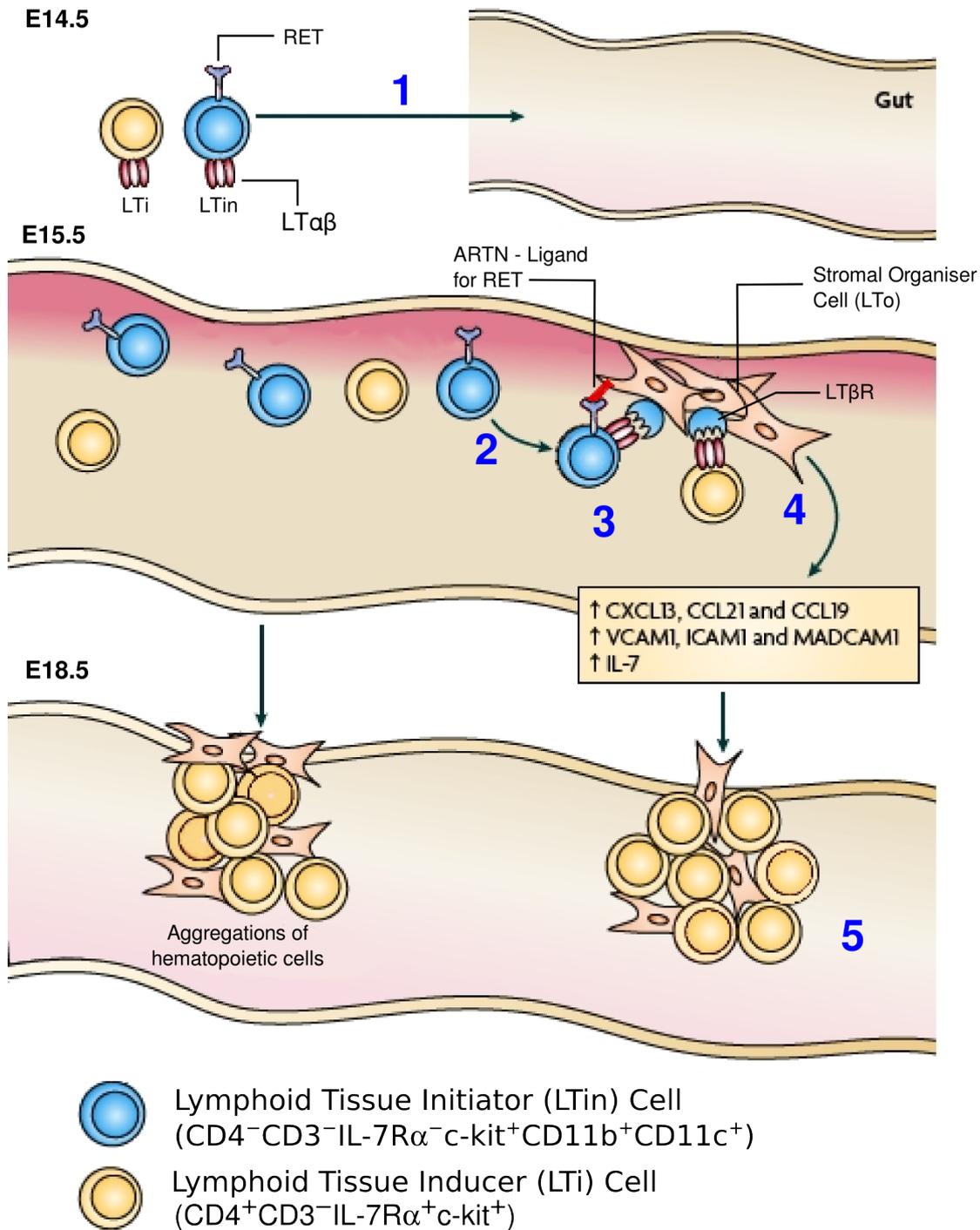
Figure 1.4: Steps involved in Peyer's Patch development, adapted from Figure 3 in van de Pavert et al (2010). 1: LTin and LTi cells migrate from the foetal liver into the gut; 2: LTin cells express RET, the receptor for ARTN, which is expressed by the LTo cell. RET/ARTN binding upregulates $LT\alpha\beta$ expression on LTin cells; 3: $LT\alpha\beta$ on LTin cells binds to $LT\beta R$ expressed by LTo cells, causing LTo cell differentiation; 4: $LT\alpha\beta$ expressed by LTi cells binds to $LT\beta R$ expressed by LTo cells, causing further LTo cell differentiation and expression of chemokines and adhesion factors; 5: Expression of chemokines causes LTi cell chemotaxis, and adhesion factors retain these cells around a primordial patch. This process continues until E18.5 where no further aggregation occurs.

1.3.2 Key Factors in Development

A number of studies have taken a reductionist approach and focused solely on the influence of individual biological components in PP organogenesis (such as particular chemokines and adhesion factors) to attempt to give some insight into the role each has in this process.

RET

Experimental work by Veiga-Fernandes et al (2007) suggests that the tyrosine kinase receptor RET, expressed by LTin cells, has a key role in PP organogenesis. LTo cells are thought to express ARTN, a known ligand for RET (Randall *et al.*, 2008; Veiga-Fernandes *et al.*, 2007). Flow cytometry reveals that genetic mutants that lack RET have a full repertoire of hematopoietic cells (LTin/LTi), yet these fail to aggregate, resulting in a lack of PP. This suggests that the RET/ARTN signalling pathway is vital in orchestrating secondary lymphoid organ development. Stimulation of RET has been suggested as the factor that then upregulates $LT\alpha\beta$ expression on LTin cells.

IL-7R and Lymphotoxin

$LT\alpha\beta$ expressing LTin and LTi cells bind to the $LT\beta R$ expressed on the surface of LTo cells, leading to LTo cell differentiation and the expresion of chemokines and adhesion factors (described in the following sections). Mice that are deficient for either $LT\alpha\beta$ or $LT\beta R$ do not form PP, suggesting this interaction is key in PP development (Banks *et al.*, 1995; De Togni *et al.*, 1994; Futterer *et al.*, 1998; Honda *et al.*, 2001; Pasparakis *et al.*, 1997).

$LT\alpha\beta$ is expressed by LTin cells through stimulation of RET (Veiga-Fernandes *et al.*, 2007). It is thought that stimulation of Interleukin 7 receptor α (IL-7Rα) induces the expression of $LT\alpha\beta$ on LTi cells. Initially it was shown that mice deficient for IL-7Rα fail to form PP (Adachi *et al.*, 1997, 1998). A further study that utilised a monoclonal antibody to block IL-7Rα signal at different time-points supported this result, showing that no PP formed when IL-7R was blocked before E16.5 (Yoshida *et al.*, 1999). However, where IL-7R signalling was blocked after E16.5, the authors still detected PP formation, suggesting that the influence of IL-7R is time-dependent, and is involved in the initiation of tissue development. The stimulant for IL-7Rα triggering remains an open question, with some suggesting this occurs through IL-7 expression by an LTo cell (Yoshida *et al.*, 2002), yet others have to date failed to detect IL-7 in the embryonic intestine (Honda *et al.*, 2001), and detected normal PP anlagen in IL-7 deficient mice (Nishikawa *et al.*, 2003). An experimental overexpression of IL-7 leads to an abnormal number of LTi cells and thus a higher number of patches, suggesting a potential role if present, although this is unlikely (Meier *et al.*, 2007). Whereas RANKL signalling triggers IL-7Rα expression in developing lymph nodes and

was thus suggested as a potential regulator of IL-7Rα, PP organogenesis still occurs in RANKL deficient mice (Yoshida *et al.*, 2002).

Chemokines

Stimulation of the LTβ receptor on LTo cells upregulates the expression of chemokines CXCL13, CCL19, and CCL21 into the localised environment around a site of patch genesis (Cupedo *et al.*, 2004; Dejardin *et al.*, 2002). These chemokines bind to chemokine receptors CXCR5 and CCR7 expressed on the surface of LTi cells, the first stimulated by CXCR13 and the latter by CCL19 and CCL21 (Luther *et al.*, 2003; Ohl *et al.*, 2003). Expression of homeostatic chemokines by an LTo cell causes LTi cell chemotaxis towards sites of patch genesis through CXCR5 and CCR7 signalling, which are retained by adhesion factors expressed in the primordial patch (expanded in next subsection). The attraction of cells to a site of patch genesis through chemotaxis promotes cellular interactions through LT$\alpha\beta$/LTβR signalling, and a further upregulation in chemokine expression, thus expanding the area around a primordial patch that is affected by chemokine expression.

PP and lymph node formation has been found to be significantly reduced in CXCR5-deficient mice, and where PP do form, these are typically smaller and lack the structural characteristics observed in wild-type mice (Ansel *et al.*, 2000). The existence of a second chemokine pathway, through CCR7 signalling, may explain why some hematopoietic cells still aggregate. Interestingly, studies suggest there is no significant difference in lymph node formation for CCR7 deficient mice, suggesting that the CXCL13 pathway could have a more dominant role in the recruitment and clustering of LTi cells (Luther *et al.*, 2003).

Adhesion Factors

The above details the interactions that promote LTo cell differentiation, cause the expression of chemokines and adhesion molecules, and ensure LTi cell chemotaxis towards a site of PP organogenesis. However none of the above would result in the formation of a PP if the aggregation was not held together by adhesion factors. LTin and LTi cells are retained by LTo expression of adhesion molecules VCAM-1 and MAdCAM, which bind to $\alpha 4\beta 1$ and $\alpha 4\beta 7$ receptors expressed on the surface of LTin and LTi cells (Yoshida *et al.*, 2001) respectively. A blockage in VCAM-1 expression has been found to show a profound reduction in cell aggregation (Finke *et al.*, 2002; Patel *et al.*, 2012) yet some aggregations do still form. This result could suggest that adhesion remains possible through the MAdCAM pathway, yet VCAM-1 is a more dominant factor.

1.3.3 Role and Dynamics of Hematopoietic Cells

Previous investigations have found that an absence of LTi cells results in a failure to form PP (Sun, 2000; Yokota *et al.*, 1999). The role and behaviour of the second population of hematopoietic cells, LTin cells, is not fully understood, although it has been suggested that this population is involved in an early phase of PP development (Fukuyama and Kiyono, 2007; Veiga-Fernandes *et al.*, 2007).

Early events in PP organogenesis have been explored through use of an *ex vivo* culture system (Patel *et al.*, 2012), in an attempt to further understand the role of LTin cells. In this study, explant cultures of developing intestines from human-CD2-GFP transgenic mice were incubated with beads soaked with ARTN, a known ligand for RET expressed by LTin cells (Figure 1.5). Immuno-staining after a 12 hour period revealed an accumulation of LTin cells in the vicinity of the bead and a strong upregulation of adhesion factor VCAM-1 in the vicinity of the bead. The detection of LTi cells in the vicinity of a bead however was rare. The same aggregation was also found to occur in LTi deficient mice , suggesting no role for LTi cells at an early stage of development. This finding supports the role of RET and RET ligand signalling as the initiator in PP development (Veiga-Fernandes *et al.*, 2007).

As LTin and LTi cells in CD2-GFP mice both express green fluorescent protein (GFP), it was possible to capture images of cell movement over the period of one hour at this early time-point in development. These images were processed using the Volocity software tool (PerkinElmer) that provides the capability of creating a sequence from time-lapse images and tracking cell motility over time, producing a distribution of cell behaviour statistics. The tracked cells were categorised into two groups, those $<50\mu$m from the ARTN-soaked bead and those further away (Patel *et al.*, 2012). A statistical analysis of cell track length, velocity, and displacement using the Volocity software tool reveals that there is a statistically significant difference between each behaviour response between cells close to the bead and cells further away (Figure 1.5). This alteration in cell motility is suggested to be mediated by the expression of adhesion factors, supported by the finding that VCAM-1 expression is strongly upregulated.

1.3.4 Open Questions

Section 1.3.2 above details how several reductionist studies have been used to determine the role of each factor in PP organogenesis. However, such an approach leaves interesting questions that cannot be addressed using this technique. This section suggests a few such explorations that have yet to be performed.

Figure 1.2 demonstrates that the formation of PP in mice is highly stochastic, in terms of location, size, and number of PP that form. Indeed Cornes (1965) suggests that no two observations will be identical. What causes such variance is yet to be understood. One could suggest that this is caused by a variation in the availability
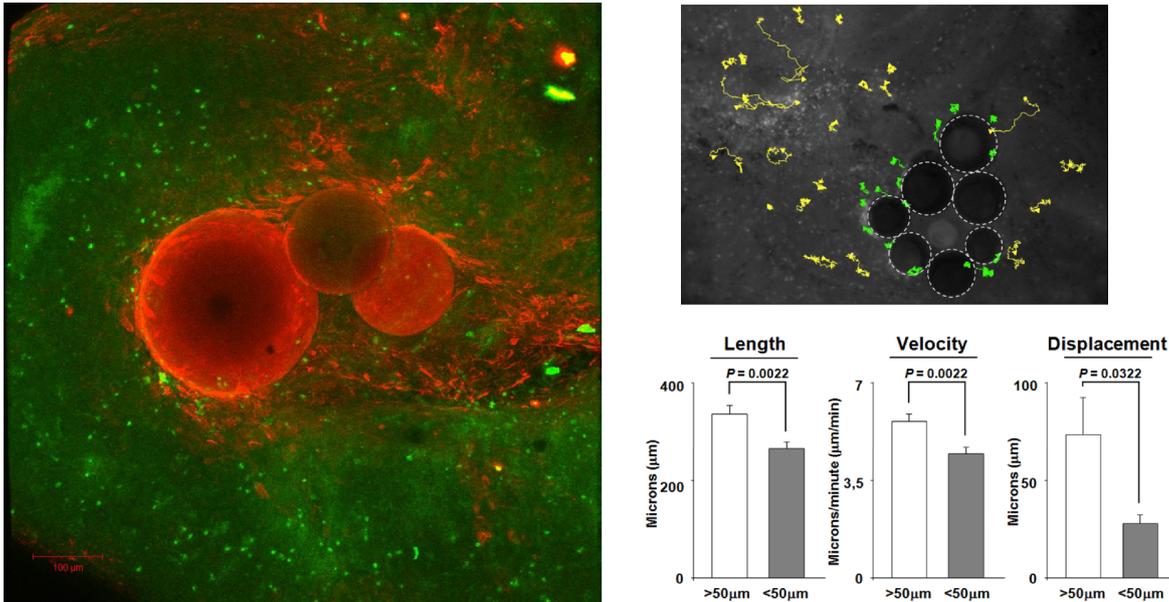
Figure 1.5: Use of an *ex vivo* culture system to explore cell behaviour. Left: Representative image of confocal microscopic analysis of an explant culture of a developing intestine from a human-CD2-GFP transgenic mouse incubated with an ARTN soaked bead. Immunofluorescence staining: green - LTin/LTi cells; red - VCAM-1 cells. Scale: 100 μm. Top right: 60-minute time-lapse cell tracks at the twelve hour time-point, produced using Volocity. Green - cells <50μm from the bead; Yellow - cells >50μm from the bead. Scale: 100 μm. Bottom right: Cell behaviour analysis of the cell tracks shown top-right. White - cells >50μm from the bead (n=27 cells); Grey - cells <50μm from the bead (n=17 cells). Length is the distance covered by the cell in that hour. Displacement is the distance between the cell location at the start of the hour and the cell position when tracking was ceased. P-Values calculated using Mann-Whitney U-Test. Image adapted from Figures 1 and 2 published in Patel et al (2012)

of key cell types and factors between each observation, by the physical geometry of the intestine, or simply that the process relies on the location and timing of contact between an LTin and LTo cell, and thus is inherently stochastic in nature. However current approaches do little to support any such conclusions.

The role of chemokines and adhesion factors has been elucidated from both gene knockout experiments and *ex vivo* culture systems as detailed previously. However, it has not yet been possible to quantify the expression of chemokine and adhesion factors during development. The existence of such data may go some way towards explaining the variance detailed above, while also suggesting if there is a limitation in the expression of these factors, and thus a limitation on the size of PP. Though techniques exist that could make this possible (ELISA, bioassys, RNA extraction (Mahajan *et al.*, 2003)), it is currently believed that no such quantitative data exists.

Finally, the majority of the investigations detailed in Section 1.3.2 all examine the end result: the formation of PP at E17.5. As noted previously in Section 1.3.1, studies have suggested that PP development does occur in three distinct phases (Adachi *et al.*, 1997). However, the authors do not fully specify the time-points these phases occur in the seventy-two hour period. The more recent *ex vivo* study does go some way to

addressing this by examining the initial period of development (Patel *et al.*, 2012), but does not identify changes in system dynamics after that point. If such distinct phases exist, different biological factors may be more influential at different time-points; an interesting result that cannot be elucidated using gene-knockout experiments.

1.4 Advancing Biological Understanding Through Modelling and Simulation

The use of models in advancing the understanding of biological systems is common place and has been used for generations, the famous example being Watson and Crick's model that defined the structure of DNA (Watson and Crick, 1953). Modelling provides a means of exploring a concept or available biological data with the aim of generating hypotheses in advance or in place of further experimental investigation. Such an approach has provided some of our fundamental understanding of immunology, an example being the Nobel Prize shared by Burnet and Medawar for their combination of modelling and experimentation to explore how the immune system discriminates between self and non-self (Chakraborty *et al.*, 2003). The use of computational modelling and simulation is a continuation of this approach. An underlying biological system is examined with the intent of generating a *model* that details the current understanding of the system, or an abstraction of it. This may then be instantiated as a *simulation* that can be executed on a computer. (Polack *et al.*, 2008; Read, 2011).

The integration of computational modelling with current experimental techniques is an important step in moving biological explorations from a reductionist, descriptive state to one that is predictive (Kumar *et al.*, 2006). Section 1.3.2 described how previously published investigations have determined the role of each factor in PP analysis by examining each factor individually. The adoption of computational modelling shifts the focus from an examination of each individual component part to that of the higher order behaviour, and how this emerges from components that lack the capability to do this alone (Germain *et al.*, 2011). This is known as the particular system's emergent property. The application of this approach has previously permitted the exploration of a range of complex biological systems, including T-cell signalling cascades (Chakraborty and Das, 2010), autoimmune disease pathology (Read *et al.*, 2009), investigating cell migration within germinal centers (Figge *et al.*, 2008), emergence of immune memory (Jacob *et al.*, 2004; Lagreca *et al.*, 2001), and system dynamics under HIV-1 infection (Sieburg *et al.*, 1990; Stafford *et al.*, 2000). However, the approach has yet to be adopted in exploring the formation of the immune system, and could have the potential to examine the open questions detailed above in Section 1.3.4.

As the use of computational modelling becomes more prevalent, the potential benefits of adopting the approach in immunology are becoming increasingly apparent. Significant advances in laboratory experimental techniques has coincided with advances in

and availability of computational power, making it possible to generate models and simulations that can explain large datasets generated using high throughput techniques. Such datasets may be derived locally through experimentation, or be a consolidation of datasets gathered from a variety of laboratories and techniques (Chakraborty *et al.*, 2003; Kirschner and Linderman, 2009). Computer simulations of a biological process may also be more amenable to experimentation than that of the natural system (Forrest and Beauchemin, 2007; Guo and Tay, 2005). Investigations that can be conducted on a computer do not have the physical or ethical constraints that may apply to investigations in the wet-lab. Such computer experimentation can aid the formation of hypotheses that explain the derived biological data and be used to evaluate these hypotheses by comparing simulation results to those in the established literature (Chakraborty and Das, 2010; Chakraborty *et al.*, 2003; Kirschner and Linderman, 2009). However, alongside its use as a tool supporting such analysis of biological data, a robust simulator can be used as a predictive tool for informing future wet-lab investigations, through performing novel *in silico* experimentation prior to or in place of laboratory work. Such investigations may reveal areas of the system where the current understanding is incomplete and areas where further wet-lab experimentation is not required. A focus on the higher-order behaviour and interactions between factors in the captured system also makes it possible to understand how quantitative alterations in individual components affects overall system behaviour.

This thesis continues by examining this technique and how it is applied in the modelling of biological systems. This includes a summary of methodologies, tools and frameworks that have found application in the creation of such simulations.

1.4.1   Modelling Methodologies

There are broadly two categories of modelling approaches, although some hybrid versions combining elements of the two are beginning to emerge. These are introduced below.

Mathematical Approaches

Mathematical models that summarise a biological system as a set of Ordinary Differential Equations (ODE) have frequently been used to provide biological insight. These capture populations of factors rather than individual instances, with each assigned a real-number variable. A set of equations is compiled that specify the impact each factor has on the size of the population of its complementary factors. Using these equations, system behaviour that emerges from interations between large populations of factors can be explored. Use of this approach has found application in explorations of innate immune responses (Hu *et al.*, 2007), immune system memory (Antia *et al.*, 2005), pancreatic cancer treatments (Haeno *et al.*, 2012), and in furthering the understanding

of influenza A infection (Baccam *et al.*, 2006; Beauchemin *et al.*, 2008; Sidorenko and Reichl, 2004; Smith *et al.*, 2011; Smith and Perelson, 2012).

A substantial amount of ODE modelling work has also been undertaken to explore the dynamics of human immunodeficiency virus 1, or HIV-1. (Perelson, 2002; Perelson *et al.*, 1996; Vaidya *et al.*, 2010), and is thus a good example to use in exploring this technique. The basic model of infection by Perelson and colleagues (2002) maintains counts of three populations: host cells free of infection (T), host cells that are infected (I), and circulating virions in the blood (V). The model describes the change in the sizes of these populations over a period of time (dt). The number of target cells in the model is dependent on parameters that specify the rate at which new cells are generated (λ), the death rate per cell (δ), and the rate at which cells become infected. The latter occurs when target cells upon interact with viral particles at a rate stated by κVT, with κ representing an infection rate constant. The population of infected cells changes at the rate stated by κVT and the rate at which infected cells die, set by parameter μ. The final population, the virions, changes at a rate at which the infected cells generate new virus particles, set by parameter p, and the rate at which virions are cleared from the blood, captured by parameter c. All of this behaviour is captured in equations 1.1-1.3 below:

$$\frac{dT}{dt} = \lambda - \delta T - \kappa VT \tag{1.1}$$

$$\frac{dI}{dt} = \kappa VT - \mu I \tag{1.2}$$

$$\frac{dV}{dt} = pI - cV \tag{1.3}$$

These equations capture the basics of HIV-1 dynamics. This model can be extended to include potential interventions that change these dynamics: the work of Perelson and colleagues (1996) and Vaidya and colleagues (2010) that capture the inhibition of the virus replication by a anti-retroviral drug and the resistance to such drugs being good examples. Once the model is generated, a parameter fitting stage is conducted where output from the calculation is set to match data that has been derived through wet-lab experimentation, often through use of least sum of squares regression analysis to minimise the squared difference between the laboratory data and that from the model (Read, 2011). With this assured, a wide variety of analytical techniques can be used to explore the captured process. The use of ODE's lends itself to such analyses as they are computationally efficient, and thus a large parameter space can be explored (Bauer *et al.*, 2009).

As may have been apparent in the above description, there is no mention of spatial considerations in the HIV-1 model. Space is assumed to be one continuous compartment, assuming a well-mixed space of elements. Thus traditional ODE's do not have the capability to represent the effect an environment may have on the populations: the focus is purely on examining the parameter values that capture interactions between the populations. This can be countered through the use of Partial Differential Equations (PDE's) that can capture changes in time and space, yet increase the complexity of a model such that the advantages of using an ODE-based approach are mitigated (Bauer *et al.*, 2009).

Agent-Based Approaches

The above technique examines how interactions between biological factors influence a population, such as how the level of viral production affects the population of target cells in HIV-1 infection. The assumption is made that each entity within the population is identical. An agent-based modelling (ABM) approach differs as each biological entity, such as a cell, is represented explicitly, and can thus maintain its own attributes and cell state (An, 2006; Bauer *et al.*, 2009). Agent behaviour is specified as rules that determine the set of states an agent, such as a cell, may exist within, and the event that must occur for an agent to change state. This event could be an interaction between another agent or the environment. Taking the HIV-1 study above as an example, if an ABM approach was taken, each target cell would thus be its own individual entity, and rules specified on how the cell would change into an infected state upon interaction with a 'viral' agent.

Modelling at an individual level rather than population level opens up an array of investigations that are not possible with ODE modelling (Bauer *et al.*, 2009). With the increased use of two-photon imaging it has become possible to visualise the interactions between individual cells and their environment: one example of this being an observation of the primary immune response for fifty hours after an antigen is encountered (Catron *et al.*, 2004). The *ex vivo* culture system described in section 1.3.3 has also utilised this technique to observe individual cells rather than a population (Patel *et al.*, 2012). Studies such as these reveal that it may not be correct to model all cells as one population, and inherent stochasticity in the biological system may be an important part of the system dynamics. As each cell would be represented explicitly using ABM, such stochasticity can be captured, making this a suitable approach if this is an important consideration (Germain, 2001; Milanesi *et al.*, 2009). This methodology has found application in modelling cancer vaccination (Motta *et al.*, 2005), experimental autoimmune encephalomyelitis (Read, 2011), vaccine design (Kohler *et al.*, 2000), and tumour growth (Alarcon *et al.*, 2005; Jiang *et al.*, 2005). It has also been applied in simulating clinical trials, linking explorations through modelling with ongoing work in the clinic (An, 2001, 2006).

The other significant difference between an ABM and traditional ODE approach is the ability to explicitly represent the environment in which the cellular interactions are taking place. This is a vital consideration for biological systems such as those taking place in foetal development where the environment has an influence over cell behaviour yet is also constantly developing, or those where cell behaviour is constrained by the environment. Efroni and colleagues' (2007) model of the cell structure within a lymphoid organ is a good example of the latter.

Choosing the Correct Approach

The choice of modelling approach is dependent on the scope of the problem being investigated, and thus there is no correct or incorrect choice of strategy. The following considerations tend to guide the decision of which approach should be adopted.

One of the advantages of using an ODE is that these tend to be more simplistic than applying an ABM (Bauer *et al.*, 2009). The equation or set of equations that are generated tend to be formulated from a lower number of parameters than required using an ABM approach, capturing less but recreating patterns observed in the domain of interest. In both cases these parameters can be set through a process of calibration, to ensure that the model correctly captures expected behaviour. Where an ABM implementation may lead to a large number of parameters, an ODE approach may be more viable (Bauer *et al.*, 2009). The set of equations generated using an ODE approach may be complicated but are unambiguous: it is these which capture the system being modelled. With an ABM approach however, the implementation of the model is much more complex, and the detail hidden within the implementation can affect the overall result. However, there are frameworks that are being utilised to make the design of ABMs more transparent and easier to interact with, which will be examined later in this chapter.

A modeller needs to consider whether the scope of the problem is to investigate organism wide cell population dynamics or whether each individual in the system needs to be examined as its own distinct entity (Germain *et al.*, 2011). As described previously in the HIV-1 model, ODE's lend themselves well to a study of host-pathogen interactions within a population, yet make the assumption that each entity within the population is identical. However if characteristics of each individual target cell were to be examined, an ABM approach is required.

The inclusion of spatial aspects is difficult to achieve with an ODE model. Where this is an important consideration, an agent-based model can be more appropriate (Germain *et al.*, 2011), and in some cases can contribute to the accuracy and meaning of the result. Strain et al (2002) compared their ABM of HIV infection with that developed using an ODE approach (Perelson *et al.*, 1997) and found that the viral infectiousness was overestimated by more than an order of magnitude in the ODE model (Bauer *et al.*, 2009), suggesting a consideration of space is important. Beauchemin and

colleagues (2006) have also utilised an ABM model to capture HIV-1 dynamics, and determined that representation of space is indeed an issue, reporting more reliable results from an ABM when compared to real-world experimental data. For studies that focus on the formation of cellular structures or tissue rather than host-pathogen interactions, for instance the formation of germinal centres (Figge *et al.*, 2008), the modelling of space, and how this affects the formation could be a key consideration.

Where choice is not constrained by either of the above, it may be necessary to consider the availability of computing resources when determining whether to use an ODE or ABM. In an ABM implementation, each individual entity is captured as an agent, with each possessing a set of properties. In the majority of computational implementations this will require each individual entity to be represented as an individual process. For systems with a large number of agents this will require a substantial computational resource. However as the availability of powerful computing resources such as computer clusters and cloud computing services increase, this drawback could be negated. Additionally, it could be possible to consider a combination of aspects from both ODE and ABM techniques, and implement a stage-structured model. This approach has been demonstrated in a model of the cytotoxic T lymphocyte response to antigen (Chao *et al.*, 2003). In this implementation, each agent's life cycle was divided into stages and deemed to be identical to all other agents currently in that stage. As this is the case, an integer value can be used to represent the population of agents in that stage, removing the need to represent each agent explicitly, thus improving the efficiency of the model by a number of orders of magnitude. Where the purpose of a model is to capture an emergent population-level phenomenon while considering the global behaviour of each agent involved in producing that phenomenon, this case study is a good example of a methodology to adopt where it is necessary to capture a very large number of agents.

1.4.2 Modelling Tools and Frameworks

As the integration of modelling and simulation with conventional wet-lab research has become more popular and prevalent, a number of frameworks and toolkits have been made available to aid simulation creation. This section summarises currently available techniques and tools that are available and that have previously been used in the modelling of biological systems, after which a brief evaluation is provided on the tool suitability.

The CoSMoS Process

Although not designed specifically for biological systems, the Complex Systems Modelling and Simulation Infrastructure (CoSMoS) Project was an EPSRC funded study that established generic tools and techniques to support the modelling, simulation, and analysis of complex systems (Andrews *et al.*, 2010), and has found application in

modelling biological systems (Garnett *et al.*, 2008; Read, 2011) as well as ecological and sociological systems (Polack *et al.*, 2010). Collaboration between an investigator constructing the model and an expert(s) in the system being modelled is encouraged, and processes developed that aid interaction between the two. A set of rigorous activities is proposed that leads to the generation of a series of models, a process that is detailed in Figure 1.6. These models underpin the understanding of the system that is to be simulated, providing a level of transparency in the manner in which a simulation has been implemented, an agreed specification of the scope of the work, and a means in which results can be interpreted in relation to the captured real-world system. This section examines each of these models generated by utilising this process.
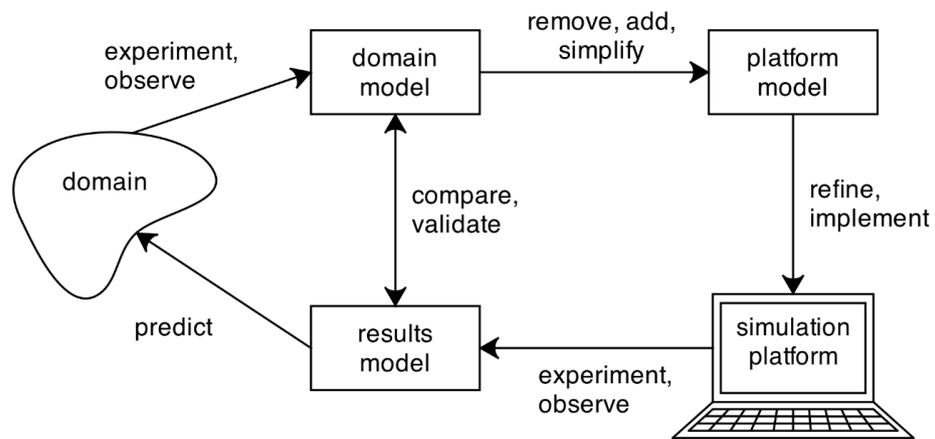
Figure 1.6: The CoSMoS Process for modelling complex systems, detailing the stages of the process and flow of information. This image of the process was taken from Read et al (2011)

The system of interest that is being modelled, for example the foetal development of PP, is termed as the *Domain*. The current scientific understanding of the system dynamics, such as that detailed in section 1.3.1 for PP development, is captured in a *Domain Model*. This process sets the scope of the exploration in collaboration with an expert in that system, and can reveal areas to be captured where current understanding is incomplete. Where this occurs, suitable assumptions are made and justified for sake of scientific transparency. All the entities that have a role in system dynamics, and the behaviours that produce an observed phenomenon, are included within the domain model. At this point no consideration is given into how this will be implemented as a simulation.

With this specification agreed, a *Platform Model* is generated that specifies how the domain model can be implemented as a simulation platform. This specifically details how each entity and its relevant behaviour described in the domain model will be implemented in a computer simulation. Critical to the process, emergent behaviour specified in the domain model, such as a statistically significant change in a cells behaviour, is removed. Such observed behaviour must emerge through interactions between entities and not be directly encoded into the model, as this invalidates the simulator as a pre-

dictive experimental tool. If this behaviour does not emerge through interactions, the platform model does not capture the domain model appropriately, or the specification of the system under study in the domain model is incorrect. In the process of generating this model, further simplifications and assumptions may be necessary and are documented accordingly for scrutiny alongside simulation results. Alongside the detail of the system being captured, this model also specifies user interfaces and data capture mechanisms that are required in order to interact with the simulator.

The specification that comprises the platform model is then implemented as a *Simulation Platform*, adopting one of the modelling approaches described in detail in section 1.4.1. The process places no constraints on modelling methodology or choice of programming language. Completion of this phase leads to the generation of an *in silico* tool through which experimentation can be performed.

The *Results Model* summarises the understanding generated from experimentation conducted using the simulator. At this point, results from the simulator can be contrasted with real-world results in the domain model to determine a level of confidence in the simulator as a suitable representation of the system being captured. Such comparisons under different simulation conditions may also generate predictions that inform future investigations.

It can be noted from Figure 1.6 that the process has no defined end point. A comparison between results generated by the simulator and those with the domain model may lead to a refinement of the domain model, upon which the process starts again. This is a process that will continue until the domain has been adequately captured. Where this is the case, further iterations could also occur if the model is extended further.

Immune System Simulation Platforms

Whereas the above provides a toolkit to aid the specification of a model, it does not specify how that model is implemented. A number of platforms have been developed that enable an investigator to specify the events that occur on an interaction between cells, to simulate a response. The objective behind their creation was to enable biologists who may lack the mathematical and computer coding skills to construct either ODE or ABM simulations (Meier-Schellersheim *et al.*, 2006). This section considers two such tools, IMMSIM (Puzone *et al.*, 2002) and Simmune (Meier-Schellersheim *et al.*, 2006), although it is noted that similar packages (Reactive Animation (Efroni *et al.*, 2005), SIS (Mata and Cohn, 2007)) are available.

IMMSIM is based on use of a cellular automator approach (Celada and Selden, 1992), and was constructed to examine discrete interactions within a set lattice grid environment (Puzone *et al.*, 2002). The platform has found application in exploring the generation of immunological memory (Celada and Selden, 1992), hypermutation (Celada and Seiden, 1996), and autoimmune responses (Celada and Seiden, 1998). The

modeller specifies the entities that have a role in the interaction, and the events that should occur on interaction. Each individual entity in the system being captured is represented as a data structure to store relevant properties, thus following an ABM implementation as described previously. Each entity is then associated with a receptor, with interactions between entities modelled as an interaction between these receptors. Simulation execution is performed in steps, with an event triggered if an interaction is detected in that time-step. This allows for a representation of time in the simulation. Results are generated that detail the impact that interactions during a simulation run has had on particular measures, which could include number of cells in a particular state, or antibody maturation.

Simmune has been developed to allow the modelling of molecular interactions without involvement in any underlying mathematics (Meier-Schellersheim *et al.*, 2006). These reactions are defined using a graphical representation where the molecules in the system are detailed and the complexes that comprise their respective binding sites stated. The modeller connects binding sites via arrows, and specifies conditions that must be met for the interaction to occur. These interactions may affect molecules both within a cell and on the membrane, and Simmune allows for the inclusion of both, thus making it possible to model the effect of that interaction on the underlying cell chemistry. Once population sizes are specified, Simmune calculates the time-course in which changes occur to the molecular makeup of the model once an event is applied, producing a quantitative measure of the binding states of all molecular mechanisms in the model. The use of the platform has been exemplified in examining the role of chemosensing in *Dictyostelium* (Meier-Schellersheim *et al.*, 2006).

Both packages are publicly available and in a process of ongoing development. Exploration of the reactions that are captured are performed by changing the interaction rules that are specified in the design of the model.

Simulation Toolkits

Platforms such as Simmune and IMMSIM have been constructed to aid the modelling of biological interactions that lead to an immune response. For applications that are not host-pathogen based, the platforms may not provide the required functionality. In these cases, it may be more appropriate to implement the simulation using one of many simulation toolkits that are available.

Two examples of publicly available simulation toolkits are MASON (Luke, 2005) and BREVE (Klein, 2002). Both provide a suite of background functionality upon which a simulation can be built, including tools to create an environment, create agents, perform collision detection between agents, and tools for visualising and capturing the simulated environment. The modeller specifies the agent types, the states the agents could reside within, and the actions that agent performs both in that state and on interaction with another agent. Simulations are then executed in discrete time-steps

where each agents performs the respective behaviour dictated by its current state. This allows for the inclusion of both space and time in the simulation, important for scenarios such as Peyer's Patch development where tissue formation is occurring in a set environment for a set time-period. MASON is open source and has been implemented in the Java programming language, whereas BREVE simulations are written in a bespoke programming language, steve. The former focuses on the modelling of multi-agent simulations, and has thus been optimised to ensure it can effectively handle a large number of agents. The latter focuses more on the environment that the interactions take place, and captures this as a continuous 3D space. The appeal of both is that these are open source and extendable, and can thus be more flexible than the immune simulators in the previous section. MASON has recently found application in the modelling of experimental autoimmune encephalomyelitis (Read, 2011), a model for multiple sclerosis, and BREVE in the modelling of global immune responses to primary and secondary exposure to antigen (Jacob *et al.*, 2004).

Toolkits and Frameworks: A Discussion

The objective behind the creation of the IMMSIM and Simmune packages is understandable: construct a package that makes simulation creation accessible to biologists who may not have the skills to implement a simulator using other methods. Using intuitive graphical interfaces, the process is straightforward, and quickly leads to results that could be used to back an hypothesis or inform future work. Although this is appealing, there are aspects that need to be taken into consideration with this approach.

It is unlikely that a simulation will be created where all the underlying biological detail is understood. This can be countered by making suitable assumptions and abstractions, although it may not be known as to the effect these have on system dynamics and behaviour. One drawback in the use of immune system simulation packages is that these do not allow for the documentation of assumptions and abstractions that have been made. It is noted that for Meier-Schellersheim et al's (2006) model of chemosensing that exemplifies the use of Simmune, these assumptions were released as supplementary material that accompanies their results. However it has to be questioned as to whether all investigators will do this with packages being promoted as drag-and-drop simulation creation tools. A lack of transparency on what has been included in the model, and what has been left out, will lead to some justifiable scepticism when judging results.

This could however be countered by using such packages alongside the CoSMoS Process (Andrews *et al.*, 2010). This could address any lack of transparency that using the platforms alone may create, through the creation of a model that specifically details the underlying biology that has been captured and any assumptions and abstractions made, and a further model that acts as a specification of how this could be implemented in IMMSIM or Simmune. This implementation would then take the role

of the Simulation Platform within the CoSMoS Process.

If this were to be the case, the modeller would have to decide whether IMMSIM and Simmune have the functionality to capture the biological system being studied. Both IMMSIM and Simmune have been constructed with the aim of examining responses to interactions, and can detail how the response affects system dynamics. For applications that examine host-pathogen or cell receptor binding affinity, such as the many examples where IMMSIM and Simmune have been applied, this is sufficient, and where the specification is robust, significant biological insight can be gained. However if the emergent behaviour under examination is not based on how these interactions affect cell chemistry or properties, the use of IMMSIM and Simmune may not be appropriate. These may also not be suitable for modelling biological systems where the environment these interactions take place within influences the behaviour that is observed. Tissue development could be an example of both, where aggregations of cells emerge from interactions between cells in the system, mediated through interactions with the environment. Where this is the case, it may be more appropriate to use one of the simulation packages available to implement the simulation. Again, for the sake of transparency, this could be paired with use of the CoSMoS process, with simulation packages acting as an aid in the creation of the simulation platform.

Whereas IMMSIM and Simmune have sought to develop tools for biologists that may lack the mathematical and computing skills to implement a simulation, the CoSMoS process takes a different stance, suggesting that simulation creation is aided by a collaboration between an expert in the field being simulated and those creating the model (Andrews *et al.*, 2010). A successful collaboration can raise important questions that the field expert may not have yet considered, while ensuring that all involved understand the strengths and limitations of the tool that has been created (Polack *et al.*, 2010). This is an important step in understanding what the results mean in terms of the system being modelled.

1.5 Confidence in Simulation as a Representation of the Biological System

Any exploration of a biological system, whether this uses current laboratory or computational techniques, can be treated with a degree of scepticism as the understanding of each underlying aspect is incomplete. In both types of exploration, this is addressed through the making of well justified assumptions. Thus the exploration is examining an abstraction of the real system rather than the full detail. Implementing a computer simulation of a biological system adds a further level of abstraction, as it is intractable to capture all aspects of the biological system in the model. However as the aim is to understand how interactions between factors lead to an observed, emergent behaviour, the simulator does not need to be a complete representation of the system (Germain

et al., 2011). For example, an interaction between two cells may lead to a higher-order effect on system dynamics and trigger pathways within each cell. With the interest being on the former and not the latter, an assumption can be made that the particular pathway is always triggered. It is critical that the assumptions and abstractions that are made are taken into consideration when scrutinising simulation results to determine their relevance to the biological system being explored.

Computational methods are finding increased application in the investigation of a variety of biological systems, and explorations using a computational immunology approach are increasingly being applied alongside current experimental techniques (Cohen, 2007; Germain *et al.*, 2011; Kleinstein, 2008). In Section 1.3.4 above, it was noted that although a reductionist approach to immune system development has provided some key insight, questions do remain that such methods cannot address. Computational methods are not introduced to replace such investigations, rather to complement them. For confidence to be retained in the use of simulation as a tool for understanding biological systems, it is important that the relationship between the simulation and the system it captures is appreciated. However, it has been noted that there are few cases where the adequacy of a simulation, in terms of representing the system it captures, is discussed alongside results generated from it (Read, 2011).

This section examines a number of methods that aim to establish confidence that a result from a simulation developed using one of the techniques described in previous sections is representative of the biological system it captures. This issue has been the topic of a recent study by Read (2011) that has examined current practices in generating confidence in simulation, alongside the development of a simulation of Experimental Autoimmune Encephalomyelitis (EAE). Read's work is discussed alongside other methods that were not included in his study. For this examination, the same definition of confidence is used as that proposed by Read: that confidence is not absolute, yet related to the simulation purpose and scope of experimentation, and established through ensuring simulation results are representative of the abstraction of the system it captures (Read, 2011).

1.5.1   Simulator Calibration

As was noted previously in this chapter, it is highly unlikely that a model will be generated where values can be assigned to all parameters identified in its creation. This may be as the underlying biological understanding is incomplete, or the method of implementing that factor does not translate back to laboratory derived values (for example the use of a probability function to capture binding affinity). Calibration is a process by which values are assigned to parameters such that simulation responses are generated that are representative of those in the system being modelled. Through this process the simulation is deemed an adequate representation of the higher-order behaviour observed in the real-system, although a number of assumptions and abstrac-

tions will have been made. Although such a process is therefore a vital stage in the parameterisation of a model, it is rare to find the process of calibration documented alongside simulation results (Read, 2011). A limited number of studies do mention that a procedure has been performed (Ray et al (2009) being one example) but not noted how this was performed or the data that was used in the process. However, procedures to fit simulation results to biologically-derived responses have been described in simulations of lymphocyte migration (Figge *et al.*, 2008) and the life cycle of *Mycoplasma genitalium* (Karr *et al.*, 2012).

It may however be difficult to access a suitable set of biological results upon which a simulation can be calibrated. Even where one exists, there may be significant variance both in the result set and in comparison to results generated elsewhere, using identical or other techniques, and it is thus questionable as to whether this is a representation of the system being captured. It is cases such as this where a collaboration between an expert in that biological system and a modeller is advantageous, as encouraged by the CoSMoS Process (Andrews *et al.*, 2010) detailed previously (Section 1.4.2). A good example of this is the calibration process applied to Read et al's (2012) model of Experimental Autoimmune Encephalomyelitis (EAE). Initial parameter values were established using a set of parameter values arrived at through informed estimation, and simulation responses generated. An iterative process was then conducted where the biological system expert, in this case a collaborator with expertise in EAE, examined the simulation dynamics and identified areas that did not fit with their understanding and experience of the biological system. This led to further parameter alterations and further discussions. Through a combination of the collaborators expertise and the modellers understanding of how the simulator captures system dynamics, a set of parameter values were identified that adequately captures the collaborators understanding of the system dynamics.

Calibration is an important process as it establishes the behaviour of the simulation under circumstances deemed as normal. This acts as the baseline, to which the results of future *in silico* investigations are compared. Without this link from the start, it is not possible to use the simulator in experimentation that aims to provide biological insight.

1.5.2   Validation Tools

Calibration may establish behaviour that produces an expected result, but producing evidence that shows a simulator generates a result representative of the biological system is not alone enough to provide confidence in the use of the simulator as an experimental tool. Simulations have previously been criticised as being opaque tools (Di Paulo *et al.*, 2000), where a result is generated yet it is not obvious why this is the case. Full transparency in the implementation is key in understanding why such a result is produced and verifying that there are no errors within the implementation that

have affected that result. A process of validation can be used to determine how fit a simulation is for its purpose (Polack *et al.*, 2011). Such a process exposes the decisions made in its design and implementation for scientific scrutiny. Where simulation is to be used as a scientific instrument for furthering the understanding of a captured system, this is critical, as these decisions and the evidence behind them may impact the result generated.

Previous studies concerning the validation of simulations of complex systems have taken inspiration from that used in the field of safety critical systems (Ghetiu *et al.*, 2009, 2010; Polack *et al.*, 2011). Computational tools used in this field are subjected to a stringent analysis of their compliance to a set of requirements before being introduced, as an error within or failure of the implementation could lead to potentially life-threatening circumstances (for example where this software was used in an aircraft). Such an analysis utilises an Argument-Based Validation (ABV) technique where the requirement is matched to evidence detailing how it has been met in the implementation, and where applicable, how relevant parameters where derived. Visual notations such as Goal-Structuring Notation (Kelly, 1999) have been developed to provide a method of structuring such an analysis, ensuring each step in the implementation is validated, the reasoning behind the inclusion or exclusion of a feature or assumption is provided, and evidence given as to why this conclusion has been drawn (Ghetiu *et al.*, 2010). For GSN, this process is conducted through the production of a flowchart, comprised of the notation in Figure 1.7a. A basic example of the structure of a Goal Structuring Notation (GSN) argument can be seen in Figure 1.7b.

It is proposed that ABV techniques can be used prior to, during, and after the design and implementation of a computer simulation (Ghetiu *et al.*, 2010). A set of requirements, or goals, are identified. The strategy used to ensure this requirement is met is clearly stated. Evidence that supports a claim that this requirement is met is specifically noted alongside that goal, with any assumptions that were required. The use of the technique was recently exemplified by Polack et al (2011) to explore the suitability of a simulation of cell division and differentiation in the prostate. The authors detail how the process revealed important biological areas that had been overlooked or inadequately addressed in the simulation and areas that could act as a starting point for future experimental or simulation work (Polack *et al.*, 2011). The former is an important consideration in scrutinising results produced by the simulator. For cases in some applications, there may not be the evidence to support a particular requirement. This may be due to incomplete biological knowledge. However, it is in these such cases where the use of this technique is advantageous, as a structure is provided that highlights these areas, showing where assumptions have had to be made, where further biological exploration is necessary, and where results could be affected by these factors. Thus, not only is a tool for providing a structured assessment of an ongoing simulation development, but it may potentially feed future explorations that verify a

simulation result.

Excluding the work cited previously (Ghetiu *et al.*, 2009; Polack *et al.*, 2011), the use and release of an argumentation strategy to support both a simulation and results from it, to build confidence in the simulation as a scientific instrument, is rare. The justification of each part of the simulator opens the design and parameterisation decisions up for scientific scrutiny and should be encouraged. However it should be noted that there is not currently a tool available to support such systematic documentation of simulation-based research (Polack *et al.*, 2011).

1.5.3 Ensuring a Simulation Result is Representative

For simulations that are implemented using an agent-based approach, the effect inherent stochasticity introduced by the implementation has on simulation response must be appreciated. Agent-based approaches are suitable for capturing inherent stochasticity in the biological system as each agent is its own entity, possessing a state and characteristics (Forrest and Beauchemin, 2007), as would a biological factor such as a cell. Behaviour of simulated biological factors may be dictated by the use of pseudo-random number generation. For example, the simulator may randomly choose a probability that a stable bind occurs when two cells come into contact. As behaviour is implemented in this manner, the simulation will generate different sets of responses for the same input parameter set. This implementation issue introduces uncertainty, termed *aleatory uncertainty*, that must be considered when establishing confidence in simulation response (Helton, 2008). If the simulation is to be used to perform *in silico* experimentation, it is critical to ensure that the result is representative of the condition on which the simulation was run, and not an effect of stochasticity in the implementation.

Previously published studies that utilise an agent-based approach have addressed this by performing a number of replicate simulation executions and taking a mean or median of the set of results produced. A recent model of lymphocyte priming in the lymph node utilised 10 executions (Linderman *et al.*, 2011), while conclusions have been drawn from a simulation of tuberculosis by performing either 4, 10, or 15 executions dependent on the analysis being performed (Ray *et al.*, 2009). However no statistical measure is provided that suggests this number of runs provides a sufficient level of accuracy.

Recent studies have established a statistical technique that can be used to mitigate the effect of aleatory uncertainty on simulation results (Read *et al.*, 2012). This establishes the number of simulation executions required to achieve a desired level of experimental accuracy while considering the computational resources available. The analysis operates by contrasting distributions of simulation responses, all generated using one fixed set of parameter values and containing an identical number of simulation samples. The number of samples required to obtain statistically consistent distributions is established by varying the number of samples within each distribution. As
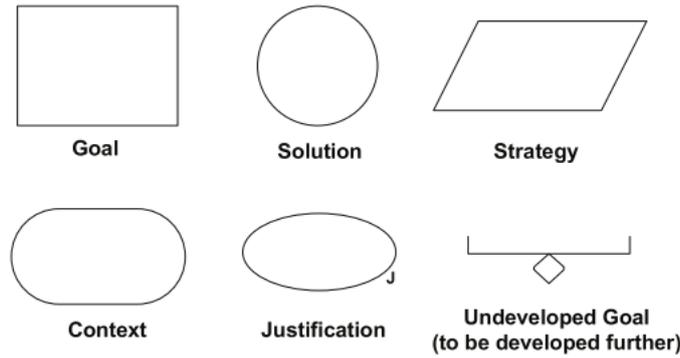
(a) Goal-Structuring Notation. Figure taken from Ghetiu et al (2010)
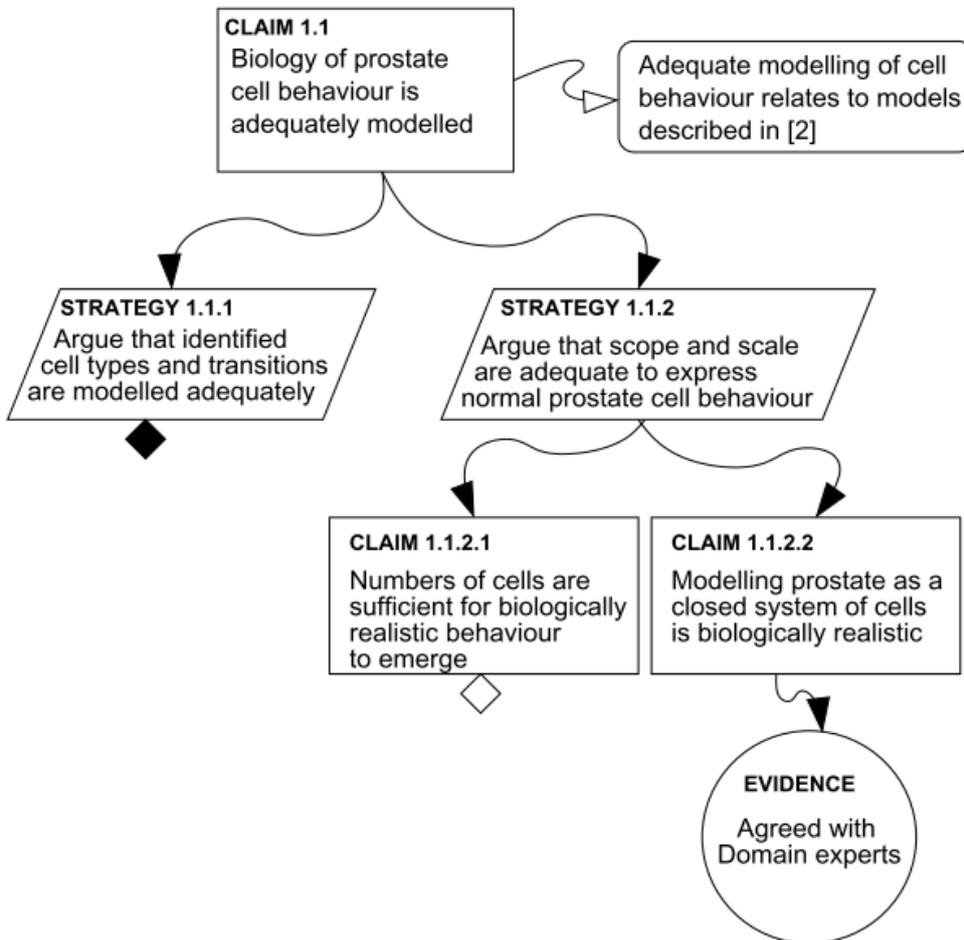
(b) A basic example of the construction of an ABV flowchart, taken from Polack et al's (2011) study on validating a model of prostate cancer cell behaviour.

Figure 1.7: Use of argument-based validation techniques to evidence how a requirement of a system has been met. Top: Goal Structuring notation used to develop the flowchart. Bottom: A basic example of the use of ABV from Polack et al (2011). Note that filled diamonds denote that the argument has been continued on a further diagram.

the sample size increases, the likelihood of producing identical distributions increases, reducing the effect of implementation specific stochasticity.

This technique was developed alongside and exemplified with the authors agent-based simulation of EAE (Read, 2011). In their case study that exemplifies the tech-

nique, 20 such distributions are generated, and the effect of stochasticity on sample sizes of 5, 10, 50, 100, 200, 350, 500, and 1000 simulation runs examined. The contrast between distributions is assessed using the Vargha-Delaney A-Test (Vargha and Delaney, 2000), a non-parametric effect magnitude test that establishes scientific significance by examining the probability that a randomly selected sample from one distribution will be larger than a randomly selected sample from the other. A result above 0.71 or below 0.29 indicates a scientifically significant difference between the populations, with 0.5 indicating no difference (Vargha and Delaney, 2000). The objective is to establish the number of samples required to reduce the effect to that deemed less than small by the Vargha-Delaney A-Test. In the case of the stated case study, this was achieved for a sample size of 500, and thus 500 simulation executions were performed for each *in silico* exploration that was performed.

The contrast between the number of executions performed in the lymphocyte priming and tuberculosis studies (Linderman *et al.*, 2011; Ray *et al.*, 2009) with that in the EAE simulation (Read, 2011) is huge, and goes someway to showing why this technique is necessary. The implementation of this technique reduces the inherent stochasticity in the simulation to a level near that observed in the biological system. Possessing a statistical measure detailing why a particular number of samples was chosen provides a strong argument that the measure generated from the distribution of results (mean or median, dependent on application) is a representation of the condition on which the simulation was run.

1.5.4 Sensitivity Analysis Techniques

A sensitivity analysis (SA) is the application of statistical techniques to examine how a system responds to an alteration in input parameter values. Through varying parameter values and analysing the resultant effect on simulation response, the parameters that have an influential effect on the simulations high-level behaviour can be identified and, for some techniques, quantified. Application of SA techniques provides a powerful tool for understanding simulation behaviour and how this affects results, aiding the scrutiny of results that are produced. Possessing this understanding helps establish the relationship between the simulation and the real-world system, to determine if a result is an affect of simulation parameterisation or a true reflection of the underlying biology.

One of the attractions of using simulation is to produce results that can further the understanding of a biological system. Sensitivity analysis techniques provide a means of performing this exploration: where a parameter is found to be highly influential, this could suggest that the biological mechanism(s) it captures are important in the behaviour of that system. Such results could then be verified using current laboratory techniques where possible.

However to retain confidence in such conclusions, uncertainty in the value of simu-

lation parameters, termed epistemic uncertainty (Helton, 2008), must be appreciated. Section 1.5.1 detailed how a process of calibration can be used to determine a set of parameters for which the simulation produces the expected result, but gives no measure of how the value assigned to each parameter influences simulation result. This measure can be provided through use of SA techniques. Where SA reveals a parameter value alteration has no influential effect on simulation response, the accuracy of the simulation is not impacted by the value set to that parameter (Read *et al.*, 2012). However if the analysis reveals that the parameter is influential, where changes in parameter value do cause a significant change in simulation response, the analysis is highlighting both the influence of and uncertainty in the value of that parameter, which must be taken into account when results are scrutinised. If uncertainty in these values were to be addressed, the simulation would produce much stronger predictions (Marino *et al.*, 2008; Read, 2011).

This section continues by examining two types of sensitivity analysis technique that provide an understanding of simulation behaviour, in the context of establishing confidence in simulation results. The first examines how robust the simulation is to a perturbation of an individual input, and are termed one-at-a-time SA techniques, whereas the second set of techniques perturb the values of two or more parameters simultaneously, to identify any compound effects between parameters.

Simulation Robustness to Parameter Perturbation

A simulators robustness to a perturbation in parameter value can be determined using a one-at-a-time approach (Read *et al.*, 2012). This technique alters the value of a single parameter at a time, with the complementary parameter set remaining at their baseline values. The simulation response under that condition is contrasted with responses generated using baseline parameter values. Where a simulation response is found to be sensitive to that parameter value, caution should be exercised when results are interpreted, as these may be artefacts of parametrisation rather than representations of the biology (Helton, 2008). Where no effect is observed, the analysis suggests that the parameter value change has little impact on simulation response. Studies that examine T-cell motility in lymphoid tissue have applied this technique with that in mind: to suggest that the simulation is robust to changes in parameter value (Zheng *et al.*, 2008). However, studies of influenza utilise the technique for the opposite effect, to examine parameter values either side of baseline values and suggest the influential parameters (Beauchemin *et al.*, 2005).

Neither of the examples stated follow a formal procedure for quantifying the effect a change in parameter value has had on simulation response. This has been addressed through the development of a technique that determines if changing a parameter value leads to a scientifically significant behavioural alteration in contrast to the baseline simulation (Read *et al.*, 2012). For each simulation parameter being explored, a range

of potential values it could take is established, and simulations performed where that parameter is assigned a value within that range. Simulation responses under those conditions are contrasted with results generated under baseline conditions using the Vargha-Delaney A-Test (Vargha and Delaney, 2000) described in the previous section. This provides a statistical measure of the effect that the change in parameter value has caused, and identifies the points in the range of values explored at which the parameter perturbation results in significant changes in simulation behaviour. Where sufficient data is available, confidence in the validity of these results can be gauged by contrasting this information with biologically accepted ranges of values. Where this is not the case, the sensitive parameters are identified, aiding the understanding of the behaviour of the simulator and potentially the underlying biological system (Read *et al.*, 2012).

Gaining Insight using Global Sensitivity Analysis Techniques

Although robustness analysis suggests the effect of perturbing single parameters, it cannot reveal compound effects that occur when two or more are adjusted simultaneously. The effect one parameter has may rely on the value assigned to another. Global sensitivity analyses perturb all parameters of interest simultaneously, to reveal whether there is a correlation between simulation response and the value assigned to a particular parameter (Saltelli *et al.*, 2000). Where this is the case, parameters that could be coupled and that have the greatest influence on simulation responses are highlighted. If the relationship between the simulation results and underlying biological foundation is strong, such conclusions can then be used to suggest the influential biological factors. Such an analysis has been applied to determine the contribution of cytokine TNF-α to an immune response against tuberculosis (Ray *et al.*, 2009), where the influential biological mechanisms have been suggested by establishing the simulation parameters which reverse the immune response.

The aforementioned tuberculosis study varies the value of a range of parameters of interest and calculates a Partial Rank Correlation Coefficient (PRCC), a statistical measure of correlation between parameter input value and response (Ray *et al.*, 2009). The result produced therefore has some link to the parameter value sampling procedure used. Recent work has taken this a stage further and considered the importance of parameter value sampling in the application of global sensitivity analysis methods (Read *et al.*, 2012). The authors technique utilises a latin-hypercube design (McKay *et al.*, 1979) that aims to select sets of parameter values that cover the entire parameter space, while ensuring no correlation is formed between the value sets that are generated (Saltelli *et al.*, 2000). This rigid approach to the selection of parameter values ensures that there is no sampling bias that could affect the PRCC that is calculated from simulation results generated from the parameter sets.

Additional studies have also examined the use of a different sampling and analysis technique, the extended Fourier Amplitude Sampling Test (eFAST) (Marino *et al.*,

2008; Saltelli, 2004; Saltelli and Bollardo, 1998), which uses fourier frequencies to generate parameter value samples. For each parameter of interest, values are chosen through use of a sinusoidal curve of a particular frequency through the parameter space, with a number of values chosen from points on the curve. The chosen frequency can then be utilised in calculation of a statistical measure revealing the proportion of variance that can be explained by perturbing the value of each factor, which suggests how sensitive the simulation and biological system is to that parameter. This technique has been applied for a number of ODE based models (King and Perera, 2007; Marino *et al.*, 2008; Zhao and Tiede, 2011), yet has not found application in statistical analysis of agent-based implementations. The combination between the inherent stochasticity of such models and the complexity of this technique may make such an analysis intractable, as the analysis requires a large number of parameter value sets. For stochastic systems, where uncertainty is addressed by performing replicate runs, this may lead to an intractable number of simulation runs, especially if the number of parameters is high (Tarantola *et al.*, 2006). There is a balance between the insight this technique provides and the computational resources available. Where the number of parameters is high, the latin-hypercube technique described in the previous paragraph is more appropriate (Read *et al.*, 2012).

Application of Sensitivity Analysis Techniques

The use of sensitivity analysis techniques may be well established in other fields (Saltelli *et al.*, 2000) but it is only recently that these techniques have found application in simulations of biological systems (Marino *et al.*, 2008; Ray *et al.*, 2009; Read, 2011; Read *et al.*, 2012). Where examples do exist, these tend to detail an application of the techniques to ODE rather than stochastic agent-based models (Read *et al.*, 2012). Although the added complexity of agent-based models means any analysis would require considerable computational resources (for reasons discussed in section 1.5.3), the insight that can be gained using the techniques, in terms of confidence in simulation result, makes the analysis worthwhile.

It was noted in section 1.4.2 that an increase in the potential and use of computational models has coincided with the development of a number of packages that aim to assist with simulation development. Although a number of sensitivity analysis techniques have been described, there currently exists no generic package for determining how representative a simulation is of its biological system and understanding how *in silico* results can be interpreted in the context of the biological domain. It could be suggested that if a package incorporating a number of techniques such as those described above was developed, a full sensitivity analysis of a simulation would be eased and thus encouraged. Increased use and exposure may then make this analysis an essential part of any study that uses simulation, for both ODE and agent-based implementations.

1.5.5 Verifying *in silico* Experimentation Results

It could be suggested that a straightforward comparison between a simulation result and a result from published literature or laboratory experiment provides a strong indication of the simulators reliability as an experimental tool. It is understandable to think that if the simulator replicates the expected result, the system dynamics have been captured. In his analysis, Read (2011) has a negative opinion of this technique, applied in work by Linderman (2011) among others, suggesting that such comparisons are open to accusations of being 'cherry-picked' to support the intended result, unless such experimentation was detailed prior to the simulators implementation. However, this is open to interpretation. For example, in a model that captures cell behaviour leading to an observed physical property, such as an aggregation of cells, the focus is on correctly capturing cell behaviour rather than the emergent property. Thus it would be sensible to compare this emergent behaviour to a number of previously observed examples. The argument is also dependent on whether the comparison is quantitative or qualitative, especially in cases where the comparison is visual rather than statistical. It is therefore difficult to rule this out as a method of verification: it is dependent on the model and the type of results being used in the comparison.

Instead Read (2011) supports a stronger approach noted as best practice, where simulation predictions are verified in the laboratory (Bauer *et al.*, 2009). Although this is definitely a strong method and one which closes the loop in terms of simulation-derived laboratory experimentation, it requires a high level of confidence in that simulation result. However the author does note that this has drawbacks: laboratory investigations are expensive, time-consuming, require specialist staff, and may require the use of animals and thus ethical certification, and thus should be performed as a final step once confidence in the simulation result is assured. However, there are successful examples where simulation predictions have been supported by laboratory investigations (Efroni et al (2007) being one), suggesting there is scope if computational modelling is integrated with laboratory experimentation from the initial design stages.

1.5.6 Simulation Availability

Performing all the steps above may make a convincing argument that a good level of confidence in the simulator as a representative tool is established. However, one additional and elementary step may aid the development of confidence in simulation as an experimental tool. Although the use of modelling and simulation is becoming increasingly popular, and results published alongside biological experimentation in high-impact journals (Kleinstein, 2008), it is rare to have access to the simulators that have been generated. Granting access to the field would provide experimental immunologists with the opportunity to examine the tool themselves, to both provide input into future iterations of the simulator and use the simulator to inform their future

investigations. This is an important consideration if this approach is to become widely accepted and confidence grow in the use of computational modelling as a technique.

1.6   Thesis Overview

This thesis details the development and use of statistical and simulation tools that enable a computational exploration of lymphoid organ development. This is driven by the following aims:

1. Development of a robust computational model that replicates emergent behaviour observed *ex vivo* and *in vivo*.

2. Development of a statistical toolkit to help determine the relationship between simulation results and the biological system.

3. An exploration of the biological factors influencing behaviours that emerge through interactions between biological factors.

4. Application of the simulation and statistical toolkit to perform novel *in silico* experimentation.

Aim 1: Development of a robust computational model that replicates emergent behaviour observed *ex vivo* and *in vivo*

Previously described examples demonstrate that computational models can provide useful insight in the continued exploration of biological systems. To date, this technique has not been utilised in furthering the understanding of immune system development. Previously published investigations have led to the generation of a basic model of secondary lymphoid formation (Section 1.3.2). This model will be used in the development on a computer simulation of Peyer's Patch development. Available biological data and published results will then be utilised to ascertain if the simulation is an adequate representation of the development of these lymphoid organs. Transparency in the design and implementation of this simulation will be a key aspect of this study, ensuring that the strengths and limitations of the simulator are understood.

Aim 2: Development of a statistical toolkit to help determine the relationship between simulation results and the biological system

This chapter has detailed the availability of a number of simulation tools and frameworks that can aid simulation development. However the simulation is developed, it is important to establish the link between the results and the real world system, to indicate whether the result is providing biological insight. Although a number of uncertainty and sensitivity analysis techniques that can help establish this link have been

described (Read *et al.*, 2012; Saltelli *et al.*, 2000), a comprehensive statistical package for the analysis of simulation results does not yet exist. This may explain why some investigators make little attempt to elucidate how representative a simulation result is (Read *et al.*, 2012). The development of a package of statistical techniques that can be applied to any simulation has the potential to address this, and thus increase overall confidence in the use of simulation as an exploratory tool.

Aim 3: An exploration of the biological factors influencing behaviours that emerge through interactions between biological factors

Using the developed simulation and statistical toolkit, a quantitative analysis of the influence of each biological factor in the behaviour that is observed can be undertaken. This provides insight that can not currently be revealed using current laboratory approaches. This analysis focuses on two emergent aspects of the system: the formation of aggregations of cells that become Peyer's Patches and hematopoietic cell behaviour observed *ex vivo* (Patel *et al.*, 2012). This furthers the understanding of the role of each biological factor in influencing that emergent behaviour.

Aim 4: Application of the simulation and statistical toolkit to perform novel *in silico* experimentation

Whereas the above aims to explore the contribution of biological factors in behaviours that are observed, computer simulation can be used to perform *in silico* experimentation. Such experimentation may be used to either support or inform current laboratory investigations where possible, or to support the generation of hypotheses which are difficult to examine in a laboratory. Two sets of *in silico* experiments are conducted: an *in silico* replication of previously published laboratory investigations, and novel *in silico* investigations that aim to address areas of biological understanding that remain incomplete.

1.6.1 Thesis Structure

This thesis addresses these aims in six chapters, organised as follows:

Chapter 2 describes the methodologies used and the creation of the tools required to perform an *in silico* analysis of lymphoid tissue formation. The chapter describes the creation of a simulation of lymphoid tissue formation, including how confidence is gained in the use of simulation as an experimental tool through calibration and validation. Next, a package of statistical tools created to analyse simulation results, to establish their relationship to the biological system, is described. Methods used to generate *in silico* experimentation results from the simulator are also detailed.

Chapter 3 utilises the simulation and statistical tools developed in Chapter 2 to examine key factors that influence hematopoietic cell behaviour during hour 12 and 72 of PP development, in conjunction with on-going *ex vivo* culture system investigations.

Chapter 4 examines the use of simulation as a tool for performing *in silico* experimentation, replicating previously performed laboratory experimentation and performing novel investigations. The latter half of the chapter utilises the statistical toolkit created in Chapter 2 to suggest the key factors that influence Peyer's Patch characteristics (e.g. size).

Chapter 5 utilises the simulator to perform a time-lapse analysis of the development process, identifying the stages of development when different biological factors may be influential. This aims to determine if PP development occurs in distinct phases.

Chapter 6 provides a critical review of the work that has been conducted in relation to the project aims identified above.

Chapter 2

Methods and Tool Development

2.1 Introduction

This chapter details the methods used in addressing the thesis objectives specified in the previous chapter. For the thesis objectives to be met, suitable tools needed to be developed. This chapter not only describes the methods that have been used in the explorations of lymphoid tissue formation that follow, but also describe the development of the tools needed to perform these explorations.

The chapter begins by describing the development of a model and simulation that replicates Peyer's Patch development. In some respects this can be thought of as an application of the CoSMoS process (Andrews *et al.*, 2010) that was described in Chapter 1. A series of models were created that describe gut-associated lymphoid tissue development in the mouse. Developed in close collaboration with experimental immunologists, these models detail the biological information that has been encapsulated, a justification on any abstractions and assumptions that were made, and a specification of how this information was encoded within a computer implementation. From these models, a simulation of the process has been implemented that captures the abstraction of the biological system specified, creating a tool through which the explorations in the following chapters can be performed. This chapter details the techniques used to develop, calibrate, and validate this tool, ensuring it is fit for the purpose of this study.

One of the motivations for implementing a computer simulation of a biological process is that is enables *in silico* experimentation to be performed. Using a well designed simulator, this has the potential to explain any underlying data on which a simulation has been constructed (Guo and Tay, 2005) and provide novel biological insight by facilitating experimentation that is impractical or impossible to perform using current laboratory methods (Andrews *et al.*, 2010; Efroni *et al.*, 2003). Many statistical methods to aid analysis of simulation results have been described (Marino *et al.*, 2008; Read *et al.*, 2012; Saltelli, 2004; Saltelli *et al.*, 2000), yet there is currently no comprehensive statistical package available that aids the analysis of simulation results. Thus a statis-

tical toolkit, *spartan*, has been developed to provide a comprehensive set of tools for understanding both uncertainty in simulation results and the relationship between the simulation and the real-world system. Each technique included within this toolkit is detailed in this chapter. The analysis tool developed has been utilised in explorations of lymphoid tissue formation in the chapters that follow.

With these tools in place, explorations that address the objectives in this thesis can be performed. The final sections of this chapter detail the methods used to perform the investigations in the following chapters.

2.2 Pairing Current Experimental Techniques with Modelling and Simulation

2.2.1 Methodology

The methodology involved in modelling and simulating lymphoid tissue development utilised the principled approach specified in the CoSMoS Framework (Andrews *et al.*, 2010), described in section 1.4.2. In this process, the biological system being explored is termed the *domain* of interest. Understanding of the functional process is captured in a series of models: *domain*, *platform*, *simulation*, and *results*. Each of these models is considered in turn below:

2.2.2 Domain Model

Overview

The initial stage is the creation of a domain model that encapsulates the current scientific understanding of the biological system, scoping both what is to be modelled and the research question the model is to address. This is a model of the biological model, one completely isolated from any thoughts on how the biological system could be translated into computer code (a simulation). This is important as at this point the details of how the simulation is to be implemented are not of concern, and may distract from the specific modelling of the biological system. The domain model may be a diagrammatic representation, detailing biological factors including cell types, factors that influence cell behaviour (e.g. chemokines, adhesion factors) and a description of the environment in which interactions between these factors take place (e.g. the foetal intestinal tract). The model will also detail any emergent properties that are observed through interactions between biological factors. Such behaviours could include the formation of cell aggregations or changes in cell behaviour.

The CoSMoS Framework (Andrews *et al.*, 2010) stresses the importance of collaboration with an expert in that biological system when generating the domain model. This aids interpreting results in the literature, with a view to including these in the

model, and potentially permits the inclusion of results from any ongoing laboratory investigations the collaborator may be involved in. In the scoping of the biological features that are to be included, many areas will be revealed where the current biological understanding is incomplete. Collaboration in domain model generation allows for suitable assumptions and abstractions to be made that have biological justification. These are documented for sake of scientific transparency.

This section presents a domain model of Peyer's Patch development that has been developed based on the literature described in Chapter 1 and in collaboration with experimental immunologists, a first stage in the development of a tool for exploring lymphoid tissue development. Through undertaking this process, a full exploration of the domain (foetal PP development) has been performed and captured using a number of diagrammatical techniques. Areas where the underlying biological understanding is incomplete have been identified and suitable assumptions made and documented where necessary.

A System-Level Overview

Figure 2.1 captures a delineation of both cell behaviour observed in *ex vivo* culture results described in section 1.3.3 and results from published literature, using a diagrammatical technique exemplified by Read et al (2009). Although this diagram has no formal schematic, it is useful in identifying the factors in the biological domain that are to be included in the domain model, the behaviours that become apparent upon interaction between them, and how these behaviours lead to the system behaviour observed.

The top of Figure 2.1 details the high level phenomena that has been observed in the system, either experimentally or detailed in the literature. The dotted line separates the observed phenomena from hypotheses that are believed to be responsible for their occurrence. It is these behaviours that will be captured in the series of models that follow. Where these hypotheses could explain an observed phenomenon, a connection is made between the two. Where there is still uncertainty in what causes the observed phenomenon, no hypothesis is included on the diagram. Connected to these hypotheses are the entities in the biological domain that are believed to be responsible for the behaviour that emerges, and are thus included in the domain model. Each is in turn connected to other entities that it may interact with to produce that hypothesis.

Observations from the *ex vivo* culture system and experimental results in the literature has led to the formation of three high level observable phenomena that are detailed on the top of Figure 2.1. The first, a small clustering of hematopoietic cells around a stromal cell after thirteen hours, can be observed in cell tracking images taken in the preceding hour, as seen in Figure 1.3.3. There is no expected behaviour linked to this observation due to the uncertainty in the factors causing this to occur. The second, an alteration in cell velocity and displacement, has become apparent through a statistical

analysis of the cell tracking data over that hour, presented by Patel et al (2012). The authors suggest that this is caused by interactions between the cells leading to an expression of adhesion factors. Thus this hypothesis is linked to LTin and LTi cells. The final observable phenomenon are the large clusters of hematopoietic cells around LTo cells that are visible along the length of the foetal mouse gut at E17.5. It is widely accepted in the literature that this emerges through LTi cell response to chemokine expression (Cyster, 1999; Luther *et al.*, 2003; van de Pavert and Mebius, 2010; Randall *et al.*, 2008). This hypothesis is in turn linked to the LTi cell, which both triggers LTo cell differentiation that upregulates expression of chemokines and expresses the receptors to respond to this expression (Luther *et al.*, 2003; Randall *et al.*, 2008).

Through a process of delineating the system, it is possible to make simplifications to make the generation of a model tractable, while ensuring the expected behaviours are reproduced. It is intractable to represent all aspects of the biological system in a model, and thus a subset are included that are sufficient to produce the phenomenon that is observed (Read, 2011). In this instance, Figure 2.1 denotes the involvement of three chemokines (CXCL13, CCL19, CCL21) and three adhesion factors (VCAM-1, ICAM-1, MAdCAM). A full quantification of the expression levels of each of these molecules has not yet been possible. As this is the case, a simplification has been made in the initial model where these are considered as one adhesion factor and one chemoattractant factor respectively. It can also be noted that IL-7R signalling is not included on the diagram. It was noted in section 1.3.2 that IL-7R signalling is thought to be responsible for causing the expression of LT$\alpha\beta$ by LTi cells, through a process that is not currently understood. The abstraction in Figure 2.1 makes the assumption that IL-7R triggering occurs and hematopoietic cells express LT$\alpha\beta$, and thus have the capability to interact with stromal cells as observed *ex vivo* and *in vivo* . The focus is on capturing the interactions that lead to the higher-order behaviour, and thus explicitly modelling the upregulation of LT$\alpha\beta$ is not necessary. The documentation of cell level assumptions and abstractions such as this is detailed in the following section.

Capturing Cell-Level Dynamics

Each of the key cell types identified in the domain is represented explicitly in the domain model. For each cell type, states (observed behaviours or gene expression profile) that the cell might exist in and the interaction(s) that must take place for that cell to change state are examined. Such descriptions are documented through the use of State Diagrams, a documentation method closely related to that included within Unified Modelling Language (UML) (Rumbaugh *et al.*, 2005). UML is a notation widely used in software engineering that has also found application in the specification of models of biological systems (Bersini and Carneiro, 2006; Read *et al.*, 2009). Through creating the domain model, biological parameters are identified and recorded. Some of these parameters have known values that have been determined experimentally or noted in

the literature, whereas the values of others are currently unknown. Where this is the case, suitable assumptions have to be made and clearly justified. This section details the dynamics for each cell type identified in the system delineation step and in the literature at an individual level.

1. Hematopoietic Cells (LTin/LTi Cells)

   Figure 2.2 depicts the domain model state machine diagrams for the hematopoietic cell populations CD4$^-$CD3$^-$IL-7Rα^-CD11b$^+$CD11c$^+$, or LTin cells, and CD4$^+$CD3$^-$IL-7Rα^+c-kit$^+$, or LTi cells, involved in Peyer's Patch development.

   Although two distinct populations, both cell types are believed to migrate from the foetal liver and can be detected in the foetal mouse intestine from E14.5 (Mebius *et al.*, 2001). Experimental work by Veiga-Fernandes et al (2007) has revealed that cell movement initially follows that of a random walk. Thus after migration, cells exist within the first state, moving randomly across the intestinal tract. This movement is at a speed within a range also determined by the authors. Transition out of this state differs dependent on cell type:

   (a) Lymphoid Tissue Initiator (LTin) Cells

   Investigations by Veiga-Fernandes et al (2007) also revealed the importance of RET in triggering lymphoid tissue formation, a receptor expressed by LTin cells. Upon contact with an LTo cell, RET binds to an appropriate ligand expressed by the LTo cell, thought to be ARTN, stimulating LT$\alpha\beta$ expression on the LTin cell. Cell contact has led to changes in the cell characteristics, thus the change in state as noted in Figure 2.2a. Cell movement from this point is then influenced by the level of localised expression of adhesion factors. Should the level of expression of these factors be sufficient, it is assumed that LTin cell movement is localised around the LTo cell, as revealed in *ex vivo* observations by Patel et al (2012) described in section 1.3.3. Where the level is insufficient, the cell moves away, returning to a state where its movement is random.

   (b) Lymphoid Tissue Inducer (LTi) Cells

   In contrast, LTi cells are thought to express receptors for chemoattractants expressed by differentiated LTo cells. An LTi cell will reside in a state mimicking a random walk until chemotaxis is triggered through localised levels of chemoattractant expression. Where this occurs, LT$\alpha\beta$ receptor expression is upregulated and the cell begins to move towards a primordial patch, thus a change in state is noted in Figure 2.2b. With chemotaxis mediating LTi recruitment, contact between an LTo and LTi cell is promoted. Such

contact leads to LTo cell differentiation through lymphotoxin receptor signalling, and the upregulation of chemokine and adhesion factor expression as detailed in section 1.3.1. A third state has been noted on Figure 2.2b to denote that this contact has occurred. Similarly to LTin cell motility after contact with an LTo cell, cell movement from this point is then influenced by the level of localised expression of adhesion factors. Should the level of expression of these factors be sufficient, cell movement remains localised around an LTo cell. Where the level is in sufficient, an LTi cell may migrate away from a a primordial patch, either returning to a state where the movement is random or being influenced by expression level of chemoattractants in its vicinity.

Through collaboration with experimental immunologists, hypotheses can also be included in the model that have yet to be published or widely accepted, with the aim of judging the affect that hypothesis has on observed emergent behaviour. In this case, collaborators assisting with the design of the model have added such a hypothesis: that LTi cells could potentially cause the differentiation of non-RET ligand expressing LTo cells within the gut surface. Such differentiation occurs through a number of contacts between these cells, deemed immature LTo cells, and LTi cells. Thus an additional state is added to the LTi state diagram in Figure 2.2b, where the LTi cell is in contact with an immature LTo. This contact is assumed to be for a brief period, after which the LTi cell returns to a state where movement is either random or being influenced by chemoattractant expression.

A series of assumptions have been made concerning LTin and LTi cell behaviour, listed in Tables 2.1 and 2.2. For LTin/LTi contact with an LTo cell, the assumption has been made that a stable contact which leads to LTo cell differentiation does not definitely occur when two cells come into contact. This introduces the need to capture binding affinity within the model. It is also assumed that the response to adhesion factor receptor signalling increases as adhesion factor expression in the vicinity is increased, thus making the LTin/LTi cell more likely to remain within a forming PP. However, some stochasticity in cell behaviour must remain, and thus there is a chance that a cell may not respond to adhesion factor expression. With regards to chemokine expression, it has not yet been possible to quantify expression level over development time. The assumption has thus been made that chemokine expression diffuses over distance, getting stronger as distance from the LTo decreases. It is also assumed that the expression does not decrease.

The process of creating a domain model of LTin and LTi cell states and interactions has led to the identification of a number of parameters that capture aspects

of these described characteristics. These parameters, and the value determined from the literature, are detailed in Figure 2.2.

2. Lymphoid Tissue Organiser (LTo) Cells

As detailed in section 1.3.1, LTin and LTi cells aggregate around non-hematopoietic VCAM-1$^+$ICAM-1$^+$ stromal cells, termed LTo cells. An analysis of foetal mice intestines taken at E15.5, using flow cytometry, determined that 20% of cells were stromal. The way these cells are distributed in the epithelium remains unknown. With the agreement of the collaborating immunologists, a normal distribution has been assumed, and initial explorations will assume that 20% of the epithelium is comprised of stromal cells. This observation raised interesting questions prior to creation of the domain model, concerning how the number of aggregations (patches) that form is limited if this were the case. If this is correct, and there were no limiting factor, it would be expected that a large number of PP would form, rather than the 8-12 previously observed (Figure 1.2). Thus a further assumption is made that only a subset of these stromal cells have the capability to differentiate into VCAM-1$^+$ LTo cells that have the capability to mediate PP development.

Figure 2.4 depicts the domain model state machine diagram for this cell type. In contrast to the hematopoietic cells in the previous section, it can be noted that there are four potential initial states (black circles). Working from left to right, the first two are states that LTo cells may exist within at E14.5, when PP development is thought to commence (Mebius *et al.*, 2001): the first for stromal LTo cells that have the capability to mediate PP development and the second for stromal LTo cells that do not. LTo cells that do have the capability express a ligand for RET. Ligation of the RET ligand upregulates LT$\alpha\beta$ expression on LTin cells and causes LTo cell differentiation, thought to trigger the process of PP development (Veiga-Fernandes *et al.*, 2007). As the RET pathway has this key role in development, it has been assumed that LTo cells that do not have the capability to mediate PP development do not express RET ligand. At present the ratio between the two subsets of LTo cells remains unknown, yet could be established through experimentation using the simulator at a later stage. Should the cell be in the second initial state, where RET ligand is not expressed, there is a theory that these cells could also differentiate under certain environmental conditions, namely a large number of contacts with LTi cells and being located within a set distance from another differentiated LTo cell. The second initial state on the diagram is included to allow this theory to be considered if necessary. The third and fourth initial states in Figure 2.4 are triggered upon LTo cell mitosis. The assumption has been made that LTo cells divide every 12 hours, and the resulting cells have the same attributes as the parent cell. Thus these initial states are required to capture that process.

Once LTo cell differentiation has occurred through the RET signalling pathway, further differentiation and thus a change in state occurs on stable contact with an LTi cell, upregulating the expression of chemokines CXCL13, CCL19, and CCL21. In this state, the cell will also continue to upregulate the expression of adhesion factors on each stable contact with an LTin or LTi cell. The assumption has been made that this continues up to a point where expression of both chemokine and adhesion factors is saturated, at which point the cell enters a final state that has been deemed as mature.

Biological parameters that have been identified in the generation of this model are documented in Figure 2.4, and again assigned values from the literature where possible. Similarly to the LTin and LTi domain models, a number of assumptions have been made concerning LTo cell behaviour and documented in Table 2.3. These mainly concern chemokine and adhesion factor expression by an LTo cell, both of which have yet to be quantified experimentally and thus need to be assumed.

Capturing Cell-Level Interactions

Within the domain model a further UML diagram, an Activity Diagram, is utilised to specify how the cells detailed in the state diagrams interact. It is through these interactions that the observed system dynamics are expected to emerge. This diagram documents the order in which cellular events and interactions take place for this emergent behaviour to be observed. Boxes represent an event that the cell may be undertaking during the process, with arrows representing a flow of actions after the event, that become possible if the condition on the arrow is met. Diamonds represent two or more actions that may be possible, with the resultant flow dependent on meeting the condition on one of the arrows.

Behaviour that has been observed both in *in vivo* imaging (Veiga-Fernandes *et al.*, 2007) and *ex vivo* culture systems (Patel *et al.*, 2012) emerges through interactions between three cell types specified in the state machine diagrams, mediated by the expression of adhesion and chemoattractant factors. The Activity Diagram for this model can be seen in Figure 2.5. This activity diagram captures two forms of emergent behaviour that is observed. The first of these is the statistically significant change in cell motility around a forming patch, as observed in the *ex vivo* culture system (Patel *et al.*, 2012) described in section 1.3.3. The second is the aggregation of cells that is characterised at the end of hour 72 as an primordial Peyer's Patch. Whereas current experimental techniques have provided a quantification of a change in cell behaviour, no current investigations have provided quantitative data on what should be considered as a Peyer's Patch at the end of hour 72, in terms of number of cells or area occupied by such an aggregation. Although this is problematic, the development of such a model can be used to help address this prior to such data becoming available.

Development Period and Spatial Dynamics

This section of the domain model is a specification of the environment in which the interactions described above are taking place, and the time-period over which this occurs. This ensures that the model considers any restrictions that are imposed by environmental effects. Having this specification aids the decision on which modelling methodology to adopt (i.e. ODE or agent-based) when this is considered in the platform model.

From the activity diagram in Figure 2.5, two forms of emergent behaviour were noted: an alteration in cellular behaviour when in the vicinity of an LTo cell and the identification of aggregations of hematopoietic cells along the gut. The process is known to commence at E14.5 with the migration of hematopoietic cells into the intestine (Mebius *et al.*, 2001), with aggregations of hematopoietic cells being visible at E17.5 (Randall *et al.*, 2008). Thus this model captures this 72 hour period. The maturation of LTin and LTi cells that occurs prior to E14.5 (van de Pavert and Mebius, 2010), and PP compartmentalisation that occurs after E17.5 (Hashi *et al.*, 2001) is outside of the scope of this model.

The emergent behaviour has been observed in one of two settings: either *in vivo* imaging of the developing gut (Veiga-Fernandes *et al.*, 2007) or in *ex vivo* culture (Patel *et al.*, 2012). The domain model generated here will focus on capturing the emergent behaviour as it would occur in the developing gut. To accurately represent the foetal intestine in which the cells interact, and capture the dynamic nature of the developing tract, images were taken of the developing mid-gut from twelve mouse embryos, six at E14.5 and six at E15.5, using stereomicroscopy (Zeiss). Measurements of the length and circumference of each were taken using ImageJ (Fiji). Flow cytometry has been used to determine the percentage of cells that are LTin, LTi, and LTo cells, at E15.5. These figures can be used to calculate a representative number of cells present in the gut at that time-point.

A representative size for each cell type has been gathered from the literature (Veiga-Fernandes *et al.*, 2007). A value range within which a cells velocity will exist has been determined by examining the behaviour characteristics of cells in Patel et al's (2012) *ex vivo* culture system that are further than 50μm from a ligand, and thus are less likely to be influenced by adhesion and chemoattractant expression. The velocities of these individual cells can be seen in Figure 2.3, revealing cell velocities can be considered to be normally distributed.
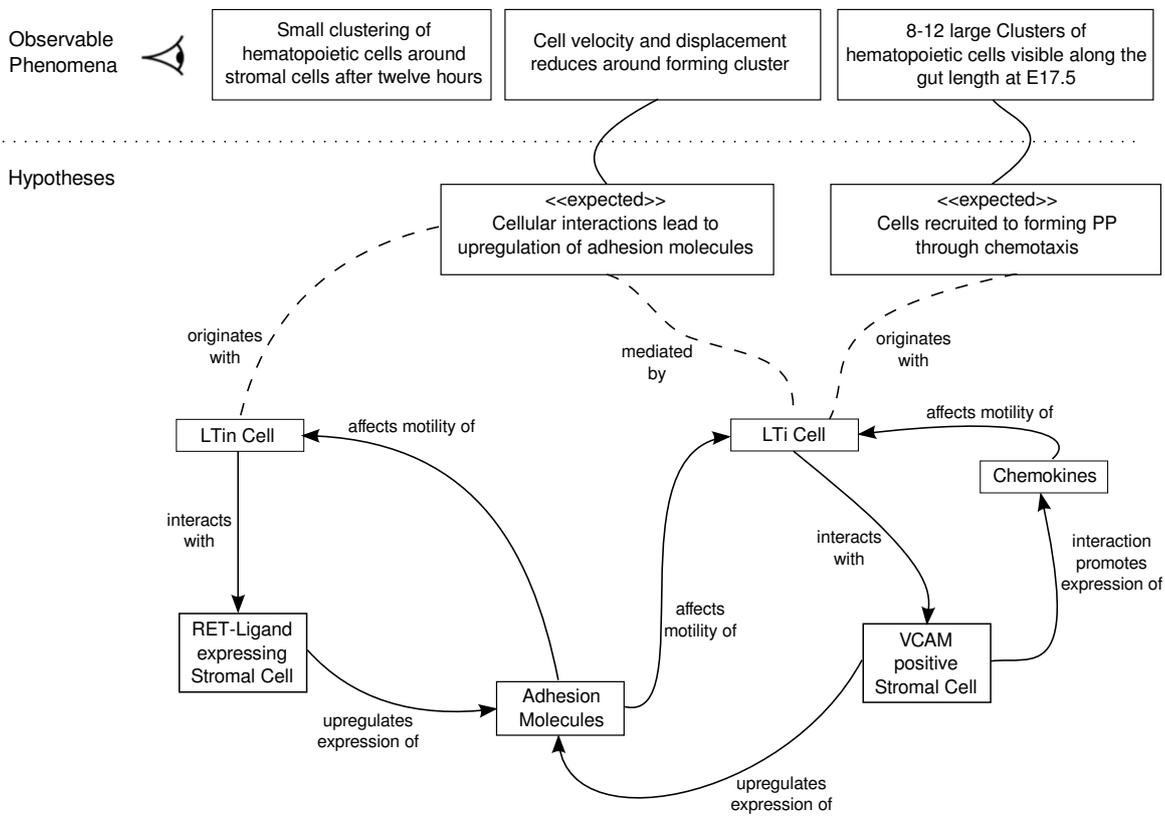
Figure 2.1: Expected Behaviours diagram, detailing the phenomena observed in the domain being modelled, and the behaviours which emerge from interactions thought to be responsible for them
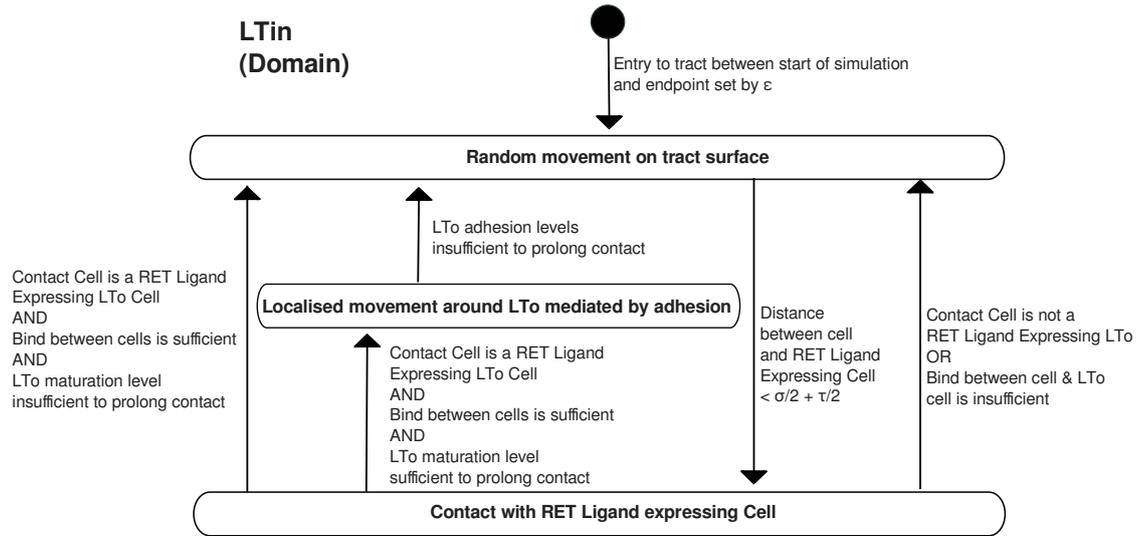
(a) LTin Cell State Diagram: Domain Model
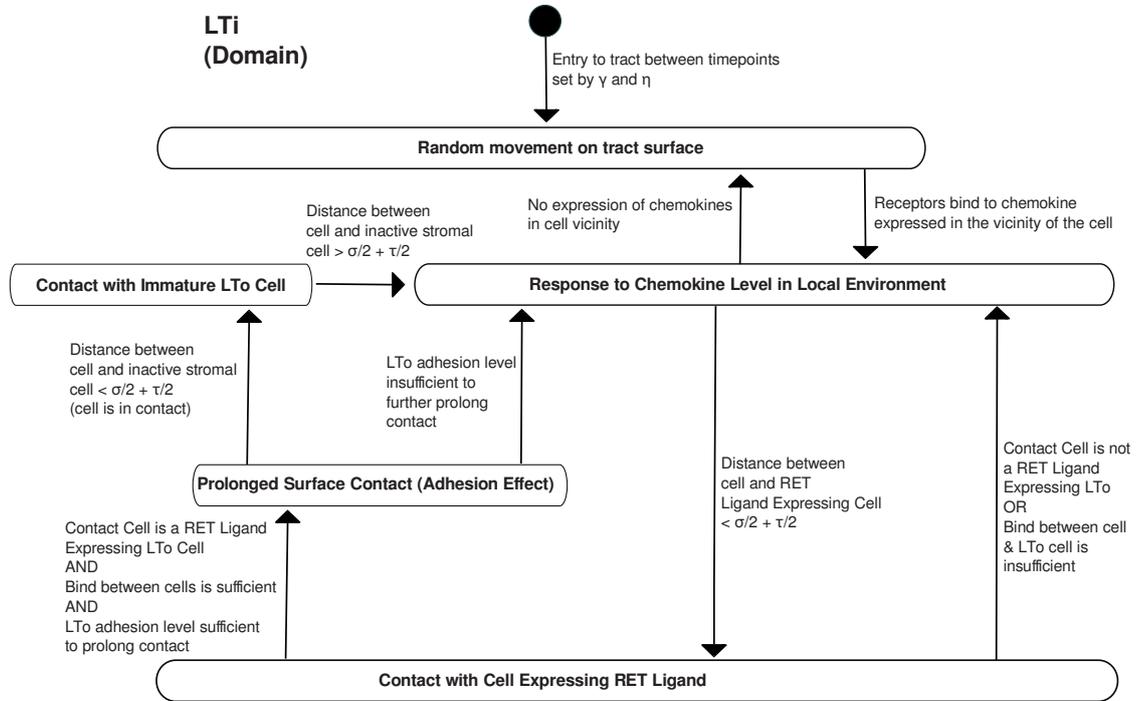
(b) LTi Cell State Diagram: Domain Model

	Parameter	Name in Model	Domain Value
τ	LTin/LTi Cell Size	HCellDiameter	$8\mu m$
σ	LTo Cell Size	LToDiameter	$24\mu m$
ω	LTin/LTi Cell Speed Lower Bound	cellSpeedLowBound	$3.8\mu m/min$
ξ	LTin/LTi Cell Speed Upper Bound	cellSpeedUpBound	$8.8\mu m/min$
ϵ	LTin Input Time	lTinInputTime	72 Hours
γ	LTi Input Delay Time	lTiInputDelayTime	0 Hours
η	LTi Input Time	lTiInputTime	72 Hours

Figure 2.2: Domain Model UML State Machine diagrams for hematopoietic LTin and LTi cells, and biological parameters identified in the creation of the model. All parameter values have been derived from laboratory explorations detailed in Veiga-Fernandes et al (2007) and Patel et al (2012).

Distribution of Cell Velocities Observed in ex vivo Culture System
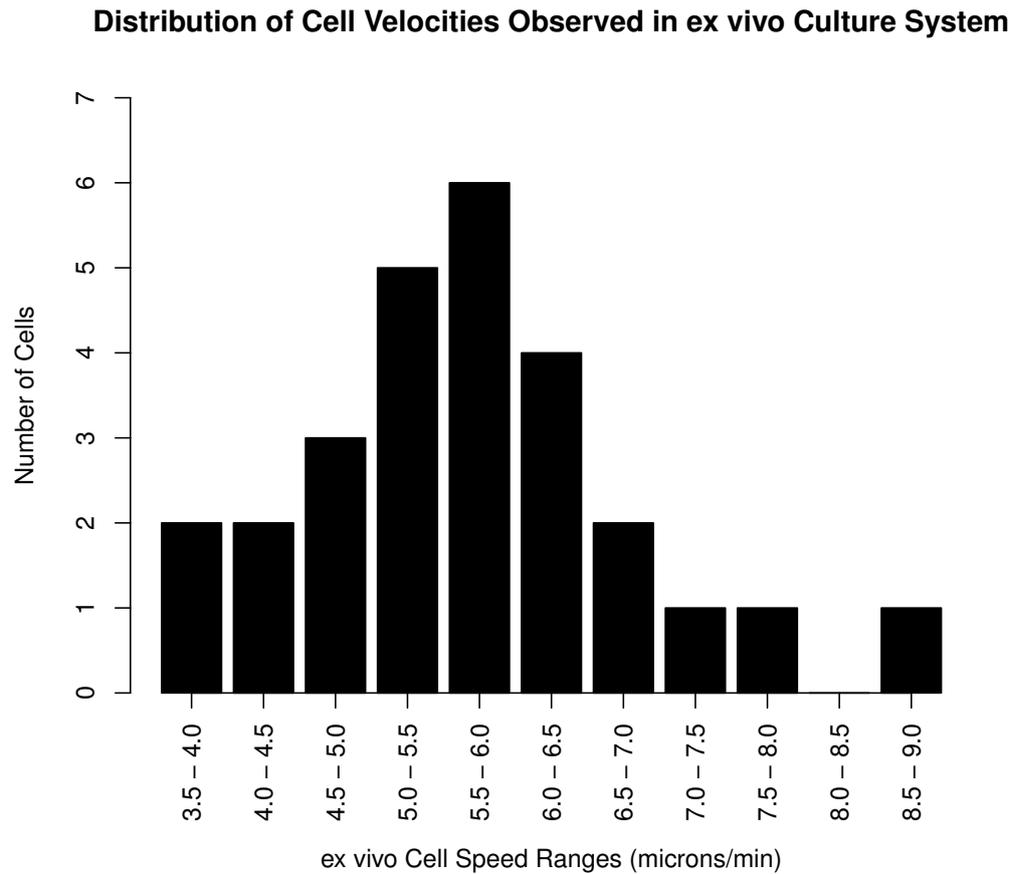
Figure 2.3: A breakdown of the cell velocities that were observed for cells in the *ex vivo* culture system. This has been calculated from tracking cells $> 50\mu$m from the ARTN-soaked bead (see Section 1.3.3), where no chemokine or attractants are influencing cell behaviour. Such individual variation is an important aspect to capture in the domain model.
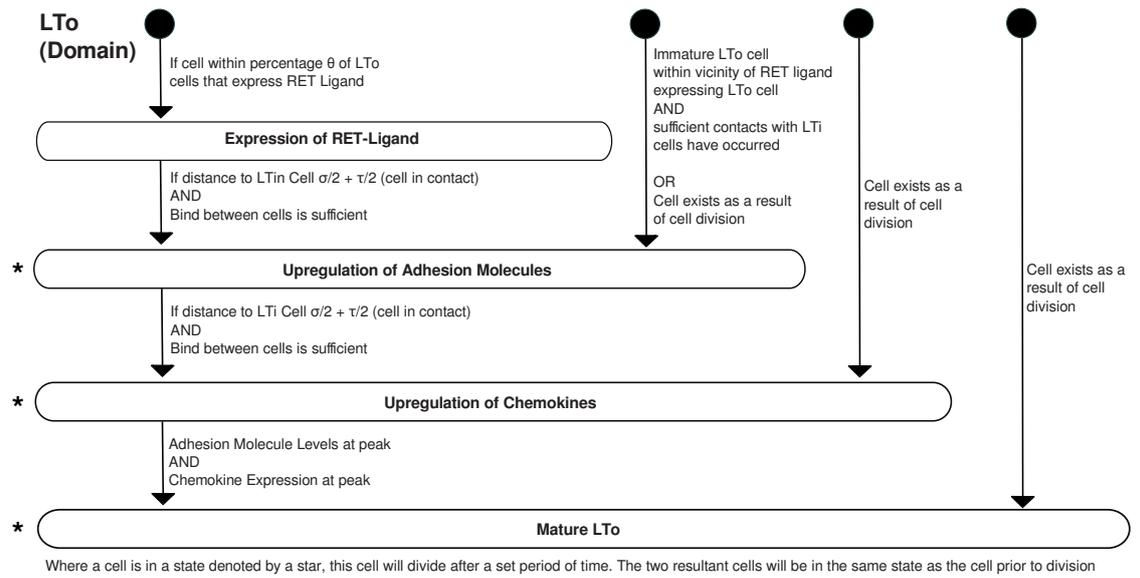
	Parameter	Name in Model	Domain Value
θ	Percentage of LTo cells expressing RET Ligand	percentStromaRETLigands	Unknown
D	Cell Division Time	lToDivisionTime	12 Hours
τ	LTin/LTi Cell Size	HCellDiameter	8μm
σ	LTo Cell Size	LToDiameter	24μm

Figure 2.4: Domain Model UML State Machine diagrams for stromal LTo cell, and biological parameters identified in the creation of this model. All parameter values have been derived from laboratory explorations detailed in Veiga-Fernandes et al (2007) and Patel et al (2012).
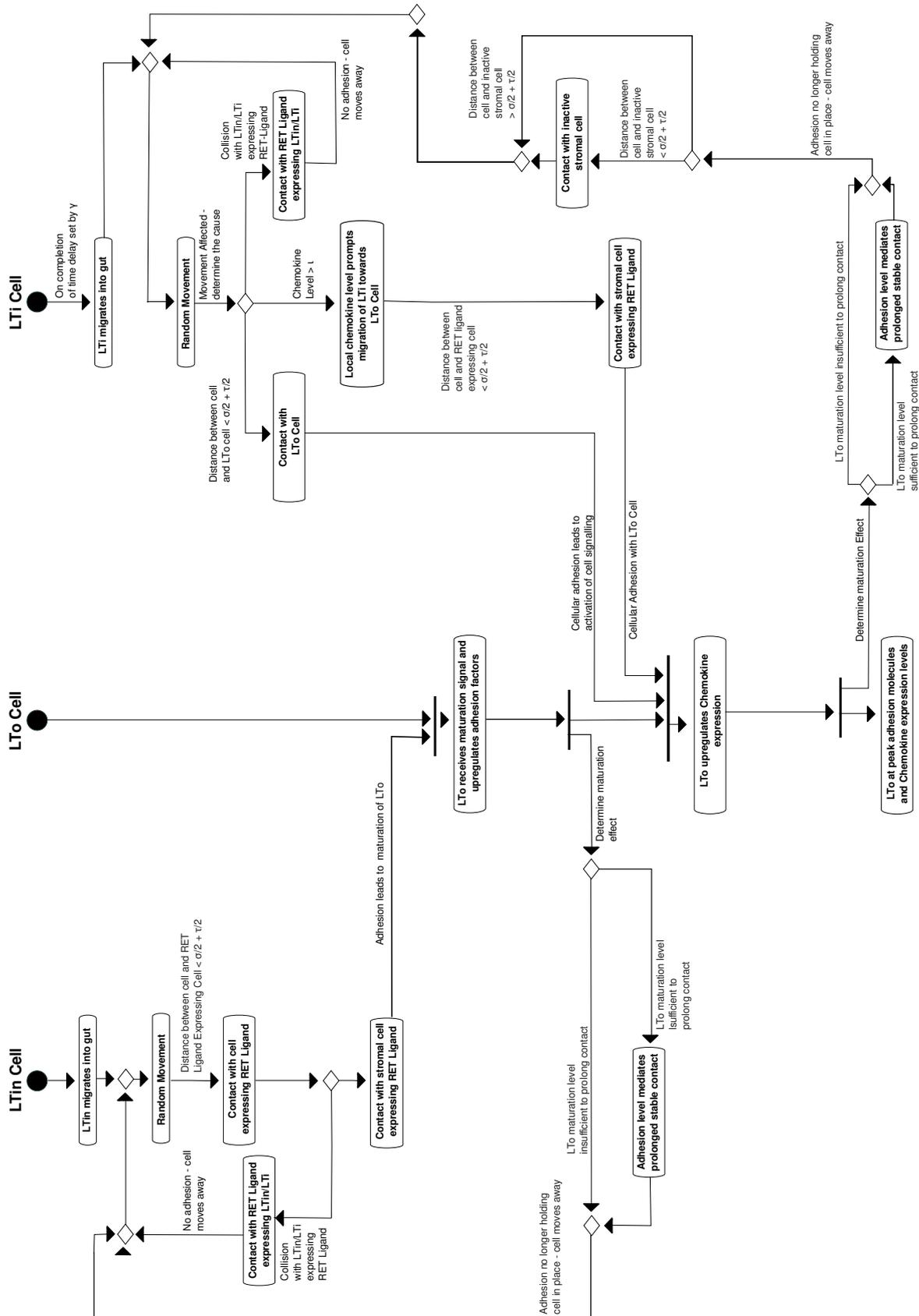
Figure 2.5: Domain Model Activity Diagram detailing the order in which interactions are thought to occur to produce the the emergent behaviour observed. Where parameters are noted on the interactions, these can be found in the parameter tables in Figures 2.2 and 2.4

2.2.3 Platform Model

Overview

The platform model acts as a specification of how the domain model could be encoded as a computer simulation. Individual cells, the environment, and interactions that lead to biological state changes are examined and translated into a form that can be captured in computer code. For example, capturing cell receptor binding affinity could be achieved through use of a probability function. Such translation therefore leads to the identification of platform specific parameters, of which the numerical value may again be unknown. Thus further assumptions are made, based on known biology or collaborator insight where possible, and documented for scrutiny alongside simulation results. Critical to the modelling process, emergent behaviour specified in the domain model is entirely removed from the platform model. Biologically observed behaviour must emerge through interactions between components and not be encoded into the model, as this invalidates the simulator as a predictive experimental tool.

Section 1.4.1 detailed a number of potential techniques available that can be applied in modelling a biological system, including use of ODE's or agent-based approaches. The detail in the domain model can now be utilised to determine which methodology is best suited to addressing the research problem identified. This section details the generation of the platform model specifying how the domain model described in section 2.2.2 could be encoded as a computer simulation.

Choosing the Modelling Methodology

Potential methodologies that can be applied in the modelling of biological systems were discussed in detail in section 1.4.1. In the Peyer's Patch domain model, one of the key emergent behaviours under examination is the change in individual cell behaviour on interaction with other cells or biological factors in the environment. Thus the scope is an individual rather than population level, focusing on the importance of interactions in mediating this change in behaviour. As this is the case, the application of an agent-based model is appropriate. Each cell type (LTin/LTi/LTo in this case) will be created as an agent within the system, with each agent possessing its own state and characteristics. This ensures it is possible to explore how quantitative changes in biological factors affect the behaviour of each cell individually.

Capturing Cell-Level Dynamics

This section details how the dynamics for each cell type identified in the domain model has been translated into a platform model description. Aspects of the detail already covered in the domain model has not been repeated here, rather this is a detail of how the aspects in the domain model will be implemented.

1. Hematopoietic Cells (LTin/LTi Cells)

Although two different cell types, thus two different agent types, the implementation of the LTin and LTi cell shares some common features. The first of these is the assignment of speed at which the cell moves. Experimental work by Veiga-Fernandes et al (2007) has established a range within which LTin/LTi cell speed resides, as was noted in Figure 2.2. Further to this work, an analysis of the cells that were tracked in the *ex vivo* culture system described in section 1.3.3 suggests cell velocities can be considered to be normally distributed (Figure 2.3). In this model, cells are assigned a random set speed within the established range using a Gaussian random number generator, and it is assumed that the cell always moves at this velocity unless affected by the adhesion or chemoattractant expression.

The second common attribute is the calculation of a respective cell input rate. The domain model contains the results of flow cytometry analyses that suggests the percentage of cells that are LTin and LTi cells at E15.5. Similarly to the assumptions made for LTo cell number (section 2.2.2), the assumption is made that these cells are normally distributed in the gut at E15.5. Using gut dimensions measured from stereomicroscopy images and these percentages, an estimation of the number of LTin and LTi cells in the gut at E15.5 has been calculated (Figure 2.10). As the model captures the time from initial migration of LTin/LTi cells into the gut, respective input rates can be calculated such that a set number of cells enter the simulated gut per time-step, ensuring the correct number is reached at the time-point that represents E15.5. In this model, it is assumed that this input rate is linear. As no cell counts have been determined at time-points after E15.5, the assumption has been made that cell input remains at that rate through to E17.5, the time-point at which this model ends.

Finally, both cells express LT$\alpha\beta$ on the cell surface, that binds to LTβR expressed on the surface of LTo cells. Section 1.3.2 detailed how lymphotoxin signalling has a key role in LTo cell differentiation, and thus the upregulation of adhesion and chemoattractant expression. For the purposes of this model, the assumption is made that lymphotoxin signalling always occurs if a stable bind is formed between an LTo cell and hematopoietic (LTin/LTi) cell. The occurrence of a stable bind is controlled through use of a probability function to mimic binding affinity. To date the value of this parameter is unknown, and thus needs to be established through a process of model calibration. For LTin cells, RET signalling is modelled the same way, with RET binding to an LTo cell expressing the required ligand if a stable bind is formed.

Similarly to the domain model, cell level behaviours are expressed as UML state machine diagrams that form a specification to be implemented in the simulation platform. Figures 2.6a and 2.6b detail LTin and LTi cell level behaviour at the

platform model level respectively. In the majority of cases, a change in LTin and LTi state is mediated through interaction with adhesion or chemoattractant factors. As the implementation of these factors is covered in detail in the latter part of this section, this has not been covered here. Additional cell level parameters that have been identified in the generation of the platform model, and will thus be included in the simulation platform, are listed in the table at the bottom of Figure 2.6. Any further assumptions that have been made are also documented in Tables 2.1 and 2.2 respectively.

2. Lymphoid Tissue Organiser (LTo) Cells

Figure 2.7 depicts the platform model UML state machine diagram for the LTo cell, capturing an implementation of the LTo domain model in Figure 2.4. Similarly to the domain model diagram, there are a number of potential initial states. Again the first two (reading from left to right) captures the hypothesis that LTo cells detected in the early stages of development may be in one of two subsets: one where cells express RET ligand and can mediate PP development, and one where RET ligand is not expressed. Where RET-ligand is expressed, a change in state is triggered through stable contact with a RET expressing LTin cell agent. A stable contact is determined through the use of a probability function. For the alternative subset, the hypothesis generated by the collaborating experimental immunologists, that the cell could differentiate after sufficient contacts with LTi cells, has also been captured. Once in this differentiated state, adhesion factors are expressed, implemented as detailed in Figure 2.9 and described later in this section.

The remaining three initial states all capture the process of LTo cell mitosis. In this model, LTo cell division occurs once the cell has been active for 12 hours. Rather than divide, a new LTo cell agent is added to the model, inheriting the same attributes as the dividing cell. These three initial states capture that inheritance.

Further change in LTo state is triggered through stable contact between the cell and an LTi cell. In this state, the cell will also express chemoattractant molecules, as well as adhesion factors. A specification of how chemokine expression can be implemented is detailed in Figure 2.8. As chemokine and adhesion factor expression increase, the number of stable contacts between the LTo cell and LTi cell increases, further upregulating chemokine and adhesion factor expression. This continues to a point where expression is deemed to be saturated, controlled by two parameters. Where this occurs, the cell changes into a further state where it is deemed to be mature.

It can be noted that it is possible to transition from differentiated LTo states to a state where RET-ligand is down-regulated. This is an experimental feature that

has been included in the model in consultation with the experimental immunologists, which has the potential to investigate the effect on patch development if RET-ligands were inhibited at certain time-points in development.

Capturing Spatial Dynamics

Figure 2.10 specifies how the environmental measurements captured from stereomicroscopy images of the developing gut have been translated into measures that can be used to create a representative *in silico* environment. Parameters identified in the creation of this specification are listed in the table at the bottom of Figure 2.10. A scale has been set such that 1 pixel represents 4μm. The platform model makes the additional assumption that the emergent behaviour dynamics can be recreated using a 2D environment, where all interactions and cell movement is occurring on the gut surface. This 2D plain will be considered semi-toroidal in that the top and bottom are connected, forming a continuous plain across the intestine width. In reality, Peyer's Patches are 3D structures and cell aggregations form on a 3D scaffold. However, forming the aggregation on a 2D plain is a suitable abstraction in this case due to the lack of quantitative data to define a Peyer's Patch. Later results in this chapter will detail a calibration process that will form a measure of what a 'patch' is in terms of the simulation, and it is this measure that can be used as a baseline when performing *in silico* experimentation to explore the process of cell aggregation.

With a scale set that translates the biological measurements to pixels, cell sizes and speeds have also be translated accordingly, ensuring a direct mapping between the real and simulated system. To capture the cell velocity distribution observed in Figure 2.3, a cell speed will be derived using a Gaussian random number generator. With this being the case, a continuous coordinate grid is preferable, allowing for greater accuracy in capturing cell movement dynamics. Thus there will be no set grid-space in the simulation.

However a grid is used to manage LTo cell location in the model. The grid overlays the 2D environment, with each grid square the size of an LTo cell agent. Section 2.2.2 noted how calculations from flow cytometry data have been used to estimate the number of stromal cells present at E15.5. This is used to calculate the number of LTo cells that should be placed in the grid at start of any simulation. A percentage of these are then set to express RET ligand, with the remaining LTo cell agents remaining in the second subset where no ligand is expressed. The value of this percentage parameter remains unknown. For LTo cell division, assumed to occur every twelve hours, a replicate of the 'dividing' LTo cell agent is created in the nearest free location on the grid.

As noted previously, cellular interactions are occurring while the intestine environment is still developing. As such, the environment starts at pixel dimensions translated from measurements taken from stereomicroscopy images taken at E14.5, and expands

at the same rate each simulation time-step until it reaches the dimensions observed from imaging at E15.5. With no further measurements available, the assumption is made that the gut continues to grow at the same rate for the remaining 24 hours of development.

Although it would be preferable, available computational resources may make it intractable to model the entire length of the intestine. Such an implementation would capture a representative number of LTin, LTi, and LTo cells. With the implementation being agent-based, and each of these existing as an individual process, this would be computationally expensive. The need to perform a replicate number of such runs (as detailed in section 1.5.3) to reduce aleatory uncertainty inherent in agent-based simulations does not make this viable. Instead, a 10% section of the length will be captured, with the number of patches that would be expected to emerge scaled down respectively.

Modelling Chemokine Expression

The domain model specifies the existence of three chemokines expressed by the LTo cell: CXCL13, CCL19 and CCL21, that in turn bind to receptors CXCR5 and CCR7 expressed by LTi cells. Expression of chemokines by an LTo cell causes LTi cell chemotaxis towards sites of patch genesis, promoting cellular interactions and further upregulation in chemokine expression, thus expanding the area around a primordial patch where LTi motility is affected. It was noted in the domain model description that a full quantitative analysis of chemokine expression levels in the mid-gut during foetal development has yet to be performed, thus it is difficult to assign a set role to each of the three chemokines involved. However, PP formation in CXCR5-deficient mice, the receptor for CXCL13, is significantly reduced (Ansel *et al.*, 2000). Although not examined in PP formation, mice deficient for CCR7, the receptor for CCL19 and CCL21, did form a normal number of lymph nodes (Luther *et al.*, 2003). This generates the hypothesis that CXCL13 expression could have a dominant role in the clustering of LTi cells.

Taking this into consideration, a simplification has been introduced in the platform model where CXCL13, CCL19, and CCL21 will be modelled as a single chemokine, that binds to one single receptor on LTi cells. This will be modelled as an attractant where level of expression by an LTo cell is directly related to the number of stable contacts that occur between the LTo cell and LTi cells. As expression increases, the distance over which this chemoattractant influences cell motility also increases, with level of expression getting greater as distance to the LTo cell reduces.

This diffusion pattern can be modelled using an inverse sigmoidal curve, as described in Figure 2.8. On initial LTo cell differentiation, chemokine expression is low, and thus only affects a relatively short area around an LTo cell. Thus the equation of the curve can be set such that the curve is tight. As cellular contacts increase,

expression increases and the curve can be relaxed by changing the respective input variable, modelling an increase in the distance that the chemokine is being diffused. Figure 2.8 then describes how this function is used to calculate chemoattractant level in the vicinity of an LTi cell, to determine if chemotaxis is induced. This calculation produces a result between 0 (no expression) and 1 (strong expression). It can be noted from the description that a threshold is used within this calculation to ensure that a sufficient level of chemokine expression is in the vicinity of an LTi cell for chemotaxis to occur. The value for this threshold is currently unknown.

This method of chemokine expression modelling introduces the parameters listed in the table at the bottom of Figure 2.8. With the current biological understanding incomplete, values for these parameters need to be determined using a process of calibration once the model is implemented as a simulation.

Modelling Adhesion Factor Expression

A study of the literature reveals the involvement of three adhesion factors in PP development: VCAM-1, ICAM-1, and MAdCAM (Yoshida *et al.*, 2001). However as noted in the domain model description, there is no quantitative data detailing the expression levels of these three factors. As this is the case, a simplification has been introduced where the effect of these three adhesion factors will be modelled as just one factor. This section details the method that will be used to encapsulate adhesion factor expression and response in the model.

Adhesion factors are expressed upon stable interaction between a hematopoietic cell and an LTo cell. This model makes the assumption that the increase in expression is identical with each cell contact, thus the level of expression is directly related to the number of cellular interactions. Figure 2.9 details this relationship and LTin/LTi response to expression. An increase in the level of adhesion factor expression influences the probability that an LTin or LTi cell will remain in the vicinity of the LTo cell for a prolonged period. This probability increases linearly up to a threshold, set to ensure some stochasticity remains in cell behaviour, capturing the small likelihood that adhesion factors may not bind to receptors and thus not influence cell motility.

This model of adhesion identifies two parameters for which a value remains unknown. These are the slope of the linear equation that captures the relationship between probability of prolonged adhesion and number of cellular contacts, and the maximum probability that an LTin/LTi cell responds to adhesion factor expression (see image in Figure 2.9). Values for these parameters would therefore need to be established through a process of calibration once this model is implemented.

User Interaction and Data Collection

Within the platform model, consideration is also given into how user interaction and data collection will be added to the simulator. Table 2.4 specifies the considerations

that have been given to user interaction and data collection in the case of the lymphoid tissue development tool being developed here. This acts as a specification of the tooling that is necessary to include in implementation.

An implementation of the Platform Model will have two interfaces. The first is a graphical user interface that provides a visual representation of the simulated intestine tract. This aids the use of any simulator as a visual experimental tool. The second captures this behaviour without any graphical interface, with the aim of generating datasets that can be used in statistical analyses. This aids the performance of *in silico* experimentation where a high number of replicate runs may be necessary. In both cases, parameter values are specified prior to simulation run in an Extensible Markup Language (XML) file. This ensures that the parameters can be adjusted without the need to access the simulation implementation.

A specification of simulation results and how these should be output is also stated in Table 2.4. These are discussed in more detail in the Results Model section of this chapter (Section 2.2.5).

LTin (Platform)

Entry to tract between start of simulation and endpoint set by ε [1]

Random movement on tract surface

Probability of escape > Probability adhesion factors prolong contact [2]

If distance between cell and stromal cell < σ/2 + τ/2 (cell is in contact)

AND

Strength of LTi/LTo bind sufficient [3]

If contact cell is not a RET Ligand Expressing LTo Cell OR Probability of escape < Probability adhesion factors prolong contact [2]

Localised movement around LTo mediated by adhesion

If contact cell is an LTo Cell AND Probability of escape < Probability adhesion factors prolong contact [2]

Stable Contact with RET Ligand expressing Cell

1: Each cell is assigned a speed using a gaussian random number generator. The cell speed limits in the domain model (parameters ω and ξ) are translated into platform model values (1 pixel = 4 microns, see Figure 2.10). This creates limits for the random number generation that are in pixels/min. As each timestep may not necessarily represent one minute (number of seconds per time-step is set in parameter Λ), the lower and upper bounds are recalculated based on the number of seconds represented per time-step, and stored as parameters Π and Θ respectively
2: Probability adhesion prolongs contact is calculated as described in Figure 2.9
3: Strength of Bind: Result of Probability Calculation where bind occurs if generated random number (0-1) < χ
Cells that are in a state denoted by a star are deemed to take no further part in the simulation if they leave the right or left side of the simulated environment

(a) LTin Cell State Diagram: Platform Model

LTi (Platform)

Entry to tract between timepoints set by γ and η [1]

Random movement on tract surface

Distance between cell and inactive stromal cell > σ/2 + τ/2

Local Chemokine Level < Φ [4]

Local Chemokine Level > Φ [4]

Contact with Immature LTo Cell

Response to Chemokine Level in Local Environment

If distance between cell and stromal cell < σ/2 + τ/2 (cell is in contact)

Probability of escape > Probability adhesion factors prolong contact [2]

If distance between cell and stromal cell < σ/2 + τ/2 (cell is in contact)

AND

Strength of LTi/LTo bind sufficient [3]

Prolonged Surface Contact (Adhesion Effect)

If contact cell is an LTo Cell AND Probability of escape < Probability adhesion factors prolong contact [2]

Contact with Cell Expressing RET Ligand

1: Each cell is assigned a speed using the gaussian random number generator, as described for LTin cells above
2: Probability adhesion prolongs contact is calculated as described in Figure 2.9
3: Strength of Bind: Result of Probability Calculation where bind occurs if generated random number (0-1) < χ
4: Response to chemokine in the vicinity is calculated as described in Figure 2.8
Cells that are in a state denoted by a star are deemed to take no further part in the simulation if they leave the right or left side of the simulated environment

(b) LTi Cell State Diagram: Platform Model

|	Parameter	Name in Model	Platform Value
τ	LTin / LTi cell size	HCellDiameter	2 pixels
σ	LTo cell size	lToDiameter	6 pixels
Λ	Seconds per simulation step	secondsPerStep	60 seconds
Π	Simulation run cell speed lower bound	cellSpeedSimLowBound	Calculated
Θ	Simulation Run Cell Speed Upper Bound	cellSpeedSimUpBound	Calculated
χ	Probability stable bind occurs on contact	stableBindProbability	Calibrated
ϕ	Threshold Value that triggers chemotaxis	chemoThreshold	Calibrated

Figure 2.6: Platform Model UML State Machine diagrams for hematopoietic LTin and LTi cells, and additional LTin/LTi cell parameters identified in the creation of the Platform Model
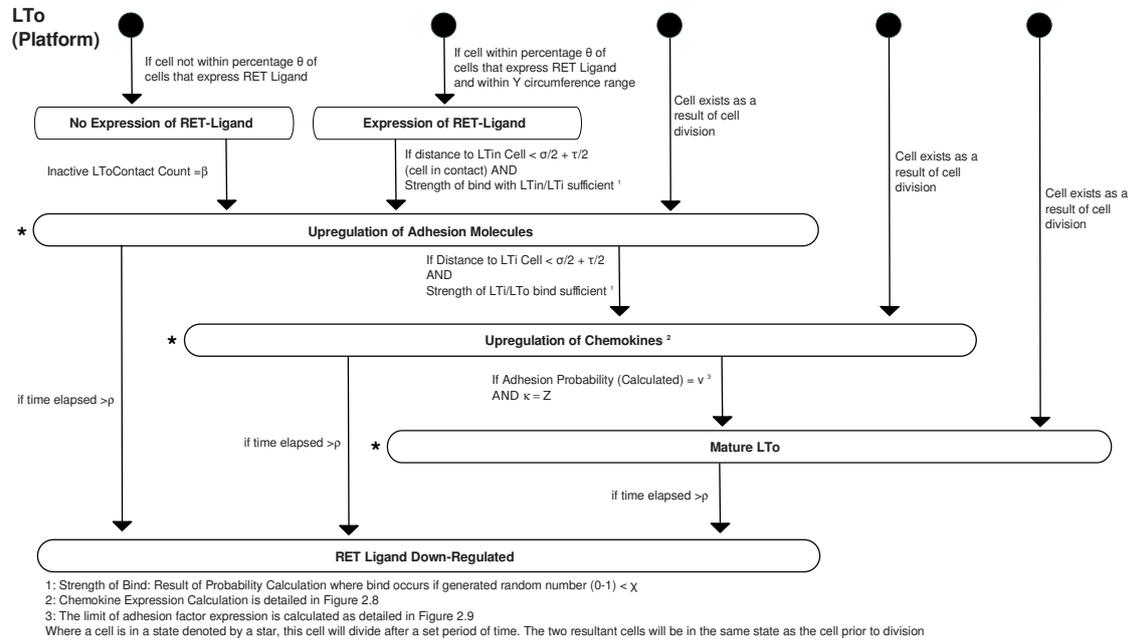
	Parameter	Name in Model	Platform Value
I	Initial Chemokine Curve Value	initialChemokineExpressionValue	Calibrated
Z	Maximum Chemokine Curve Value	maxChemokineExpressionValue	Calibrated
κ	Chemokine Expression Level	chemoExpressionLevel	Calculated
A	Adhesion Factor Expression Level	adhesionExpression	0
π	Hours immature LTo cell remains active	imLToActiveTime	Unknown
ρ	Hours RET Ligand Expressed	numHoursRETLigandActive	72 Hours
χ	Probability stable bind occurs on contact	stableBindProbability	Calibrated
β	Number of Stable Contacts required to activate immature cell	immatureContacts	Unknown

Figure 2.7: Platform Model UML State Machine diagrams for stromal LTo cell, and parameters identified in the creation of this model. It can be noted that there are five starting points on this diagram, and only four on the LTo Domain Model diagram (Figure 2.4). This captures the abstraction made in the Platform Model that only a percentage of LTo cells in the simulator will express a ligand for RET (expressed by the LTin cell). Reasons behind the need for an abstraction are detailed in the domain model, section 2.2.2.

.

Modelling Chemokine Expression and Response
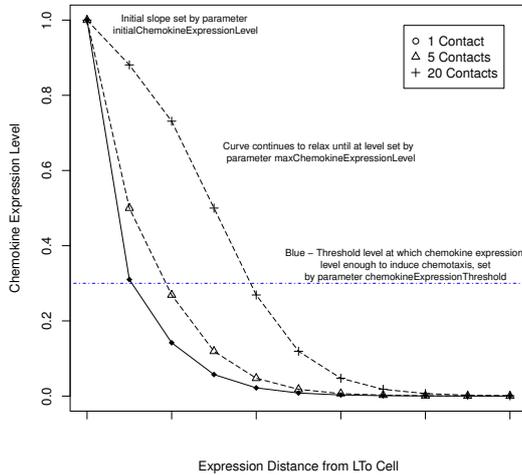
The chemokine expression effect has been modelled using an inverse sigmoid curve, adjusted such that the top of the curve meets the top of the y axis (at 1), by setting a constant within the sigmoid function to 3.

The initial curve is tight, calculated through using the initial chemokine expression level assigned to parameter initialChemokineExpressionLevel. This models expression over a limited distance, but one which strengthens as distance to the LTo reduces. With each stable contact between an LTo and LTi cell, the curve is relaxed by adjusting the parameter increaseChemoExpression, representing an increase in expression. With this increase, diffusion affects LTi cells over a greater distance. This expression increases until a maximum level of expression is reached, set by parameter maxChemokineExpressionLevel.

With each time-step, an LTi cell performs a move. To determine whether chemotaxis is triggered during this move, a virtual grid is drawn around the cell, with the centre of each grid space being placed at the distance the cell is to move (denoted by that particular cells speed). The simulator then evaluates the chemokine level in each square using the formula:

$$\frac{1}{1 + e^{-(-LToChemokineExpressionLevel + distanceToLTo + sigmoidCurveAdjust)}} \tag{2.1}$$

where the sigmoid curve adjustment is 3, the constant used to make the curve meet the y axis. Should the level be over a threshold level at which chemotaxis is induced (parameter chemokineExpressionThreshold), the chemokine level has an effect on that cells behaviour.

The calculated level of expression becomes the probability that the cell will move in that direction. Therefore it is most likely the cell will follow the level of chemokine expression as it gets closer to the LTo cell, but there remains some possibility that the cell may not respond to the level of expression at a greater distance.

	Parameter	Name in Model	Simulator Value
ϕ	Chemokine Expression Threshold	chemoExpressionThreshold	Range: 0-1, Calibrated
B	Sigmoid Curve Adjustment	chemoCurveAdjust	3
I	Initial Curve Value	initialChemokineExpressionValue	Calibrated
Z	Maximum Curve Value	maxChemokineExpressionValue	Calibrated
ι	Increase in expression on contact	increaseChemoExpression	Calibrated

Figure 2.8: Description of how the chemokine factors are included in the Platform Model. This details how the LTo cell increases chemokine expression with each stable contact and how the probability that LTi chemotaxis is induced is calculated. The table details the simulation parameters that have been identified in this process. This figure has been adapted from that included in Alden et al (2012b)

Modelling Adhesion Factor Expression and Response
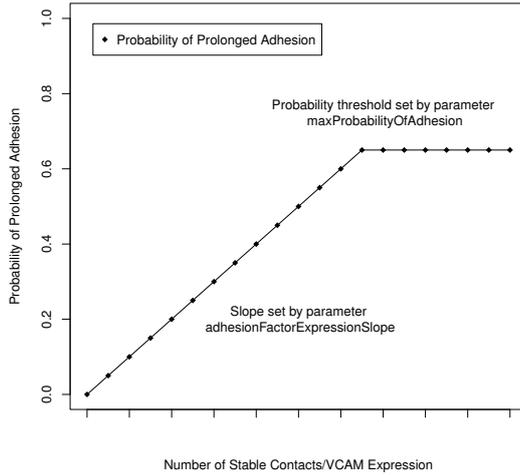
The probability of prolonged cellular contact mediated by adhesion factors is modelled using a linear equation, with a slope parameter value set through a process of calibration (parameter: adhesionSlope)

Each LTo cell has the same initial adhesion factor expression level (set by parameter initialAdhesion). With each stable contact, the level of adhesion factor expression increases (parameter: adhesionIncrement). This increases the probability that a cell remains in the vicinity of the LTo cell for a prolonged period. This probability increases until a threshold is reached (parameter: maxProbabilityOfAdheison). This threshold exists to ensure some stochasticity remains, and although adhesion factor expression may be high, there is a chance that an LTin/LTi cell may move away from the forming primordial patch

Upon stable contact between an LTo and LTin/LTi cell, the probability of prolonged adhesion is calculated using the formula:

$$adhesionSlope * LToAdhesionExpressionLevel \qquad (2.2)$$

where adhesion slope is set during a process of calibration. Should this probability be higher than the maximum probability threshold (also set during calibration), the probability is set to the threshold value.

	Parameter	Name in Model	Simulator Value
M	Initial Expression of Adhesion Factors	initialAdhesion	0
Ξ	Linear Equation Slope	adhesionSlope	Calibrated
E	Increase in Expression on Contact	adhesionIncrement	Calibrated
ν	Maximum Probability of LTi Response	maxProbabilityOfAdhesion	Range 0-1

Figure 2.9: Description of how the adhesion factors are included in the Platform Model. This details how the LTo cell increases adhesion factor expression with each stable contact and how the probability that an LTin or LTi cell remains in the vicinity of an LTo cell is calculated. The table details the simulation parameters that have been identified in this process. This figure has been adapted from that included in Alden et al (2012b)

Modelling the Intestine Environment

Mid-Gut as Extracted

Gut Straightened for Measuring
Average Length 28.8mm, Width 0.976mm
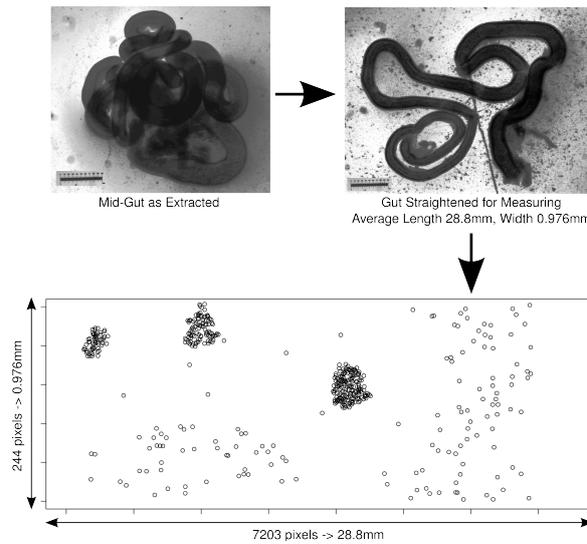
Biology: Stereomicroscopy (Zeiss) and ImageJ (Fiji) used to measure length and circumference of small intestines from twelve mice, six at E14.5 and six at E15.5. Top: One mouse intestine sample at E15.5. Average intestine length and circumference at E14.5 and E15.5 were calculated from these measurements. Intestine dimensions at E17.5 were estimated from the difference in these values.

Simulation: The biological measures inform the generation of the simulation environment: a 2D representation of the small intestine (represented by the graph above). 1 pixel = 4μm. X axis: small intestine length; Y axis: small intestine circumference. Dots represent LTin and LTi cells. Cells that leave the top/bottom appear on opposite side. Cells that leave right/ left are removed from the simulation

Cell Counts: Estimations of LTin and LTi numbers at E15.5 have been determined using flow cytometry. To ensure the correct number of cells are created, a linear input rate is used to create the required number of cells per time-step. This cell input rate that continues through until the end of the simulation run.

$$\varsigma = (((\Gamma/\tau) * (\Delta/\tau))/100) * \delta$$
$$\Upsilon = (((\Gamma/\tau) * (\Delta/\tau))/100) * \psi$$

(2.3)

	Parameter	Name in Model	Domain Value	Simulator Value
Γ	Initial Circumference	initialGridHeight	0.966mm	244 pixels
Δ	Initial Length	initialGridLength	28.80mm	7203 pixels
K	Maximum Circumference	upperGridHeight	1.016mm	254 pixels
P	Maximum Length	upperGridLength	29.22mm	7303 pixels
ζ	LTo Cell Density	stromalCellDensity	20%	20%
Υ	Intestine Growth Time	growthTime	72 Hours	72 Hours
δ	Percentage of cells that are LTin at E15.5	percentLTinfromFC	0.45%	0.45%
ψ	Percentage of cells that are LTi at E15.5	percentLTifromFC	0.37%	0.37%
ς	LTin Input Rate	lTinInputRate		Calculated
λ	LTin Input Rate Function	lTinInputRateFunction		linear
Ξ	LTin Input Rate Constant	lTinInputRateConstant		Not Used
Ψ	LTi Input Rate	lTiInputRate		Calculated
Υ	LTi Input Rate Function	lTiInputRateFunction		linear
Σ	LTi Input Rate Constant	lTiInputRateConstant		Not Used

Figure 2.10: A description of how the simulation environment relates back to the developing mid gut, and parameters identified in the creation of this representation

State	Model	Assumption
Random Movement on Tract Surface	Domain	There is no attractive influence on an LTin cell any contact with RET ligand-expressing cells will occur randomly.
	Platform	Each cell is assigned a speed between the lower limit set by parameter Π and upper limit set by parameter Θ. This is chosen randomly from a Gaussian random number generator.
Contact with RET ligand-expressing cell	Domain	For lymphotoxin signalling to occur, the bind between the two cells must be of sufficient strength. If the bind affinity is sufficient, we assume that cell signalling always occurs. If contact is with a cell expressing RET ligand yet not an LTo, and a stable bind occurs, the cells will bind briefly but no signalling occurs.
	Platform	Whether LTin and LTo cells bind will be determined by a probability function. If a chosen probability is > parameter χ then a stable bind is formed.
Localised movement around LTo mediated by adhesion	Domain	An LTin cell will remain in contact with an LTo cell if there is a sufficient expression level of adhesion factors.
		As expression level increases, the LTin cell is more likely to remain in contact
		Though there may be sufficient expression level of adhesion factors, there is still a possibility that the LTin cell may move away from the LTo.
		Though the cell remains in contact with the LTo, LT signalling and up-regulation of adhesion factors and chemokines only occurs on initial contact.
	Platform	LTin cell will remain in close contact with the LTo cell making small movements around it.
		Prolonged adhesion is decided through use of a probability function, detailed in Figure 2.9.
Other Assumptions	Domain	LTin cells migrate into the tract throughout the whole period being modelled. All LTin cells are the same size, 8 μm.
	Platform	Flow cytometry has helped estimate the number of LTin cells that should be present at E15.5 in development. A linear input rate is used to ensure this is reached. This rate remains constant throughout the simulated period.
		The environment is modelled as a 2D plane on which all movement and interactions occur (Figure 2.10). Should an LTin leave the left/right of the screen, this cell is removed from the simulation.

Table 2.1: List of assumptions made at both domain and platform model level concerning the behaviour of and interactions with an LTin cell.

State	Model	Assumption
Random Movement on Tract Surface	Domain	Cells move randomly until the level of chemokine expression in the vicinity is above a threshold.
	Platform	To ascertain chemokine level, the expression level in each gridsquare around the LTi is calculated (Figure 2.8). If none of these values is above ϕ, the cell moves randomly.
Response to chemokine level in local environment	Domain	Three chemokines are known to play a part in the process CXCL13, CCL19, and CCL21. However as an abstraction we will assume these can be modelled as a single chemokine. IL-7, which could stimulate IL-7 receptor signalling and regulate chemokine receptor expression levels of LTi cells, has not been included in the model. The assumption will be made that there is always sufficient IL-7 present for chemokine receptor expression to be upregulated. There is always a small chance that the cell may not respond to the level of chemokine.
	Platform	Chemokine expression is modelled using an inverse sigmoid curve (see Figure 2.8). As some stochasticity must remain, the chance that the cell will move in the direction of the strongest level is determined by probability function.
Contact with RET ligand-expressing cell	Domain	For lymphotoxin signalling to occur, the bind between the two cells must be of sufficient strength. If the bind affinity is sufficient, we assume that cell signalling always occurs.
	Platform	Whether LTi and LTo cells bind will be determined by a probability function. If a chosen probability is > parameter χ then a stable bind is formed.
Prolonged surface contact (adhesion effect)	Domain	An LTi cell will remain in contact with an LTo cell if there is a sufficient expression level of adhesion factors. As expression level increases, the LTi cell is more likely to remain in contact. Although there may be sufficient expression level of adhesion factors, there is still a possibility that the LTi cell may move away from the LTo. Although the cell remains in contact with the LTo, LT signalling and upregulation of adhesion factors and chemokines only occurs on initial contact.
	Platform	The LTi cell would remain in close contact with the LTo cell making movements around it. Prolonged adhesion is decided through use of a probability function (Figure 2.9)
Other assumptions	Domain	LTi cells migrate into the tract throughout the whole simulated period. All LTi cells are the same size 8 μm.
	Platform	Flow cytometry has helped estimate the number of LTi cells that should be present at E15.5. A linear input rate is used to ensure this is reached. This rate remains constant throughout the simulated period. The environment is a 2D plane where all movement and interactions occur (see Figure 2.10). A cell is removed should it leave the left/right of the screen.

Table 2.2: List of assumptions made at both domain and platform model level concerning the behaviour of and interactions with an LTi cell.

State	Model	Assumption
No expression of RET ligand	Domain	Although we are aware that 20% of the intestine tract contains stromal cells, we assume only a percentage of these have the potential to become patches.
	Platform	Where only a percentage of LTo cells are active, all are still placed on the intestine tract, but interactions only occur with LTo cells which have the potential to become patches (that express RET ligand).
Expression of RET Ligand	Domain	Cell will remain active throughout the time period, irrespective of whether the cell changes state or not.
	Platform	All LTo cells which express RET ligand have the potential to express adhesion factors and chemokines (thus form patches).
Upregulation of adhesion molecules	Domain	Adhesion molecules are upregulated with every contact where the strength of the bind is sufficient (Figure 2.9). Upregulation only occurs on initial contact with the cell; prolonged contact due to adhesion does not lead to further upregulation. Cells in this state will divide after a set number of hours.
	Platform	Expression of adhesion factors does not degrade over time. With each stable contact, a counter representing adhesion factor expression is increased. This determines the strength of adhesion and probability the cell will remain in contact (see Figure 2.9).
Upregulation of chemokines	Domain	Chemokines are up-regulated with each LTi/LTo contact where the strength of the bind is sufficient (see Figure 2.8). Upregulation only occurs on initial contact with the cell; prolonged contact due to adhesion does not lead to further upregulation. Cells in this state will divide after a set number of hours.
	Platform	Chemokine expression does not degrade over time. With each stable contact, a constant that is used to calculate chemokine expression in the sigmoidal curve function is adjusted (Figure 2.8). This determines the distance over which the chemokine has an effect.
Mature LTo	Domain	
	Platform	Both adhesion molecules and chemokines must have reached their peak of expression to reach this state.
Other Assumptions	Domain	LTo cells in all bar the top two states will divide after 12 hours. On division, the cells will possess the same attributes as the original cell prior to division. It is assumed that other pathways, such as the NF-κB pathway, are always activated.

Table 2.3: List of assumptions made at both domain and platform model level concerning the behaviour of and interactions with an LTo cell.

Consideration	Platform Specification
Simulation Interface	**Graphical User Interface:** • Enabled with use of MASON Toolkit • Environment and cell movement displayed in MASON window, settings can be varied on simulation control console **Non-GUI Simulator:** • Interaction via XML parameter file read by simulator when started
Instrumentation	**Simulation results output as CSV files:** • Tracking results: cells in vicinity of LTo cell between two stated time-points • Tracking results: cells $>50\mu$m from LTo cell between two stated time-points • Cell Positions: X and Y positions of all LTin and LTi cells at a stated time-point • LTo Statistics: Value of Chemokine and Adhesion Factor expression parameters for all LTo cells, at a stated time-point **Images:** • Screenshots every timestep during tracking (for time-lapse movie generation) • Screenshots at every 12 hour time-point • Screenshot at end of simulation
Quantifying Data	**Stored by Simulation:** • Cell Position (x,y) • Position when cell tracking commenced • Position when cell tracking period elapsed • Distance covered by cell in tracking period **Calculated by simulation:** • Cell track length covered in tracking period • Cell velocity in tracking period • Cell displacement in tracking period **Calculated from simulation CSV file output:** • Number of aggregations of LTin/LTi cells formed (patches) • 2D area of the patches that form

Table 2.4: List of simulation design considerations included within the platform model. These detail how interaction with the simulation will occur, the instrumentation that is required, and the quantifying data that will be produced, detailed further in the Results Model. Whereas the previous figures detail how the biological system is captured, this is the first time the use of the simulation as a software tool is considered

2.2.4 Simulator

A computer simulation is created from the specification in the platform model. In this case, the platform model has been implemented as a computer simulation using the Java programming language and MASON simulation environment, a cross-platform toolkit for the creation of multi-agent simulations (Luke, 2005). Each of the cell types detailed in section 2.2.2 have been implemented as a Java class and methods created that match the transition events detailed in the state machine diagrams. Expression of chemoattractants and adhesion factors has been captured as specified in Figures 2.8 and 2.9. A virtual environment has been created that matches the specification in Figure 2.10. As MASON simulations are executed in steps where each active agent performs a behaviour set by its current state, developmental time is incorporated by setting each timestep to represent one minute. This however is a default figure and can be changed if an exploration requires it. Where an alteration is made, the simulator adjusts relevant parameters, such as cell speeds accordingly.

2.2.5 Results Model

The results model provides a structure to interpret the results that arise from *in silico* experiments performed using the simulator. Having a structure with which to contrast simulation results with the domain model provides a level of confidence that the simulator is a fair representation of the system being modelled. A specification is created that documents the output obtained from the simulation, what domain knowledge this is compared to, and the statistical methods used to generate this result. Simulator output is in the form of simulation responses that are deemed to be of biological interest. They may include cell movement properties, factor expression levels, or space measurements.

For the PP simulation platform, output captured falls into three categories:

Cell Behaviour Responses

The domain model specifies two forms of emergent behaviour, the first of these concerning a change in cell behaviour around the site of a forming Peyer's Patch. For a judgement to be made on whether the simulation correctly captures this behaviour, cell responses need to be captured and contrasted with cell distributions observed *ex vivo* (Patel *et al.*, 2012), where cell behaviour was tracked for a period of one hour, and three responses calculated: path length, velocity, displacement. In the platform model, a scale was set such that environmental measurements and cell speeds can be translated back into a form that can be directly compared with the biological measures. Thus, the simulation platform contains the functionality to track cells for a set period and produce these three statistical responses for each cell tracked (Figure 2.11). These responses are output as a comma separated value (CSV) file at the end of the

tracking period. This produces a distribution of cell behaviour responses that can then be directly compared to those captured *ex vivo* .

Although the *ex vivo* investigation only considered one time-point, for a length of one hour, functionality has been included in the results model for cell tracking to be completed at multiple time-points in simulation, for any length of time required.

Patch Characteristic Responses

As has been noted previously in this chapter, characterisation of a PP in simulation is difficult for two reasons. Firstly, little biological data exists that can be used to classify what a patch is, in terms of size or area, and the existence of patches tends to be noted visually rather than through use of any statistical measure. Secondly, PP can form as 3D cell aggregations rather than the 2D aggregations that this model generates. Thus, even if quantitative data did exist, the two structures are not directly comparable. However, though these form in 2D, conclusions can still be drawn on how biological factors are influencing the formation of these aggregations.

The simulator produces four responses that are deemed 'Patch Characteristics' in future experimentation in this thesis. The first two are the 2D area of the aggregation of LTin/LTi cells, and the number of such aggregations that form. These are calculated using the R statistical package once the simulation run is complete, using X and Y coordinates of LTin and LTi cells that are output from the simulator as a CSV file. Potential patches are identified and counted using k-means hierarchical clustering, and agents that are a greater distance than double the diameter of an LTi cell away from another cell removed. Cluster size is generated by calculating the area within eight coordinates selected from the perimeter of the cluster, as detailed in Figure 2.12. Being in possession of these measures makes it possible to determine the effect quantitative changes in simulated biological factors has on cell aggregation. The other two measures detail the level of adhesion and chemoattractant factors expressed by the LTo cell at the end of the simulation. As these measures are directly related to cell interactions, this gives an indication of how such quantitative changes in parameter value affected interactions between the LTo and LTin/LTi cells.

Similarly to the cell behaviour characteristics, cell X and Y coordinates can be output from the simulation at any time-point, making it possible to examine influences on cell aggregation over time.

Simulation Snapshots

Alongside the statistical measures generated above, conclusions can also be drawn from visual simulation output. The simulation platform includes the functionality to automatically produce snapshots of the simulated gut environment, either at twelve hour intervals and at the end of simulation, or for each simulation time-step. The latter enables movies to be generated from simulation runs, or cell tracking analysis to

be performed using the Volocity software package that is used for the performing the same analysis for *ex vivo* and *in vivo* images.

in silico Cell Tracking

Agents (LTin and LTi cells) can be tracked at any time-point in the simulation, for any duration of simulation time-steps. In the *ex vivo* work, (Patel *et al.*, 2012), three cell responses were calculated using the software Volocity package and cell tracking images. These responses, cell path length, displacement, and velocity, are also calculated for the simulator, using the equations below.

Agent position at the start of the tracking period is recorded in agent variable *trackingStartPosition*. At the end of the period, the end position is recorded in agent variable *trackingEndPosition*.

Distance Between Two Coordinates:

The distance between two points, used in these calculations, would be calculated as follows:

$$diffX = new\_X\_location - original\_X\_location$$

$$diffY = new\_Y\_location - original\_Y\_location \tag{2.4}$$

$$distance = \sqrt{diffX^2 + diffY^2}$$

However, using this formula does not allow for the fact that a cell may roll around the top and bottom of the screen. To take this into account, an adjuster is introduced. Initially the calculation is performed with an adjustment of 0. Should a distance be returned that is greater than half the width of the simulated tract, a roll-around is detected. If the end Y coordinate is less than the original Y, the calculation is performed again, with the height of the tract used as the adjuster in the calculation. If the end Y coordinate is more than the original Y, the calculation is performed with the height of the tract as the adjuster, but this time the adjustment is negative.

$$distance = \sqrt{diffX^2 + (diffY + adjuster)^2} \tag{2.5}$$

Cell Path Length:
Cell path length is calculated during each time-step that tracking is performed. This starts as 0, and is incremented with each time-step as follows:

$$pathLength = pathLength + distance\_from\_old\_to\_new\_coordinates \tag{2.6}$$

Cell Displacement:
Cell displacement is simply the distance between the start and end coordinates:

$$diffX = trackingEndPosition\_X - trackingStartPosition\_X$$

$$diffY = trackingEndPosition\_Y - trackingStartPosition\_Y \tag{2.7}$$

$$displacement = \sqrt{diffX^2 + (diffY + adjuster)^2}$$

Cell Velocity:
Cell Velocity is calculated at the end of the tracking period:

$$cellVelocity = pathLength/tracking\_time\_steps\_completed \tag{2.8}$$

Figure 2.11: Equations detailing how the simulated cell behaviour characteristics are calculated in the Platform Model

.
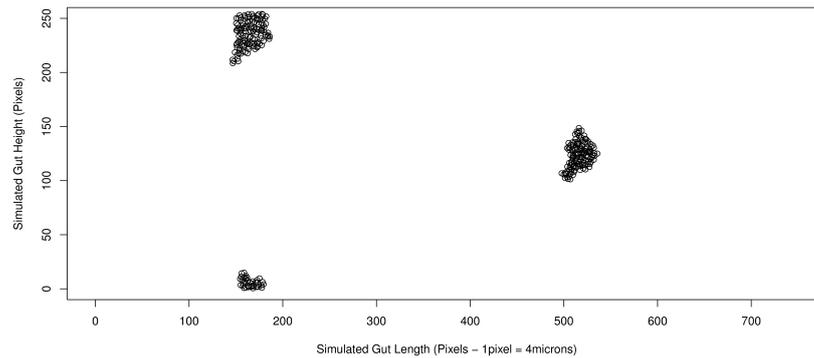
(a) Individual patches
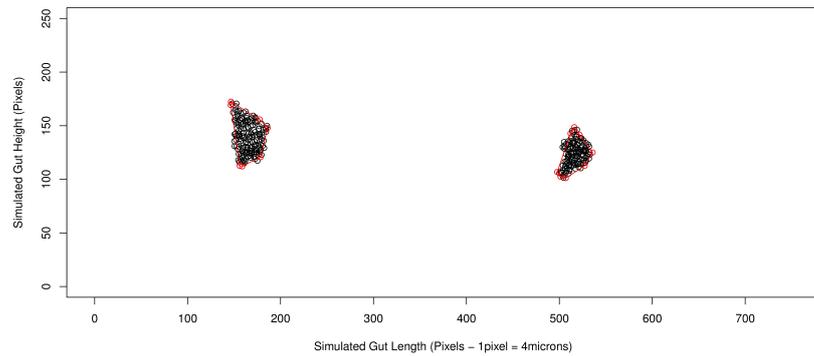
(b) Approximating patch size

Figure 2.12: Identifying Peyer's Patches and approximating the size. In (a), patches are identified using k-means hierarchical clustering. Any patches that roll around the top and bottom of the screen, as on the left hand side, are deemed one patch. In (b), patch size is approximated by choosing eight points on the perimeter of the cluster. The value returned for the patch size is the area of the shape formed by joining these eight points (as shown by the red line in (b)). For patches that roll around the top and bottom, the coordinates are normalised to join the patch as one, and the area calculated.

2.2.6   Calibration to Establish Baseline Behaviours

The simulation platform is an implementation of the platform model specification, that is in turn a specification of the biological system captured in the domain model. However, the domain model does not capture every detail of the biological system, rather it captures an abstracted view of it. There is no guarantee that the simulation platform will produce the cell behaviours observed in *in vivo* imaging and *ex vivo* culture. As can be seen in the domain and platform model descriptions above, there are a number of parameters for which a biological value remains unknown, and parameters introduced in the platform model which capture a biological process (and thus do not translate back to the biological system) for which values are also unknown.

Calibration is the process by which values are obtained for any parameters for which a numerical value is unknown, with the objective to ensure that the simulator produces behaviour responses observed in previously published studies or ongoing experimentation where available. Where biological data is available, a comparison between a simulation result and available data can be made using statistical tests such as the Mann-Whitney U-Test, which will indicate statistical similarity between two sets. Using a structured trial and error approach, values are assigned to parameters for which a value is unknown, and altered until no statistical difference exists between the available data being used for comparison and simulation result distributions. Where there is more than one emergent behaviour observed, the simulation should be calibrated to ensure each emergent behaviour is reproduced. This section examines the calibration of the Peyer's Patch simulation developed in this chapter.

Cell Behaviour Baseline Simulation

Analysis of *ex vivo* cell tracking images (Patel *et al.*, 2012) has provided distributions of cell behaviour responses upon which simulation results can be compared. As described in section 1.3.3, cell behaviour was tracked for an hour at the twelve hour time-point, with results capturing two distributions: cells within 50μm of a RET ligand expressing cell, and cells further away. With this data available, the simulation can be run and cell responses for hour twelve of development captured, and the respective distributions compared in order to judge how suitable the simulator has captured this behaviour.

In this case, there are six parameters for which a value remains uncertain:

(a) Probability at which an LTin/LTi cell will form a stable bind with an LTo cell upon contact. A stable bind is defined as contact that leads to LTo cell differentiation and an increase in expression of adhesion and chemoattractant factors.

(b) Initial level of chemokine expression upon LTo differentiation.

(c) Saturation limit of chemokine diffusion.

(d) Level of chemokine required in a cells local environment to induce LTi cell chemotaxis.

(e) Level at which surface adhesion factors are expressed with each stable contact between an LTin/LTi and LTo cell.

(f) Probability that the level of adhesion factors expressed on the surface of an LTo cell will restrict LTin/LTi cell movement.

The goal of the calibration process is to find a potential set of values for these six parameters where the simulation produces cell behaviour responses that are statistically similar to those observed in the *ex vivo* culture system. The simulation has been run and cell velocity and displacement responses captured for hour twelve of development. With no guarantee the data is normally distributed, the Mann-Whitney U-Test has been used to compare the *in silico* and *ex vivo* distributions, and dot-plots produced to aid visual comparison of cell behaviour responses. This has been performed using a structured trial and error approach, tweaking parameter values until behaviour is produced that is statistically similar to that observed *ex vivo*. A selection of the dot-plots produced in this process are included in Figure 2.13.

Figure 2.14, produced at the end of this process, is a visual comparison of cell behaviour responses *ex vivo* against those from the calibrated simulator. It can be noted that there is no statistical difference between the two sets of results. This provides a level of confidence in the simulators use in exploring the cell behaviour it has captures.

Cell Aggregation Baseline Simulation

Calibrating the simulator with respect to the second emergent behaviour, the formation of cell aggregations after 72 hours that mature to form PP, is however more complex for two reasons previously noted in this chapter: the unavailability of biological measurements, and relating the 2D abstraction used in simulation to the real-life 3D structures. Explorations in Chapter 4 of this thesis seek to examine the influence that the six parameters listed in the previous section have on patch characteristics, thus a baseline result needs to be established. In this instance, a baseline behaviour has been found using the simulation parameter values established above. This establishes the patch characteristic responses of aggregations that form under parameter conditions that replicate emergent cell behaviour observed *ex vivo* . This baseline can then be used to examine the effect an alteration in these parameters has on these simulation patch characteristics.

Key factors in patch formation can be determined by examining the formation of one patch. To do this, one RET-ligand expressing LTo cell agent is placed at the centre coordinate of the simulation environment. The analysis is examining the parameters influence on *a* patch rather than the effect of LTo cell position, and thus this is not important in the analysis. Restricting its position ensures there is no variability in

result caused by the simulators choice of LTo cell position. The simulation can then be run to produce the patch characteristic responses detailed in Section 2.2.5. A representative baseline result can then be established by generating a number of replicate results. The method by which this number is established is discussed later in this chapter. This forms a distribution of responses under normal conditions for use as a comparator when conditions are changed.

Calibration based on Number of Patches

It was shown in Figure 1.2 that between eight and twelve PP develop along the length of the mouse intestine, variability that cannot currently be explained. Flow cytometry results have been used to estimate the number of LTo cells to include in the model (Section 2.2.2). However if patches formed around each of these cells, the number of patches would be in three figures. Thus the assumption was made in the domain model that only a certain number of these cells would have the capability to form PP.

As was noted in Section 2.2.3, the simulation captures 10% of the length of the mouse intestine. Thus one would expect no more than 2-3 patches form in this short section. The simulation randomly fills 20% of the environment with stromal cells. A set percentage of these are then chosen at random to have the capability to form PP. Calibration here has thus focused on establishing the percentage of cells that have the capability to differentiate into LTo cells, and thus form PP.

A range of percentage values was set and 100 simulation runs performed for a set of values in that range. Using a snapshot taken at the end of the simulation run, number of patches that form was determined visually, by experts in the field manually counting what they determine to be a patch. Averages were then taken for each percentage used, as shown in Figure 2.15. It was determined that the correct number of patches form where only 0.25% of the 20% of cells express RET-ligand, having the potential to mediate PP development. Thus, this is used as the value for that parameter in the relevant analyses in the chapters that follow.

2.2.7 Argument-Driven Validation

Argument-Driven Validation is a technique that can be used to structure an assessment of the model in such a way that each step in the construction process was validated, the reasoning behind the inclusion or exclusion of a feature or assumption provided, and evidence given as to why this conclusion has been drawn (Ghetiu *et al.*, 2010; Polack *et al.*, 2011). Thus, in this case, features included from the domain and platform models are both assessed and scrutinised. The overall objective is to go through the model in steps, linked together by the available evidence to support that step, leading to increased confidence that certain parts of the model are correct, while identifying areas open to further examination. This may identify features where assumptions have been made which need further investigation, or identify clarifications needed from the

biological experts. The latter may then feed into wet-lab experimentation in order to assess a feature, or lead to an assessment of the reliability of biological results gathered from external sources such as published literature.

This process is captured in diagrammatical form using Goal-Structuring Notation (Ghetiu *et al.*, 2010; Kelly, 1999), the results of which can be seen in Figures 2.16-2.19. The argumentation has been split into three figures for ease of reading. The first, Figure 2.16, is the top level of the argument. This states the main claim being argued, in this case that the simulation is an adequate representation of the biology. This in turn is broken down into four sub-claims which, if solved, provide a argument that supports the main claim. This section considers each of these subclaims in turn.

For ease of comparison, claim 1.1.3 is considered first, as this is detailed on the same diagram as the main claim (Figure 2.16). The claim is made that the simulation produces cell behaviour statistically similar to that observed *ex vivo*. The previous section of this chapter examined the use of calibration to ensure that this is the case, and these results are noted on the diagram as evidence that the claim can be verified.

Claim 1.1.1 is more complex, and examines whether there is adequate biological data included in the model. The argument is detailed in Figure 2.17. In this case, the claim is examined using four strategies, each of which concerning a subset of the parameters that have been derived from biological data. The first considers the representation of LTin and LTi cells, in terms of number, size, and cell velocity; the second the representation of stromal LTo cells; the third the representation of the biological environment; and the final strategy data related to Peyer's Patch characteristics. Where relevant, sources of the biological data are noted, and where necessary any assumptions that have been made based on this data documented. This gives an overview of how the data has been obtained and how this has then be utilised in the creation of the simulation. This helps improve confidence in the design of the model as parameter generation from biological data is more transparent. One claim, under strategy 1.1.1.4, is noted with a blank diamond. This concerns biological data that can be used to compare PP generated *in silico* to that *ex vivo*. As has been noted previously in this chapter, such quantitative data is not currently available, making it difficult to support a claim that PP are generated that are of a representative size. Thus the claim needs to be developed further, noted by the presence of the diamond. This however should not be viewed negatively: one of the objectives in performing ABV is to identify such areas where current understanding is lacking.

The second claim made examines the abstractions made, and how these are justified. This argument is detailed in Figure 2.18. It is vital that the use of abstraction is transparent, as these may affect the meaning of the results generated, thus in turn affecting any hypotheses developed from them. Again this claim is examined in four strategies: the first considers the implementation of chemokines, justifying why the implementation of one chemokine rather than three is a suitable abstraction; the sec-

ond the implementation of adhesion factors; the third the 2D implementation of a 3D environment; and the final strategy abstractions linked to cell signalling. In the majority of cases, subclaims generated from these strategies are supported by available evidence or the insight provided by collaborating experimental immunologists. There are three claims however that need to be developed further. The first two claim that simulation chemokine and adhesion factor expression levels are representative of that in the biological system. No previous laboratory studies have generated quantitative data on which simulation expression levels can be compared, and thus this is an area where further experimental work is required. The third claim is that the physical shape of the gut (e.g. the bends) has no impact on PP formation. As no previous study has examined the role that the physical geometry of the environment, the assumption has been made that this has no affect on the process, mainly due to the difference in scales between the individual cells and the environment.

Claim 1.1.3 examined the first of the emergent behaviours observed, the change in cellular behaviour in the vicinity of a forming PP. Claim 1.1.4 (Figure 2.19) examines the second, the development of aggregations of cells after 72 hours, that later become PP. The claim is made that the simulation adequately captures this emergent behaviour. This is justified using three strategies: determining whether a representative number of PP are formed along the intestine length; that previously published experimental results that examine PP formation in different physiological conditions are replicated; and that the simulation correctly captures the characteristics of a PP observed *in vivo*. The first strategy uses evidence from the calibration process in section 2.2.6, and results in Figure 2.15, to demonstrate that a representative number of patches do form. It is noted that this is based on the assumption that only a certain percentage of LTo cells can express RET ligand and thus mediate PP development. The second strategy, replicating previously published results, is noted as needing to be developed further. Experimentation using the simulator to examine this claim is addressed in Chapter 4 of this thesis. Finally, the last strategy examines the size of these emergent aggregations. As noted previously, there is no biological data on which simulation PP size response can be contrasted. However, the method by which PP are identified is justified alongside the claim.

When considered together, these diagrams provide a detailed flow of the decisions that have been made in the course of simulation development. In this case, the process may come across as static: claims have been made and supported and areas of further development identified, and a final document produced. In practice however, the development of these arguments would not end here, yet would continue as simulation results are analysed and the simulation developed further.

2.3 Making the Simulation Tool Publicly Accessible

Making the simulator freely available enables immunologists to engage with the tool for their own research and provide critical feedback on any future iterations of the simulator. The simulator and its underlying source code are freely available to run online and for download (http://www.cs.york.ac.uk/immunesims/frontiers).
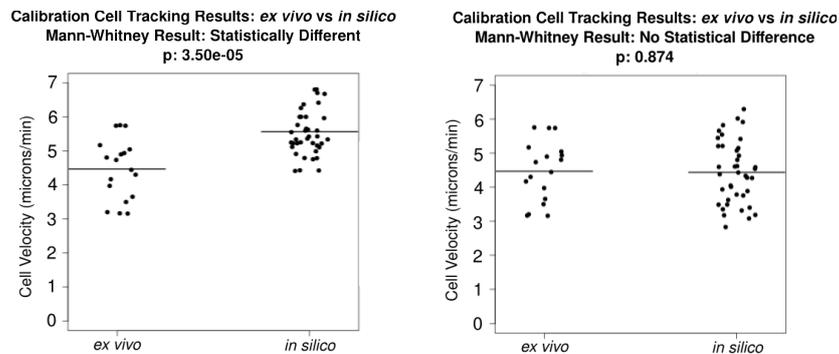

Figure 2.13: Two example dot-plots produced during the process of calibration. In both cases, the right hand side of the plot shows the average cell velocity for a particular simulation run, for cells tracked in during hour twelve of development. The left hand side is the cell velocity distribution that was observed *ex vivo* . The statistical result produced by the Mann-Whitney U-Test when the two distributions are compared is noted in the graph header. For the scenario on the left, cell velocity is too fast, and thus the parameters chosen have not correctly captured cell behaviour. For that on the right, the parameters are much more suitable, and there is no statistical difference between the two distributions.
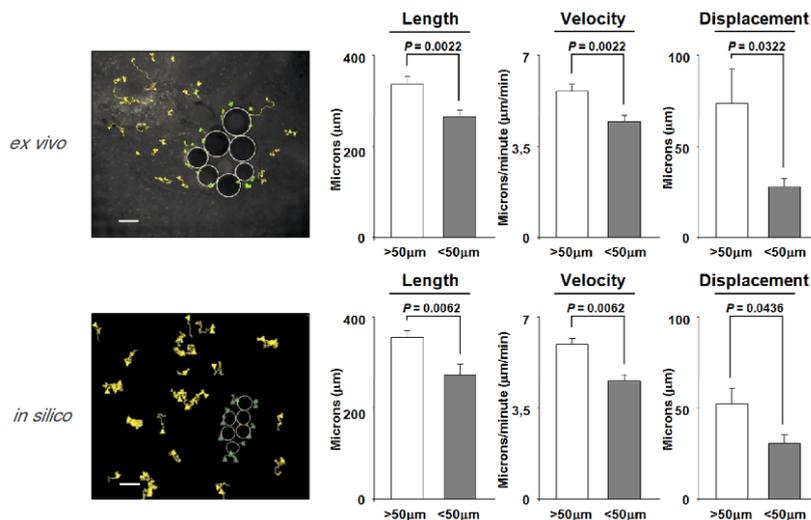

Figure 2.14: A comparison of the behaviour of the calibrated simulation platform with cell behaviour responses observed *ex vivo* . Top row: *ex vivo* cell responses; Bottom row: *in silico* cell responses. The first column contains cell tracking images for hour twelve of development. Both were produced using Volocity (PerkinElmer), the *ex vivo* image by producing an sequence from images captured each minute, and the *in silico* image by producing a sequence from screenshots captured at each time-point representing one minute. This figure has been adapted from that published in Patel et al (2012).
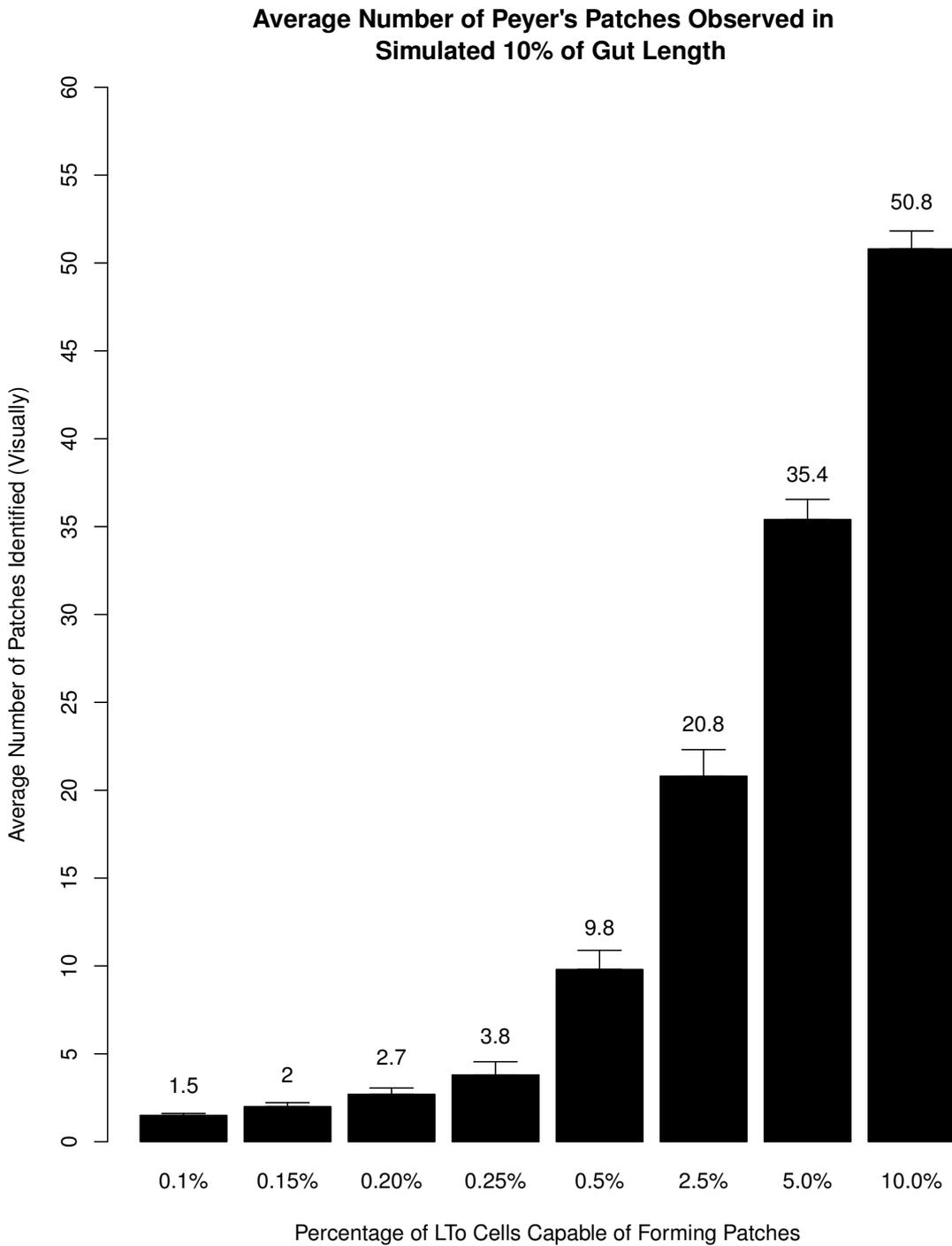
Figure 2.15: Average number of PP that form where the percentage of LTo cells that express RET ligand, and thus have the capability to form PP, is adjusted. Averages are taken from 100 runs of the simulation under each condition. Average number of patches that form in the simulated 10% of the gut length is noted above each bar. Patches are determined visually in conjunction with experts in lymphoid tissue development. As 10% of the gut has been captured, between 2-3 patches would be expected.
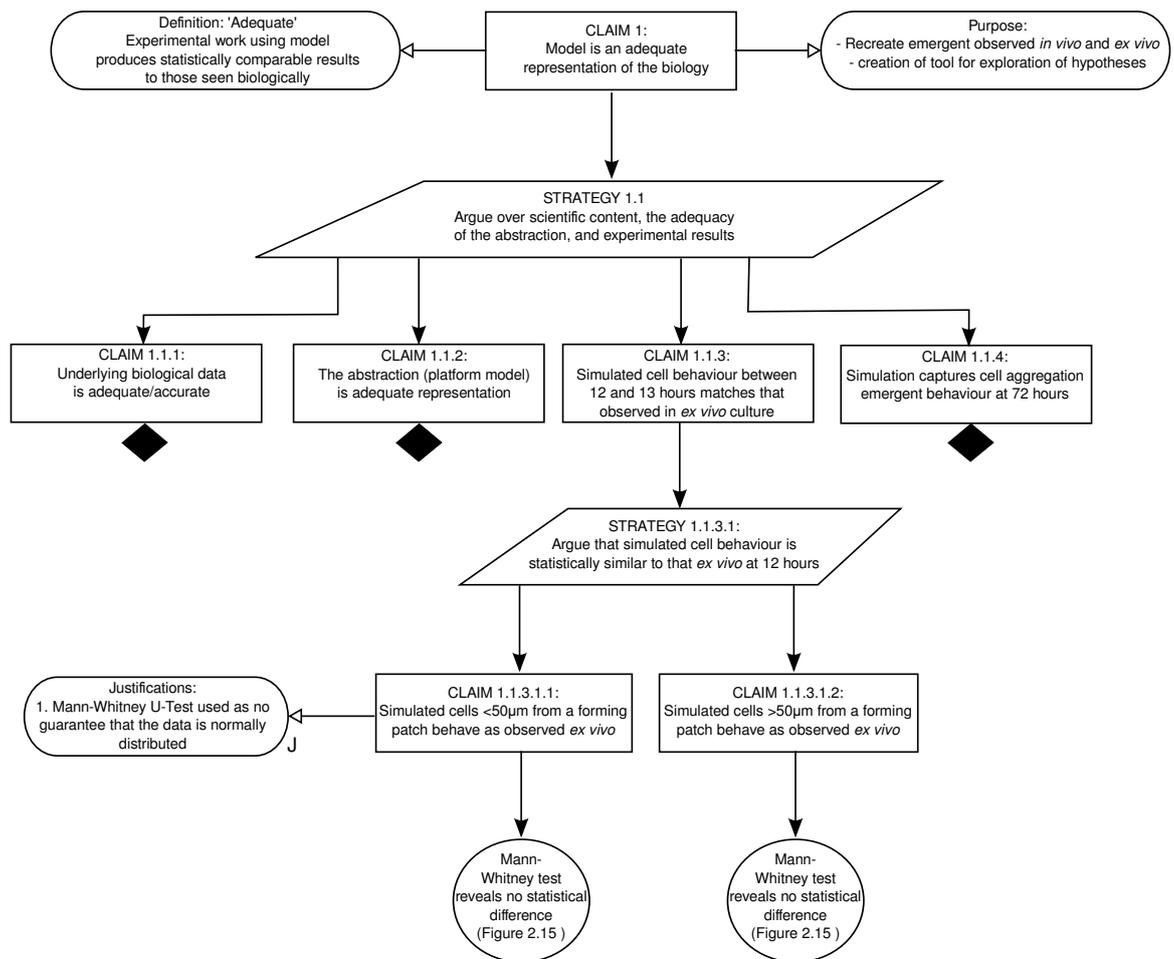
Figure 2.16: Argument-Based Validation for the development of the Peyer's Patch Simulation. This is the top level. A claim is made that the simulation is an adequate representation of the biology, and arguments put forward to support this where possible. This is broken down into four subclaims. A black diamond shows that the claim has been developed in another figure due to limitations on space. These follow on the next pages. Claim 1.1.3 has been developed in this figure, noting the evidence that simulated cell behaviour at the twelve hour time-point is statistically similar to that observed *ex vivo*, and where this result can be found.
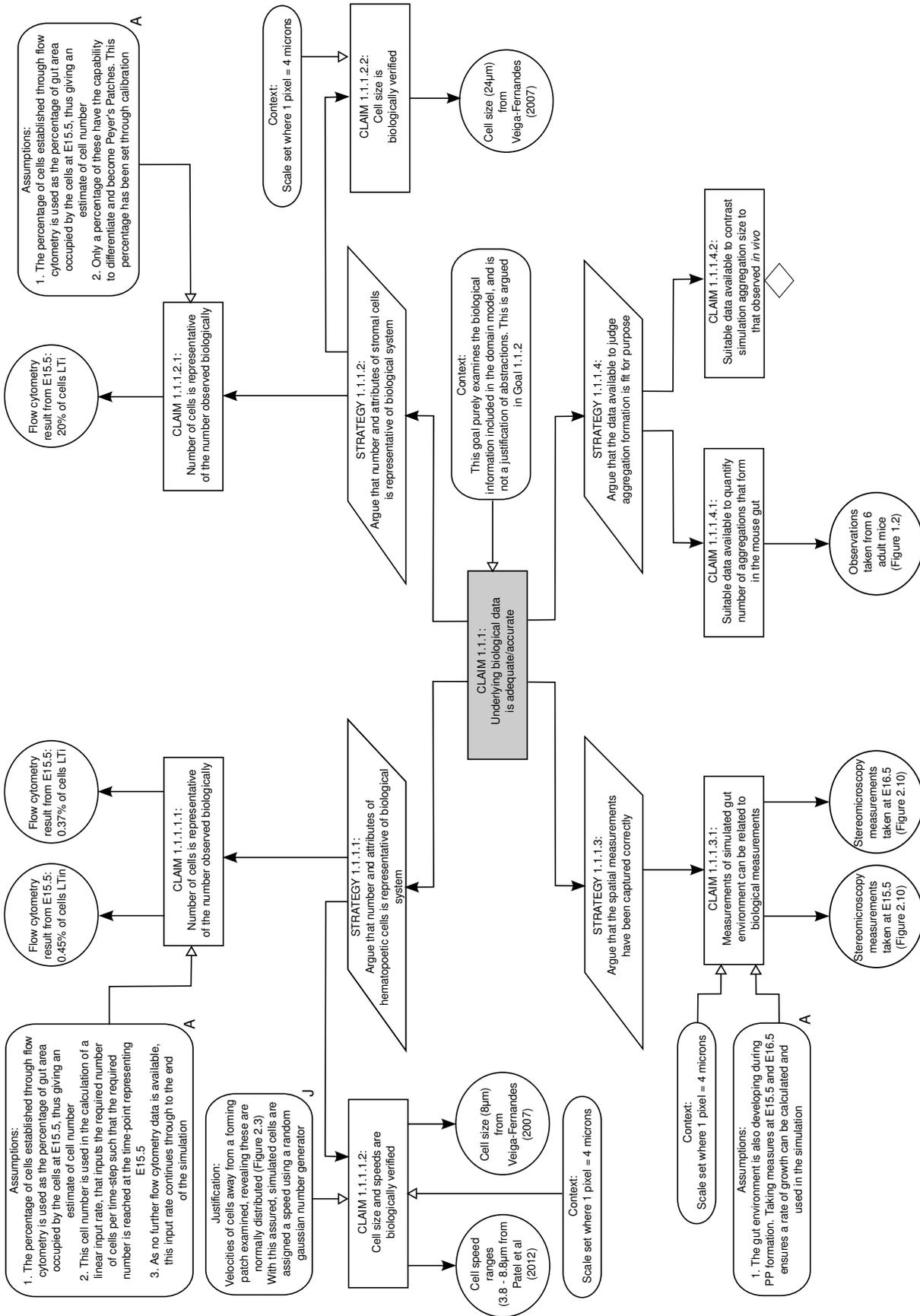
Figure 2.17: Argument-Based Validation for the development of the Peyer's Patch Simulation - Claim 1.1.1. This claim argues that the biological data included within the model and on which the simulation is judged is adequate. In the majority of cases evidence is provided for subclaims of the main claim. Where data is unavailable, the claim cannot be met and needs more development, and is noted with a blank diamond.
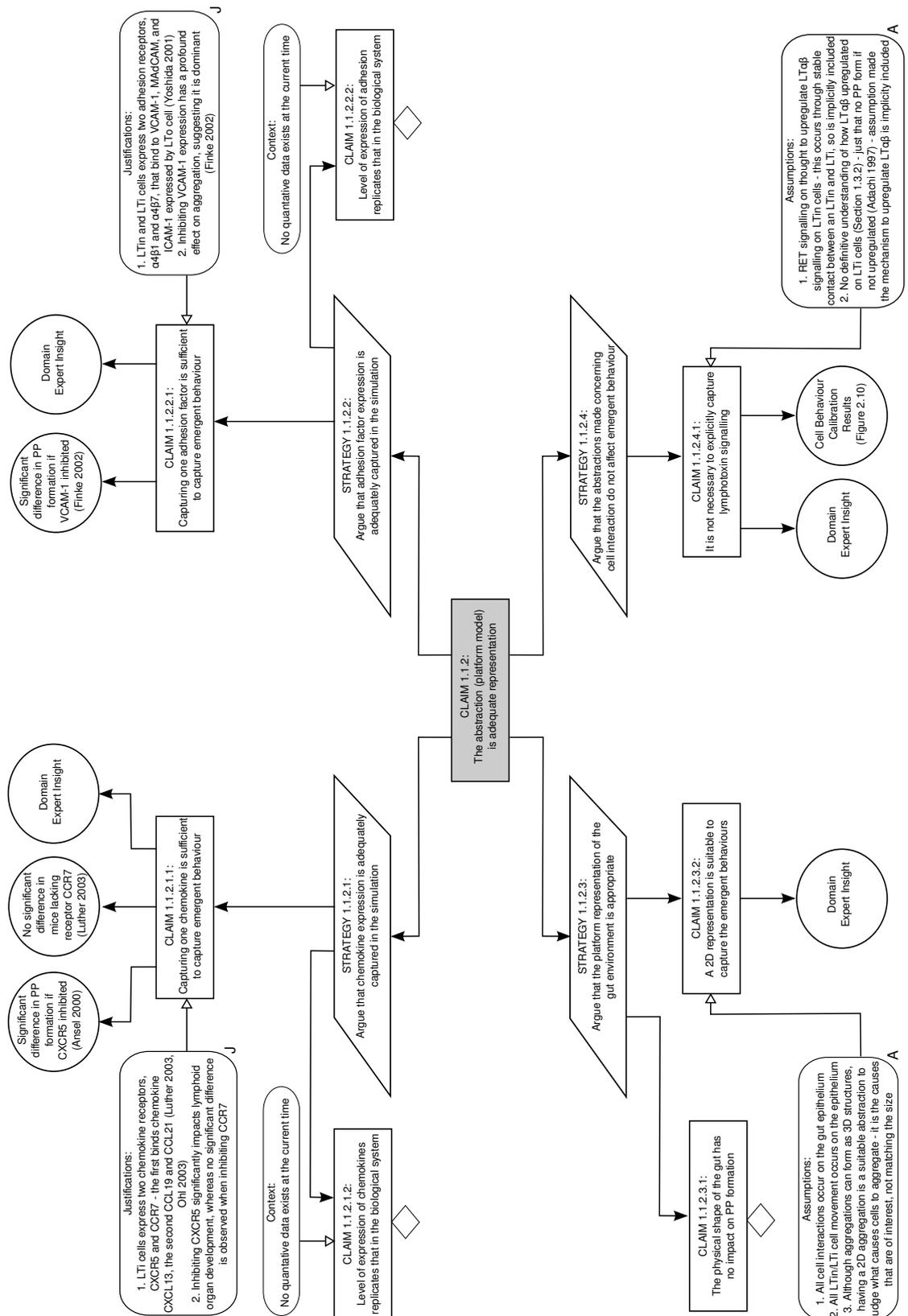
Figure 2.18: Argument-Based Validation for the development of the Peyer's Patch Simulation - Claim 1.1.2. This claim examines the abstractions that have been made, and whether these are suitable. Thus in this case the implementation of chemokines, adhesion factors, the environment, and cell signalling is explored. In the majority of cases evidence is provided for subclaims of the main claim. Where data is unavailable, such as chemokine and adhesion factor expression levels, the claim cannot be met and needs more development, and is noted with a blank diamond.
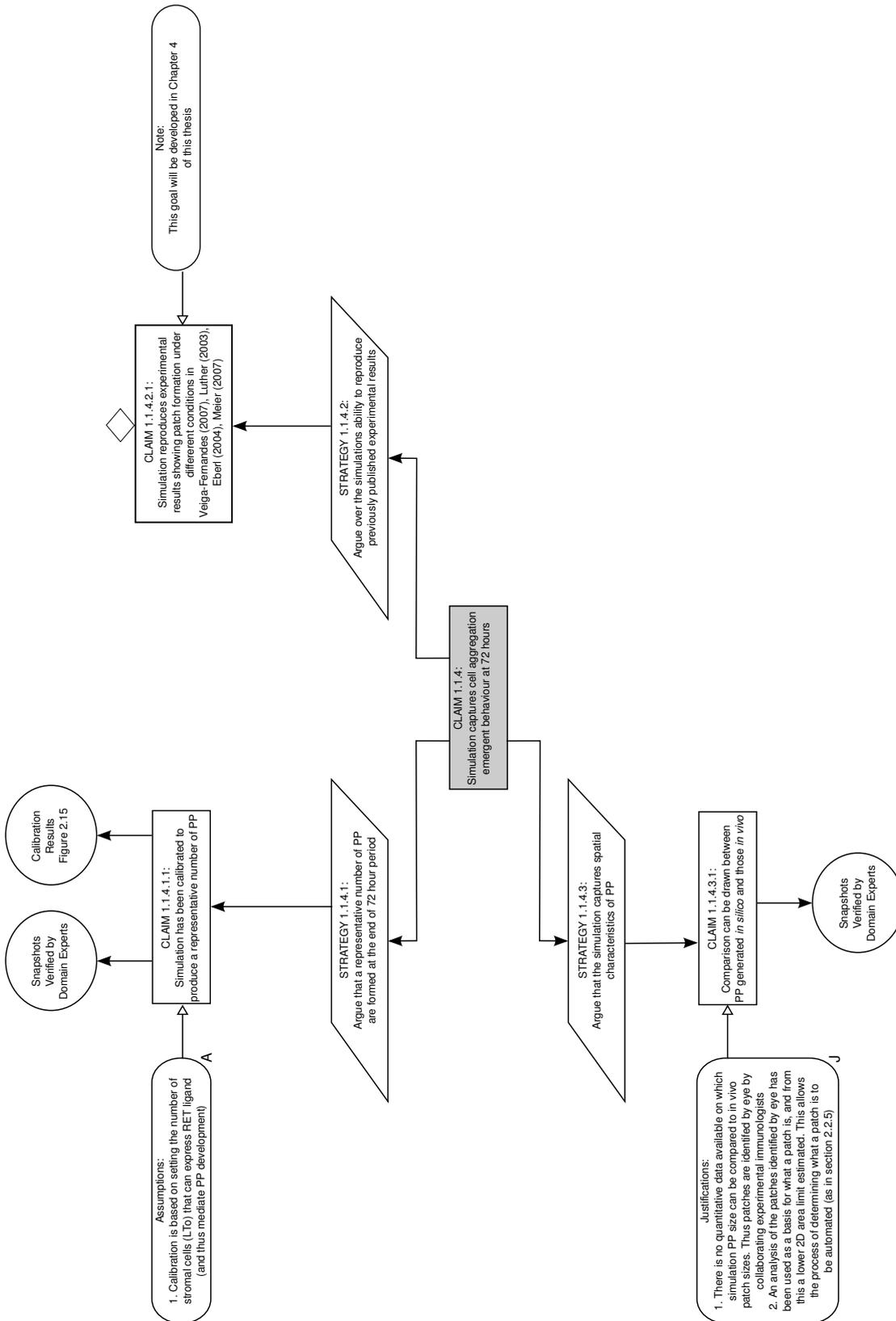
Figure 2.19: Argument-Based Validation for the development of the Peyer's Patch Simulation - Claim 1.1.4. This claim states that the simulator correctly replicates the second observed emergent behaviour, cell aggregations that become PP. It is noted that one of the claims, that previously published results are replicated, will be examined in this thesis. Where data is unavailable, the claim cannot be met and needs more development, and is noted with a blank diamond.

2.4 Developing the *spartan* Statistical Analysis Toolkit

Results generated through simulation may be affected by uncertainty caused by aspects of the biological system that are currently unknown and need to be assumed, and by uncertainty introduced in the implementation of the simulator (Helton, 2008). Such uncertainty may be present in two forms: *aleatory* uncertainty that arises through stochasticity inherent in both the biological and simulated system, and *epistemic* uncertainty arising as the values of some simulation parameters cannot currently be defined (Helton, 2008). Although integrating computer simulation with current experimental techniques has become a popular approach in furthering the understanding of biological systems (Germain *et al.*, 2011), in many cases where this approach is applied little attempt is made to reveal how representative the simulation result is in terms of the biological system it has captured (Read *et al.*, 2012).

The increase in popularity of computer simulation as an tool for exploring the dynamics of biological systems has led to the development of a number of packages that aid simulation development, as detailed in section 1.4.2. However there is no comprehensive statistical analysis package available to help determine how representative a simulation is of the biological system it has been constructed to represent and understand how results generated from the simulation can be interpreted in the context of that biological system. Uncertainty and sensitivity analysis techniques have recently found application in exploring results from biological simulations in order to appreciate these factors and the effect of uncertainty on simulation results (Marino *et al.*, 2008; Ray *et al.*, 2009; Read *et al.*, 2012). An application of these techniques provides the means to understand the relationship between the simulation and the real system and to provide some biological insight.

This section details the creation of a package of statistical techniques to aid the understanding and analysis of results generated through simulation. This package has been called *spartan* (**S**imulation **P**arameter **A**nalysis **R** **T**oolkit **A**pplicatio**N**) and provides implementations of previously described uncertainty and sensitivity analysis techniques (Marino *et al.*, 2008; Read *et al.*, 2012; Saltelli *et al.*, 2000) that when compiled as one package provide a comprehensive toolkit to explore the effect of uncertainty on simulation results. *Spartan* has been developed in the open-source R statistical environment and is freely available from the R package repository or for download from http://www.cs.york.ac.uk/spartan. This should help encourage simulation developers to perform such analyses to reveal how representative a simulation result is in terms of the biological system being captured, results that can then be published alongside their results for full scrutiny.

The *spartan* toolkit is utilised in conjunction with the simulator developed above in the chapters that follow to further understand the development of secondary lymphoid organs in the gut. The package contains four techniques, all of which are utilised in this thesis, each providing a different method of analysing results from the simulation

with the aim to understand the effect of uncertainty on results and provide some novel biological insight. The remaining part of this section examines each of the techniques that has been included.

2.4.1   Mitigating Aleatory Uncertainty

In agent-based simulations such as the lymphoid tissue formation simulation developed in this chapter, agent behaviour is affected by use of pseudo-random number generation. Thus, different results are produced, although input parameter values remain constant. Prior to the simulators use as a predictive tool, it is critical that the effect inherent stochasticity has on results is understood (Helton, 2008). To mitigate the effect of this aleatory uncertainty and achieve a representative result, replicate simulation runs are necessary. To determine the number of replicates required (n) that reduces the uncertainty to level at which the result can be considered representative of the condition on which the simulator is being run, while considering computational resources, the technique described by Read et al (2012) has been included within *spartan* .

To establish n, a number of replicate run sizes (sample sizes) are chosen. Taking a sample size of five as an example, twenty simulation result sets are obtained, with each of the twenty sets containing the results from five simulation runs. From the results of each simulation run, medians are calculated for each of the simulation responses. These are collated to form a set of median responses for each of the twenty subsets. Thus, in this case we have 20 sets of median responses, each of which contains the medians of each response of five simulation runs. The effect of uncertainty between the 20 sets of results is quantified using the Vargha-Delaney A-Test (2000), a non-parametric effect magnitude test that establishes scientific significance by contrasting two populations of samples and returning the probability that a randomly selected sample from one population will be larger than a randomly selected sample from the other. Results above 0.71 or below 0.29 indicate a scientifically significant difference between the populations, and 0.5 indicates no difference (Table 2.5) (Vargha and Delaney, 2000). The responses from the first set are compared with the remaining response sets in turn. Repeating this procedure for different sample sizes determines how many simulation samples should be compiled in generating averaged results to reduce the scientific effect of stochasticity to an acceptable level. To achieve a representative result, there should be no statistical difference in all twenty comparisons. The *spartan* package produces a plot for each sample size, detailing the A-Test result for each of the twenty comparisons. A summary plot is then produced revealing the maximum A-Test result for each sample size, helping determine the number of simulation runs required to mitigate aleatory uncertainty.

Effect Size	Large	Medium	Small	None
A-Test Score	<0.29 & >0.71	<0.36 & >0.64	<0.44 & >0.56	0.5

Table 2.5: A-Test magnitude effect sizes as specified by Vargha and Delaney (2000). An A-Test score is between 0 and 1, with 0.5 representing no difference between two distributions. As this is a magnitude the result has a direction, and thus there are small, medium, and large boundaries either side of 0.5

2.4.2 Parameter Robustness Analysis

The simulation features parameters for which values are unknown or cannot currently be determined. This may be as currently available techniques cannot determine the biological value, or caused through the translation of biological information into a format that can be implemented within a simulation. Robustness analysis examines the implications of biological uncertainty or parameter estimation on simulation results. Where a simulation is found highly sensitive to the value of these parameters, caution must be exercised in interpretation of results; they may be artefacts of parametrisation rather than representations of the biology (Helton, 2008).

Robustness to parameter perturbation can be explored using a one at a time approach (Read *et al.*, 2012). A set of simulation parameters of interest is determined. Taking each in turn, the value of that parameter is adjusted, with all other parameters remaining at their calibrated value. The Vargha-Delaney A-Test described previously (Vargha and Delaney, 2000) is employed to determine if changing the parameter value has led to scientifically significant behavioural alteration in contrast to the baseline simulation. This indicates how robust the simulator is to an alteration in the value of each parameter, and can indicate the validity of results produced by the simulator when considering results over a biologically accepted range of values.

For an agent-based implementation such as the simulator developed here, replicate runs are required to mitigate the effect of aleatory uncertainty. *Spartan* takes this into account where this is the case and compares the distribution of simulation responses from a number of replicate runs with that from the number of replicate runs of the baseline simulation. The robustness to parameter change is thus being judged on a result that is representative of the condition on which the simulator was run. For each parameter examined, *spartan* produces a plot detailing the A-Test score for each parameter value, in comparison to a result from the baseline. The plot thus reveals the statistical change in simulation response caused by a change in parameter value, and any statistical affect that becomes apparent as the parameter value is increased or decreased.

2.4.3 Global Sensitivity Analysis: Sampling-Based Approach

Though robustness analysis elucidates any affects of perturbing single parameters, it cannot reveal compound effects that become apparent when two or more are adjusted

simultaneously. Global sensitivity analyses reveal such effects, and can indicate the parameters that have the greatest influence on simulation response. To identify such parameters, a sampling-based technique that perturbs the values of all parameters of interest simultaneously has been included within *spartan* (Read *et al.*, 2012; Saltelli *et al.*, 2000). A set of simulation parameters of interest is determined, and for each, a range of parameter values to explore. A number of simulation parameter value sets are then created through use of a latin-hypercube sampling approach (McKay *et al.*, 1979). This selects values for each parameter from the parameter space, aiming to reduce any possible correlations while ensuring efficient coverage of the space over a minimal number of samples (demonstrated in Figure 2.20).

Simulations are then performed for each set of parameter values generated. Where the simulation is agent-based, a number of replicate runs are performed for each set to mitigate the effect of aleatory uncertainty as described in section 2.4.1. The *spartan* package includes functionality to process these replicates and calculate the median of simulation responses observed for simulations run under the conditions specified by that parameter set.

Each parameter is then taken in turn, and simulation responses ordered by the value assigned to that parameter. A plot is produced for each simulation response detailing the value of the response observed against the parameter value. This eases the identification of any relationship between the value of that parameter and the simulation response, although a number of parameters are being perturbed simultaneously. For example, the trend of points on the graph may suggest that a simulation response, such as velocity, decreases as the value of a particular parameter increases. The plot may also reveal if such a trend only becomes apparent when that parameter is in a specific value range. A statistical measure is also provided through calculation of a Partial Rank Correlation Coefficient (PRCC), a robust measure for quantifying non-linear relationships between an input and output (Marino *et al.*, 2008), and the calculated value stated in the plot header. Correlations that become apparent can be attributed to the value of the parameter, and the parameters that have significant impact on simulation behaviour determined by the size of the effect identified.

2.4.4   Global Sensitivity Analysis: Variance-Based Approach

In this approach, simulation parameters are varied, and resultant variation in simulation response partitioned between those parameters. The extended Fourier Amplitude Sampling Test (eFAST), developed by Saltelli et al (Saltelli, 2004; Saltelli and Bollardo, 1998) is also a global sensitivity analysis technique, and has proven one of the most reliable methods among variance-based techniques (Marino *et al.*, 2008). This has been included in *spartan* to provide an alternative global analysis technique that can be contrasted to results generated by the technique above. A set of simulation parameters of interest is determined, and for each, a parameter value range to explore.
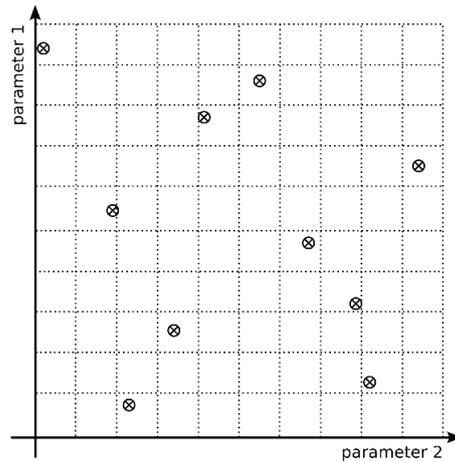
Figure 2.20: Image taken from Read et al (2012), exemplifying use of a latin-hypercube sampling approach for two parameters. The parameter space is split into subdomains, indicated by dotted lines. Ten samples have been taken, with one coming from each subdomain to ensure the parameter space is fully explored

Taking each in turn, values are chosen for all parameters through the use of sinusoidal functions of a particular frequency through the parameter space, with the frequency of the parameter of interest being much different to that used for its complementary set. This is demonstrated in Figure 2.21. A number of parameter values are selected from points along each of these curves. This creates a set of simulation parameters for each parameter of interest. Due to the symmetrical properties of sinusoidal functions, it is probable that the same parameter value sets could be selected. To address this, a resampling scheme is encouraged where a phase shift is introduced into each frequency, and sampling repeated (Marino *et al.*, 2008; Saltelli *et al.*, 2000). Thus, a number of parameter value sets are created for each parameter of interest. This process is repeated for an extra parameter, the dummy, which has an arbitrary value range but no impact on simulation behaviour. This enables a comparison between the impact of each parameter and one known to have no effect on simulation response. As an example of sampling using this approach, for 7 parameters, plus a dummy, three resample curves, and 65 parameter values from points along the curves, 1,560 sets of parameters would be produced. For analyses where a large number of parameters are explored, this technique could be computationally expensive (Ratto *et al.*, 2007; Tarantola *et al.*, 2006).

Simulations are performed for each set of parameter values generated. As described with the latin-hypercube technique above, *spartan* includes functionality to process replicate runs required to mitigate aleatory uncertainty in agent-based simulations, through the calculation of medians for each simulation response under the conditions set by that parameter set.

Results generated are analysed taking into account the frequencies that were used to generate that parameter set. Through Fourier analysis using these frequencies, variation in output can be partitioned between the parameters, giving an indication of

the impact each has on simulation response. Using the equations given in Marino et al (2008), two sensitivity indexes are calculated for each parameter: an eFAST First-Order Sensitivity Index (Si) and eFAST Total-Order Sensitivity Index (STi). The first indicates the fraction of output variance that can be explained by the value assigned to that parameter. The latter indicates the variance caused by higher-order non-linear effects between that parameter and the others explored. The *spartan* package produces plots of these measures for each simulation response. To determine whether a parameter has a significant impact on simulation response, these sensitivity indexes are compared to those calculated for the Dummy using a two-sample t-test.
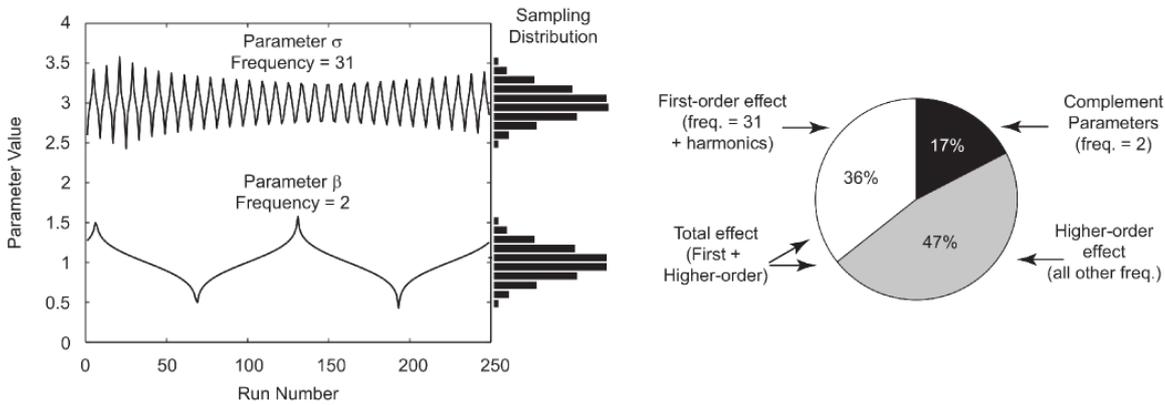

Figure 2.21: Parameter sampling and analysis using the extended Fourier Amplitude Sampling Test (eFAST). Left: Input parameter sampling - each parameter is varied through use of a sinusoidal curve of a particular frequency, and values chosen from points along the curve. The image shows sampling for two parameters. The parameter of interest is assigned a significantly different frequency. Right: Chart detailing the sensitivity indexes for each parameter. The first-order sensitivity index Si, or the fraction of output variance that can be explained by the value assigned to that parameter, is in white; the higher order effects between parameters STi is in grey; the remaining variance SCi is explained as variance accounted for by the parameters complementary set. Image taken from panels A and D from Figure 3 of Marino et al Marino *et al.* (2008).

2.5   Use of the Simulator and *spartan* to Explore Lymphoid Tissue Development

The simulation and *spartan* package provide the tools necessary to perform an *in silico* exploration of lymphoid tissue development. The following chapters make use of these tools to explore how changing the conditions under which the simulator is run can provide additional biological insight that could inform future laboratory experimentation. This section details the methods that are used in producing these results and hypotheses.

2.5.1   Analysing Changes in Cell Behaviour

Section 2.2.6 described how the simulation has been calibrated such that emergent cell behaviour replicates that observed *ex vivo* (Patel *et al.*, 2012). The simulation provides the functionality to perform this analysis through tracking simulated cells for a period of one hour at the twelve hour time-point, to produce output files that can be processed by statistical tools. Thus it is possible to run the simulation under different parameter value conditions and utilise *spartan* to assess the impact this has on cell behaviour responses. Additionally, the simulator provides functionality to produce this output for any time point in the simulated 72 hour period. This makes it possible to examine if cell behaviour at the end of the simulated period differs from that at the twelve-hour time-point, and note if the influence of simulated biological factors changes over time. This is the focus of Chapter 3.

Chapter 5 takes this a stage further, and utilises a combination of the simulator and *spartan* package as a tool to perform a time-lapse analysis of cell behaviour responses. This involves running the simulation under different conditions and capturing cell behaviour responses at twelve-hour intervals. Techniques in the *spartan* package can then be utilised to determine if and when the influence of simulated biological factors changes, providing biological insight difficult to obtain in the laboratory.

2.5.2   Contrasting Simulator With Published Results

With regard to the second emergent behaviour captured by the simulation, the formation of PP, comparisons between a simulator result and published experimental results have been made visually. It was noted in section 2.2.5 that X and Y coordinates of all LTin and LTi cells are output from the simulation at the end of the run. These cell co-ordinate files are processed in the R statistical environment to produce a plot showing the formation of PP across the simulated gut length. In Chapter 4, these visual images are contrasted with phenotypes in the relevant publications observed through use of antibody staining of LTi and LTo cells, with the help of collaborating experimental immunologists.

2.5.3   Simulating Gene-Deficient Mice Experiments

To simulate gene-knockout experiments, with the aim to replicate previously published results *in silico*, boolean parameters have been included in the simulator. A gene knockout is indicated by setting the relevant knockout parameter to true. Three boolean parameters have been included that model a knockout of RET, chemokines, and adhesion factors respectively.

2.5.4   Simulating Experiments that have Examined Reduced and Over Expression of Biological Factors

To simulate changes to level of expression or cell numbers, values for relevant parameters in the platform model can be adjusted as required prior to simulation run. Simulation parameter values are specified in an XML file, that is read by the simulator when the run starts. A reduced or overexpression is modelled by changing the values within that file. For example, experimentation in Chapter 4 examines the impact of a change in LTin cell number on PP formation. This study is performed by simply running the simulation with different values for that parameter.

Chapter 3

Factors Influencing Hematopoietic Cell Behaviour in Peyer's Patch Development

Laboratory explorations of hour 12 of Peyer's Patch development have suggested there is a statistically significant difference between hematopoietic cell behaviour near a forming Peyer's Patch and cell behaviour further away. The previous chapter detailed how a computer simulation has been implemented that replicates this emergent behaviour, producing results that are statistically similar to cell behaviour observed ex vivo. This chapter applies the techniques in the spartan toolkit developed in the course of this study to analyse simulation results and suggest the biological factors that could be causing this change in cell behaviour during hour 12. Furthermore, the same analysis techniques are applied to examine simulated cell behaviour during the final hour of organ formation, to determine if the influence of biological factors changes over the course of development.

3.1 Introduction

The previous chapter describes the development of a set of tools that can be used to perform an *in silico* exploration of lymphoid tissue formation. This detailed how, through a process of calibration, it has been ensured cell behaviours that emerge at an early time-point in simulation are statistically similar to that seen in *ex vivo* observations at the same time-point (Patel *et al.*, 2012). The *ex vivo* cell tracking analysis reveals a statistically significant change in hematopoietic cell velocity and displacement (LTin/LTi cells) when in the vicinity of an ARTN-soaked bead, placed to model cellular behaviour around a stromal (LTo) cell. Such behavioural alterations are thought to occur through interactions between the cells, mediated by adhesion and chemoattractant factors expressed by LTo cells (van de Pavert and Mebius, 2010; Randall *et al.*, 2008; Veiga-Fernandes *et al.*, 2007). As calibration and validation results have provided confidence that the simulator is a suitable representation of the development process, explorations in this chapter use the simulator as a tool to suggest the factors that could be influencing this change in cellular behaviour at the twelve hour time point.

The simulation is however an abstraction of the biological system that it captures, and this separation must be appreciated when interpreting *in silico* explorations with respect to the biological system under study. Such simulation results may be affected by uncertainty arising from aspects of the biological system that are currently unknown and thus needed to be assumed, and by uncertainty introduced in the implementation of the simulator (Helton, 2008). In Chapter 2, a set of statistical techniques to appreciate the effect of uncertainty in simulation results was described. In this chapter the Simulation Parameter Analysis R Toolkit Application (spartan) package of statistical techniques compiled in completion of this thesis (Alden *et al.*, 2012a) is utilised to appreciate the effect of uncertainty on results generated from this simulation and to explore the factors influencing cell behaviour in the vicinity of an LTo cell. Techniques available within *spartan* can be used to determine how representative the simulation is of its biological system and understand how *in silico* results can be interpreted in the context of the biological domain. When brought together, these techniques provide a comprehensive set of tools that work towards establishing the relationship between the simulation and the biological system, enabling the use of the simulator as a tool for providing such novel biological insights. Using *spartan*, the number of simulation samples required to mitigate stochastic effects and attain a desired level of experimental accuracy is determined, confidence is built that results are representative of biology as opposed to parameterisation artefacts resulting from epistemic uncertainty, and valuable biological insight is gained through rigorous statistical analysis of simulation results.

This chapter begins by examining the effect of aleatory uncertainty on simulated cell behaviour at the twelve hour time-point (Section 3.3). As the model has been implemented using an agent-based approach, each cell is represented as an individual

entity, and thus may behave differently to other agents of the same cell type. This captures the stochasticity in cell velocity and displacement that was observed in the *ex vivo* culture system (Patel *et al.*, 2012) described in section 1.3.3, which tracked cell behaviour for an hour at the twelve-hour time-point, data on which this simulator has been calibrated (Section 2.2.6). It was noted in section 2.4.1 that such stochasticity does however imply that different simulation runs under the same parameter conditions will produce differing results. In that section a method by which this uncertainty, termed aleatory uncertainty, could be mitigated was described. This is utilised in this chapter to ensure results are produced that are representative of the condition on which the simulation was run.

With the above technique ensuring the effect of aleatory uncertainty is mitigated, the parameters of the simulation can now be perturbed to examine the effect each has on simulated cell behaviour. As noted previously, the simulation is comprised of a number of parameters for which a value is not yet known. Although suitable values have been obtained through a process of calibration, there remains some uncertainty in the true value of these parameters, termed epistemic uncertainty (Helton, 2008). This chapter continues by using techniques within the *spartan* package to explore how robust simulated cell behaviour during hour 12 is to an alteration in the values of these unknown parameters (Section 3.5.1). Where cell behaviour at this time-point is found to be highly sensitive to parameter value, it must be considered whether this sensitivity is caused by parameterisation or whether this is a true representation of the biology. Such parameters have been identified using a technique that perturbs their value independently of all other parameters over a set range (Read *et al.*, 2012), an approach described in detail in section 2.4.2.

However, the effect one parameter has may rely on the value that is assigned to another. Further statistical techniques included within *spartan* and described in section 2.4 have been utilised to examine the effect on simulated cell behaviour of changing the value of all unknown parameters simultaneously (Sections 3.5.2 and 3.5.3). Such statistical approaches are called global sensitivity analysis techniques, and can be used to identify compound effects that occur although the values of all parameters in a subset are being perturbed. The identification of compound effects can indicate parameters that are highly influential in affecting simulated cell behaviour at this time-point, and in identifying such relationships, this analysis can offer unique biological insight.

Considering the results gained from each statistical technique together provides a means of suggesting the biological factors influencing the change in cell behaviour during hour twelve of development observed *ex vivo* (Patel *et al.*, 2012). However, the process of lymphoid tissue formation in the gut is known to continue for another 60 hours after that time-point (Mebius, 2003; Randall *et al.*, 2008). Thus, it would be interesting to reveal if the same conclusions concerning the impact of each factor are to be drawn at hour 72 as drawn for hour 12, or whether different biological factors

become influential at different time-points of development. The latter conclusion could potentially extend Adachi et al's (1997) hypothesis that PP formation can be split into distinct phases. The authors determine these phases to be the appearance of VCAM-1^+ stromal cells in the gut, the identification of clusters of LTi cells around VCAM-1^+ expressing stromal cells, and the recruitment of lymphocytes from E18.5. However, an exploration of the period between the appearance of LTo cells and clustering of LTi cells up to E17.5 where this simulation stops, through an analysis of cell behaviour, may suggest additional phases between Adachi et al's first and second phase, with different biological factors becoming influential at different time-points.

This chapter continues by exploring this hypothesis, and examining simulated cell behaviour after 72 hours of PP development, or E17.5. To do this, the simulation was run, and as performed previously, cells that are within 50μm of a primordial PP tracked for a period representing one hour, but this time during the 72nd hour of development. The factors that influence simulated cell behaviour were then determined using the same statistical techniques as those described above for the twelve hour time-point. Conclusions from these results can be contrasted with those produced after twelve hours of development, to determine if different factors are influential at the end of the PP development period. Should this be the case, the hypothesis that there could be different development phases within the 72 hour development window will hold.

3.2   Aims

Explorations in this chapter utilise the developed simulator and statistical techniques within the *spartan* package to achieve the following aims:

1. To determine the number of simulation runs required per run condition that attains a desired level of experimental accuracy, mitigating aleatory uncertainty.

2. To examine the implications of biological uncertainty or parameter estimation on simulation results at both hour 12 and 72 of development, ensuring a result is representative of the biology rather than parameterisation artefacts, and thus biological insights can be drawn.

3. To identify any compound effects present between two or more simulated biological factors at both hour 12 and 72 of development, indicating the pathways and components that have a substantial effect on simulation behaviour at that particular time-point.

4. To partition the variance in results caused by perturbing the values of each factor, and thus establish how sensitive the simulation and biological system is to the value of each factor at both the 12 and 72 hour time-point.

5. To contrast the results gathered in achieving the above aims, to determine whether the impact of a biological factor changes over time.

3.3 Mitigating the effect of Aleatory Uncertainty

The effect of aleatory uncertainty can be mitigated by performing a number of simulation runs under identical conditions. To determine the number of simulation runs required to obtain a representative result, sample sizes of 1, 5, 50, 100, 300, 500, and 800 simulation runs were analysed using the technique described by Read et al (2012) detailed in section 2.4.1. The objective is to reduce the variance in simulation output response measures, in this case the cell behaviour measures of Velocity and Displacement between hours twelve and thirteen. Figures 3.1a, 3.1b, 3.1c show the A-Test scores for these simulation output responses in each of the 20 result sets, for 5, 100, and 500 samples respectively. Figure 3.1d shows the maximum A Test score for each simulation response over the 20 result sets, for all sample sizes analysed. The latter indicates that reducing the effect magnitude of aleatory uncertainty on simulation results to less than small (the desired level, as defined by Vargha-Delaney (2000) and listed in Table 2.5) requires 500 simulation runs. Thus, 500 runs should be performed for each investigation conducted using this simulator, where the focus is on examining changes in cellular behaviour.

3.4 Investigating the Impact of Factors for Which No Value is Currently Known

In this simulation there are six parameters for which the value is uncertain:

(a) Probability at which an LTin/LTi cell will form a stable bind with an LTo cell upon contact. A stable bind is defined as contact that leads to LTo cell differentiation and an increase in expression of adhesion and chemoattractant factors.

(b) Initial level of chemokine expression upon LTo differentiation.

(c) Saturation limit of chemokine diffusion.

(d) Level of chemokine required in a cell's local environment to induce LTi cell chemotaxis.

(e) Level at which surface adhesion factors are expressed with each stable contact between an LTin/LTi and LTo cell.

(f) Probability that the level of adhesion factors expressed on the surface of an LTo cell will restrict LTin/LTi cell movement.

The remaining analyses in this chapter explore the effect uncertainty in the value of these parameters has on simulation response. A change in response is measured by examining alterations in two simulation output responses: velocity and displacement of cells within a 50μm distance of an LTo cell. These responses are captured for each cell within that distance over a period representing one hour, at both the simulated 12 and 72 hour time-points. The range of values explored for each of these parameters is specified in Table 3.1.

Parameter	Baseline Value	Lower Limit	Upper Limit
stableBindProbability*	0.5	0.0	1.0
chemokineExpressionThreshold*	0.3	0.0	1.0
initialChemokineExpressionValue	0.20	0.10	0.50
maxChemokineExpressionValue	0.04	0.015	0.08
adhesionFactorExpressionSlope	1	0.25	5.0
maxProbabilityOfAdhesion*	0.65	0.1	1.0

Table 3.1: The six simulator parameters for which a value is not currently known, the value each has been set in calibration, and the ranges explored using sensitivity and uncertainty analysis techniques. * denotes the parameters for which a full range of potential values has been explored.

3.5   Examining Hematopoietic Cell Behaviour During Hour 12 of Development

With the number of replicate runs required per condition established, the statistical techniques available within the *spartan* package can be used to explore the impact of each parameter in Table 3.1 on simulated cell behaviour during hour twelve of development.   This hour is considered first as this is the time-point at which the simulator has been calibrated, using *ex vivo* data from the same hour of development.

3.5.1   Simulation Robustness to Parameter Perturbation

One-a-time analysis (Read *et al.*, 2012) was used to determine how sensitive the simulation behaviour during hour twelve is to the value of each parameter in Section 3.4. Each parameter was examined in turn, and its values perturbed over the range of values specified in Table 3.1. Five-hundred simulation executions were performed for each parameter value in accordance with the aleatory analysis in Section 3.3. The distribution of response values obtained for each parameter value is contrasted with a distribution obtained using baseline parameter values using the Vargha-Delaney A-Test (2000), as detailed in section 2.4.1.

(i) **Chemokine Related Parameters**

Figures 3.2a and 3.2b show the effect of adjusting the initial level of chemokine expression upon LTo cell differentiation and the saturation limit of chemokine diffusion respectively. Ranges of values were chosen such that the sigmoidal curve used to model the distance at which the chemokine is diffused (explained in Figure 2.8) explores a significant range either side of the calibrated value. For both parameters, this analysis indicates that perturbing the expression level of chemokines at this early time point has no statistical effect on the behaviour of cells in the vicinity of a forming patch.

Figure 3.2c shows the effect of adjusting the probability that an LTi cell will not respond to chemokine expression in its locality. All potential values for this parameter have been explored, and a similar conclusion to that above is drawn, that altering this parameter has no statistical effect on cell velocity at this time-point. However there is a small effect on cell displacement when set to the parameters extreme upper value, where the cell will never respond to chemokine expression.

Thus early in PP formation the model predicts that chemokines are unlikely to be the key force driving tissue formation and causing the statistically significant change in cellular behaviour.

(ii) **Cell Binding Probability Parameters**

Figure 3.2d reveals an alteration in the probability that a stable interaction occurs when an LTin/LTi cell is in contact with an LTo cell has no effect on cell behaviour for all values except when set to the lower extreme. Again all possible values have been explored for this parameter. As the lower extreme value is a probability of zero, no stable binding would occur, meaning no LTo differentiation, and thus no expression of adhesion or chemoattractant factors that influence cell behaviour. As such this effect is an expected result.

(iii) **Adhesion Factor Related Parameters**

In contrast to Figures 3.2a-d, Figure 3.2e reveals a trend in cell velocity A-Test response when the maximum probability that adhesion factor expression affects cell motility is perturbed. This suggests that simulated cell behaviour is sensitive to the value of this parameter, and its value may only lie in a small range either side of its calibrated value, after which the simulation would produce results that are, on the basis of data seen in calibration, biologically implausible. A change in parameter value does however have no significant effect on displacement unless set to the extreme upper value, where adhesion factors will always retain a cell within a primordial patch.

An alteration in the level of adhesion factors expressed by an LTo cell upon stable contact also reveals a significant change in cell velocity response (Figure 3.2f). Although not classified as a large difference by the A-Test bounds set by

Vargha and Delaney (2000), results either side of parameter values 0.5 and 2.25 do fall outside the category that the authors deem as a 'small' difference and into that deemed a 'medium' difference. Again these results suggest that the value of this parameter is restricted within a certain window of values, between 0.5 and 2. Surprisingly, there is little statistical difference between adhesion expression values that are two, three, and four times the calibrated value, suggesting that once an initial overexpression has occurred, increasing this further has no impact on simulated cell behaviour. In contrast, the value of adhesion factor expression parameters seems to have no statistically significant effect on cell displacement at this time-point.

3.5.2   Identifying Compound Effects at 12 Hours Through Simultaneously Perturbing all Unknown Parameter Values

Using the latin-hypercube sampling approach described in section 2.4.3, 500 sets of simulation parameter values were generated, with each parameter being assigned a value within the ranges specified in Table 3.1. For each set of values, 500 simulation runs were performed to mitigate the effect of aleatory uncertainty. Median cell velocity and displacement measures were calculated for each run, thus generating a distribution of 500 median cell behaviour responses at hour 12 for each parameter set. In contrast to the above, where distributions of results were being compared using the A-Test, in this case the median of this set of medians is calculated for both cell behaviour measures, and assigned as the simulation behaviour response under the conditions specified in that parameter set. Taking each of the six parameters specified in Table 3.1 in turn, simulation responses were ordered by the value assigned to that parameter, and plots generated for each output response (Figures 3.3 and 3.4), detailing the responses observed for all values assigned to that parameter. Compound effects are noted by the identification of any trends on the plot, and value of the Partial Rank Correlation Coefficient (PRCC) specified in the plot header. Both the generated plot and the PRCC value can be used to suggest the parameters that are highly influential on simulation behaviour, and provided unique biological insight into the factors that are important at this stage of tissue development.

(i) **Chemoattractant Related Parameters**

Both the plots in Figures 3.3a and 3.3b and the respective PRCC values reveal no trend between simulation cell velocity responses and the value assigned to the parameters that capture expression of chemoattractant molecules. For the cell displacement measure, the same result is found for the initial level of chemokine expression (Figure 3.4a), but a small trend does become apparent when altering the parameter that limits chemoattractant expression (Figure 3.4b)

Similarly, an alteration in the probability that an LTi cell does not respond to chemokine expression in the environment (Figure 3.3c) reveals no compound effects when examining cell velocity. The same can be said for the cell displacement measure, unless the parameter is assigned a value near the upper extreme. At this point a trend does emerge, suggesting a significant effect on cell behaviour although all other parameter values are also being perturbed. This supports the previous finding in one-a-time analysis, where changing the same parameter independently produced no significant effect on displacement unless set to its extreme value, where the cell never enters a phase of chemotaxis towards a forming patch (Figure 3.2c).

(ii) **Cell Binding Probability Parameters**

For both cell velocity and cell displacement measures (Figures 3.3d and 3.4d), there is no trend between results obtained for the 500 parameter sets and the value assigned to the probability that a stable bind occurs between two cells that are in contact.

(iii) **Adhesion Factor Related Parameters**

Figure 3.3e reveals a very strong trend between the simulation cell velocity responses and the maximum probability that an LTin/LTi cell remains in prolonged contact with an LTo cell. There is a marked reduction in velocity as this probability increases, although the values of the other five unknown parameters (Table 3.1) are also being perturbed. Such a strong correlation suggests that cell response to adhesion factor expression is a key factor in affecting cell velocity. For the related parameter, that captures level of adhesion factor expressed with each stable LTin/LTi cell contact with an LTo cell (Figure 3.3f), a visual effect is apparent, but no significant correlation between the level of expression and velocity. At this early time-point of development, few stable contacts may have occurred, and combining this with a low level of expression with each contact means that not enough adhesion factors are expressed to impact cell velocity. An increase in the parameter value however does begin to reveal the retention affect. However, the absence of a clear trend between parameter value and velocity does suggest that there is a large amount of uncertainty in the value of this parameter. This conclusion is drawn as similar affects on cell velocity can be observed for both high and low parameter values.

For the cell displacement responses, there is no trend between simulation response and the level of adhesion factor expression with each stable contact (Figure 3.4f). Similar to the displacement results described in the chemoattractant section, a trend does become apparent for extreme values assigned to the parameter that captures the maximum probability that an LTin/LTi cell remains in prolonged contact with an LTo cell (Figure 3.4e).

3.5.3   Partitioning Variance in Simulation Response Between Parameters

This analysis examined the six parameters specified in section 3.4, to determine the proportion of variation in simulation response during hour 12 of development that can be explained by perturbing the value of each parameter. Through use of the eFAST approach (Marino *et al.*, 2008; Saltelli, 2004) the sensitivity of the simulation to each parameter is determined and quantified, and thus suggests the impact of each biological factor on tissue development.

Using the sinusoidal curve sampling approach, 500 parameter value sets were generated, with each parameter being assigned a value within the ranges specified in Table 3.1. To determine the parameters that have a significant influence on simulation output, the eFAST analysis technique requires an additional 'dummy' parameter that has no effect on simulation result. This 'dummy' parameter was assigned a value range of 1 to 10. Although the 'dummy' has no influence on simulation output, the algorithm will determine that a small proportion of variation in simulation response can be accounted for by the 'dummy'. This proportion is used for statistical comparison purposes, with the variance accounted for by each of the six parameters of interest being compared with that assigned to the dummy, to determine if a simulation parameter is more influential than one known to have no effect. With the 'dummy' added, there are seven parameters (six plus the dummy), with 65 parameter values taken from each curve, and three re-sampling curves employed, producing 1,365 parameter value sets, or 195 per parameter, using the sampling procedure detailed in Section 2.4.4. For each parameter value set, 500 runs were performed to mitigate aleatory uncertainty and median simulation response values calculated.

Simulation responses were analysed using the Fourier frequency approach described in section 2.4.4 (Marino *et al.*, 2008; Saltelli, 2004). Plots are created for each simulation output response (velocity and displacement), detailing the median First-Order (Si) and Total-Order (STi) sensitivity indexes calculated for each parameter of interest (Figures 3.5a and 3.5b), generated from the Si and STi values calculated for each re-sample curve. Sensitivity indexes and measures of significance in comparison to the 'dummy' parameter are detailed in Table 3.2.

Cell Velocity

Contrasting the first-order sensitivity indexes (Si) for each parameter with those of the dummy reveal that five of the parameters account for a statistically significant portion of simulation variance ($p < 0.05$). Of these, the probability that stable contact occurs between an LTin/LTi cell can be discounted as, due to the effect at the extreme value (as seen in Figure 3.2d and explained in 3.5.1) this is an expected result. A similar percentage of the variance is also explained by the chemoattractant expression

parameters, a result contrasting those in analyses performed previously that revealed chemoattractant expression has no appreciable effect on cell behaviour. Of the remaining two, which both capture adhesion factor effect and expression, the maximum probability that adhesion factors prolong cell contact is significant at the 1% level, supporting the inference in the analyses above that adhesion factors have a key role in affecting cellular velocity.

When using the total-order sensitivity indexes (STi) to identify any non-linear effects between the parameter and its complementary set, all STi values are deemed to be statistically significant. As each includes the significant variance caused by adjusting the maximum probability adhesion factors prolong contact, a value that the 'dummy' STi value does not include, this is to be expected.

Cell Displacement

When contrasted to the first-order sensitivity index of the 'Dummy' parameter, only two of the six parameters are found to have a statistically significant effect ($p < 0.05$) on cell displacement (Figure 3.5b). The first of these is the probability that stable contact occurs between an LTin/LTi and LTo cell on contact (stableBindProbability). For the reasons specified above in 3.5.3, this result is expected and thus can be overlooked. The second is the initial level of chemoattractant expressed on LTo cell differentiation (initialChemokineExpressionValue), further supporting the small compound effect identified in latin-hypercube analysis (3.4b).

T-Test results contrasting each parameter total-order sensitivity index (STi) with that of the dummy parameter reveals that none of the STi values are statistically significant. This suggests that there are no significant compound effects occurring between the parameters, in terms of influencing cell displacement. This supports the majority of the findings revealed in latin hypercube analysis (Section 3.5.2), with the exception of a small trend that occurs when two of the parameters are at extreme values (chemokineExpressionThreshold and maxProbabilityOfAdhesion).

3.6 Examining Hematopoietic Cell Behaviour During the Final Hour of Development

The analysis in the previous section suggests a key role for adhesion factor expression during hour 12, with chemokine expression having no appreciable effect on cell behaviour. This section repeats this analysis but for a simulation time period that represents hour 72, the final hour of PP development. The same techniques as used above are applied, and the same parameters examined (as specified in Table 3.1). This analysis has been performed to reveal if the conclusions drawn above hold for a later time-point of development, or whether the influence of each simulated biological factor changes over time.

Parameter	Si	P-Value	STi	P-Value
stableBindProbability	0.125	0.048*	0.217	0.029*
chemokineExpressionThreshold	0.124	0.063	0.269	0.010*
initialChemokineExpressionValue	0.123	0.048*	0.268	0.009**
maxChemokineExpressionValue	0.118	0.027*	0.271	0.049*
adhesionFactorExpressionSlope	0.342	0.048*	0.551	0.012*
maxProbabilityOfAdhesion	0.163	0.009**	0.310	0.008**
dummy	0.007		0.057	

(a) Cell Velocity Response

Parameter	Si	P-Value	STi	P-Value
stableBindProbability	0.148	0.022*	0.322	0.268
chemokineExpressionThreshold	0.122	0.061	0.313	0.288
initialChemokineExpressionValue	0.107	0.019*	0.331	0.246
maxChemokineExpressionValue	0.134	0.084	0.376	0.198
adhesionFactorExpressionSlope	0.039	0.212	0.362	0.197
maxProbabilityOfAdhesion	0.100	0.087	0.434	0.097
dummy	0.023		0.256	

(b) Cell Displacement Response

Table 3.2: Median sensitivity indexes and measures of statistical significance for each parameter examined using the eFAST technique, for both simulation cell behaviour responses during hour 12. Si: First-Order Sensitivity Index; STi: Total-Order Sensitivity Index. Both are calculated for each re-sample curve and the median value taken. P-Value calculated using two-sample t-test to the distributions comprised of the results from each re-sample curve. * indicates statistical significance at 0.05 level, ** indicates statistical significance at 0.01 level.

3.6.1   Robustness to Parameter Perturbation

Each parameter in Table 3.1 has been considered independently, and assigned a value within its set range prior to the simulation run. To mitigate aleatory uncertainty, 500 runs were performed for each value assigned to the parameter. The analysis uses the same one-a-time analysis technique (Read *et al.*, 2012) included within the *spartan* package as used to explore hour twelve of development in the previous section.

(i) **Chemokine Related Parameters**

Figures 3.6a and 3.6b show the effect of adjusting the initial level of chemokine expression upon LTo cell differentiation and the saturation limit of chemokine diffusion respectively. The analysis indicates that the value which captures the initial level of chemokine expression, through use of the sigmoidal curve function as described in Figure 2.8, has no impact on either cell behaviour measure at this time-point. However, a change in the maximum level of chemokine expression has a significant effect on both cell behaviour measures. A small change in this parameter value produces behaviour that is statistically significantly different, or to use the terminology in the A-Test, a 'large' difference (Vargha and Delaney, 2000). The calibrated value for this parameter is 0.04, and these results suggest

that there is only a small window +/- this figure before there is a change in cell behaviour that Vargha and Delaney classify as 'medium difference.' Thus unlike the twelve hour timepoint, the level of expression of chemoattractants, and thus the distance over which this diffuses, is highly influential at this time-point.

An alteration in the probability that an LTi cell does not respond to chemokine in the vicinity also significantly effects both cell behaviour measures. In terms of displacement, a change in the probability of just 0.2 produces cell behaviour that is significantly different to that observed *ex vivo*. The results suggest that the range of values that this probability can take can only fall within the range of 0.2-0.4. Thus the simulation is suggesting in this case that there is only a small possibility that an LTi cell may not respond to chemokine expression.

(ii) **Cell Binding Probability Parameters**

A stable bind between an LTo and LTi cell upregulates expression of chemokines and adhesion factors through lymphotoxin signalling. At the twelve hour time-point, the simulation was not sensitive to this parameter (bar the case where set to its extreme value of zero), however this is not the case during the last hour of development (Figure 3.6d). Similarly to the chemokine parameters above, cell behaviour at this time-point is highly sensitive to the value of this parameter. Again if the Vargha-Delaney 'medium difference' figures were to be taken as a guide, significantly different cell behaviour emerges outside a window +/-0.2 of the parameters calibrated value, and further increases outside this window produce significantly different cellular behaviour.

(iii) **Adhesion Factor Related Parameters**

Figure 3.6f reveals that an increase in adhesion factor expressed on stable contact between an LTo and LTin/LTi cell has no significant difference on either cell velocity or displacement. The value can also be halved with no impact on cell behaviour, however a reduction of more than half reduces the amount of adhesion factor in the environment to a level at which cell behaviour is significantly different.

3.6.2 Identifying Compound Effects on Cell Behaviour at E17.5 Through Simultaneously Perturbing all Unknown Parameter Values

The same 500 simulation parameter value sets that were generated for analysis at hour twelve were used in this analysis to ensure the two sets of results could be compared. Five-hundred simulation runs were performed for each parameter set to mitigate the effect of aleatory uncertainty on cell behaviour results, a figure established from analyses in section 3.3. Median cell velocity and displacement results were calculated for each

run at the 72 hour time-point, generating a distribution of 500 median cell behaviour responses, for which the medians were again calculated and considered representative simulator behaviour under the conditions specified in that parameter set. Taking each parameter in turn, simulation responses were ordered by the value assigned to that parameter, easing the identification of compound effects on cell behaviour and thus the identification of parameters that highly influence cell behaviour in the final hour of development.

(i) **Chemokine Related Parameters**

The analysis reveals there is no correlation between the value assigned to the initial level of chemokine expressed by an LTo cell upon differentiation and either cell velocity or displacement responses (Figures 3.7a and 3.8a). An alteration in the maximum level of chemokine expression (Figure 3.8b) does reveal a trend between the value of this parameter and cell displacement, although the other five parameters are also being perturbed simultaneously. This supports the result observed in one-a-time analysis, where a change in this parameter alone significantly affected cell displacement (section 3.5.1). Interestingly, although a significant affect on cell velocity was observed when this parameter was adjusted independently, no correlation is revealed between this parameter and cell velocity when all are adjusted simultaneously (Figure 3.7b). This suggests that, in terms of influencing cell velocity, the affect of this parameter is very dependent on the value of others.

A similar conclusion can also be drawn for the parameter that captures the probability that an LTi cell will not respond to localised chemokine expression. An independent alteration of this parameter significantly affects cell velocity (Figure 3.7c), yet no trend becomes apparent when the value is adjusted at the same time as the other five under examination. For the cell displacement measure, a trend is apparent close to the extreme upper value, where an LTi cell will never respond to chemokine expression, yet this is not continued through the lower 90% of the value range, again suggesting that the affect this parameter has may be dependent on others in the simulation.

(ii) **Cell Binding Probability Parameters**

When examining both cell behaviour measures at this time-point, no correlation is found between either measure and the probability an LTo and hematopoietic cell bind upon contact (Figures 3.7d and 3.8d). A similar conclusion is again drawn as detailed in the chemokine parameter analysis above: an alteration in the value of this parameter alone causes a significant alteration in cell behaviour, yet as no effect is apparent when a subset of parameters are all being perturbed simultaneously, this effect must be dependent on the value of other parameters.

(iii) **Adhesion Factor Related Parameters**

The results of this analysis for both cell behaviour measures closely match those revealed by tracking cells during hour twelve of development (Hour 12: Figures 3.3 and 3.4, Hour 72: Figures 3.7 and 3.8). There is a very strong correlation between cell response to adhesion factors and cell velocity at both time-points. The analysis reveals that this is the only compound effect on cell velocity at during hour 72 of development, suggesting that response to adhesion factors is the key factor in affecting cell velocity. This is in contrast to adhesion factor expression, where no correlation between parameter value and velocity is observed. In the previous section, for hour 12, it was noted that an effect becomes apparent for the latter parameter, where there are no results in the bottom left-hand corner of the graph (Figure 3.3f), and it was noted that this was expected due to the nature of the time-point being observed. In this case, no such effect is apparent, as a larger number of stable contacts will have occurred, mediated by chemokine levels that are greatly higher than that at twelve hours. Thus the level of expression is deemed to not have a significant effect on cell velocity response, as no trend is observed.

For the cell displacement response, a correlation is apparent between the maximum probability that a cell is retained in a forming patch and displacement in the final hour of development, with displacement decreasing steadily as the probability increases (Figure 3.7e). This trend becomes stronger close to the extreme value, where a cell would constantly be retained within a forming patch. The same effect was apparent during hour twelve (Figure 3.4e). No correlation was found between the level of adhesion factor expressed with each stable contact and cell displacement (Figure 3.8f), again matching the result seen in analysis of hour 12.

3.6.3 Partitioning Variance in Simulation Responses Between Parameters

The eFAST approach (Marino *et al.*, 2008; Saltelli, 2004) has again been used to partition variance in simulation results to quantify the sensitivity of the simulation to each parameter at hour 72, and thus suggest the impact each has on cell behaviour. Similarly to the latin-hypercube analysis above, the same 500 parameter sets that were generated for analysis at the twelve hour time-point were used here, to ensure that results being compared have been generated from the same parameter conditions. For each set of parameter values generated, 500 simulation runs were performed to mitigate aleatory uncertainty in cell behaviour responses as established in section 3.3. Simulation responses were analysed using the Fourier frequency approach described in Section 2.4.4. Plots are created for each simulation output response (velocity and displacement) for

hour 72, detailing the median First-Order (Si) and Total-Order sensitivity indexes calculated for each parameter of interest (Figures 3.9a and 3.9b). Sensitivity indexes and measures of significance in comparison to the dummy parameter are detailed in Table 3.3.

Cell Velocity

A statistical comparison of the variance caused by each parameter of interest with that of the dummy reveals that one parameter is statistically significant, the maximum probability that a cell responds to adhesion factor expression. As can be noted from Figure 3.9a and Table 3.3, the technique reveals that a perturbation in this probability accounts for 81% of the total variance. This supports latin-hypercube sampling findings presented above, where it was revealed that this was the only parameter where a correlation in value and cell velocity became apparent.

Cell Displacement

Four of the six parameters are identified as accounting for a statistically significant amount of variance in comparison to that of the dummy. The probability that two cells form a stable bind is a result that becomes apparent due to the effect the lower extreme value has in inhibiting LTo cell differentiation, and will thus be disregarded. The analysis suggests that significant variance in cell displacement is influenced by the probability an LTi cell responds to chemokine expression and the maximum level of chemokine that can be expressed by an LTo cell. This further supports the role of chemokine expression and response during hour 72, a factor deemed to have little influence during hour twelve. Also, the maximum probability an LTi cell is affected by adhesion factors is shown to have a significant impact, accounting for 40% of the variance when taken as the parameter of interest. Analyses for hour twelve in section 3.5 suggested adhesion to be the key biological factor at the twelve hour time-point, and this result suggests this still has a significant role in the later stages of development. The total-order, or STi results, reveal that these three parameters also have significant interactions with their complementary set. This may suggest that cell displacement may influenced by a combination of these factors working together, rather than one or two that are highly influential.

3.7   Discussion

3.7.1   A Representative Simulation Result

It has recently been noted that in many cases where modelling and simulation is applied in the exploration of biological systems, little attempt is made to show how representative that simulation is of the biological system it captures (Read *et al.*, 2012), and this

Parameter	Si	P-Value	STi	P-Value
stableBindProbability	0.071	0.059	0.150	0.119
chemokineExpressionThreshold	0.013	0.085	0.086	0.272
initialChemokineExpressionValue	0.003	0.569	0.051	0.486
maxChemokineExpressionValue	0.022	0.106	0.056	0.472
adhesionFactorExpressionSlope	0.012	0.173	0.047	0.546
maxProbabilityOfAdhesion	0.818	0.004*	0.896	0.001**
dummy	0.004		0.051	

(a) Cell Velocity Response

Parameter	Si	P-Value	STi	P-Value
stableBindProbability	0.078	0.01*	0.255	0.019*
chemokineExpressionThreshold	0.099	0.020*	0.199	0.038*
initialChemokineExpressionValue	0.005	0.302	0.059	0.297
maxChemokineExpressionValue	0.308	0.042*	0.414	0.031*
adhesionFactorExpressionSlope	0.018	0.179	0.108	0.197
maxProbabilityOfAdhesion	0.397	0.038*	0.471	0.036*
dummy	0.023		0.256	

(b) Cell Displacement Response

Table 3.3: Median sensitivity indexes and measures of statistical significance for each parameter examined using the eFAST technique, for both simulation cell behaviour responses obtained during hour 72. Si: First-Order Sensitivity Index; STi: Total-Order Sensitivity Index. Both are calculated for each re-sample curve and the median value taken. P-Value calculated using two-sample t-test to the distributions comprised of the results from each re-sample curve. * indicates statistical significance at 0.05 level, ** indicates statistical significance at 0.01 level.

was one of the motivations behind the development of *spartan* (Section 2.4). This chapter utilises *spartan* to provide a level of confidence in the simulator as a representative tool and suggest the degree of confidence in the values assigned to parameters. Presenting these alongside statistical analyses that have aimed to provide some biological insight reveals a full picture of simulation dynamics, aiding the drawing of conclusions as to what a simulation result actually means in terms of the real system.

In Chapter 1 it was noted that the process of Peyer's Patch development is highly stochastic (Figure 1.2) and differs significantly from mouse to mouse. Simulating this process does therefore involve capturing this biological uncertainty. However, as noted in the introduction of this chapter, an agent-based simulation of the process adds a further level of uncertainty that must be considered when simulation results are scrutinised in terms of the biological system captured. The objective is to produce a simulation where uncertainty caused by implementation is mitigated (Helton, 2008), and that inherently within the biological system matched. Through the use of the aleatory analysis technique developed by Read et al (2012), it has been shown that 500 simulation runs are necessary to meet this requirement, and thus this number of runs has been performed for all experiments in this chapter. With the exception of analyses presented here and studies of Experimental Autoimmune Encephalomyelitis

(EAE) by Read et al (2012), the presentation of such an analysis is rare, although
this is important in judging the relationship between the simulator and the real world
system. The development of these techniques and tools should hopefully encourage
more simulation developers to consider presenting such results alongside the biological
insight they claim the simulation provides.

The remaining sections of this discussion examine how the tools developed in this
completion of this thesis have been used to provide some biological insight into the
development of secondary lymphoid organs in the gut.

3.7.2 Simulated Cell Behaviour at the Twelve Hour Time Point is Highly Influenced by Adhesion Factor Expression

A minute-by-minute time-lapse analysis of cell behaviour during hour 12 of PP devel-
opment, explored using an *ex vivo* culture system, reveals that there is a statistically
significant difference in behaviour of cells that are within 50μm of their respective lig-
and and those that are further away (Patel *et al.*, 2012). The initial sections of this
chapter have sought to use the simulator as a tool to explain the biological factors that
could be causing this change in behaviour.

The data presented here have shown that a statistically significant change in cell
velocity is observed when the parameters that model the expression of adhesion factors
are adjusted. This is apparent both when the value of adhesion factor expression is
adjusted independently of all other parameters (Figures 3.2e and 3.2f), and when the
values of all unknown parameters are adjusted simultaneously (Figure 3.3e). Analysis
using an eFAST approach (Marino *et al.*, 2008; Saltelli *et al.*, 2000) also infers that a
large percentage of variation in simulation output can be accounted for by a change
in cell response to adhesion factors. In contrast, a change in the expression level of
chemoattractant molecules by an LTo cell has no significant effect on cell velocity using
either analysis technique.

The robustness to a change in expression value suggests that, at this early stage of
development, chemoattractants have no role in affecting cell velocity. In contrast, the
simulation is less robust to a change in adhesion factor expression, implying that these
factors have a key role in influencing cell behaviour at this phase of development, and
that there is a higher degree of uncertainty in the value of adhesion factor expression
as these directly influence simulation output.

Previously published observations have suggested that a blockage of VCAM-1 ex-
pression leads to a profound reduction in PP formation (Finke *et al.*, 2002), with the
assumption made that the remaining patches that do form are mediated by ICAM-1
and MAdCAM expression. If this was the case, it would be safe to assume that no
PP would form in the absence of adhesion factors. With data here suggesting no role

for chemokine expression in the early stages of PP development, this provides evidence that there is a phase in development where adhesion expression is the major factor. This suggested phase would commence when the process is triggered by RET$^+$ LTin cells interacting with RET ligand expressing LTo cells (Veiga-Fernandes *et al.*, 2007).

Such a phase would have been very difficult to determine in a laboratory. As noted in Chapter 2, it has not yet been possible to quantify the expression of adhesion factors VCAM, ICAM, and MAdCAM by an LTo cell. Instead the simulator has been used as a tool to explore the effect of changing adhesion factor expression, producing the results that has led to this hypothesis. This shows the strength and potential of simulation as an experimental tool. Future immunofluorescence staining investigations could potentially examine this hypothesis and detect adhesion expression levels in an *ex vivo* culture system, to both confirm or deny this hypothesis, while informing future development of the tool.

3.7.3 A High Level of Chemoattractant Expression Would Be Required to Influence Cell Displacement at the Twelve Hour Time-Point

The data presented in this chapter suggests that a change in chemoattractant expression parameters has no significant effect on cell displacement during hour twelve unless set to, or near to, the upper extreme value. This result becomes apparent when the values of all six unknown parameters (Section 3.4) are altered sequentially (Figure 3.4b), suggesting the emergence of an effect at and near to chemokine expression saturation.

If this extreme value effect had not been elucidated, it would be reasonable to draw the conclusion that chemoattractant expression has no significant influence on cell displacement at this time-point in development. As this is not the case, it is important to consider what this result means in terms of the biological system that has been captured.

As the process of calibration in section 2.2.6 has established values that produce cellular behaviour statistically similar to that seen *ex vivo*, assigning the chemokine parameters values at which this effect becomes apparent represents the simulation of an over-expression of chemoattractant factors. Although the expression of chemoattractant factors has yet to be quantified experimentally, the setting of these parameters to such extreme values can be deemed as not biologically plausible. This hypothesis is based on the statistics that cells further than 50μm from a primordial patch behave statistically differently, a statistical difference that would not be apparent if chemoattractant expression was this high. One would also expect to see more cells within a primordial patch if this was the case. However, this result suggests that an over expression of chemoattractant molecules at this early time point of development could be influential, and is a result that would be very difficult to establish biologically.

3.7.4 Chemokine Expression and Response a Key Factor During Hour 72

The chemokine result at twelve hours is interesting as previously published studies suggest a key role for chemokine expression in PP development (Cyster, 1999; Luther *et al.*, 2003; Ohl *et al.*, 2003). For this reason, it made sense to examine the factors affecting cell behaviour during hour 72. Should the same result be repeated, this would call into question whether the simulator correctly captures the emergent behaviour that was specified in the domain model.

Performing the same analyses for the final hour of development does reveal a key role for chemokines in influencing cell displacement. An individual alteration in any of the three parameters that capture chemokine expression and response has a significant impact on cell behaviour, suggesting that at this time-point, simulation behaviour is no longer robust to a change in these parameter values. Statistical responses generated through use of the eFAST technique (Marino *et al.*, 2008; Saltelli *et al.*, 2000) reveal that the maximum level of chemokine expression and the probability an LTi cell responds to chemokine expression are two of three parameters that account for a statistically significant proportion of variance in results. Analysis using a latin-hypercube approach (Read *et al.*, 2012) also suggests there is a clear trend between cell displacement and the level of chemokine expressed by the LTo cell at this time-point. Considering the results of all these analysis together concludes that chemokine expression is influencing simulated cell behaviour at this time-point, in agreement with published studies (Cyster, 1999; Luther *et al.*, 2003; Ohl *et al.*, 2003). The contrast in results between twelve and 72 hours suggests there could be a point in the 72 hour period where the main biological factor that influences cell behaviour changes. The next section of the discussion examines that hypothesis in more detail.

Results from both the eFAST and latin-hypercube analysis techniques suggest that a change in adhesion factor response remains the key factor in affecting cell velocity, as found at the twelve-hour time-point. Statistical responses from the eFAST technique suggest that a large proportion of variation in displacement can be accounted for by a change in the value of the parameter that captures adhesion response, with this parameter being the only one to have a statistically significant contribution. This suggests that adhesion factor response is key throughout the whole time-period rather than just at 12 hours. The same parameter is also found to explain a significant proportion of the variance for the cell velocity response (being the third of the three that were noted as significant in the previous paragraph). Thus, in contrast to suggestions in the previous paragraph, there may not be a phase where the biological factor that affects cell behaviour changes from adhesion to chemokine, rather a time-point where chemokine expression becomes influential.

3.7.5 The Potential for Phases of Development Between E14.5 and E17.5

This chapter set out to suggest the key biological factors influencing cell behaviour during hours 12 and 72 of development, and examine whether conclusions drawn at hour 12 differed from hour 72. The simulation captures the period between phases one and two of Adachi et al's (1997) three phases of Peyer's Patch development, the first being the appearance of VCAM-1$^+$ stromal cells, and the second the identification of clusters of hematopoietic cells. Data presented here does show that the influence of simulated biological factors does change between hour 12 and 72.

These data could be used to suggest further phases of PP development, occurring after the appearance of VCAM-1$^+$ stromal cells. The first of these is a period where hematopoietic (LTin/LTi) cell behaviour is influenced by adhesion factor expression. Statistical analysis results show that this is the case at the twelve hour time-point. This could be deemed a 'triggering' phase, where LTo cell differentiation has occurred through RET signalling (Veiga-Fernandes *et al.*, 2007) and adhesion factor expression is upregulated, affecting the behaviour of cells around a forming PP. Use of the same statistical techniques reveals that by the final hour, chemokine expression and response has become influential. This influences cell displacement, mediating recruitment of LTi cells towards a forming patch (Cyster, 1999; Luther *et al.*, 2003). Thus this could be referred to as a 'clustering' or 'aggregation' phase, after which Adachi's second phase occurs: the visual identification of hematopoietic cell clusters (Adachi *et al.*, 1997).

The *ex vivo* studies produced cell behaviour results that informed the calibration of the simulated cell behaviour at hour 12 (Patel *et al.*, 2012). The resultant simulation has now produced cell behaviour statistics that could be verified in further *ex vivo* work that examines hour 72. This would verify whether cell behaviour during the final hour of development has been captured correctly, and thus go some way to supporting the phases hypothesis generated above. Experimentation using the simulator can take this further. It is much easier to examine cell behaviour at a number of time-points in simulation than it would be in the laboratory. Performing the same analyses undertaken in this chapter at a number of time-points in development would reveal where the suggested 'triggering' phase ends and the 'clustering' phase starts, providing further insight into each phase. Use of simulation as a time-lapse tool to perform such investigations is examined in a later chapter of this thesis.
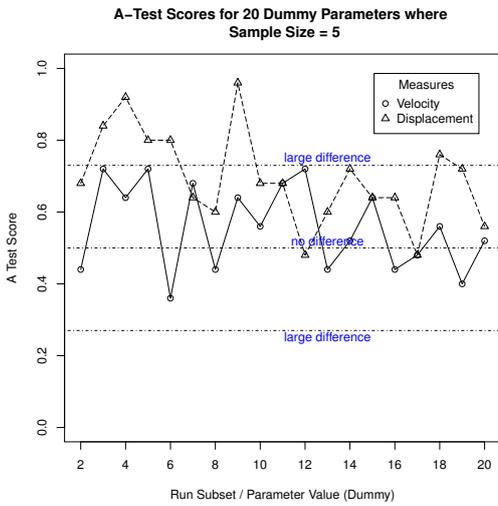
(a) A test scores for 20 sets of runs, using a sample size of 5.
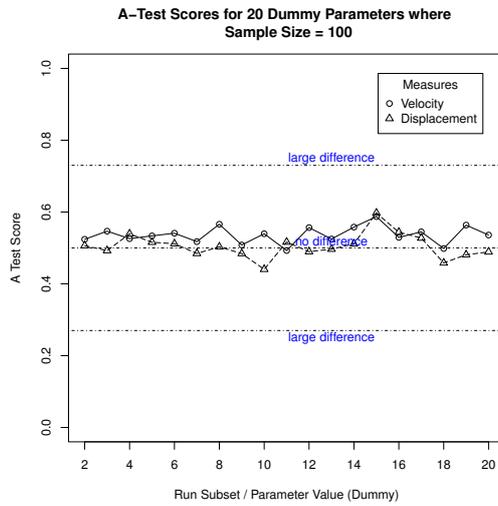
(b) A test scores for 20 sets of runs, using a sample size of 100.
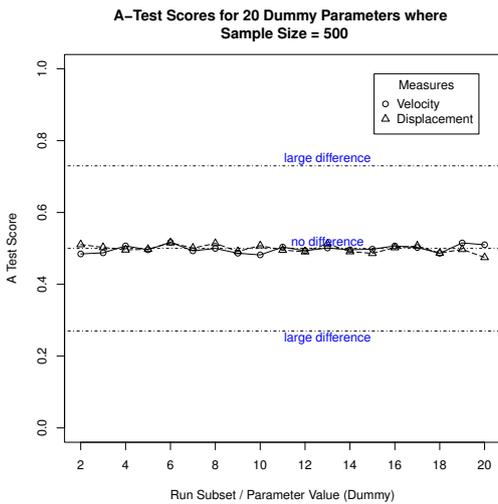
(c) A test scores for 20 sets of runs, using a sample size of 500.
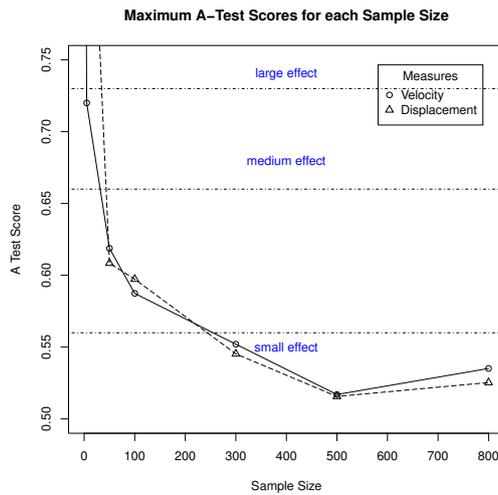
(d) Summary of the maximum A test score for both cell tracking measures when the sample size is varied. Where the maximum score was less than 0.5, the corresponding value above 0.5 has been assigned. This has been done as the magnitude of the effect is of more interest than the direction

Figure 3.1: Examining the effect of aleatory uncertainty on the results of the simulation for a variety of sample sizes. For each sample size, twenty runs are performed, with no parameters changed each time. Difference between the distributions is determined using the Vargha-Delaney A-Test (2000), with difference categorised by the limits set in Table 2.5. Using this technique, the number of runs necessary to produce a robust result can be ascertained. As can be seen in Figure 3.1d, there is little reduction from 500 runs onwards, making this a representative sample size to use
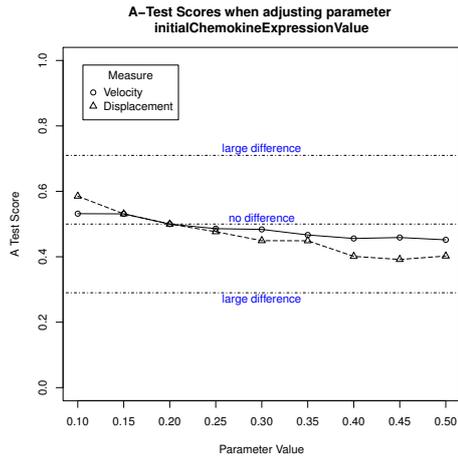
**(a)** Initial level of chemoattractant expression at LTo differentiation
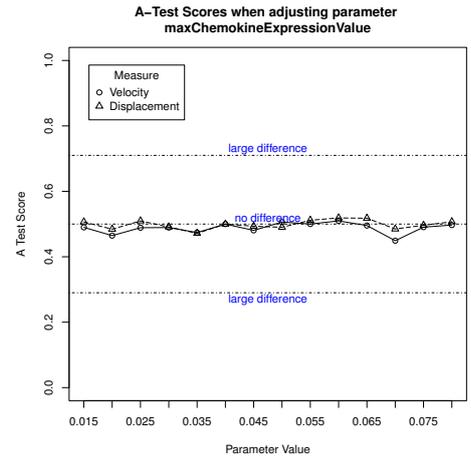
**(b)** Saturation limit of chemoattractant expression
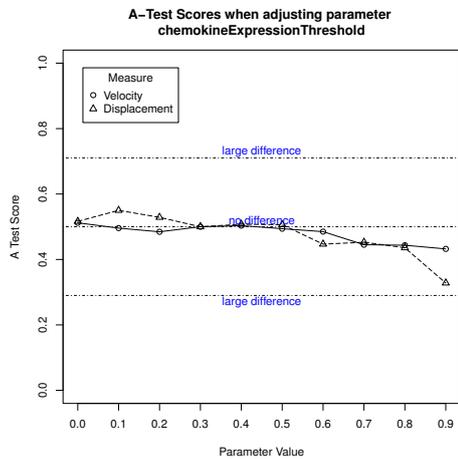
(c) Chemokine Level at which LTi chemotaxis occurs
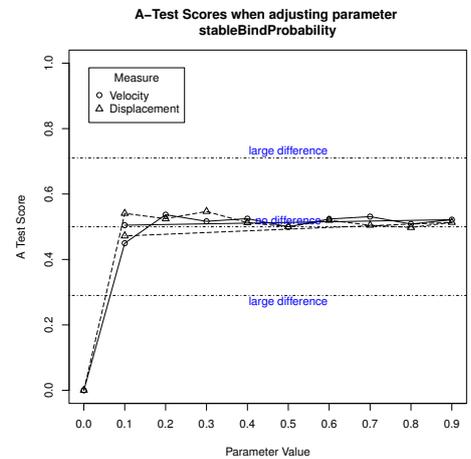
**(d)** Probability a LTin/LTi and LTo cell form stable bind on contact
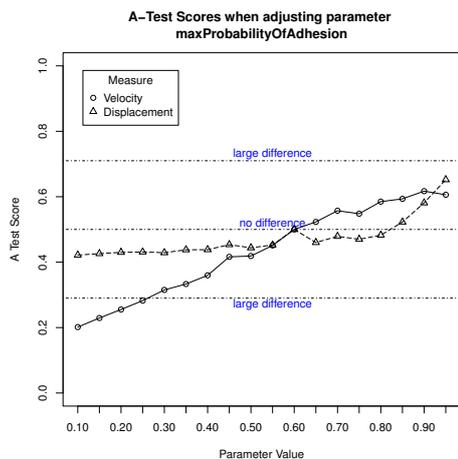
**(e)** Maximum probability adhesion factors prolong cell contact
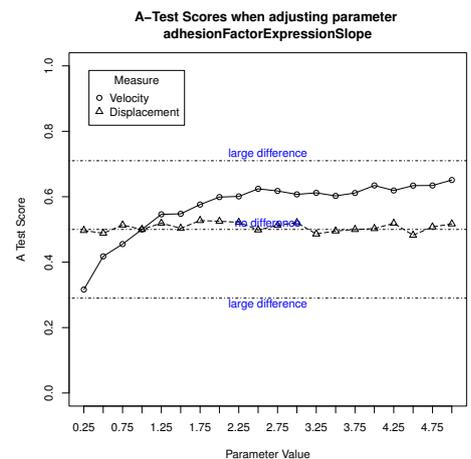
(f) Level of adhesion factor expression per stable contact

Figure 3.2: Determining the robustness of simulated cell behaviour responses during hour 12 of development. The six parameters for which a value is unknown were examined in turn, and the value of each perturbed within a specified range. Simulation results were compared to those generated during calibration, using the Vargha-Delaney A-Test, to determine if a significant change in cell behaviour has occurred
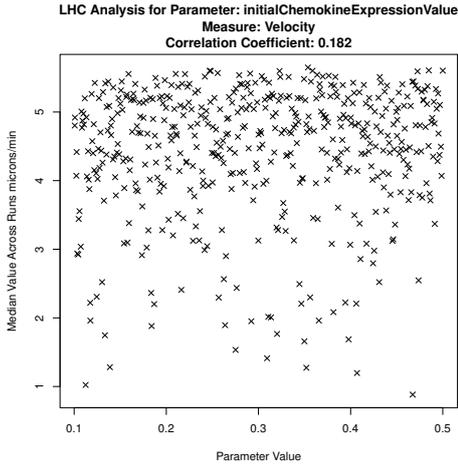
(a) Initial level of chemoattractant expression at LTo differentiation


(b) Saturation limit of chemoattractant expression
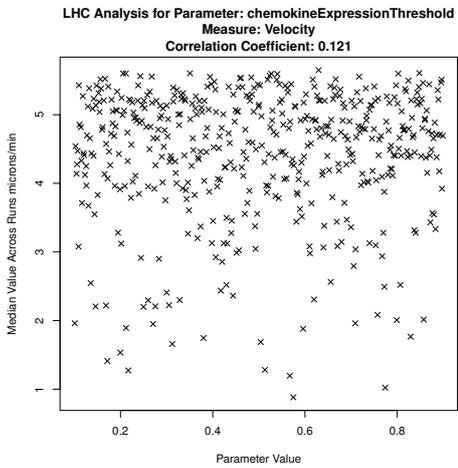
(c) Chemokine Level at which LTi chemotaxis occurs
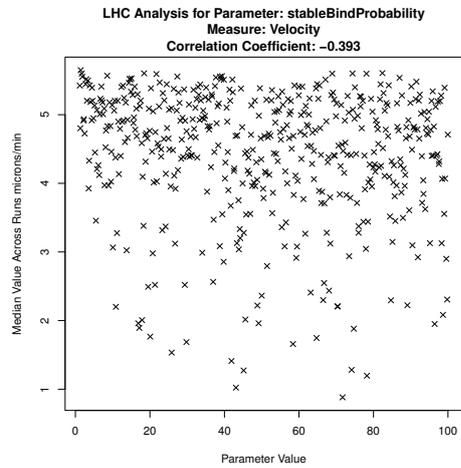
(d) Probability an LTin/LTi and LTo cell form stable bind on contact
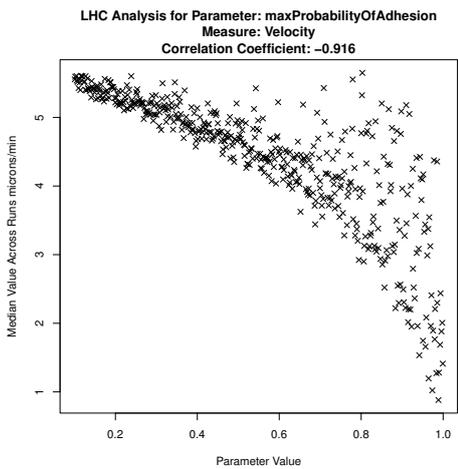
(e) Maximum probability adhesion factors prolong cell contact
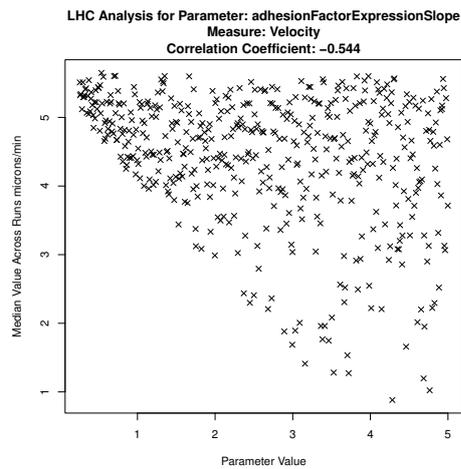
(f) Level of adhesion factor expression per stable contact

Figure 3.3: Identifying any compound effects on cell velocity at hour 12 through latin-hypercube sampling, a technique that perturbs the value of all parameters simultaneously. Influential parameters can be identified through trends of points within each plot and through the value of the Partial Rank Correlation Coefficient specified in the graph header, as described in section 2.4.3.

(a) Initial level of chemoattractant expression at LTo differentiation
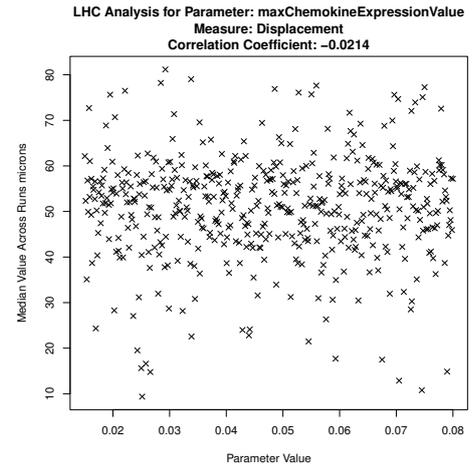
(b) Saturation limit of chemoattractant expression
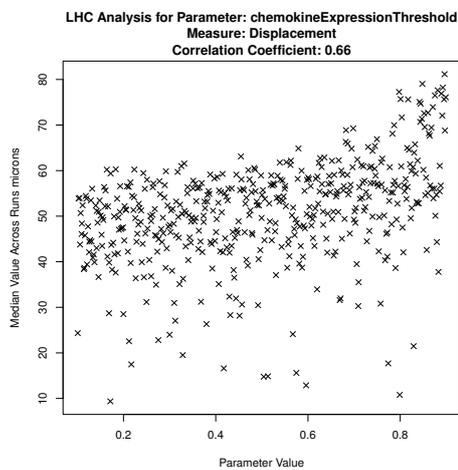
(c) Chemokine Level at which LTi chemotaxis occurs
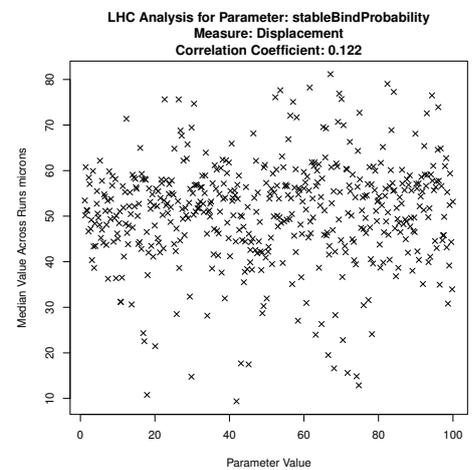
(d) Probability a LTin/LTi and LTo cell form stable bind on contact
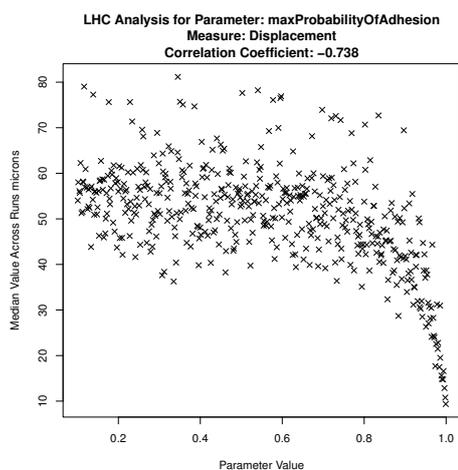
**(e)** Maximum probability adhesion factors prolong cell contact
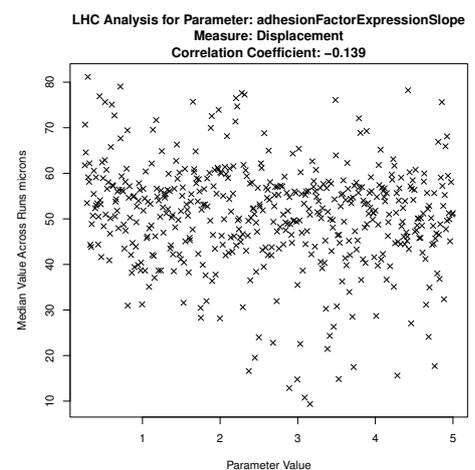
(f) Level of adhesion factor expression per stable contact

Figure 3.4: Identifying any compound effects on cell displacement at hour 12 through latin-hypercube sampling, a technique that perturbs the value of all parameters simultaneously. Influential parameters can be identified through trends of points within each plot and through the value of the Partial Rank Correlation Coefficient specified in the graph header, as described in section 2.4.3.
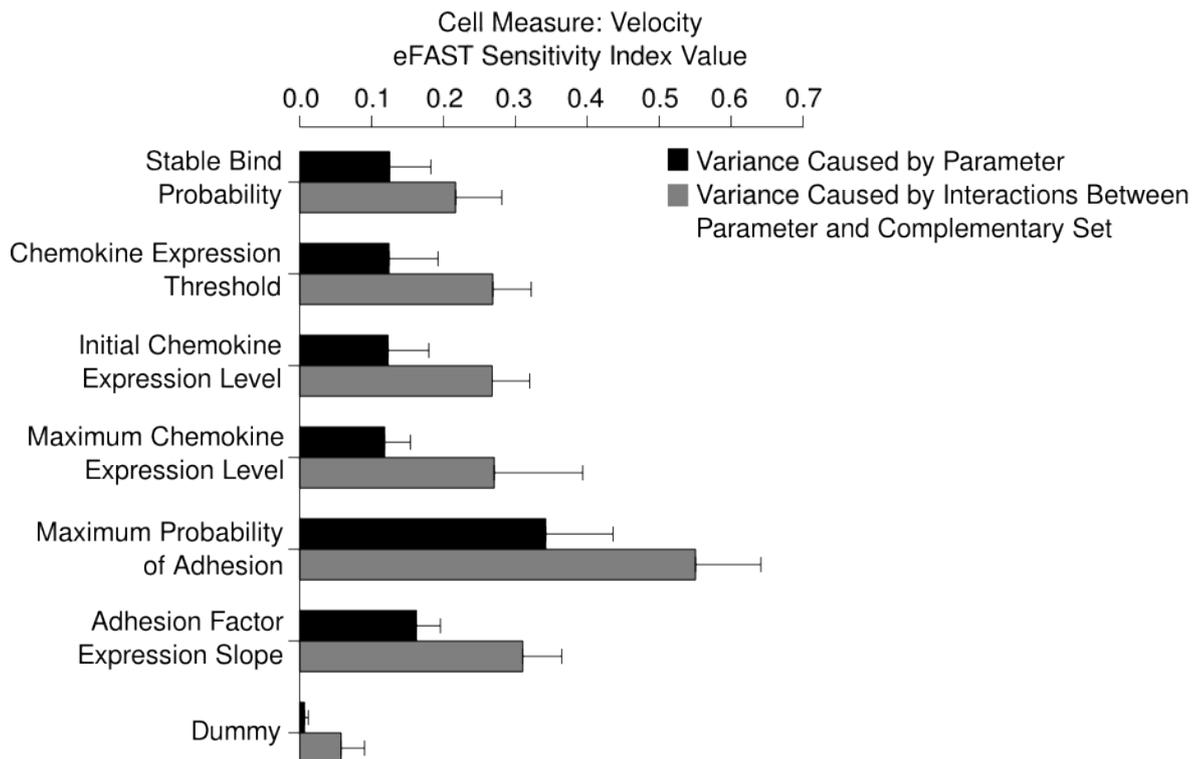
(a) eFAST Sensitivity Indexes for Cell Velocity Response
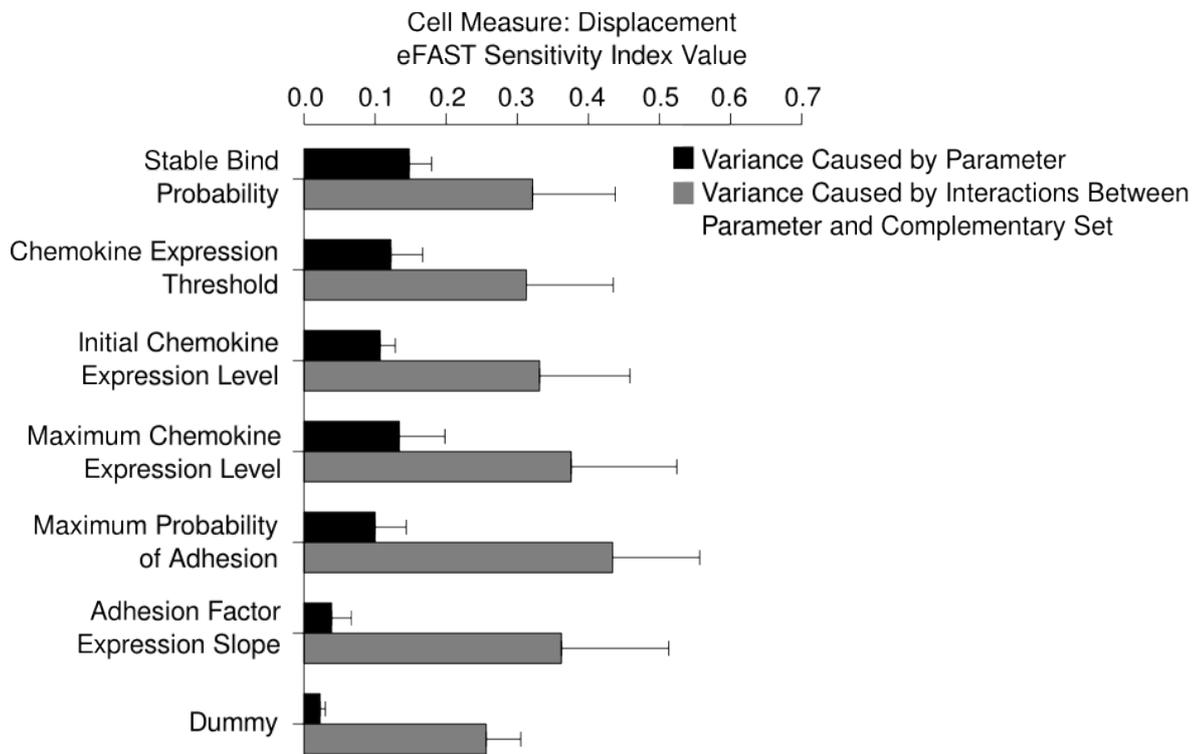

**(b)** eFAST Sensitivity Indexes for Cell Displacement Response

Figure 3.5: Hour 12 sensitivity indexes for each parameter where a value is unknown, calculated using the Extended Fourier Amplitude Sampling Test (eFAST). Black: the fraction of model output variance accounted for by a variation in the value of that parameter; Grey: remaining variance accounted for by higher-order interactions between this parameter and its complementary set.

**(a)** Initial level of chemoattractant expression at LTo differentiation
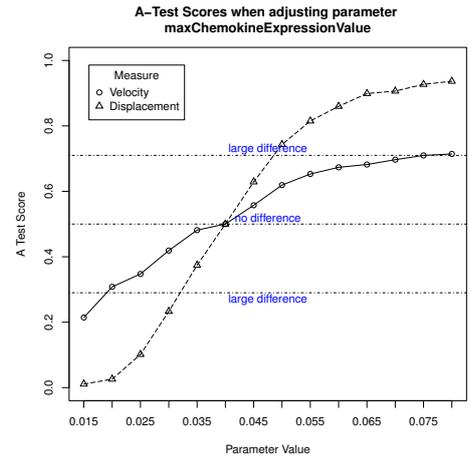
(b) Saturation limit of chemoattractant expression
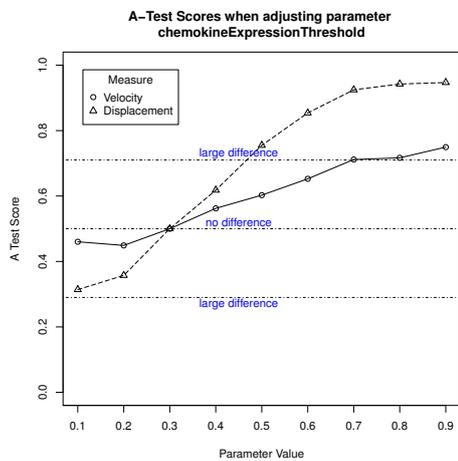
(c) Chemokine Level at which LTi chemotaxis occurs
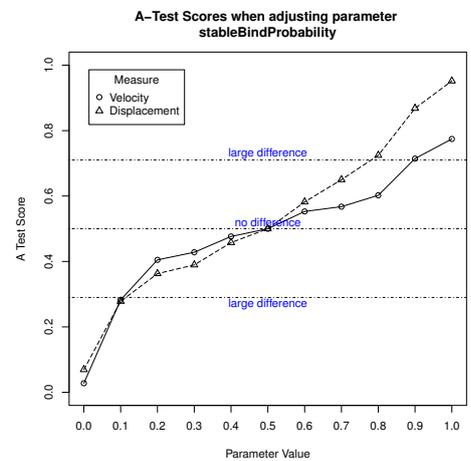
(d) Probability a LTin/LTi and LTo cell form stable bind on contact
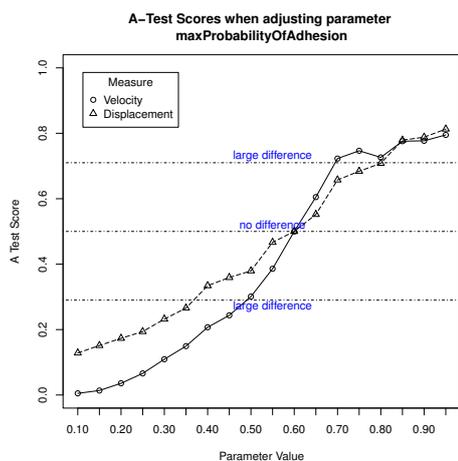
(e) Maximum probability adhesion factors prolong cell contact
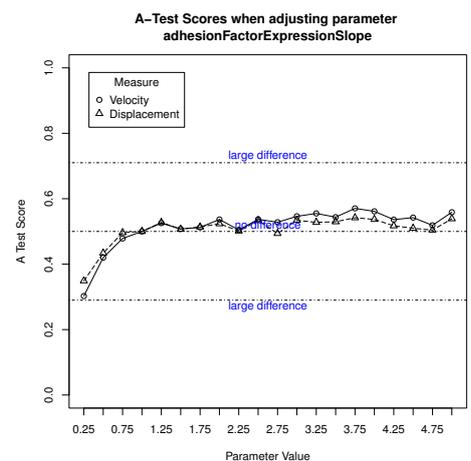
(f) Level of adhesion factor expression per stable contact

Figure 3.6: Determining the robustness of simulated cell behaviour response at the end of the development period. The six parameters for which a value is unknown were examined in turn, and the value of each perturbed within a specified range. Simulation results were compared to those generated during calibration, using the Vargha-Delaney A-Test, to determine if a significant change in cell behaviour has occurred.

(a) Initial level of chemoattractant expression at LTo differentiation
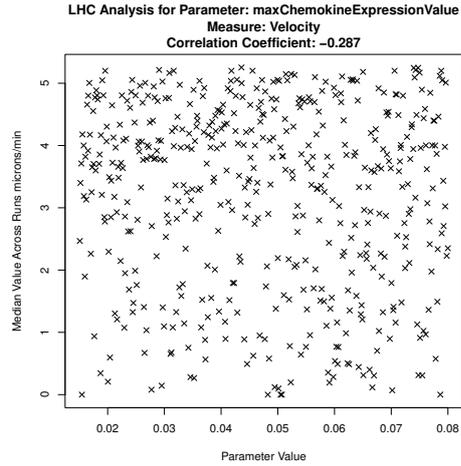
(b) Saturation limit of chemoattractant expression

(c) Chemokine Level at which LTi chemotaxis occurs

**(d)** Probability a LTin/LTi and LTo cell form stable bind on contact

**(e)** Maximum probability adhesion factors prolong cell contact
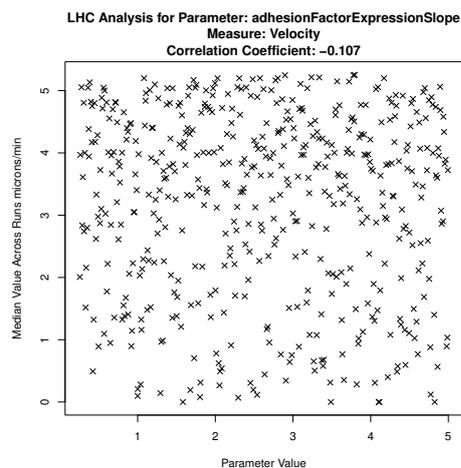
**(f)** Level of adhesion factor expression per stable contact

Figure 3.7: Identifying any compound effects on cell velocity at hour 72 through latin-hypercube sampling, a technique that perturbs the value of all parameters simultaneously. Influential parameters are identified by trends that emerge in the simulation results, and through the Partial Rank Correlation Coefficient specified in the graph header.

(a) Initial level of chemoattractant expression at LTo differentiation
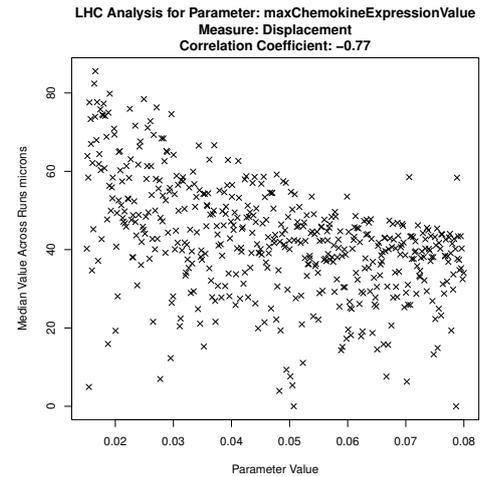
(b) Saturation limit of chemoattractant expression
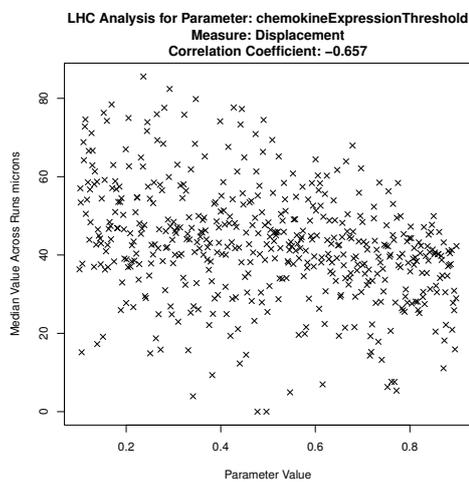
(c) Chemokine Level at which LTi chemotaxis occurs
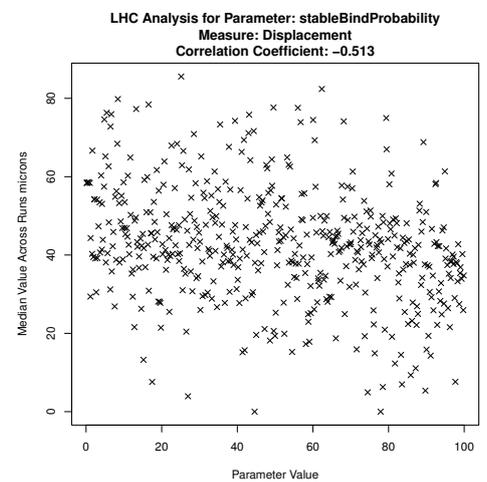
(d) Probability a LTin/LTi and LTo cell form stable bind on contact
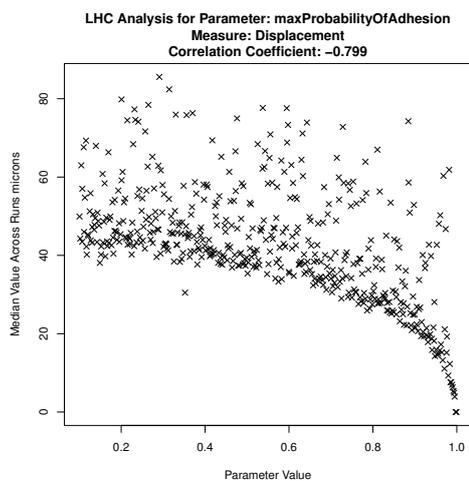
(e) Maximum probability adhesion factors prolong cell contact
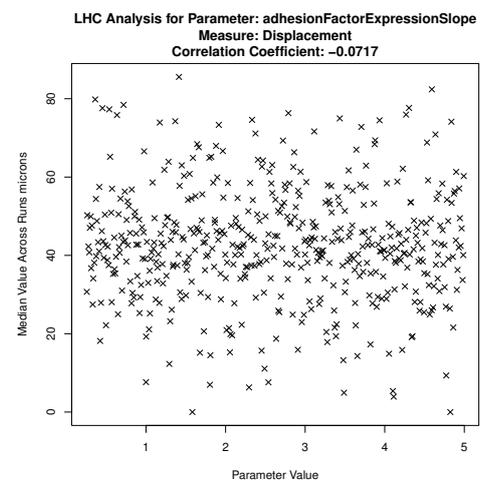
(f) Level of adhesion factor expression per stable contact

Figure 3.8: Identifying any compound effects on cell displacement at hour 72 through latin-hypercube sampling, a technique that perturbs the value of all parameters simultaneously. Influential parameters are identified by trends that emerge in the simulation results, and through the Partial Rank Correlation Coefficient specified in the graph header.
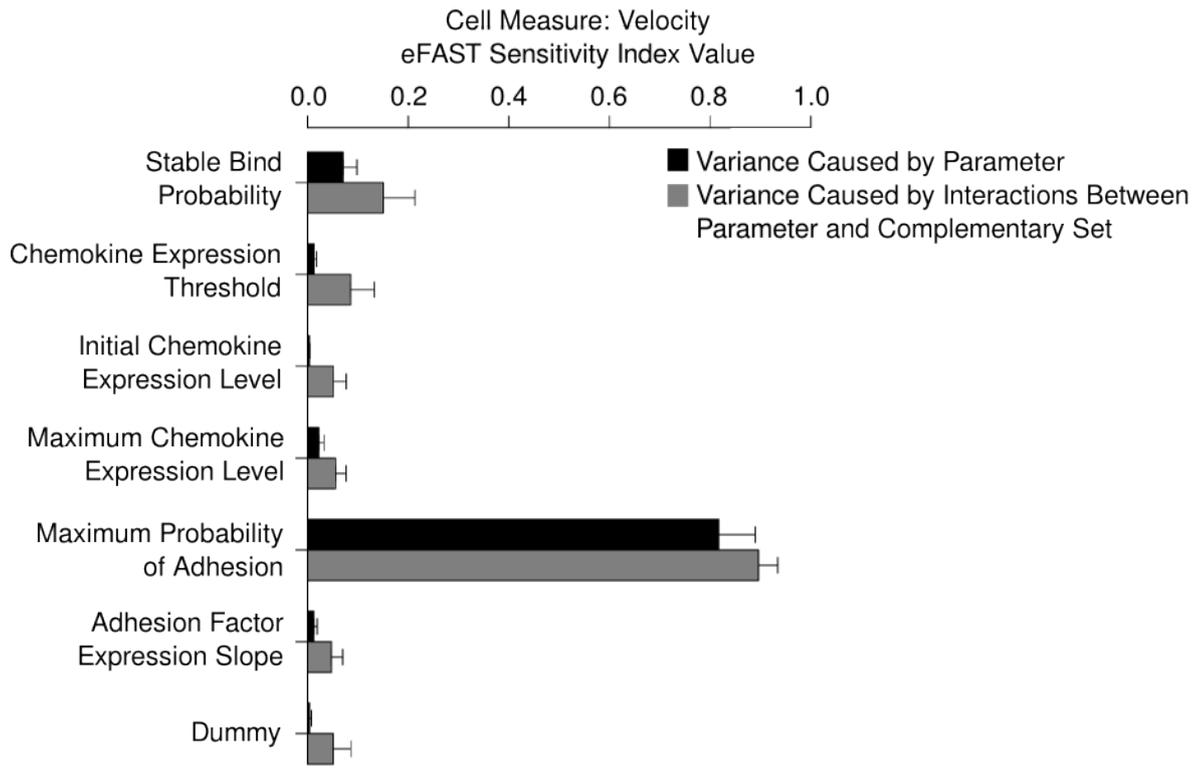
(a) eFAST Sensitivity Indexes for Cell Velocity Response
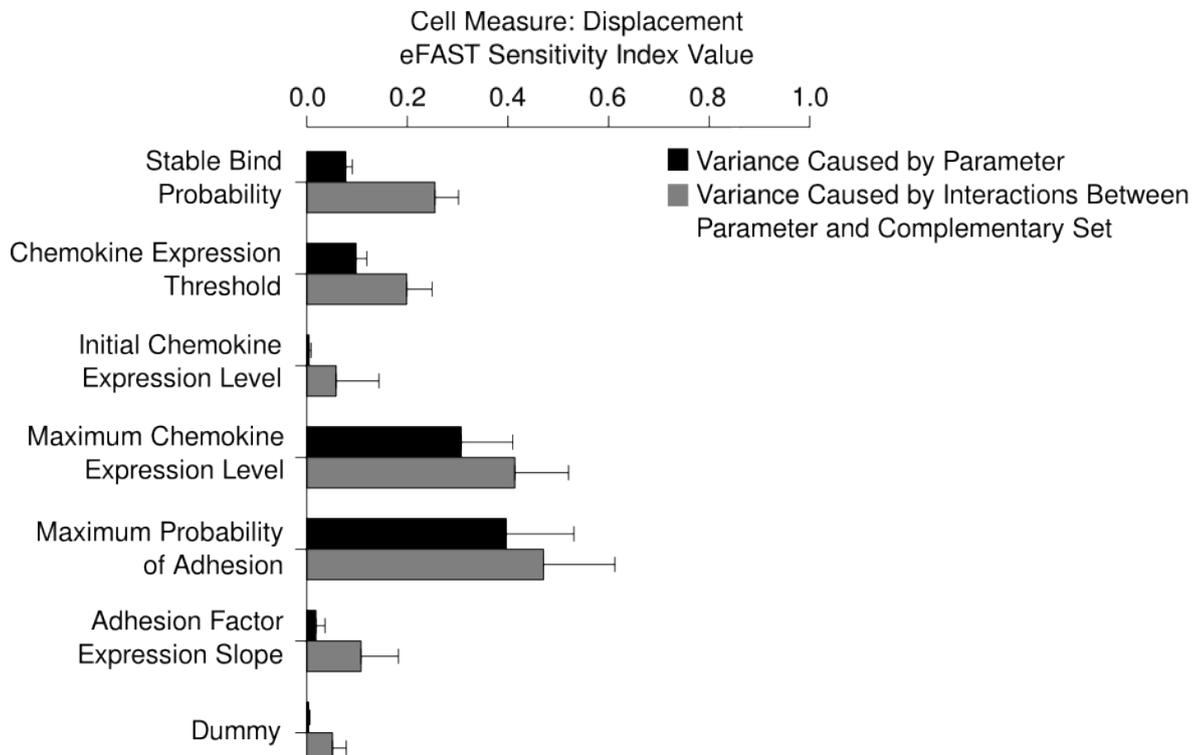
(b) eFAST Sensitivity Indexes for Cell Displacement Response

**Figure 3.9:** Hour 72 Sensitivity indexes for each parameter where a value is unknown, calculated using the Extended Fourier Amplitude Sampling Test (eFAST). Black: the fraction of model output variance accounted for by a variation in the value of that parameter; Grey: remaining variance accounted for by higher-order interactions between this parameter and its complementary set.

Chapter 4

Exploring Factors Affecting Peyer's Patch Characteristics Through Simulation

A combination of a number of previously published laboratory studies has led to the formation of a basic model describing Peyer's Patch development. However these investigations have left a number of interesting questions that are difficult to address. Simulating the process makes it possible to perform in silico experimentation that can attempt to address these open questions. In this chapter, the simulator is utilised as a tool for replicating previously published explorations and performing novel explorations that cannot be performed in a laboratory. Furthermore, the techniques in the spartan statistical toolkit developed in the course of this study have been utilised to analyse simulation results and attempt to quantify the role of each biological factor in influencing Peyer's Patch physical characteristics. The results of these explorations may then inform future laboratory studies that attempt to further understand lymphoid organogenesis.

4.1 Introduction

The analyses performed in the previous chapter examined the first of the emergent behaviours identified in the domain model (Figure 2.1), the change in cell behaviour in the location of PP genesis. This chapter examines the second, the formation of aggregations of LTin and LTi cells in the vicinity of LTo cells by the end of hour 72 (E17.5), aggregations that mature from E18.5 into secondary lymphoid organs.

The calibration of the simulator and conclusions drawn from analyses in the previous chapter are aided greatly by the availability of cell behaviour responses, obtained from the *ex vivo* culture system explorations (Patel *et al.*, 2012). Statistical techniques have been used to ensure that there is no statistically significant difference between cell behaviour produced by the simulator and the measures observed *ex vivo*. However, there is currently little biological data available that provides a quantification of what a 'patch' is at this time-point of development, for example in terms of size or number of cells in the aggregation. The existence of PP at this time-point has, to date, been noted visually rather than through use of any statistical measure. What is known is that the aggregations can form in 3D, protruding away from the mucosal epithelium (Jung *et al.*, 2010; van de Pavert and Mebius, 2010). This is in contrast to the simulator developed in Chapter 2, where aggregations of LTin and LTi cells form on a 2D plane. It could be suggested therefore that it would be difficult to compare the result from the simulator to any biological data that was available. However, the emergent behaviour of interest here is the aggregation of cells, with the aim to understand how this is affected by other biological factors, rather than replicating the real characteristics of a developing PP. Thus a 2D plane is a suitable abstraction. With this in mind, the simulator produces three responses deemed to be patch characteristics: area of the aggregation that is formed, the level of chemokine expressed, and the level of adhesion factor expressed. The first gives an indication of the ability of the patch to recruit LTi cells through chemotaxis (Cyster, 1999; Luther *et al.*, 2003), whereas the latter two give an indication of the number of cellular interactions that have occurred, resulting in the upregulation of adhesion and chemokine expression (van de Pavert and Mebius, 2010; Randall *et al.*, 2008). Also, similar to the traditional method of identifying PP at this development phase, the simulator produces a visual representation of the gut tract, showing the location of any cell aggregations that are formed by hour 72. These simulation responses and output snapshots can thus be used as a basis for analysing aggregation development, to understand how this is affected by each biological factor captured in the simulation. Baseline behaviour, to which the simulation results under differing conditions is compared, has been set using the calibration methods detailed in section 2.2.6. A comparison between the baseline and a simulation run under different parameter conditions can be used to suggest the factors influencing the number of patches that form and the three patch characteristic responses noted previously.

The first section of this chapter utilises the visual output to judge how successfully

the simulator replicates previously published *in vivo* and *in vitro* experiments. The key factors in the development of PP, detailed in section 1.3.2, have been established using current laboratory techniques. The objective of the initial section of this chapter is to ensure that the simulator produces results that match phenotypes in the relevant publications that were observed through use of antibody staining of LTi and LTo cells. These include replicating the gut of developing mice deficient for certain genes (Eberl *et al.*, 2004; Luther *et al.*, 2003; Veiga-Fernandes *et al.*, 2007) and replicating an over-expression of key factors (Meier *et al.*, 2007). The results of these investigations were taken into consideration when the domain model was constructed. Replicating previously published results provides a further degree of confidence that the simulator correctly captures this domain model upon which it is based.

With this assured, the simulator can act as a tool through which novel *in silico* experimentation can be performed. Such investigations can help generate new hypotheses on open questions that have yet to be explored (a range of which were discussed in section 1.3.4), potentially feeding future laboratory investigation (Andrews *et al.*, 2010; Germain *et al.*, 2011). This chapter details three such investigations. The first of these explores the role of the LTin cell population in PP development. It was noted in the section 1.3.3 that the role of LTin cells has largely remained unknown, although the *ex vivo* work by Patel et al (2012) has suggested a role in early stages of development. Investigations in this chapter use the simulator to examine if patch characteristics change when the LTin cell population is increased and decreased. This change is simulated by changing the parameter that specifies the number of LTin cells at E15.5, a figure that has been estimated from flow cytometry results gathered at this time-point. With data from other time-points unavailable, a linear input rate has been implemented where a certain number of cells migrate into the gut at each time-step, ensuring the correct number of cells are present at the time-point representing E15.5. The second investigation in this chapter uses the simulator to examine this input rate, changing this from a straight line rate to an exponential and square root function that passes through the same cell number point at E15.5, to determine if a change in the migration rate of LTin cells has an impact on PP characteristics. Performing these two *in silico* investigations may provide further hypotheses on the role of LTin cells in PP development. Additionally, the simulator has been used to investigate the geography of PP development. In the current model, LTo cells can be expressed anywhere on the gut surface, and a percentage of these express the ligand for RET, and have the potential to form PP (Veiga-Fernandes *et al.*, 2007). However, some studies suggest that PP tend to form on the intestines antimesenteric border (Hoorweg and Cupedo, 2008; Randall *et al.*, 2008), suggesting this is the area where LTo cells that have the potential to form PP reside. In simulation terms, the anti-mesenteric border would represent a small band across the length of the simulated gut. This has been simulated by restricting RET ligand expressing LTo cell placement to a small band across the

centre, then comparing the number of PP that are formed under this condition with that of the current calibrated simulator.

Finally, the chapter returns to a theme examined in the previous chapter, the effect of parameters in the simulation for which a value is currently unknown (Table 3.1). In the previous chapter, the *spartan* statistical toolkit developed during the course of this study was used to examine how changes in the values of these parameters affected cell behaviour in the vicinity of a forming PP. This chapter uses the same techniques, examines the same parameters, but this time examines the effect a change in these parameters has on the formation of cell aggregations that mature to become PP. In section 2.2.6, it was detailed how a baseline behaviour for the formation of one patch has been established using an experimental set up that restricted LTo cell placement such that the simulator forms one patch, always in the same location. With this in place, the six parameters for which a value is unknown are perturbed using the techniques available in the *spartan* package, to examine how a change in these factors affects the three characteristics of that patch described earlier in this section (patch area, level of chemokine expressed, and level of adhesion factors expressed). This statistical investigation provides an overall picture of the role of each parameter in the producing the two emergent behaviours observed.

4.2 Aims

The examinations in this chapter use the simulation platform as an experimental tool to achieve the following aims:

1. To perform previously published experiments on the simulator and determine if the same result is replicated.

2. To perform *in silico* experimentation that could provide novel insights into lymphoid tissue development.

3. To determine the key biological factors that influence the size and characteristics of a primordial Peyer's Patch.

4.3 Replicating Previously Published Experimental Results

The simulation has been run under the following conditions from the established literature: Figure 4.1a - Normal parameter setting (control wild type mice); Figure 4.1b - RET deficiency (RET$^{-/-}$ mice, (Veiga-Fernandes *et al.*, 2007)); Figure 4.1c - chemokine knockout (CXCL13$^{-/-}$, CCL19/21$^{-/-}$ mice, (Luther *et al.*, 2003)); Figure 4.1d - no LTin cells (ROR$^{-/-}$ mice, (Eberl *et al.*, 2004)); Figure 4.1e - doubling number of LTi cells (IL-7^{Tg} mice, (Meier *et al.*, 2007)).

Consistent with established results, no PP are observed in either RET, chemokine or LTi deficient mice. In mice with increased numbers of LTi cells in the simulation, more, larger PPs were observed to develop consistent with published results.

4.4 Novel *in silico* Experimentation

4.4.1 Producing a Representative Result that Minimises Aleatory Uncertainty

As examinations in this chapter explore patch characteristic simulation responses and not the same cell behaviour responses conducted previously, the aleatory analysis technique by Read et al (2012) described in section 2.4.1 has been repeated. To determine the number of replicate runs required to reduce aleatory uncertainty in patch characteristic measures, the same sample sizes were examined as previously (1, 5, 50, 100, 300, 500, 800). Figure 4.2 details the maximum A-Test response for each sample size examined. This indicates that for analyses where the focus is on patch characteristics, 300 runs is sufficient to reduce the effect of aleatory uncertainty, in contrast to the 500 required for cell behaviour measures. Thus 300 simulation runs were performed for each investigation conducted using the simulator in this chapter.

4.4.2 Investigating the Impact of LTin Cell Number on PP Formation

LTin cell numbers have been estimated using results from flow cytometry experiments, as explained in Chapter 2. The estimate of cell number at E15.5 has been used to establish a cell migration rate, and thus is directly linked to the number of LTin cells in the simulation. It has been suggested that this cell population has a role in the early initiation of PP development (Patel *et al.*, 2012; Veiga-Fernandes *et al.*, 2007), with LTi cells having the main role in cell aggregation (van de Pavert and Mebius, 2010; Randall *et al.*, 2008). As yet the effect of a reduction or increase in LTin cell numbers on PP formation has not been established. This section examines the use of the simulator to suggest the effect on PP formation when the cell population is decreased and increased.

To examine the effect of a decrease in LTin cell numbers, simulations have been run for gut surface area percentages of 0.05% to 0.45%, with an increment of 0.05. From this percentage, the simulator calculates how many cells should be present at the twenty-four hour time-point. For each value examined, 300 simulation runs were performed to mitigate the effect of aleatory uncertainty. Median values for patch area and patch number were calculated for each of the 300 runs, producing a distribution of 300 results for each LTin percentage value examined. Figure 4.3a is a plot of the median calculated from the 300 median patch areas for each value, revealing a decrease

in patch area at 72 hours as LTin cell number decreases. To gain a statistical measure of the effect the decrease has had, the distribution of 300 patch characteristic responses was compared to a distribution generated from 300 runs of the simulator at baseline values, using the Vargha-Delaney A-Test (Vargha and Delaney, 2000). A-Test scores for each LTin percentage value have been plotted in Figure 4.3b. This reveals that the percentage of area occupied by LTin cells at E15.5 can be reduced by 0.15% (0.20%) before a change in response is observed that Vargha-Delaney classify as 'small'. In terms of patch area, a reduction in cell number approaches the statistical effect classified as 'medium' when the percentage is reduced further than 0.20%, however never meets that classification. For patch number however, a trend emerges between the distribution of number of patches formed and the number of LTin cells in the simulation, with the difference between distributions becoming more statistically significant as LTin cell numbers decrease. Thus it could be concluded that LTin cell number could have a role in controlling the number of PP that form.

To examine the effect of an increase in LTin cell number, simulations have been run that model a 2, 3, 4 and 5 fold increase in cell number. For each percentage examined, 300 simulation runs have again been performed and the medians taken for each patch characteristic for each run. The median patch area for each percentage is plotted in Figure 4.3c. Patch area increases as cell number increases, however this begins to stabilise after a 3 fold increase in LTin cell number. The distributions for each patch characteristic response have again been compared using the Vargha-Delaney A-Test (Vargha and Delaney, 2000) (Figure 4.3d). This reveals a statistically 'small' difference by Vargha-Delaney's criteria when the cell population is doubled, an effect that increases to 'medium' for a 3 fold increase. However, the statistical difference stabilises after this point, suggesting that a further increase in LTin cell numbers has no effect on the patch characteristics observed.

4.4.3   Investigating LTin Cell Migration Rate

It was noted in the previous section that flow cytometry has been used to establish an estimate of the number of LTin cells present in the gut at the 24 hour time-point. In the current simulation, a linear input rate is used to create the number of LTin cells at each time-point such that the required number are present at 24 hours, using equation 2.3 specified in Figure 2.10. This is exemplified by the red line on the plot in Figure 4.4a. However, the assumption that cells migrate into the gut in such an orderly manner is questionable. With no further biological data available, the simulator has been used to examine the effect on PP characteristics if this migration rate was changed. Two alternative rates have been investigated that replace equation 2.3: an exponential rate where cell migration is initially slow and increases rapidly (green line in Figure 4.4a), and a square root function rate that models the opposite effect (Black line in Figure 4.4a). The replacement equations calculate the number of LTin

cells that should be present at the current simulation time-step. The number of cells currently within the simulation is subtracted from this to generate the input rate per step. In these calculations, constants have been calculated to ensure that the function creates the correct number of LTin cells at E15.5.

Square Root:

$$\text{Required LTin Cells In This Time-Step} = (27.8 * \text{Number of Time-Steps Elapsed})^{0.5}$$
$$(4.1)$$

Exponential:

$$\text{Required LTin Cells In This Time-Step} = 1.00368^{\text{Number of Time-Steps Elapsed}} \quad (4.2)$$

For each migration rate function, 300 simulation runs have been performed to mitigate any aleatory uncertainty in patch characteristic results. For each run, median patch number and patch area at the 72 hour time-point have been calculated and plotted in Figures 4.4b and 4.4c respectively. To gain a statistical measure of the impact this change has had, the distribution of patch characteristic medians calculated for the exponential and square root functions has been compared to results from the calibrated simulation using the Vargha-Delaney A-Test (Vargha and Delaney, 2000). Results from this comparison are noted above the respective bar within the plot.

Simulation responses reveal that a change in LTin migration rate to that of a square root function has no statistically significant effect on the area of PP generated. An effect classified by the test as 'small' is observed when contrasting number of patches generated for the same migration rate function, with a square root input function producing slightly fewer patches. When modelling cell input through use of an exponential function, larger statistical differences are observed for both patch area and number, with the Vargha-Delaney test resulting in a 'medium' difference for both patch characteristics, and fewer and smaller PP. This could suggest that LTin cell migration could potentially have a role in restricting the number and size of aggregations of cells that form by hour 72.

4.4.4 Investigating the Geography of PP Formation

In the current model, LTo cells that have the potential to form PP can be placed at any coordinate in the simulated tract. With some studies suggesting PP formation is localised to the anti-mesenteric border (Hoorweg and Cupedo, 2008; Randall *et al.*, 2008), LTo cell placement was restricted to a 15% band across the centre of the simulated gut, a band which has been used to represent that region. This percentage represents a best estimate as no biological measure was available. The calibrated simulation was run 300

times for both a restriction of 15% and no restriction. Under position restriction, the model predicts that few PP would develop in contrast to normal conditions (For 100% of the gut length - 15% restriction: $3.8+/-0.75$ patches; No restriction: $11.4+/-0.8$ patches; Mann-Whitney test p=$7x10^{-9}$).

The current calibrated simulator produces a representative number of patches when only 0.25% of the available LTo cells are set to have the potential to form PP (revealed in calibration in section 2.2.6). As the above result suggests that fewer patches are formed when cell placement is restricted to the anti-mesenteric border, simulation runs were performed to determine the number of patches produced if this percentage was increased. Results from this analysis have been plotted in Figure 4.5, which details number of patches along the whole intestine length rather than 10% as examined previously. When considering the whole length, experimental results suggest 8-12 PP are formed (Figure 1.2). The results in Figure 4.5 suggest that under a restricted condition, a greater number of LTo cells could have the potential to form PP, with the potential for this value to be near 2%.

At the current stage of biological understanding, it is not possible to conclude which theory is correct, either that there is a large restriction on the LTo cells that can potentially form patches, or there is a restriction in the geographic area where PP formation can occur. However, the simulation has suggested that both methods do form potential biological methods by which patch number is controlled.

4.5   Determining the Role of Simulation Parameters in Aggregation Size and Formation

The analyses in the previous chapter explored the impact that uncertainty in the value of six simulation parameters (Table 3.1) has on simulated cell behaviour during hour twelve and seventy-two of development . The remaining part of this chapter explores the impact that uncertainty in the value of these parameters has in the aggregation of hematopoietic cells, or the formation of a PP.

This analysis has been conducted by fixing one LTo cell at the centre of the simulated intestine tract at the beginning of the simulation. This restriction is in place to ensure that the analysis results reflect a change in parameter value, and are not a reflection of any effect caused by a change in location of the LTo cell. These experiments can be recreated in the publicly available simulator by setting the relevant simulation parameter controlling LTo cell positioning accordingly. Statistical methods within the *spartan* toolkit compiled in the completion of this study have been used to perturb the value of six parameters of interest and assess the impact on the cell aggregation that forms. The statistical techniques examine three patch characteristics that are output as simulation responses: the 2D area of the aggregation (calculated as detailed in Section 2.2.5) and the final expression levels of chemokine and adhesion factors.

4.5.1 Parameter Robustness

One-a-time analysis (Read *et al.*, 2012) was used to determine how sensitive the formation of PP is to the value of the six parameters detailed in section 3.4. Each was examined in turn, over the same values used in previous analyses and described in Table 3.1. Three-hundred simulation runs were performed for each parameter value to mitigate the effect of aleatory uncertainty. The distribution of patch characteristics obtained for that parameter value were then contrasted with a distribution of patch characteristics obtained from the simulator at baseline values using the Vargha-Delaney A-Test (Vargha and Delaney, 2000).

(i) **Chemokine Related Parameters**

Chemokine expression level is modelled by a use of a sigmoidal function as detailed in Figure 2.8, where two constants are used to set the initial level of expression on LTo cell differentiation and a maximum level of expression. An alteration in the constant that captures initial level of chemoattractant expression (Figure 4.6a) upon LTo cell differentiation has no significant effect on the size of the patch that is formed or the level of adhesion factor expression at the end of the development period. In contrast, an alteration in the initial chemokine expression level significantly effects the level of chemoattractants expressed by the LTo cell at the end of the 72 hour period. As the final expression level is dependent on both this initial level and cellular interactions that upregulate the expression of chemoattractants, changing the curve, this result could suggest that the patch characteristics observed in the baseline simulation occur through a restricted number of cellular interactions. A restricted number of interactions could make it impossible to recreate the baseline final chemokine level, thus explaining why there is a significant difference if the initial value is changed. In terms of assessing the current value assigned to this parameter, it becomes apparent that there is a large degree of uncertainty in what this value should be.

When examining Figure 4.6b, it should be noted that one of the patch characteristic cell responses, chemokine expression value at the end of the 72 hour period, is directly affected by a change in this parameter. For this reason, a trend is revealed where there is a significant difference in chemokine expression level at the end of the development period for any value other than the calibrated baseline. This suggests that for the baseline simulation, the maximum chemokine expression level is always reached, through the increase in chemokine expression brought about by stable cellular contacts. Thus if this threshold is changed, the new threshold is also reached, and thus the A-Test reports a significant change in result. This would suggest that determining a value for the maximum level of chemokine expression using the simulator would be difficult. Unsurprisingly, an increase in the maximum level of chemoattractant that can be expressed does

cause an effect on both patch area and adhesion factor expression that Vargha and Delaney deem a 'medium effect'. An increase in this parameter affects LTi cells over a greater distance from the LTo cell, thus promoting LTi cell chemotaxis and more cellular contacts, further upregulating adhesion factor expression and thus the number of cells retained in a forming patch. For this parameter, the whole range has not been analysed, as this represents a constant that is used in the sigmoidal curve calculation that captures the distance over which the chemokine is diffused. It would be assumed that a further increase in this parameter, and thus range of chemokine expression, would lead to a continuation of the trend seen for both these output responses, and a significant difference between the baseline patch characteristics.

Figure 4.6c is the same analysis for the parameter that captures the probability an LTi cell will not respond to chemokine level in its vicinity. An increase in this probability thus has a significant effect on all patch characteristic measures, as impairing LTi cell chemotaxis results in less cellular contacts and therefore less upregulation of adhesion and chemoattractant factors.

(ii) **Cell Binding Probability Parameters**

Figure 4.6d details the affect of changing the probability that a stable bind occurs when an LTo is in contact with an LTin/LTi cell. Such stable binds promote further upregulation of chemokines and adhesion factors. The results suggest that an alteration in the value of this probability has no significant effect on the formation of the PP. A significant effect is noticed when the parameter is assigned a value of zero but, as noted in analyses in the previous chapter that examine this parameter, this is an expected result, as a probability of zero would mean that a stable bind never occurs, and thus the LTo will never differentiate and produce the adhesion factors and chemokines required for PP formation.

(iii) **Adhesion Factor Related Parameters**

This analysis reveals that a change in adhesion factor expression by an LTo cell (Figure 4.6e) or the maximum probability an LTin/LTi cell will remain in prolonged contact with an LTo cell (Figure 4.6f) has no effect on any of the three patch characteristic simulation responses. This suggests that other biological factors are key to the formation of the aggregation, as a reduced or over expression still results in the same patch characteristics. This is in contrast to previous analyses that reveal adhesion factors play a key role in simulated cell behaviour (Section 3.7.2).

4.5.2 Compound Effects between Parameters

Using the latin-hypercube sampling approach described in section 2.4.3, 500 sets of simulation parameter values were generated, with values for each of the six parameters of interest (detailed in Section 3.4) chosen from within the ranges specified in Table 3.1 such that any potential correlations are negated. For each parameter set, 300 simulation runs were performed to mitigate the effect of aleatory uncertainty. Each run produces a set of output responses that describe the patch formed. Thus for each set of parameters, a distribution of 300 patch characteristics is generated. For each patch characteristic simulation measure, the median is taken of this distribution, and assigned as the simulation result under the conditions set by that parameter set. Each of the parameters was taken in turn and plots generated for each patch characteristic response, detailing the median responses observed for all values assigned to that parameter. With this resulting in eighteen separate plots, many of which reveal no compound effect, only a selection of these have been included in this thesis within Figure 4.7, but each result is detailed in the section below.

(i) **Chemokine Related Parameters**

No correlation is apparent between the initial level of chemokine expression on LTo cell differentiation and the resultant patch area (plot not shown). The same conclusion can also be drawn for the chemokine and adhesion factor expression level measures.

However a significant trend does emerge when the parameter that captures maximum chemokine expression is adjusted (Figure 4.7a), supporting evidence in one-a-time analysis that this is a key parameter in controlling patch area, and further supporting concerns in the uncertainty of this parameter value. A small trend is also apparent between this parameter and the level of adhesion factor expression reached by the LTo cell (Figure 4.7b). As an increase in maximum chemokine expression increases the distance over which LTi cell behaviour is affected, more cells would be recruited and more stable cellular contacts should take place, explaining the increase in adhesion factor expression. Finally, a large trend was observed for the chemokine expression level patch characteristic response. As a change in this parameter directly effects this simulation response, as discussed in one-a-time analysis (Section 4.5.1), this is an expected result of this analysis, and can thus be discounted (plot not shown).

When ordering results by the probability that an LTi cell does not respond to chemoattractant expression in its vicinity, a correlation between the parameter value and patch area does become apparent (Figure 4.7c), and becomes stronger as this probability is increased. When examining the effect of perturbing this parameter on adhesion factor expression (not shown), a similar effect is observed, with a correlation appearing at upper extreme values. As this probability affects

the recruitment of LTi cells towards an aggregating patch, this suggests the probability is an important factor in the size of the patch generated. The result also suggests that as the correlation is stronger towards extreme values, there could potentially be a large window of values beneath this extreme range that the true value for this parameter could fall within. No correlation is apparent between the value assigned to this parameter and the level of chemokine expression at the end of the development period.

(ii) **Cell Binding Probability Parameters**

This analysis reveals no trend between the probability that a stable bind occurs between an LTo and an LTin/LTi cell and the resultant patch area (not shown). The same conclusion was also drawn for chemokine expression response (not shown). When considering the level of adhesion factor expression, a small trend is apparent for probabilities between 0 and 0.1 (Figure 4.7d). As noted when this parameter was examined previously, this small trend can be explained by the fact that setting this parameter to zero results in no cell binding and no upregulation of the factors that cause patch aggregation, and thus such a small trend is expected.

(iii) **Adhesion Factor Related Parameters**

Figures 4.7e and 4.7f support the conclusion drawn in one-a-time analysis above; that the value of adhesion factor expression does not have a key role in the formation of a PP. No trend becomes apparent for either parameter. This is also the case for both the adhesion and chemokine expression responses (not shown).

4.5.3   Partitioning of Variance

The extended Fourier Amplitude Sampling Test (eFAST) (Marino *et al.*, 2008; Saltelli *et al.*, 2000), detailed in section 2.4.4, has been used to determine the proportion of variance in patch characteristic responses that can be explained by each of the parameters investigated. Five-hundred sets of parameter values were generated, with each parameter being assigned a value within the ranges specified in Table 3.1. The 'dummy' parameter used for statistical comparison was again set a value range of 1 to 10 and included in the sampling procedure. For each parameter value set, 300 runs were performed to mitigate aleatory uncertainty that could affect patch characteristic results. Simulation responses are analysed using the Fourier Frequency approach described in 2.4.4 (Marino *et al.*, 2008; Saltelli, 2004), and plots produced detailing the median First-Order (Si) and Total-Order (STi) sensitivity indexes for each parameter. The sensitivity indexes and measures of significance in comparison to the dummy parameter are detailed in Table 4.1.

(i) **Patch Area**

Contrasting the first-order sensitivity indexes (Si) for each of the six parameters with those of the dummy (Figure 4.8a) reveal that three parameters account for a statistically significant proportion of the variance (where *p < 0.05*). Similar to the previous analyses performed using this approach, one of these parameters is the probability that a stable bind is formed on contact with an LTo and hematopoietic (LTin/LTi) cell, a result that will become apparent due to the effect the lower extreme value has on both cell behaviour and patch formation. The remaining parameters are chemokine related; the probability an LTi cell responds to chemokine in the vicinity and the maximum level of chemokine expression. This supports the trends observed in the latin-hypercube sampling analysis performed previously, that the key pathway in patch size is the expression of and response to chemokines.

When examining the total-order sensitivity indexes (STi) in Figure 4.8a, that identify compound effects between the parameter and its complementary set, the probability that an LTi cell responds to chemokine is again revealed as being significant. This result suggests that the response to chemokine expression by an LTi cell is key to the aggregation of patches, rather than the amount of adhesion factor expression that is retaining these cells in the vicinity of the LTo.

(ii) **Adhesion Factor Expression Level**

This analysis technique determines that three of the six parameters examined account for a significant proportion of variance in adhesion factor expression caused by parameter perturbation (Figure 4.8b). Two are the same chemokine parameters that account for a significant proportion in patch area as described above. The third is an alteration in the expression level of adhesion factors on each stable contact. As this is directly linked to the response being measured (the final level of adhesion factor expression), this is a result that is to be expected. However the analysis has given further support to the identification of the chemokine expression and response pathway as the key reason for variance in simulation results.

Significant compound effects (STi values) are revealed for four of the parameters, the three that capture the chemokine expression and response pathway and the expression level of adhesion factors. As noted above, one would expect a change in the expression level of adhesion factors to influence the final level of adhesion factor expression within a forming patch, and thus it is no surprise this is revealed as a parameter that strongly interacts with the complementary set. Again chemokine expression and response parameters are revealed as key to variance in adhesion levels, although the values of all parameters are being perturbed simultaneously.

(iii) **Chemokine Expression Level**

An initial examination of the graph in Figure 4.8c reveals that a large percentage of the variance in final chemokine expression level can be explained by compound effects between each parameter and its complementary set (the STi measure). In this case, as observed in the latin-hypercube analysis above, the final level of chemokine expression achieved is directly influenced by changing the value of the maximum chemokine expression parameter. As this is the case, this analysis is affected by this link, and no further meaningful result has been produced. Thus these results can be disregarded.

Parameter	Si	P-Value	STi	P-Value
stableBindProbability	0.043	0.049*	0.226	0.097
chemokineExpressionThreshold	0.278	0.003**	0.501	0.015*
initialChemokineExpressionValue	0.043	0.074	0.311	0.068
maxChemokineExpressionValue	0.422	0.033*	0.311	0.068
adhesionFactorExpressionSlope	0.031	0.133	0.246	0.142
maxProbabilityOfAdhesion	0.025	0.117	0.142	0.288
dummy	0.011		0.099	

(a) Patch Area Response

Parameter	Si	P-Value	STi	P-Value
stableBindProbability	0.144	0.101	0.216	0.135
chemokineExpressionThreshold	0.403	0.002*	0.568	0.003**
initialChemokineExpressionValue	0.033	0.120	0.167	0.006**
maxChemokineExpressionValue	0.274	0.028*	0.441	0.018*
adhesionFactorExpressionSlope	0.025	0.001**	0.185	0.013*
maxProbabilityOfAdhesion	0.077	0.087	0.288	0.099
dummy	0.006		0.058	

(b) Adhesion Factor Expression Level Response

Parameter	Si	P-Value	STi	P-Value
stableBindProbability	0.108	0.120	0.750	0.130
chemokineExpressionThreshold	0.105	0.088	0.762	0.088
initialChemokineExpressionValue	0.108	0.082	0.766	0.084
maxChemokineExpressionValue	0.336	0.070	0.906	0.048*
adhesionFactorExpressionSlope	0.096	0.247	0.667	0.266
maxProbabilityOfAdhesion	0.093	0.266	0.661	0.269
dummy	0.072		0.518	

(c) Chemokine Expression Level Response

Table 4.1: Median sensitivity indexes and measures of statistical significance for each parameter examined using the eFAST technique, for all patch characteristic responses. Si: First-Order Sensitivity Index; STi: Total-Order Sensitivity Index. Both are calculated for each re-sample curve and the median value taken. P-Value calculated using two-sample t-test to the distributions comprised of the results from each re-sample curve. * indicates statistical significance at 0.05 level, ** indicates statistical significance at 0.01 level.

4.6 Discussion

4.6.1 Simulation as a tool for hypothesis generation

The pairing of computer simulation with current laboratory approaches is becoming increasingly prevalent as it provides a means to perform investigations that either cannot currently be performed by other means, or that can inform future laboratory investigations (Andrews *et al.*, 2008; Germain *et al.*, 2011). Investigations that apply this technique can be termed *in silico* experimentation. This approach has yet to be utilised to explore secondary lymphoid organ development. However the simulation platform developed as described in Chapter 2 makes such explorations possible.

Three novel *in silico* explorations have been detailed in this chapter. Although much of the underlying detail of PP development is now well understood (van de Pavert and Mebius, 2010; Randall *et al.*, 2008; Veiga-Fernandes *et al.*, 2007), there remain interesting questions that have yet to be addressed. The variation in patch number, size, and location between the different intestine observations in Figure 1.2 for mice and in studies by Cornes (1965) is one example. The three explorations conducted in this chapter suggest hypotheses that could explain some of this variability, as well as addressing further uncertainty surrounding the role of LTin cells (Patel *et al.*, 2012).

Firstly the impact of changing the LTin cell population size was examined (Figure 4.3). Data presented here suggests that the number of PP that form by the end of hour 72 is related to the number of LTin cells in the gut. A reduction in cell number leads to a statistically significant change in the number of patches, although not a significant reduction in patch area. The opposite effect on patch number is observed when the number of LTin cells is increased by 2 and 3 fold, however no further increase in patch area or number occurs when LTin cell population is increased by 4 and 5 times. Previous experimental work has shown that LTin cells express RET, which binds to an LTo cell expressing ARTN, initiating LTo cell differentiation and the process of PP formation (Patel *et al.*, 2012; Veiga-Fernandes *et al.*, 2007). Explorations using the simulator suggest that not only do LTin cells trigger the process, they could potentially control PP number. Small variations in LTin cell number between samples may then explain the variation seen in Figure 1.2. However, the trend between patch number and LTin cell number does not continue for a vast overexpression of LTin cells. This is potentially explained by limited RET ligand availability, and thus a continued overexpression has no impact, especially as LTin cells are not thought to express receptors for chemoattractants and are rarely recruited into PP (Randall *et al.*, 2008). This is in contrast to an overexpression of LTi cells, which has been done experimentally and does lead to a significant difference in PP size (Meier *et al.*, 2007).

Results that help form the above hypothesis are based on the assumption that LTin cells migrate into the gut at a set rate that does not change through the course of the simulation. Whereas flow cytometry results have been used to estimate LTin

cell count at E15.5, no further time-point data was available, explaining the need to make this assumption. The second *in silico* exploration examined the effect on PP characteristics if LTin migration rate was not constant, but either high initially and tails off (a square root function), or low initially and increases rapidly (an exponential function). These input functions were both investigated and results presented in Figure 4.5. Input rate functions were generated such that each method of setting the rate led to the creation of the number of cells estimated using flow cytometry results at E15.5. Changing the migration rate to a use a square root function had no significant impact on the number and area of patches observed (Figures 4.4b and 4.4c). However it can be observed in the same figures that patch number and area did reduce if cell migration is modelled using an exponential curve, where migration is initially slow and rapidly increases. This could suggest that it is not only LTin cell number that is controlling patch number, but the rate of LTin migration into the gut could also be influential. Future laboratory investigations should test this hypothesis, by performing flow cytometry on guts taken from human-CD2-GFP transgenic mice at different time-points, through which estimations of cell counts can be made and used to set a realistic LTin input curve. If this data was made available, the simulator could have a key role in furthering the understanding of the role of LTin cells, providing more weight to the hypotheses generated.

The final *in silico* exploration performed in this chapter considered an alternative method of controlling the number of PP that develop over the 72 hour period. The current model makes the assumption that LTo cells that express RET ligand could be present at any point on the gut surface. Flow cytometry results were used to estimate that 20% of the surface area of the gut contains cells that have the potential to form PP, but simulation calibration results suggested that the correct number of PP are formed if only 0.25% of this 20% express RET ligand (Section 2.2.6). However, it has been observed that PP tend to form opposite the mesentery, known as the anti-mesenteric border (van de Pavert and Mebius, 2010; Randall *et al.*, 2008). The mesentery attaches the small intestine to the abdominal cavity, and contains vessels through which hematopoietic cells migrate into the intestine (Eberl *et al.*, 2004). One theory concerning why PP form in this particular region could be that the expression of RET ligand is restricted to LTo cells opposite the mesentery. A restriction was implemented such that only LTo cells placed within a set region could express RET ligand. Unsurprisingly when run under these conditions, the number of patches formed by the calibrated simulator fell sharply (Figure 4.5). With this result taken into consideration, the simulator was used to investigate how much the percentage of active LTo cells in the simulation could be increased with this restriction in place. A representative number of patches (8-12 in the mouse, Figure 1.2) were observed when increasing this percentage to 2%. The exploration does suggest RET ligand expression restriction could be a further method by which patch formation is controlled. As RET

is known to be a key initiator in the development process (Veiga-Fernandes *et al.*, 2007) this hypothesis could hold. However, there is still little biological evidence to suggest that such a geographic restriction occurs, and some patches do form that are not on the anti-mesenteric border. It is not known how plausible it is as to whether only 0.25%, or indeed 2%, of the large number of LTo cells on the gut surface possess the ability to form PP. If this is plausible, this itself could be the factor controlling PP number, and it would be assumed that there is some unknown factor causing an early differentiation of this selected set of LTo cells. If these percentages were implausible, this suggests patch number is being controlled another way, potentially by the physical geometry of the intestine or environmental changes as the gut develops. The simulator can produce hypotheses but there is currently little geographic data on which these can be based and thus supported.

4.6.2 Statistical Analysis Reveals Chemokine Expression Dominant Factor in Patch Aggregation

It has previously been suggested that the process of PP development is chemokine driven (Luther *et al.*, 2003; Randall *et al.*, 2008). This conclusion has been drawn from studies of mice deficient for genes encoding a particular function (Eberl *et al.*, 2004; Luther *et al.*, 2003; Meier *et al.*, 2007; Veiga-Fernandes *et al.*, 2007). However no study has yet been undertaken to quantify the effect of changing the levels of each biological factor involved in organogenesis. The statistical techniques compiled within the *spartan* package have been utilised to reveal the key factors in causing cell aggregation throughout the simulated time. This has been achieved by examining the six biological factors for which a value remains unknown (Table 3.1), and exploring the effect on cell behaviour when the value of these parameters is perturbed.

Application of latin-hypercube (Read *et al.*, 2012) and eFAST (Marino *et al.*, 2008; Saltelli *et al.*, 2000) techniques provides an indication of the influence of a particular factor. Analyses using these approaches has revealed a significant role for chemokine expression and response, suggested by Partial Rank Correlation Coefficients calculated from latin-hypercube simulation results (Figure 4.7) and by eFAST sensitivity indexes (Table 4.1). Previous analyses in this thesis have suggested that adhesion factors have a key role in influencing cell behaviour, however a change in the value of adhesion factor expression has no impact on the three patch characteristics the simulation produces. Thus the simulation agrees with the hypotheses in the literature, that the process is chemokine dependent.

Previous studies by Meier et al (2007) examined the role of LTi cells in tissue development using IL-7 transgenic mice. This revealed that an overexpression of IL-7 led to a larger number of LTi cells in the gut and larger Peyer's Patches. In contrast, mice deficient for LTi cells do not have the ability to form PP (Sun, 2000; Yokota *et al.*, 1999). A number of studies have revealed that LTi cells express receptors for

chemokines CXCL13, CCL19 and CCL21 expressed by the LTo cell (Luther *et al.*, 2003; Ohl *et al.*, 2003). In the study in this thesis, an assumption was made in the domain model that an LTi cell may not always respond to chemokine in its vicinity. In the platform model, this is captured as a probability that the cell may not respond to chemokine expression. Statistical analyses of a change in this probability reveals it is influential in affecting patch area and chemokine expression when perturbed individually (Figure 4.6c), and is an influential pathway in affecting both these measures when the six parameters of interest are being perturbed simultaneously (Figure 4.7c). This suggests it is not sufficient to state that the process of PP development is dependent on chemokine expression: the result is dependent on both the expression of and LTi response to chemokines. During calibration, this probability was determined to be a low value (0.3%), suggesting it is unlikely an LTi cell will not respond to chemokine expression in the vicinity. It is unknown as to whether this is biologically plausible, or whether the affinity between the receptor and chemokine expressed needs to be stronger. If this was the case, this would also imply that LTi cells have a role in controlling Peyer's Patch size, through limiting response to environmental factors.

4.6.3 An Interaction Focused Rather Than Reductionist Approach

The use of simulation to study cellular interactions shifts the focus from an examination of each individual component part to that of the higher order behaviour and how this emerges from components that lack the capability to create this phenomena alone (Germain *et al.*, 2011). A large number of published studies exist that examine the role of particular biological components involved in PP development (Eberl *et al.*, 2004; Luther *et al.*, 2003; Meier *et al.*, 2007; Veiga-Fernandes *et al.*, 2007). However, the patch characteristic analyses in this chapter provides good examples of why taking such a reductionist approach does not reveal the full picture concerning the system dynamics. This section discusses one such example.

One-a-time analysis was performed for each parameter using the technique described by Read et al (2012), and revealed that a significant change in the level of chemokine expression reached by the LTo cell when the initial level of expression on cell differentiation was altered (Figure 4.6a). In some respects this mimics a reductionist approach that considers the effect of each component individually, and draws conclusions concerning how the component affects system dynamics just from this examination. Thus in this example, it would be assumed that the initial level of chemokine expression is an important factor. However, perturbing the value of initial chemokine expression level with the other five parameters being examined simultaneously, using both latin-hypercube sampling (Read *et al.*, 2012) and eFAST techniques (Marino *et al.*, 2008; Saltelli *et al.*, 2000) reveals no trend in simulation response. The parameter may be influential when examined on its own, but this influence is largely

dependent on the value assigned to other parameters in the model.

It is examples like this that highlight reasons for integrating modelling, simulation and statistical analysis techniques with current laboratory techniques. The quantitative analyses that have been performed on six simulated biological components in this and the previous chapter cannot be performed in a lab, and have the potential to offer biological insight that cannot be revealed in other methods (Germain *et al.*, 2011). The development of structured methods in the creation of models and simulations, such as that of the CoSMoS process (Andrews *et al.*, 2010), and the availability of statistical toolkits such as *spartan*, is making it possible to generate models where although some detail is abstracted from system dynamics, the level of confidence in the simulator as a representative tool can be established (Andrews *et al.*, 2008; Polack *et al.*, 2010; Read *et al.*, 2012). Establishing a level of confidence is vital if meaningful results are to be produced that shift the field towards utilising an interaction based approach.

## 4.6.4 A Simulation Approach Can Have Limitations

This discussion has raised the prospect of studies being conducted where a robust model and simulation is produced, statistical analyses used that establish the confidence in results generated, and *in silico* experimentation performed to reveal additional biological insight. Previous sections of this discussion detail areas of the biological understanding where taking this approach in examining PP development has produced some hypotheses that could address some unanswered questions concerning system dynamics. However, it must be noted that this should not be considered the end of the process, and explorations using simulation can raise additional questions.

Previously published studies have detailed a role for chemokine expression and response in the development of PP (Ansel *et al.*, 2000; Luther *et al.*, 2003). This chapter has taken this forward and provided a quantitative examination of the effect of changing chemokine expression levels, a feature that has been determined to be a key pathway in the simulation. However it has been difficult to set an exact value for the maximum level of chemokine expression by an LTo cell, and there is a large degree of uncertainty in the value of this parameter. It has been demonstrated that a change in the maximum level of expression always changes the level of chemokine expressed reached by the LTo cell, suggesting this threshold is always met (Figure 4.6b).

This does not detract from the identification of this as a key pathway, more emphasises the fact that simulation can provide some key insight, but does have some limitations. In this case, it has not proved possible to reveal the maximum level of chemokine expressed by an LTo. However this could be achieved with more biological information, which could come in one of three forms. The first is a quantitative study of chemokine expression at different time-points in development, although this is very difficult to achieve in the lab. This data could then inform the setting of the constants

controlling the sigmoid curve that models chemokine expression. Secondly, and much more tractable, is a study of LTi cell behaviour at different time-points in development. This could be done in a similar way to the *ex vivo* work produced by Patel et al (2012), and cells tracked in an explant culture system during a variety of hours of development. Changes in LTi cell behaviour over time, and at certain distances from an LTo cell, would provide a means of calculating how chemokines are diffused over time, and thus the simulator could be recalibrated based on cell behaviour. Finally, reevaluating the measures that define a patch could also be beneficial. If it was possible to determine the number of LTi cells within an aggregation at certain time-points, chemokine expression could be modelled in such a way that aims to recruit LTi cells in such a way that recreates that dynamic.

The approach to simulation construction in this thesis has followed the framework in the CoSMoS process (Andrews *et al.*, 2010), which defines the generation of a series of models, from which a simulation is implemented. One of the key points in this framework is that the process has no end point: results from simulation may then feed further work in developing the model, either by examining the abstractions and assumptions that were made or through performing work in the lab that then feeds a further iteration of the model. In some respects, work in this thesis is the end of one iteration. Experiments have been suggested that could improve the model, and if these were completed, the process would then begin again, and potentially the limitation identified above could be addressed.
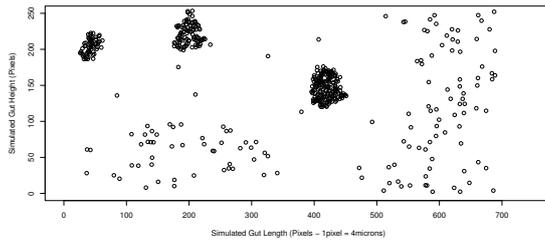
**(a)** Control: Wild-Type Mice (Simulation under baseline conditions)
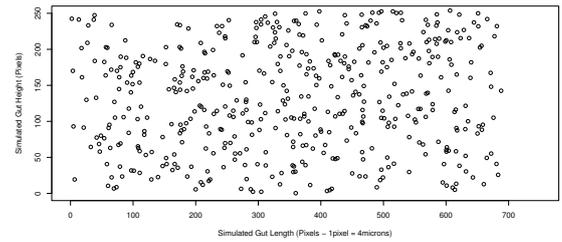
(b) RET Deficient Mice (Veiga-Fernandes *et al.*, 2007)
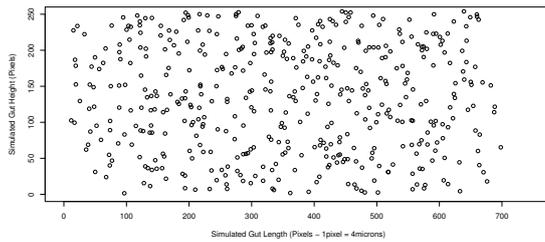
(c) Chemokine deficient mice (Luther *et al.*, 2003)
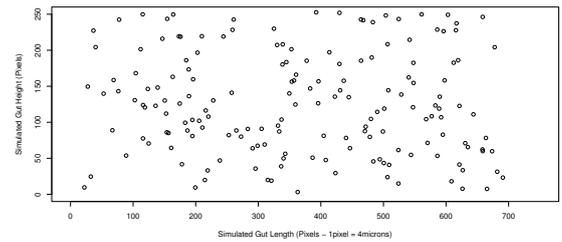
(d) RORγ deficient mice (Eberl *et al.*, 2004)
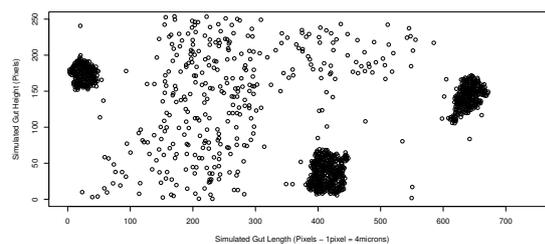
(e) IL-7^{Tg} mice (Meier *et al.*, 2007)

Figure 4.1: Use of the simulator to reproduce previously published results. Graph represents 10% of the foetal intestine length. Circles represent the locations of LTin and LTi cells. This figure has been adapted from that included within Alden et al (2012b)
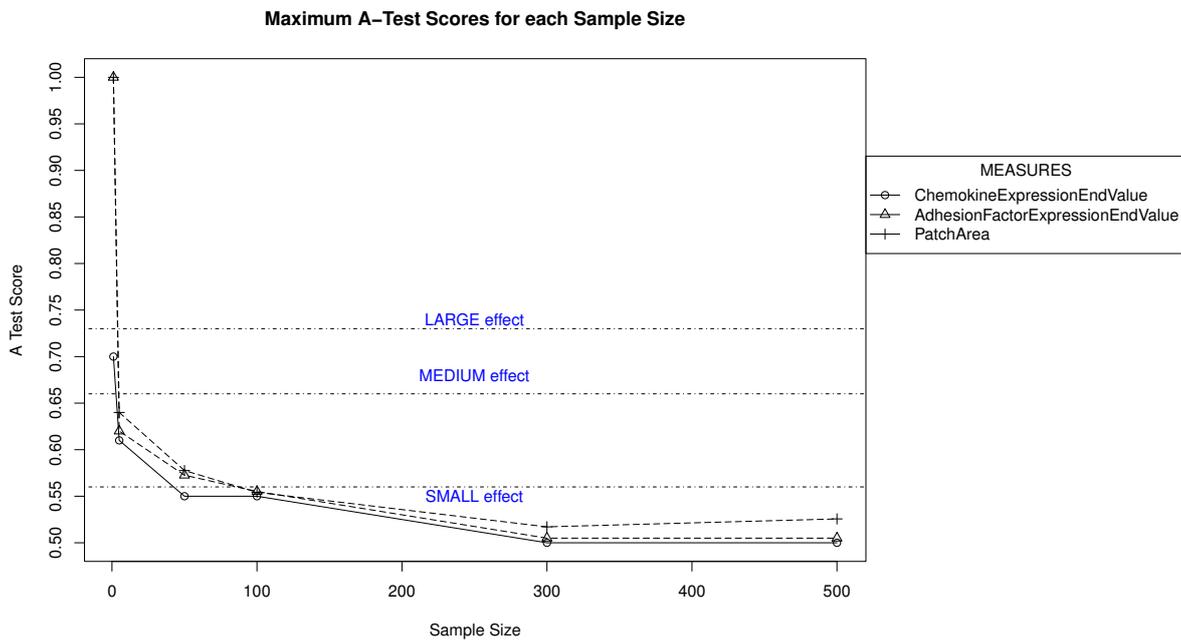
Figure 4.2: Examining the effect of aleatory uncertainty on the results of the simulation for a variety of sample sizes. This summarises the maximum A test score for patch characteristic measures when the sample size is varied. Where the maximum score was less than 0.5, the corresponding value above 0.5 has been assigned. This has been done as the magnitude of the effect is of more interest than the direction. For each sample size, twenty runs are performed, with no parameters changed each time. Using this technique, the number of runs necessary to produce a robust result can be ascertained. In this case, there is little reduction from 300 runs or greater.

(a) Patch Area on LTin Cell Decrease

(b) Statistical Measure of Difference when LTin Cell Number Decreased

(c) Patch Area on LTin Cell Increase

(d) Statistical Measure of Difference when LTin Cell Number Increased
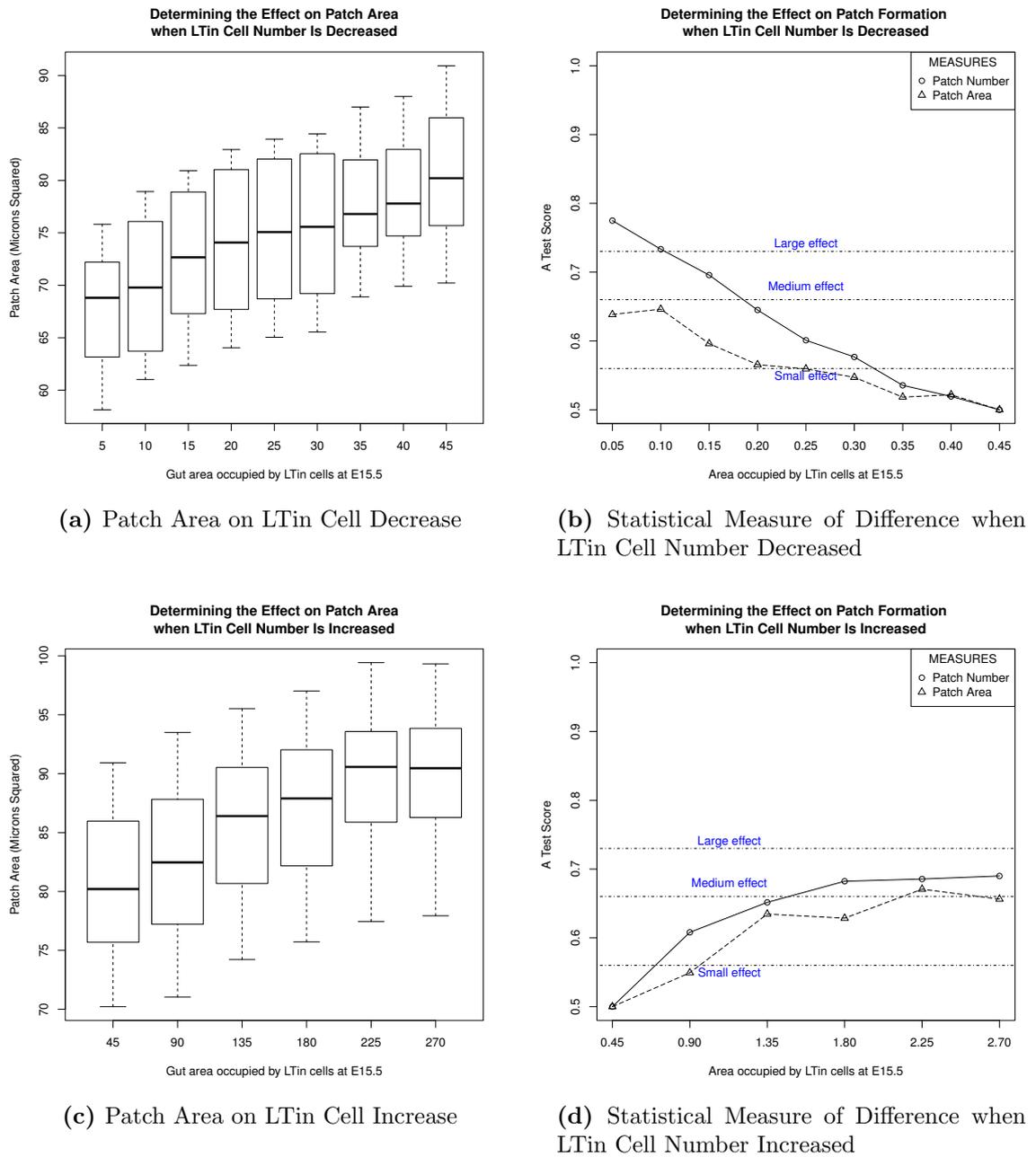
Figure 4.3: Use of the simulator to investigate the impact of LTin cell number to simulation result. Top: investigating a decrease in LTin cell number; Bottom: investigating a 2,3,4, and 5 fold increase in the number of LTin cells. Simulations were run 300 times for each LTin cell parameter value and medians taken for both Patch Area and Patch Number. The left column contains a boxplot of the patch area for each value the parameter has been assigned. The right column contains the result of a comparison between patch characteristics observed at baseline values and those observed when the parameter is perturbed, using the Vargha-Delaney Test (Vargha and Delaney, 2000).
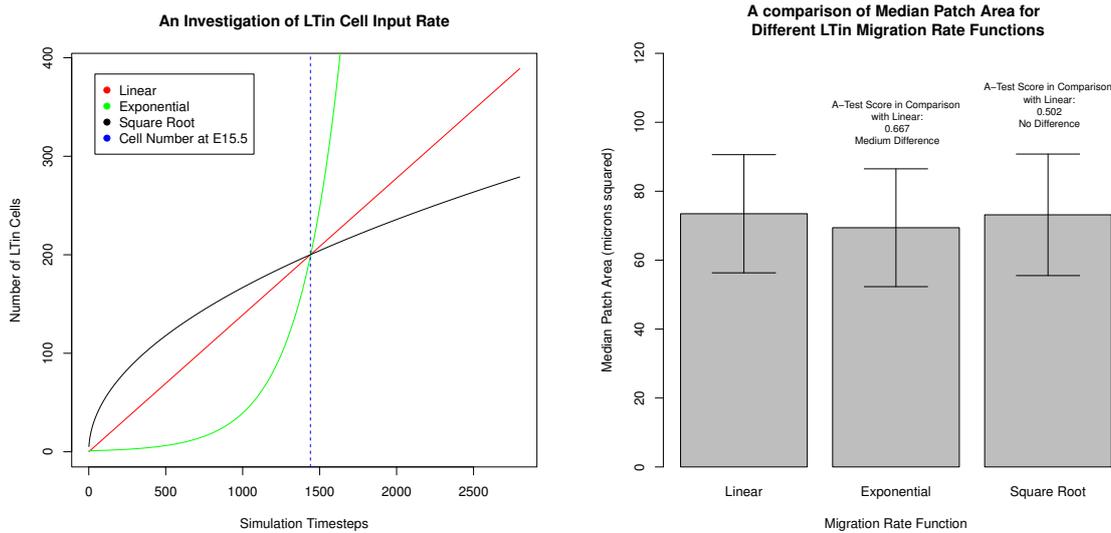
(a) Migration Rate Functions Investigated

(b) A comparison of median patch area at hour 72 for three LTin cell migration rate functions
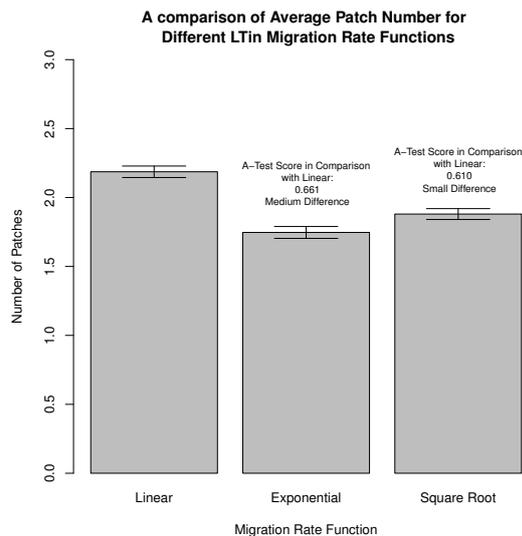
(c) A comparison of median number of PP at hour 72 for three LTin cell migration rate functions

Figure 4.4: Investigating LTin cell migration rate by changing the input rate function within the simulator. (a): Flow cytometry data has been used to estimate the number of LTin cells present in the simulator at E15.5 (blue line). With cell counts at other time-points unavailable, the simulator assumes a linear input rate, that meets the number of LTin cells observed at E15.5, and continues at the same trajectory until E17.5 (red line). However this may not be the case. To investigate the effect that cell input rate has on results, two alternative input rates have been investigated, an exponential (green line) and square root (black line) function. These three lines converge at E15.5, matching the number of cells observed by flow cytometry. Three hundred simulation runs have been performed for each input rate function and median patch characteristics calculated. (b): A comparison of the median patch area observed for each input rate function. (c): A comparison of the median number of patches for each function. Results for the exponential and square root functions have been contrasted to the linear input rate using the Vargha-Delaney A-Test (Vargha and Delaney, 2000), the result of which is noted on the plot. Error bars are s.e.m.
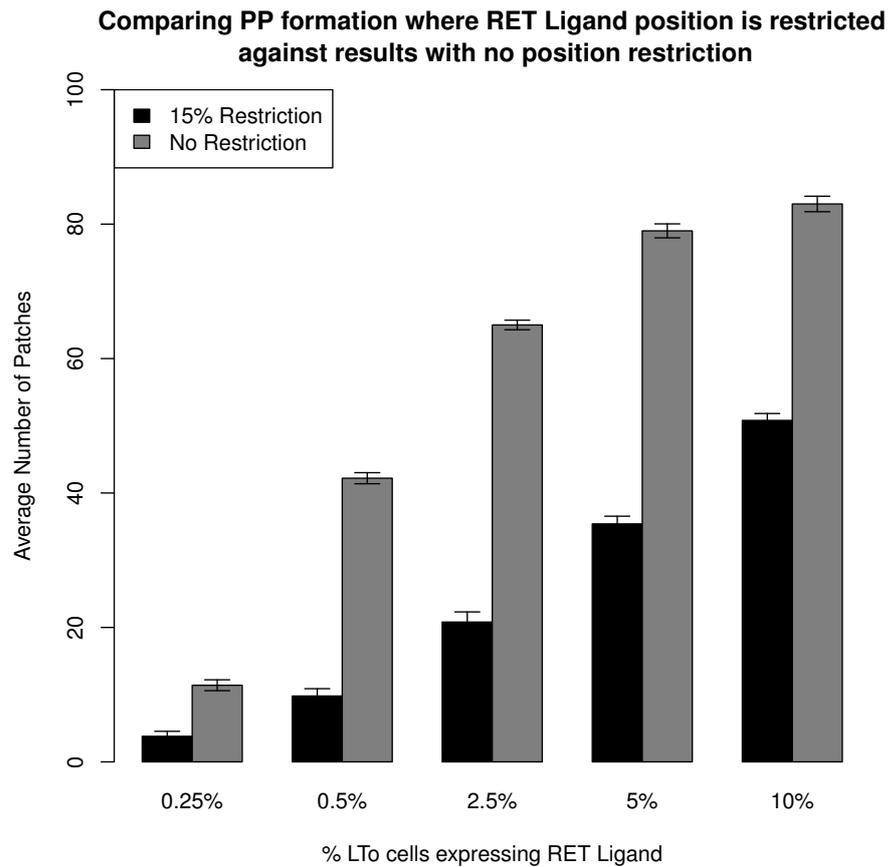
Figure 4.5: An investigation into how geographic restriction of RET ligand expression affects PP formation. Simulations were run with active RET ligand expressing LTo cell position restricted to a 15% band across the centre of the tract, and the number of patches formed compared with the case where there is no restriction. The percentage of LTo cells capable of forming PP was also investigated for both conditions. 300 runs performed for each condition. Black: 15% position band restriction; Grey: no restriction. Error bar: s.e.m.
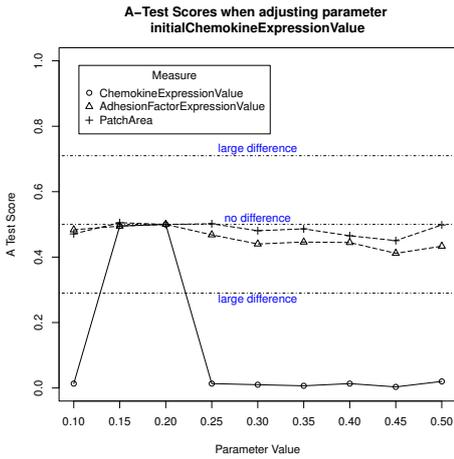
(a) Initial level of chemoattractant expression at LTo differentiation
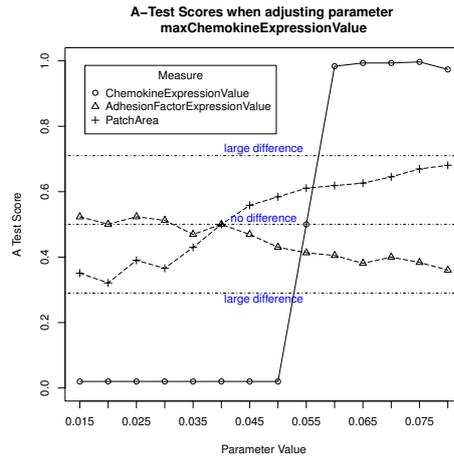
(b) Saturation limit of chemoattractant expression
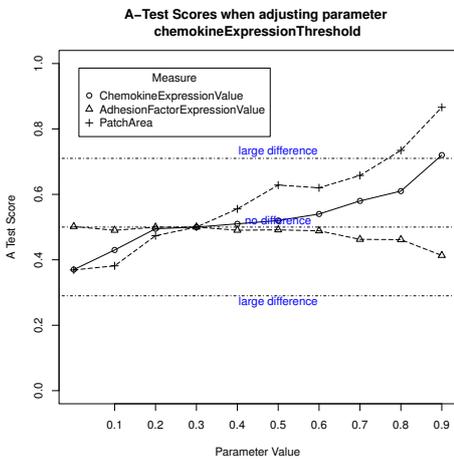
(c) Chemokine Level at which LTi chemotaxis occurs
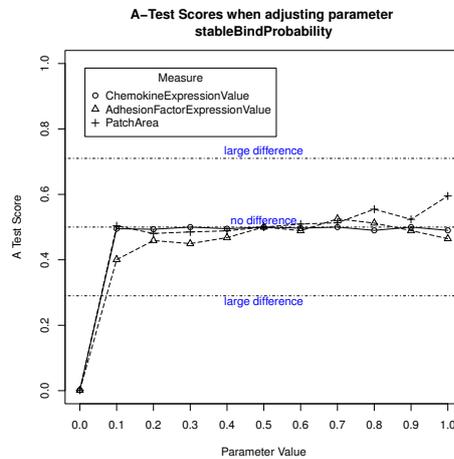
(d) Probability a LTin/LTi and LTo cell form stable bind on contact
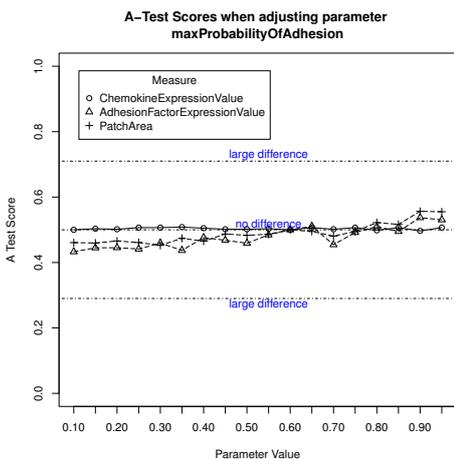
(e) Maximum probability adhesion factors prolong cell contact
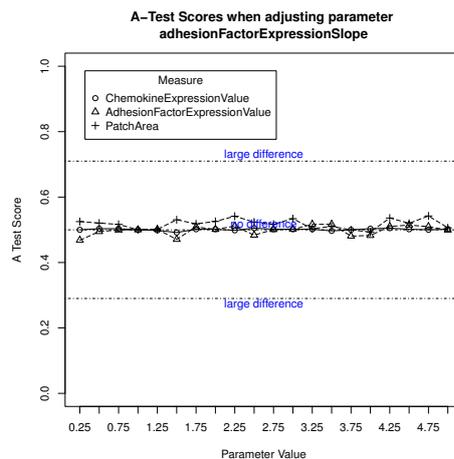
(f) Level of adhesion factor expression per stable contact

Figure 4.6: Determining the robustness of patch characteristic responses at the early time-point in development. The six parameters for which a value is unknown were examined in turn, and the value of each perturbed within a specified range. Simulation results were compared to those generated during calibration, using the Vargha-Delaney A-Test, to determine if a significant change in cell behaviour has occurred
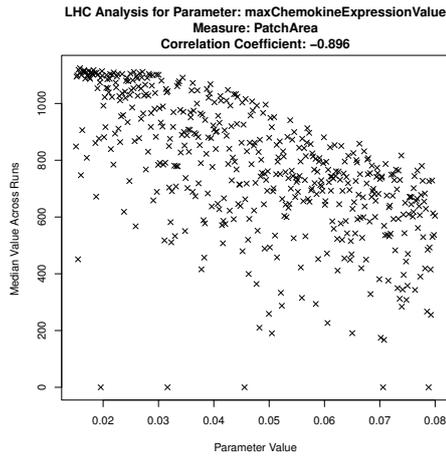
(a) Measure: Patch Area. Parameter: Chemokine saturation limit
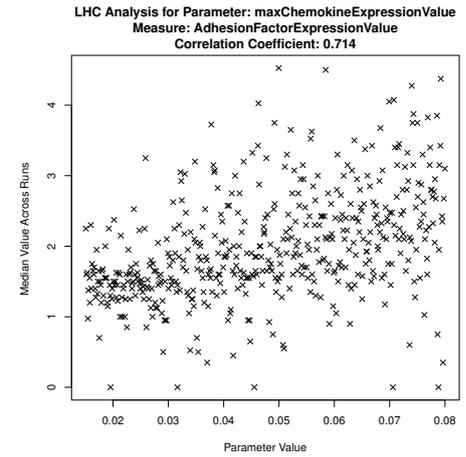
(b) Measure: Adhesion expression level. Parameter: Chemokine saturation limit
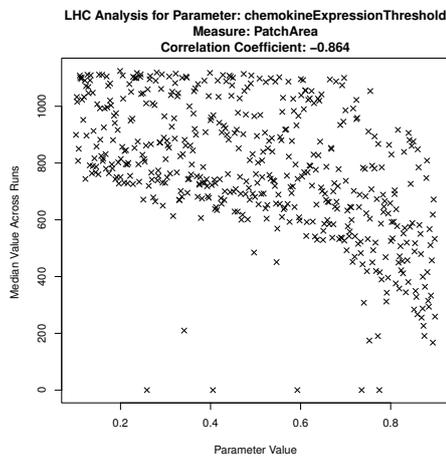
(c) Measure: Patch Area. Parameter: Probability LTi cell does not respond to chemokine
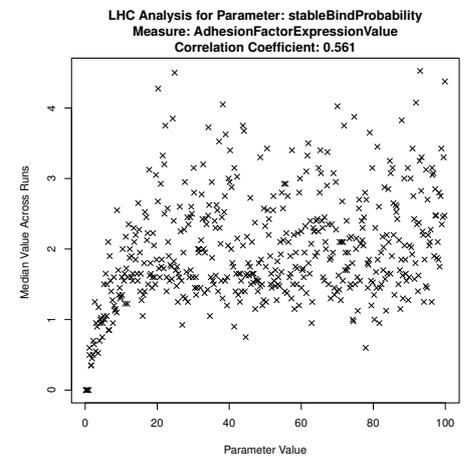
(d) Measure: Adhesion expression level. Parameter: Stable bind probability
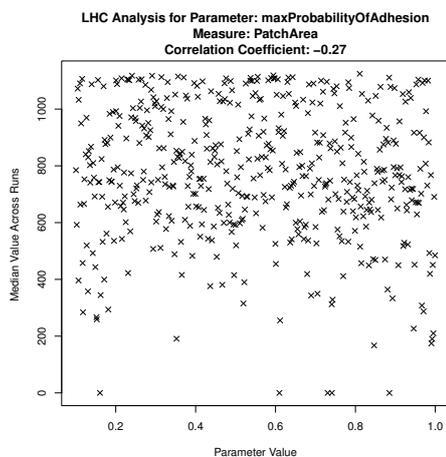
(e) Measure: Patch Area. Parameter: Maximum probability adhesion prolongs contact
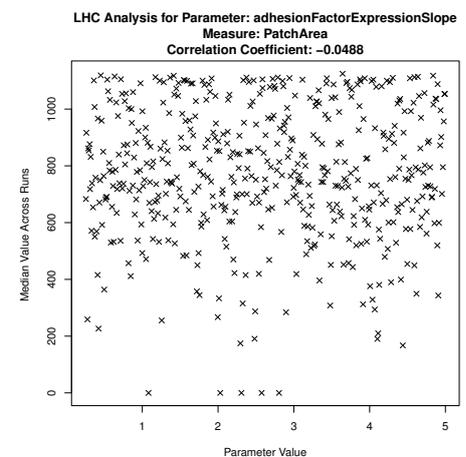
(f) Measure: Patch Area. Parameter: Level of adhesion expression per stable contact

Figure 4.7: Identifying any compound effects on patch characteristics through latin-hypercube sampling, a technique that perturbs the value of all parameters simultaneously. Influential parameters are identified by trends that emerge in simulation results, and through Partial Rank Correlation Coefficients specified in the graph header.
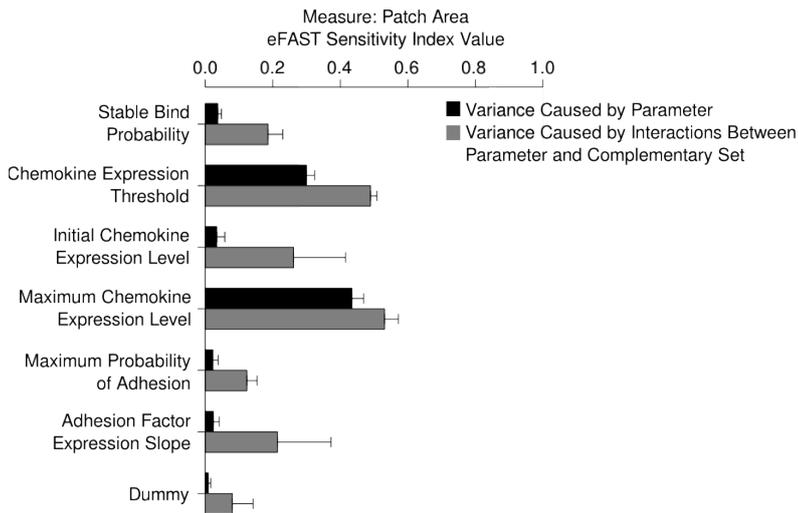
Measure: Patch Area
eFAST Sensitivity Index Value

(a) Initial level of chemoattractant expression at LTo differentiation

Measure: Adhesion Factor Expression Level
eFAST Sensitivity Index Value

(b) Saturation limit of chemoattractant expression

Measure: Chemokine Expression Level
eFAST Sensitivity Index Value
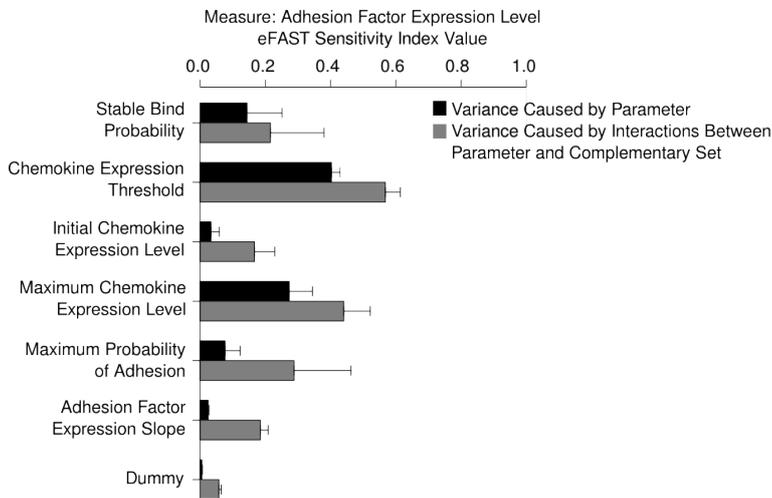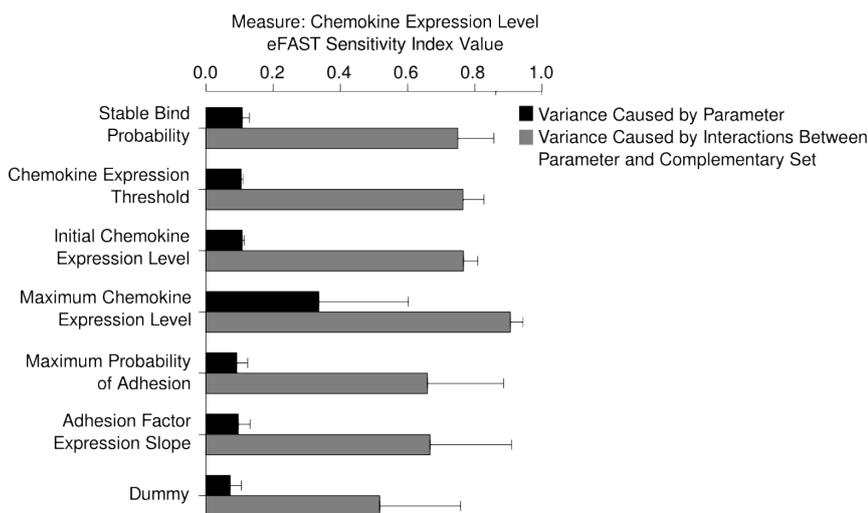
(c) Chemokine Level at which LTi chemotaxis occurs

Figure 4.8: Using the Extended Fourier Amplitude Sampling Test (eFAST) to explore the influence each parameter for which a value is unknown has on PP formation. Black: the fraction of model output variance accounted for by a variation in the value of that parameter (Si); Grey: remaining variance accounted for by higher-order interactions between this parameter and its complementary set (STi).

Chapter 5

Time-Lapse Analysis through Simulation

Data presented in this thesis suggests that Peyer's Patch development may be biphasic, the first phase mediated by adhesion factor expression and the second by chemokine expression and response. This hypothesis is drawn from a statistical analysis of simulated cell behaviour during two distinct time-points, hour 12 and hour 72. This chapter demonstrates a novel use of simulation and statistical techniques to perform a time-lapse analysis of cell behaviour, to suggest the point in the development period where a change in phase occurs.

5.1   Introduction

The analyses in the previous chapters of this thesis have utilised statistical tools to suggest which biological parameters are influencing cell behaviour during hour twelve and seventy-two of development. It has been determined that, in simulation, adhesion factors are influential in causing the emergent cellular behaviour observed at the twelve hour timepoint, and that chemoattractant factor expression has no role in influencing cell behaviour. However, performing the same analyses for cell behaviour during hour 72 of development reveal that chemoattractant expression is influential at the end of the process, as suggested in the literature (Adachi *et al.*, 1997; Eberl *et al.*, 2004). This could suggest the existence of two distinct phases of Peyer's Patch development within the 72 hour period; the first mediated by adhesion and the latter mediated by chemoattractant expression, and raises the important question of when a biological factor such as an adhesion factor or chemokine becomes influential in tissue development.

Current confocal microscopy techniques allow for the behaviour of cells to be tracked over a set time period, a technique utilised to generate the cell behaviour results on which this simulator was calibrated (Patel *et al.*, 2012). Although such time-lapse analysis has proven useful in examining cell behaviour and how this changes, it does not further the understanding of the role each biological factor has in the behaviour that is observed. Although it is possible to use the technique to examine cell behaviour under different conditions, for example within an *ex vivo* culture system, such an experimental set up can be time consuming and expensive, especially when examining the system under a number of conditions.

Although the integration of computer simulation with current experimental results is becoming a popular approach in furthering the understanding of biological systems, the application of simulation as a tool to perform time-lapse experimentation has been limited. The only example of such an application to date is in studies of granuloma performance for controlling Mycobacterium tuberculosis infection (Ray *et al.*, 2009), where agents within the system have been tracked constantly through the simulation time period and sensitivity analysis techniques used to examine behaviour at regular time intervals. Utilising such an approach has the potential to not only suggest the biological factors that are highly influential, but suggest the time-points where system dynamics are influenced by particular parameters, potentially revealing that the behaviour that emerges through these system dynamics occurs in distinct phases.

This chapter examines the use of computer simulation as a tool to perform a time-lapse analysis of cell behaviour responses in PP development. The simulator has been configured such that cell behaviour responses are recorded at twelve hour intervals, thus it is possible to examine changes in cell behaviour over time. This can be achieved using the same sensitivity analysis techniques that were utilised in the previous chapters and included within the *spartan* package. Firstly, simulation robustness to parameter perturbation is examined using the one-at-a-time technique (Read *et al.*, 2012).

Each of a set of parameters of interest is examined in turn, and its value perturbed individually within a set range. This technique has been used in previous chapters to determine the effect that this parameter value change has on cell behaviour. A time-lapse analysis will reveal if and when such a parameter change becomes influential. Secondly, the latin-hypercube sampling (Read *et al.*, 2012; Saltelli *et al.*, 2000) and eFAST global sensitivity analysis techniques (Marino *et al.*, 2008; Saltelli, 2004) have been used to perturb the values of a subset of parameters simultaneously and examine any compound effects that become apparent. Both of these techniques produce a statistical measure that determines how influential a particular parameter is, a Partial Rank Correlation Coefficient for latin-hypercube and a First-Order Sensitivity Index for eFAST. Examining how these statistical values change over time gives an indication of the time-points where each parameter becomes influential.

This chapter concludes by examining a set of cell responses that have yet to be considered within this thesis. All analyses to date have considered responses of simulated cells that are near to a forming patch, as these behave statistically differently to those that are further away. This statistical difference was observed at twelve hours, and cell behaviour calibrated accordingly. The calibrated simulator has been run for the full seventy-two hour period and behaviour of cells away from a forming patch captured at twelve hour intervals and compared with behaviour at twelve hours observed *ex vivo* . This indicates the time-point when the interactions of cells within a forming patch affect the behaviour of cells further away, through the upregulation of chemoattractants expressed on interactions with LTo cells.

5.2   Aims

Explorations in this chapter utilise the simulator and statistical tools within the *spartan* toolkit to perform a time-lapse analysis that addresses the following aims:

1. Determine whether there are set time-points in development where a biological factor becomes influential, by examining changes in cellular behaviour over time.

2. Determine if and when the behaviour of cells over a distance of 50μm is affected by cellular interactions within a primordial patch

5.3   Parameter Value Robustness over Simulation Time

Sections 3.5.1 and 3.6.1 examine how robust simulation behaviour is to a perturbation in the value of each parameter detailed in Table 3.1, with analyses focusing on hours twelve and seventy-two of development respectively. This reveals the impact that each parameter has individually on cellular behaviour, while also further supporting

the identification of parameters for which there is a high level of uncertainty in their true value. In this section the same analysis is performed, but for twelve hour intervals throughout the seventy-two hour run. Each parameter was perturbed individually, and assigned a value within the range set in Table 3.1. For each value the parameter was assigned, 500 simulation runs were performed to mitigate the effect of aleatory uncertainty. Medians of the cell velocity and displacement responses were then compiled for the 500 runs, creating a result distribution where the simulation was run under that criteria. This was contrasted to a distribution generated from 500 runs of the simulation at calibrated values using the Vargha-Delaney A-Test.

(i) **Chemoattractant Related Parameters**

A change in the initial level of chemokine expression on LTo cell differentiation has previously been seen to have no effect on cell behaviour at either hour twelve and seventy-two of development. It is thus of little surprise that performing the same analysis at twelve hour intervals reveals the same result. This could suggest that there is an large degree of uncertainty in the actual value of this parameter.

In contrast, previous analyses revealed that although changing the maximum level of chemokine adhesion had no impact on cell behaviour at twelve hours (Figure 3.3b), there was a significant effect observed on cell behaviour during hour seventy-two (Figure 4.6b). Performing the same analysis for data captured at twelve hour intervals reveals that a perturbation in parameter value causes a significant alteration in cell behaviour from hour 36 onwards, with a large difference between both cell displacement and velocity responses between hour 24 and hour 36 (Figures 5.1b and 5.2b). The same effect is seen for the parameter that captures LTi response to chemokine (Figures 5.1c and 5.2c). As this parameter is perturbed individually, the analysis reveals that there is a large degree of uncertainty in the true value of chemokine expression and response parameters, and this has to be taken into account when understanding cell behaviour after 36 hours.

(ii) **Cell Binding Probability Parameters**

It has been discussed in previous analyses that cell binding probability is a difficult parameter to assess due to the effect observed when assigned a value of zero: no LTo cell differentiation and thus no chemokine and adhesion factor expression. A significant difference in cell behaviour would therefore be expected where this parameter is zero, and this is observed for all time-points examined (Figures 5.1d and 5.2d). Other than this effect, the same observation is made as that drawn with the chemokine parameters above, that increasing just the value of this parameter has a significant effect on cell behaviour after the 36 hour time-point. Increasing this probability alone should lead to more stable contacts between hematopoietic cells and LTo cells, thus an increase in chemokine and adhesion factor expression,

influencing the behaviour of cells close to a patch. This further supports a growing conclusion that chemokine expression and response is important after the 36 hour time-point, and uncertainty in parameter values that are related to this pathway are an important consideration in analysing behaviour from that point onwards.

(iii) **Adhesion Factor Related Parameters**

One-a time analysis in Chapter 3 suggests that LTin/LTi cell response to adhesion factors has a key role in influencing cell velocity at the twelve hour time point. An examination of the robustness of simulation behaviour to this parameter over time reveals that this effect increases as simulation time elapses (Figure 5.1e), suggesting a strong relationship between the value of this parameter and cell velocity. For displacement, previous analysis revealed that a perturbation in parameter value had no effect on cell displacement during hour twelve, and thus there was a larger degree of uncertainty in the true value of this parameter. However a time-lapse analysis reveals a perturbation in expression would have a significant effect from hour 36 onwards (Figure 5.2e). It can be suggested that this affect becomes apparent due to the influence of chemoattractants later in the process. It has been revealed in previous analyses that chemoattractants have no role early on in development, but do become influential throughout the process. As this happens, more cells will be brought towards a forming patch, resulting in more contacts and a further upregulation of chemoattractants and adhesion factors. As the level of adhesion factors expressed increases, the probability an individual cell responds to this expression increases until the threshold set by this parameter is hit. From 36 hours onwards, a larger number of cellular contacts will have occurred, resulting in a high level of adhesion factor expression. Thus a change in this parameter has a direct impact on cell displacement when adhesion factor expression is high.

In contrast, for the parameter that captures expression level of adhesion factors with each stable contact, no significant change in cell behaviour is observed throughout the simulation time period for each parameter value studied (Figures 5.1f and 5.2f). As noted in previous discussions concerning this parameter, a small effect is seen for a low level of adhesion factor expression, yet a large increase after that point yields no difference, suggesting that too low a level of adhesion is influential, yet an overexpression has no significant effect.

5.4 Identifying the time-point at which a parameter becomes influential

Analyses in Chapter 3 used two global sensitivity analysis techniques (latin-hypercube sampling and eFAST (Read *et al.*, 2012; Saltelli *et al.*, 2000)) to identify parameters

that have a highly influential effect on simulated cell behaviour, thus suggesting the key biological pathways in the development process. Here, these two analyses have been performed for cell responses captured at twelve hour intervals for each of the runs described in Chapter 3, making it possible to determine if and when statistical measures generated by the technique change over time.

5.4.1   Parameter Value Sampling using Latin-Hypercube Approach

Time-lapse results have been generated using the same 500 parameter value sets that were created using a latin-hypercube approach (detailed in Figure 2.20) to examine behaviour at hours twelve and seventy-two (sections 3.5.2 and 3.6.2). For each value set the simulation was run 500 times to mitigate the effect of aleatory uncertainty, and behaviour measures for cells within 50μm of a forming PP captured at twelve hour intervals. For each time-point in each run, the median of each cell behaviour response was calculated, producing a distribution of 500 medians for each response. The median was again taken for this distribution, producing cell behaviour responses that summarise the behaviour of the simulation under a particular parameter value set at a particular time-point. Taking each parameter in turn, and each time-point, a Partial Rank Correlation Coefficient (PRCC) was generated that gives a statistical measure of any compound effects between the parameter being examined and simulation result at that time-point. For each parameter, the PRCC values calculated for each time-point were plotted, making it easier to visualise changes in PRCC value over time, and thus any emergence in compound effects between parameters over simulation time (Figure 5.3).

For the cell velocity response, no increase in correlation value appears over simulation time for the parameters that capture the initial and maximum levels of chemokine expression by an LTo cell (Figures 5.3a and 5.3b). The same conclusion can be drawn for the parameter that captures the probability an LTi cell does not respond to chemokine expression in its locality (Figure 5.3c). This supports previous conclusions put forward in this thesis that chemokine expression has no significant effect on cell velocity. When considering Figure 5.3d, the probability two cells form a stable bind upon contact, it can be noted that the PRCC values are higher than those for the chemokine parameters, but no trend in correlation value over time is observed. The higher PRCC values are expected due to the effect that setting this parameter to its extreme value of zero has on simulation response (that LTo cells cannot differentiate and express adhesion factors and chemokines). Thus this effect has been discounted throughout this study. In contrast, there is a strong correlation between the value assigned to the parameter that captures the probability that an LTin or LTi cell is retained by adhesion factor expression and cell velocity, and this remains the case for each time-point, suggesting this is the key factor in influencing cell velocity (Figure

5.3e). Any correlation between the level of adhesion factor expressed with each stable cellular contact and cell velocity does decrease between 12 and 36 hours (Figure 5.3f), potentially suggesting that adhesion factor expression is influential in early stages of development but not through the whole time period.

When considering cell displacement, the response to and expression of chemoattractants is identified as a key pathway in affecting cellular behaviour. Although no correlation between maximum level of chemokine expression and cell displacement was identified at hour twelve, a stronger correlation does become apparent in the following 24 hours, after which there is a clear trend between the value of this parameter and cell behaviour (Figure 5.3b). An LTi cell response to chemokine expression follows a similar pattern (Figure 5.3c). A combination of this and analyses in previous chapters provides further support to the hypothesis that the development period may be split into two phases, and now provides an indication of when these phases may change. However whereas the influence of the chemokine pathway increases as the simulation time elapses, the correlation between a cells response to adhesion factor expression and cell displacement is initially stronger and steadily increases (Figure 5.2e), suggesting an influence on cell displacement throughout the whole development period. Although this is the case, interestingly no correlation becomes apparent between the level of adhesion factor expression and cell displacement, at any time-point in development.

5.4.2 Parameter Value Sampling using eFAST Approach

Time-lapse responses were generated using the same eFAST parameter value sets generated to explore cell behaviour at 12 and 72 hours (sections 3.5.3 and 3.6.3). For each parameter value set, 500 simulation runs were performed to mitigate the effect of aleatory uncertainty. Median cell response measures were calculated for each time-point in each run, producing sets of 500 median cell behaviour responses. The medians of these distributions was calculated to give a median cell response for the simulator under those conditions, at that time-point. The eFAST approach was utilised to analyse these results for each time-point, producing first-order (Si) and total-order (STi) sensitivity indexes that determine the proportion of variance in simulation response that can be accounted for by that parameter, at that time-point. For each cell response, the first-order sensitivity indexes for each parameter, and each time-point, have been plotted (Figure 5.4.2) to reveal if the proportion of variance that can be accounted for by each parameter changes over simulation time.

For cell velocity responses, the analysis supports the conclusions drawn in the section above (Figure 5.4a). The expression level of adhesion factors accounts for a significant amount of the variance in simulation results at twelve hours, yet this impact reduces in a similar way to that observed in latin-hypercube analysis, again suggesting an influence in early PP development but one that does not continue throughout the development period. In contrast the proportion of variance explained by the probabil-

ity an LTi cell responds to adhesion factor expression quickly increases after 12 hours, and from 24 becomes the only significant parameter that influences cell velocity. There is no significant change in Si value for all other parameters of interest.

The cell displacement responses again support results observed in latin-hypercube analysis above, revealing a significant change in LTi adhesion response and maximum level of chemokine expression between 12 and 36 hours, after which the value stabilises (Figure 5.4a). In contrast to the results in the previous section, no significant increase is observed in the Si value for LTi response to chemokine, a value that remains close to constant throughout. For all time-points the proportion of variance accounted for by that parameter is statistically significant in comparison to the dummy parameter, suggesting the parameter has an effect but is not the major influence on cell displacement. Again there is no significant change in Si value for all other parameters.

5.5 Time-Lapse Analysis of Cells Away From a Developing Peyer's Patch

Statistical analyses throughout this thesis have focused on the behaviour of cells close to a forming patch, to understand how each biological factor influences these cells such that they behave differently to those further away ($> 50 \ \mu$m). However, the behaviour of simulated cells both close to and far from a forming patch has been calibrated from *ex vivo* data captured in the twelfth hour of development (Patel *et al.*, 2012). Previous studies in the literature suggest that chemoattractant expression has a key role in the recruitment of LTi cells to a forming patch (Cyster, 1999; Luther *et al.*, 2003), and thus there may be a point in the simulation at which the interactions close to a patch begin to affect the behaviour of those further away. To examine this, the simulator was run 500 times, and the behaviour of cells that are further than 50μm from an LTo cell tracked at four hour intervals. The distribution of cell behaviour responses for each time-point $>=24$ hours were then compared with those generated at the calibrated twelve hour time-point using the Vargha-Delaney A-Test, to determine if cell behaviour does become statistically significantly different to that at 12 hours.

The results of this analysis can be seen in Figure 5.5. Cell displacement remains statistically similar to that seen at the 12 hour time-point until hour 36, after which point there is a significant change in cell behaviour. In fact by hour 48, the two distributions are statistically completely different (A-Test score = 1.0). Such a result would suggest that under conditions created in calibration, chemokine expression does not begin to affect cells further than 50μm from a forming patch until hour 36. As expression increases, the distance over which chemokines become influential increases, and thus displacement becomes significantly different. However it is difficult to draw a firm biological conclusion based on this as the level of chemokine expression remains unknown, and thus such an effect may have become apparent through parameterisation.

5.6 Discussion

5.6.1 Implementing a Time-Lapse Approach Through Simulation and Sensitivity Analysis

This chapter has shown how a combination of a computer simulation and the range of sensitivity analysis techniques in the *spartan* package (Marino *et al.*, 2008; Read *et al.*, 2012; Saltelli *et al.*, 2000) can reveal alterations in agent behaviour over time, and suggest the biological factors that may be causing this change in behaviour. Although biological explorations are increasingly being paired with computational modelling and simulation, the application of the resultant simulator as a tool for performing a time-lapse analysis of the captured process has not been widely utilised. Such an analysis can then be used to target particular time-points to explore in the future laboratory experiments.

The one-a-time parameter robustness technique (Read *et al.*, 2012) has been extended in this chapter such that parameter robustness throughout the simulated time-course can be examined. Currently the use of this technique in determining a level of confidence in parameter value over time is novel, and aids establishing a degree of confidence in the value assigned to that parameter over time. This is an important consideration when assessing the effect of parameter value uncertainty on results. For example, Figure 5.2b reveals that there is a large degree of uncertainty in the value of the parameter that captures maximum chemokine expression by an LTo cell at 12 and 24 hours. However 12 hours later in the simulation, this changes completely, and results suggest that just a small change in just this parameter value has a significant effect on simulation result. What the results in this chapter mean biologically is detailed in the next section of this discussion, but from a technique point of view, one could now try to establish what has caused this effect to become apparent at that time-point. If cell behaviour statistics were available from other time-points in the development process, a robustness analysis could be used to ensure each simulation parameter is capturing the correct biological dynamics at that time-point.

When considering use of global sensitivity analysis techniques over time, an exploration of changes in Partial Rank Correlation Coefficients is not novel, and has been demonstrated in determining correlations between simulated biological factors and extracelluar bacterial load over a simulated time period in a model of TNF in controlling tuberculosis in a granuloma (Ray *et al.*, 2009). However no study has yet examined the first-order sensitivity indexes, generated using the eFAST technique (Saltelli, 2004; Saltelli *et al.*, 2000), over simulation time in the manner that has been presented here.

The remaining sections of this discussion examine the biological insight that results using this approach suggest in relation to the formation of Peyer's Patches through use of the simulator as a time-lapse tool.

5.6.2 Phases of Peyer's Patch Development

Experimental work has generated the generally accepted hypothesis that there are three phases of patch development (Adachi *et al.*, 1997): the appearance of VCAM-1 stromal cells in the gut, the appearance of clusters of LTi cells around stromal cells, and the recruitment of lymphocytes from E18.5. The simulation developed in this thesis captures the first two phases. Cell behaviour analysis in Chapter 3 suggest that different biological factors influence cell behaviour responses at different time-points. Thus, through use of simulation, it could be suggested that there may be additional development phases inside the first of the two generally accepted phases.

Analyses in Chapter 3, to determine the factors that cause the statistically significant change in cell behaviour observed *ex vivo* when a cell is in the vicinity of a forming patch (Patel *et al.*, 2012), determined that cell velocity during hour twelve of development is influenced by the level of adhesion factor expression per stable contact between hematopoietic and stromal cells. However analyses of behaviour responses at hour 72 suggest no influential role for adhesion factors. Use of global sensitivity analysis techniques at twelve hour intervals suggests that the influence at twelve hours becomes statistically insignificant by hour 36 (Figure 5.3). This could suggest that an initial phase could exist, mediated by cell adhesion factors, covering the first 36 hours of development, after which point the effect of a change in adhesion factor expression level reduces amid a growing influence of other factors.

In contrast, previous analyses revealed no role for chemokine expression at twelve hours, and a significant role at seventy-two as suggested in the literature (Cyster, 1999; Luther *et al.*, 2003; Ohl *et al.*, 2003). Performing a sensitivity analysis over time has been useful in determining the time-point in development when chemokine expression and response becomes significant. Each of the analyses performed, whether perturbing expression parameters individually (Figures 5.1 and 5.2) or with other parameters simultaneously (Figures 5.3 and 5.4), reveals a statistically significant change in cell displacement between hours 24 and 36, an effect that then stabilises through until hour 72. The parameter that captures LTi response to chemokine also becomes influential on cell behaviour after hour 24. As chemoattractant expression promotes cell migration towards a forming patch (Cyster, 1999), it could be suggested that a new phase of development begins between this time-point, one that moves from the triggering of adhesion and chemoattractant expression to the aggregation of cells that comprise the primordial PP observed at hour 72, mediated by chemokine expression and response.

Although the level of adhesion factor expression has been determined to only have a significant influence on cell behaviour for the first 36 hours, the parameter that captures LTin/LTi cell response to adhesion factors has been shown to be highly influential throughout the time period. When the parameter is perturbed individually to the set examined, a clear trend emerges between the simulation time and an alteration in cell behaviour (Figures 5.1c and 5.2c). Thus the system dynamics captured within

the simulation are highly sensitive to LTi cell response to adhesion, a sensitivity that increases over time. Global sensitivity analysis results support the emergence of this trend although all parameters are being perturbed sequentially. This result suggests that although there may be two distinct phases apparent in development, LTi response to adhesion does have an influential role in both phases.

5.6.3 Uncertainty in Maximum Chemokine Expression Affects Interpretation of Cells Far From Forming Patch

Tracking of cells *ex vivo* at the twelve hour time-point has produced results that determine how cells behave both close to and away from a forming PP (Patel *et al.*, 2012), and thus cell behaviour in the simulation has been calibrated based on those results. A time-lapse analysis of the behaviour of cells that are 50μm or further from a forming patch reveals a significant change in cell displacement after hour 36. This result could be used to suggest that chemoattractant expression becomes sufficient after that time-point to induce chemotaxis across a large distance from a forming patch. However, this effect is determined by the value assigned to the parameter that captures the maximum level of chemoattractant expression. The current simulator has been calibrated against behaviour of these cells at twelve hours, and with no further experimental data available, there is a great deal of uncertainty in the true value of this parameter. Thus it is difficult to accept this result as representative of the biological system, rather it is an artefact of parameterisation. For the simulator to be useful in determining the influence of chemoattractant expression on these cells, further biological experimentation is required, by performing either the same cell tracking experiment performed *ex vivo* but at further time points and then calibrating chemoattractant parameters to replicate the behaviour observed, or by obtaining a quantification of chemokine expression over time.
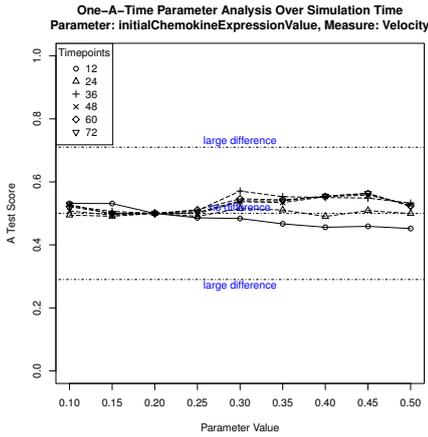
(a) Initial level of chemoattractant expression at LTo differentiation
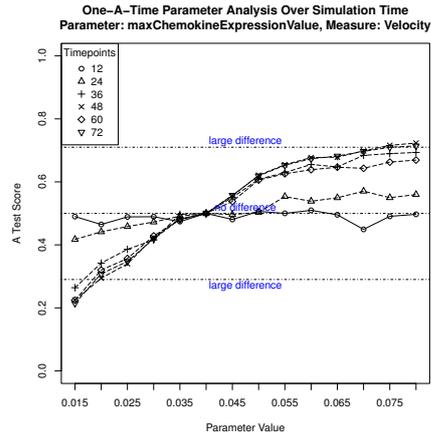
(b) Saturation limit of chemoattractant expression
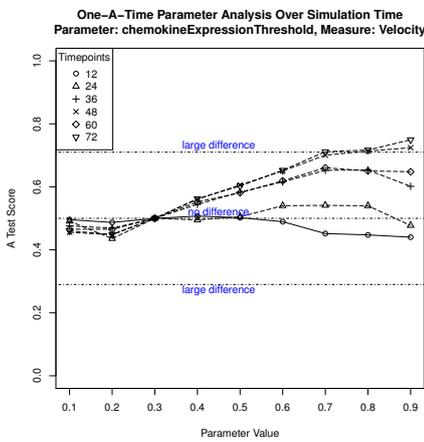
(c) Probability an LTi cell does not respond to chemokine expression
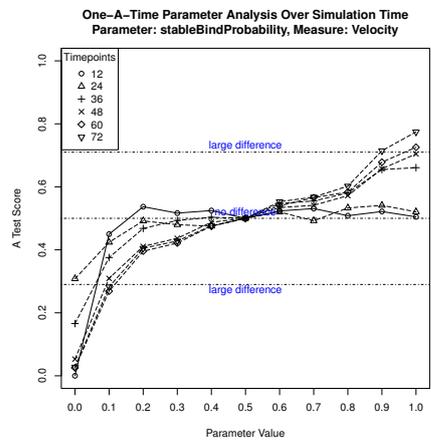
(d) Probability a LTin/LTi and LTo cell form stable bind on contact
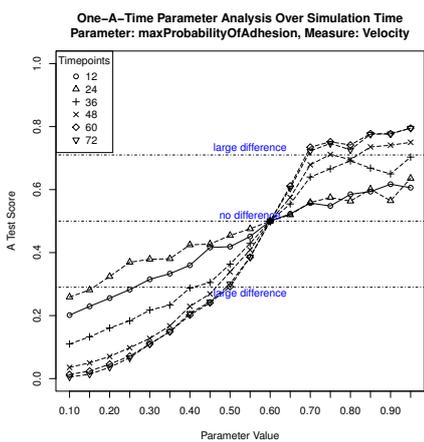
(e) Maximum probability adhesion factors prolong cell contact
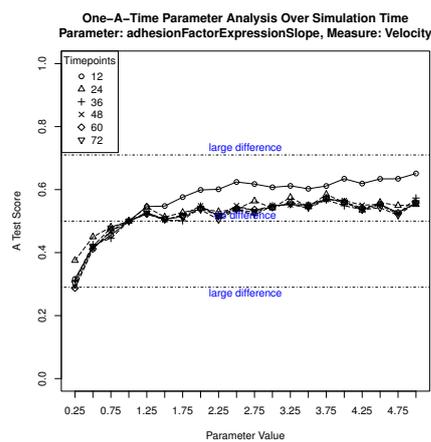
(f) Level of adhesion factor expression per stable contact

Figure 5.1: An examination of parameter robustness over simulation time: Cell velocity. The parameter values were perturbed independently as detailed in section 2.4.2, and cell behaviour results captured at twelve hour intervals. These results were then compared to the baseline simulation using the Vargha-Delaney A-Test to determine the effect a change in parameter value has had on cell velocity. Performing this analysis at twelve hour intervals reveals if the change in parameter value has an effect at a certain timepoint.

(a) Initial level of chemoattractant expression at LTo differentiation
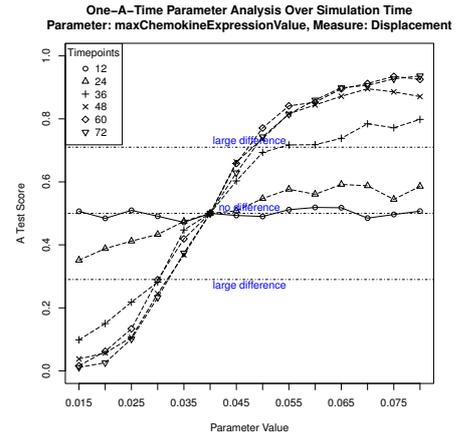
(b) Saturation limit of chemoattractant expression
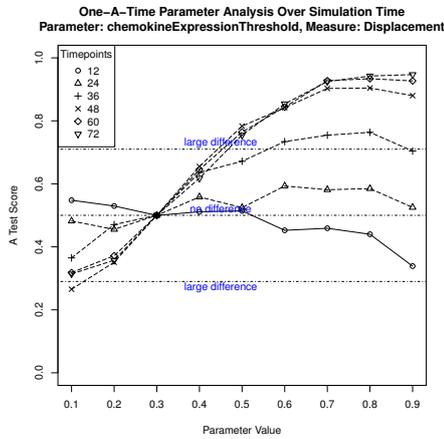
(c) Chemokine Level at which LTi chemotaxis occurs
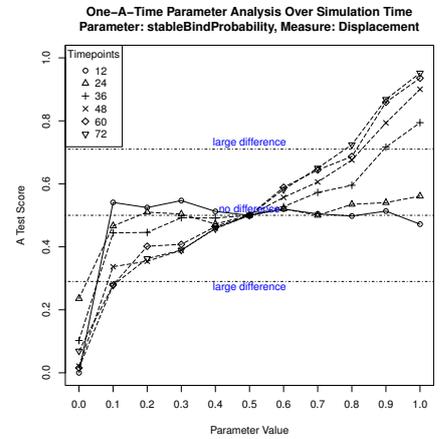
(d) Probability a LTin/LTi and LTo cell form stable bind on contact
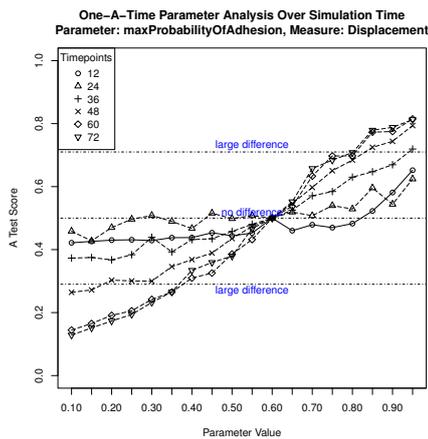
(e) Maximum probability adhesion factors prolong cell contact
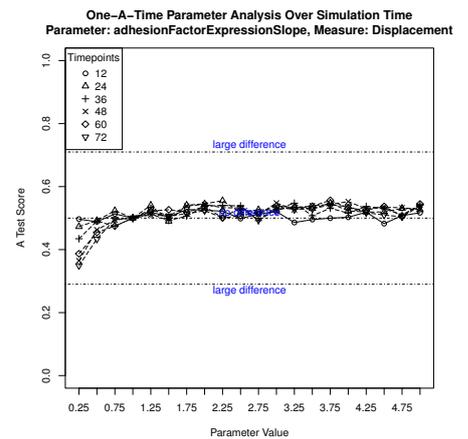
(f) Level of adhesion factor expression per stable contact

Figure 5.2: An examination of parameter robustness over simulation time: Cell displacement. The parameter values were perturbed independently as detailed in section 2.4.2, and cell behaviour results captured at twelve hour intervals. These results were then compared to the baseline simulation using the Vargha-Delaney A-Test to determine the effect a change in parameter value has had on cell displacement. Performing this analysis at twelve hour intervals reveals if the change in parameter value has an effect at a certain timepoint.

(a) Initial level of chemoattractant expression at LTo differentiation
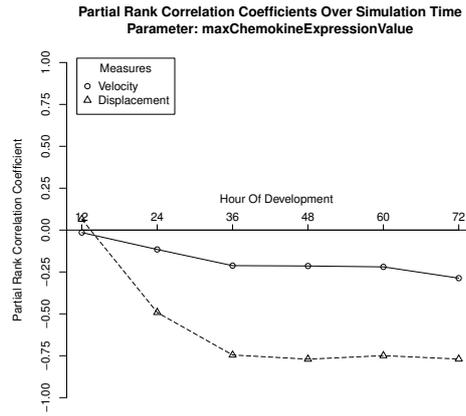
**(b)** Saturation limit of chemoattractant expression
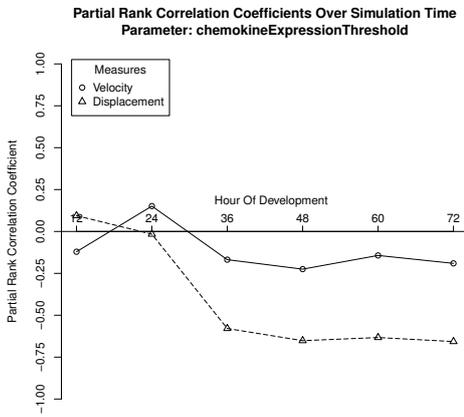
(c) Chemokine Level at which LTi chemotaxis occurs
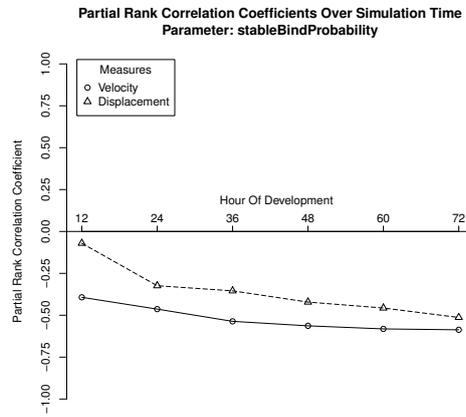
**(d)** Probability a LTin/LTi and LTo cell form stable bind on contact
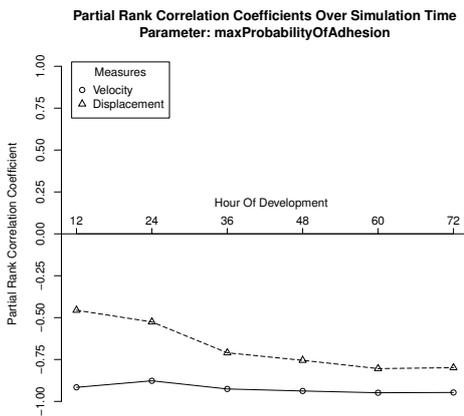
(e) Maximum probability adhesion factors prolong cell contact
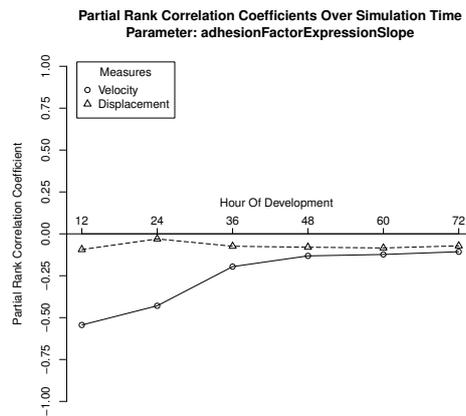

**(f)** Level of adhesion factor expression per stable contact

Figure 5.3: Partial Rank Correlation Coefficients (PRCC) for each parameter under examination, calculated at twelve hour intervals using the latin-hypercube analysis approach. Examining how the PRCC changes over time gives an indication of when a parameter begins to become influential in affecting cell velocity and displacement.
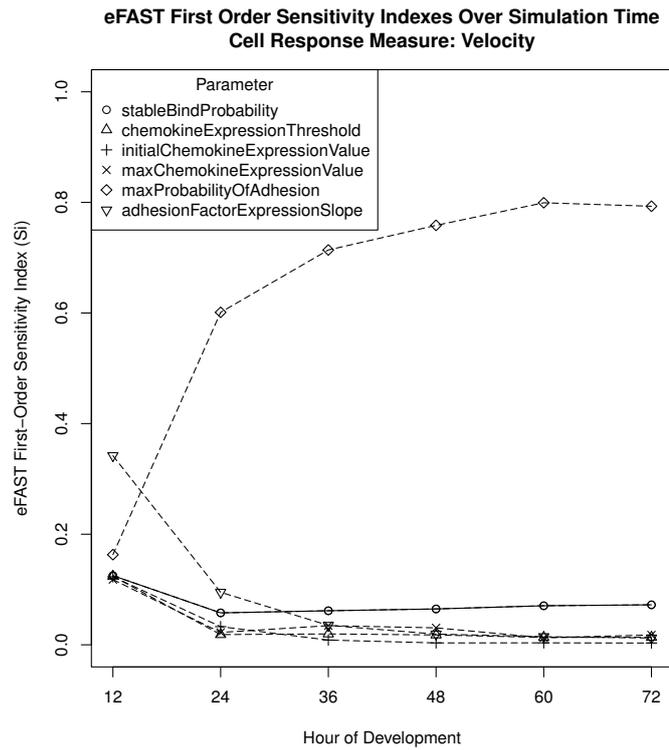
(a) Si measures for each parameter for the cell velocity response
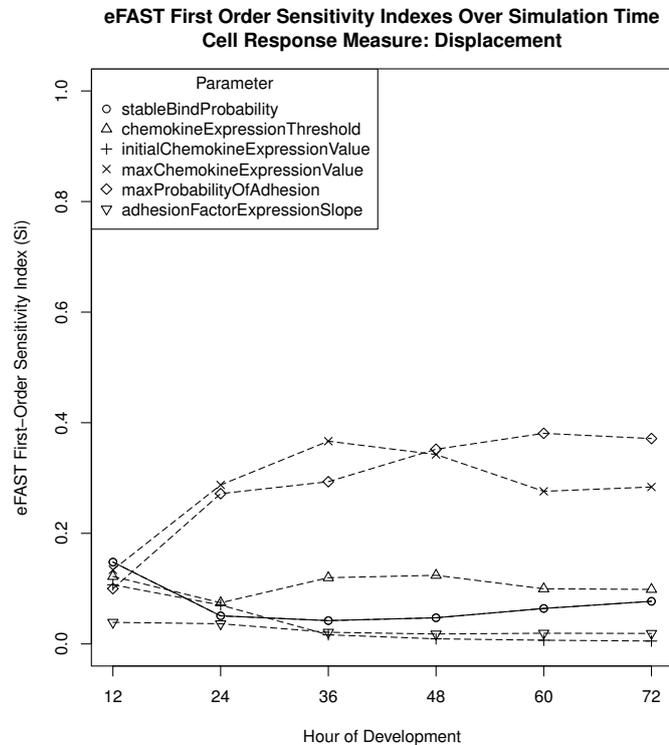
(b) Si measures for each parameter for the cell displacement response

Figure 5.4: eFAST First-Order Sensitivity Index (Si) for each parameter of interest, calculated at twelve hour intervals. This shows the percentage of variance in simulation result at that time-point can be explained by a particular parameter.
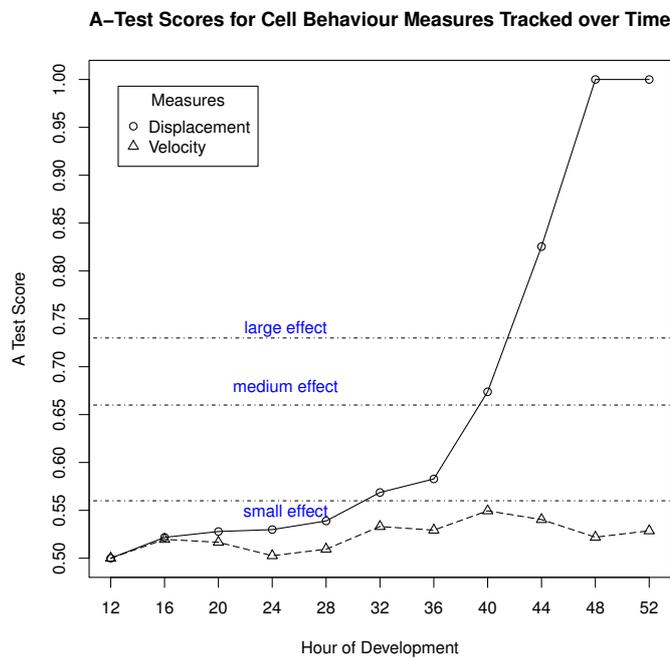
Figure 5.5: An analysis of the behaviour of LTin and LTi cells away from a forming Peyer's Patch. The simulator has been calibrated such that the behaviour of the simulated cells at 12 hours matches that observed for the same time-point *ex vivo*. This analysis reveals how the behaviour of these cells changes over time, a statistical comparison drawn using the Vargha-Delaney test (Vargha and Delaney, 2000).

Chapter 6

Discussion

The aims behind the work presented in this thesis were to construct and utilise computer simulation and statistical tools to further understand lymphoid tissue organogenesis. This chapter provides a reflection of how this study addresses these aims. However, the contribution of this study extends beyond furthering biological understanding, to a contribution to the field of computational modelling and simulation, with no discrimination towards the discipline in which modelling is being utilised.

6.1 Simulation as a Tool for Exploring Lymphoid Tissue Organogenesis

Chapter 2 of this thesis details the development of a simulation of Peyer's Patch formation, for use as a scientific instrument to explore lymphoid tissue organogenesis. Through use of the CoSMoS framework reviewed in section 1.4.2 (Andrews *et al.*, 2010), a principled approach has been adopted in the design and implementation of the simulation, in collaboration with experimental immunologists with expertise in lymphoid tissue formation. This collaboration has proven to be a vital aspect in a project to model a system where the biological understanding is incomplete and where it is not viable to include every aspect of the biological system that is understood. Constructing the domain and platform models with the insight of experimental immunologists can be considered as the first method by which confidence in the simulators representation of the domain can be judged.

The domain and platform modelling work in Chapter 2 is the end result of detailed discussions with collaborating immunologists. It may not have been possible to document this fully in that chapter, but this is the final model of a large number that were generated, and this is a time-consuming process. However, the modelling process should be seen as time well spent, as the process highlighted areas where the biological understanding was ambiguous or incomplete, and areas where simplifications could be made. For example, an initial model included a cell response to two different chemokines expressed by an LTo cell, until it was decided that modelling one was suf-

ficient. This conclusion was drawn based on experimental results that suggest PP do not form correctly in mice deficient for CXCR5, the receptor for CXCL13, although chemokines CCL19 and CCL21 are still present (Ansel *et al.*, 2000). The specification of the biological system in the domain model, and the platform model specification of how that will be encoded as a computer simulation, has been verified by the domain experts and deemed to appropriately capture the biological system under study. Thus, where assumptions and abstractions have been made, these have the support and justification from people with expertise in the field.

The simulation has been constructed from the specification in the platform model (section 2.2.3), and through a process of calibration (section 2.2.6), parameter values set such that cell behaviours emerge that are statistically similar to those observed in Patel et al's (2012) *ex vivo* investigations detailed in section 1.3.3. With these results having a direct mapping back to the biological system, there is a greater level of confidence that the simulation is an adequate representation of the biology. The second emergent behaviour captured in the model, the development of aggregations of cells that mature to become PP, has been more difficult to calibrate due to the environmental representation used in the model (2D rather than 3D, see Figure 2.10) and the lack of biological data that can be used to determine PP characteristics. Although there is no direct link to the biology in this case, the simulation can still be used as a tool for drawing qualitative conclusions on what influences patch formation, by examining how aggregations that are formed by the simulation calibrated based on cell behaviour alter when run under different conditions.

The end result of the process of modelling, simulation implementation, and calibration, is a tool that can be used to perform *in silico* experimentation. This can be performed by altering the values that have been assigned to particular parameters, or by setting boolean flags that simulate a knock-out of a particular factor. The simulation platform implemented in this thesis, described in Chapter 2, attempts to create a visual look and feel that experimental immunologists are comfortable with using, in a format that those performing the *in silico* experiment can relate to. Current laboratory explorations utilise real-time imaging techniques to take snapshots or movies of a biological system under investigation, which can then be processed using specialised image analysis software such as Volocity (PerkinElmer). This software has the capability of identifying individual cells in each image and tracking cell behaviour across a number of images from different time-points, producing statistics that describe the behaviour of that cell. The *ex vivo* data upon which this simulation is calibrated (Patel *et al.*, 2012) was generated in this way. To relate this simulation to such techniques, it is possible to perform the same analysis, by specifying time-points in the simulation when snapshots of the simulated gut should be taken: images that can then be processed by Volocity. The *in silico* cell tracking image in Figure 2.14 was generated using that technique. The results of *in silico* experimentation to replicate previously published laboratory

investigations, seen in Figure 4.1, is a further example of the use of simulation snapshots to analyse results. For each of the investigations being replicated, the result of the laboratory investigation was determined visually, and thus the same visual method was applied to determine whether the simulation replicates that result.

For a more detailed, statistical analysis of the biological system captured in the simulation, a variety of cell behaviour and patch characteristic responses are output from the simulation as comma-separated value (CSV) files, that can be processed using a variety of statistical techniques. The use of such techniques to provide biological insight is examined later in this chapter.

Chapter 1 detailed numerous examples of the use of simulation as a tool to provide insight into a particular biological system under study. However, it is unusual that the tool developed is then released to the academic community for full scientific scrutiny. The simulation developed in this thesis, and the domain and platform models associated with it, are available on the internet at http://www.cs.york.ac.uk/immunesims/frontiers, enabling the use of the tool to inform laboratory investigations of others and allowing the community to critically comment on the design and implementation. Thus, this thesis is not simply an exercise in implementation and use of simulation as a tool for drawing conclusions on a biological system, rather it is hoped that the tool developed contributes to ongoing investigations of others in the field.

6.2 *Spartan*: Statistical Techniques to Analyse Simulation Behaviour

It has been noted in this thesis that no generic comprehensive toolkit exists to aid the understanding of simulation results, in an attempt to ground a result in the domain that it captures and thus inform future laboratory investigation. This may explain why in a large number of cases in the literature, little attempt is made to reveal how representative a simulation result is in terms of the captured biological system (Read *et al.*, 2012). The *spartan* toolkit developed in the course of study aims to fill the void in the availability of ready to use statistical analysis techniques, providing a mechanism to aid the integration of simulation into wet-lab research, with the aim of ensuring *in silico* results are interpreted with rigour. Although *spartan* was constructed alongside the development of an agent-based simulation, the statistical techniques can also be applied to simulation results generated using an ODE approach.

Figure 6.1 is a schematic of a generic simulation process. A domain of interest is identified, in this case the development of secondary lymphoid organs. From this, a process of modelling is adopted (such as the CoSMoS Process) resulting in the generation of a computer simulation that captures an abstraction of that domain. This computer simulation provides the capacity to produce responses under a variety of pa-

rameter value conditions, results that are analysed by the techniques included within *spartan*. These analyses help determine the number of simulation runs required per condition to generate a representative result, how robust the simulation is to a change in individual parameter values, and can identify the biological pathways and components that have a statistically significant effect on simulation behaviour by identifying highly influential parameters. The conclusions drawn from these analysis may then inform future laboratory investigations, and thus the next iteration of the model. Section 1.4.2 noted that there are modelling frameworks and simulation packages available to aid performing the first two stages. The release of *spartan* provides a tool to perform the final stage, thus ensuring there are tools available for each component part of this generic modelling schematic.

Spartan is open-source, has been developed and runs within the platform-independent R statistical environment, and can be freely downloaded from the Comprehensive R Archive Network (CRAN); thus, no restrictions have been placed on allowing others to utilise or extend the functionality of the package. Utilising sample Peyer's Patch simulation data gathered in the course of this study, comprehensive tutorials have been developed that exemplify how the techniques included within *spartan* can be advantageous in exploring the operation of a developed simulation. With all this taken together, it is hoped that this provides a comprehensive set of tools that encourage simulation developers to publish a full statistical analysis alongside any hypotheses generated through simulation. *Spartan* remains in a state of ongoing development and further analysis methods will be added when appropriate. It was noted in Section 1.4.2 that a number of software packages have been developed that ease simulation development through use of drag and drop interfaces that remove some of the complexity involved in simulation development. Similarly, a wizard interface for the *spartan* package is under development, a feature that will remove the complexity of running the analysis in R, thus widening the potential usage of the package further.

6.3   Biological Hypotheses Generated Through Simulation

Through a process of calibration (section 2.2.6), it has been ensured that the simulator produced responses that were statistically similar to that observed in the laboratory. Explorations in preceding chapters utilised the simulator and the *spartan* statistical toolkit to generate biological hypotheses. This section reflects on two predictions that have been generated.
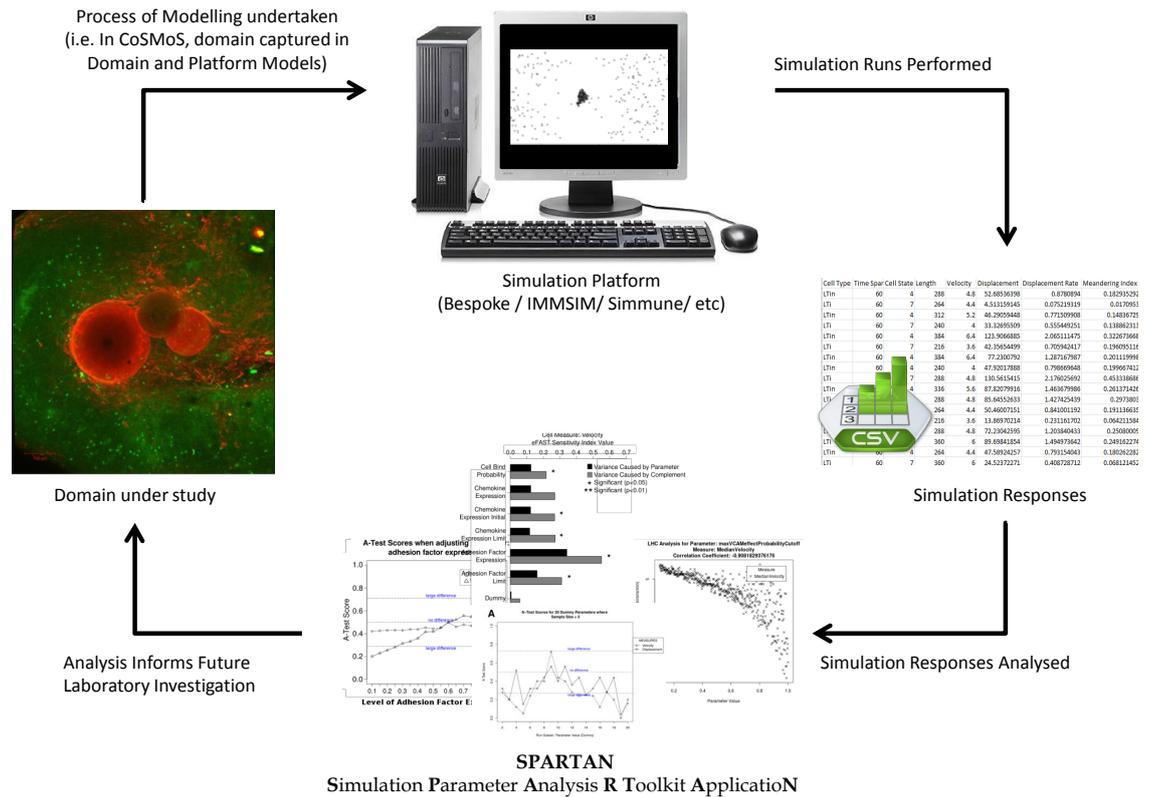
Figure 6.1: Placing *spartan* in Simulation Development. The domain being captured in the simulation is on the left. In this case, this is cell behaviour around a primordial Peyer's Patch (Image taken from Patel et al (2012)). A simulation is implemented, potentially through a process of modelling (such as the CoSMoS Framework (Andrews *et al.*, 2010) adopted in this thesis), using either available simulation platforms or a bespoke implementation. This produces simulation results that can be analysed using the statistical techniques available within *spartan* . Results of this analysis can then potentially inform future laboratory investigations that focus on the captured domain.

6.3.1 PP Development From E14.5 to E17.5 Could Be Biphasic

The preceding chapters highlighted the currently accepted view that PP development occurs in three distinct phases (Adachi *et al.*, 1997). Chapter 3 utilised the simulation and *spartan* statistical toolkit to examine cell behaviour during hours 12 (E15.0) (Section 3.5) and 72 (E17.5) (Section 3.6), and suggested that different biological factors are influential at these two different stages of development. Use of simulation as a time-lapse tool in Chapter 5 took this a stage further, examining cell behaviour at 12 hour intervals, in an attempt to identify the time-point where a change in the influence of simulation parameters occurs. Through a combination of both analyses, this thesis has proposed that the 72 hour period between the accepted first and second phases of PP development could in turn be split in to two phases, a first influenced by adhesion factor expression and response, and a second mediated by chemokine expression and response, which becomes influential between hours 24 and 36. This prediction is

counter to the accepted claim that the 72 hour period is chemokine dependent (Cyster, 1999; Luther *et al.*, 2003; Ohl *et al.*, 2003).

This result demonstrates the advantages of pairing simulation and ongoing laboratory investigations. The *ex vivo* culture system study was conducted to examine cell behaviour in the early stages of PP development (Patel *et al.*, 2012), and has developed biological hypotheses concerning the RET signalling pathway. The data produced has directly informed the development of the simulator, resulting in a tool which replicates cell behaviour that emerges *ex vivo*. From this, a range of *in silico* experimentation has been performed to explore cell behaviour at a number of other time-points in development, generating additional hypotheses to those formed in the laboratory. Thus, through simulation, the original biological study has had a wider impact than that originally intended, with the potential to counter established views in the literature. In turn, these predictions can inform future laboratory investigations that verify whether the hypotheses generated hold. Methods to verify the hypothesis that PP development is biphasic are examined later in this chapter.

6.3.2   Variation in Peyer's Patch Development

Data presented in Figure 1.2 suggests there is a large variation in location, number, and size of PP in genetically identical mice, reasons for which remains unknown. Investigations in Chapter 4 utilised the simulator to perform novel *in silico* experimentation to examine factors that could be causing this variation.

It was noted in Chapter 4 that although it has been suggested that LTin cells have a role in early initiation of PP development (Patel *et al.*, 2012; Veiga-Fernandes *et al.*, 2007), the role of LTin cells in PP formation and physical characteristics is not fully understood. Simulation results have suggested that a change in the number of LTin cells and the rate of LTin cell migration could limit the number of PP that develop and influence the size of the aggregations that do form (sections 4.4.2 and 4.4.3). Thus it could be suggested that variation observed between different mice is caused by variation in the number of LTin cells produced in the foetal liver.

As LTin cells are known to express RET (Veiga-Fernandes *et al.*, 2007) and LTo cells express RET ligand, the initial interaction between an LTin and LTo cell has previously been suggested as the trigger of PP development (Patel *et al.*, 2012; Veiga-Fernandes *et al.*, 2007). However, flow cytometry results have been used to estimate that at E15.5 (24 hour time-point in development), 20% of the gut surface could be occupied by LTo cells, thus there must be a factor that is limiting the triggering event. The findings noted in the previous paragraph suggest that LTin cell migration rate and cell number could be the limiting factor, as an alteration in cell number has been shown to affect patch size and number. Yet it is also unclear as to whether each of the LTo cells in the gut has the capability to differentiate and interact with LTin and LTi cells. For example, a restriction in expression of the ligand for RET could also be the

limiting factor. Such a restriction was examined *in silico* in section 4.4.4. It has been suggested that PP tend to form on or near the anti-mesenteric border (van de Pavert and Mebius, 2010; Randall *et al.*, 2008), leading to the hypothesis that an LTo cell could receive a signal, potentially from the blood stream, that causes an upregulation of RET ligand expression, limiting where PP form and thus the number that do form. Laboratory work would have to be undertaken to verify whether this was the case.

Potentially variation in PP development could be explained by a combination of these two hypotheses. A restriction of RET ligand expression by a yet unknown mechanism could control PP number, while variation in the number of LTin cells that migrate from the foetal liver could then influence the size of the PP that develops.

6.4 Factors Affecting These Hypotheses

In his study of confidence in simulation results, Read (2011) notes that confidence is not boolean but established to varying degrees. Confidence in the biological hypotheses that have been generated through simulation, detailed in the previous section, could be affected by a number of factors.

As noted in the first section of this discussion, the model on which the simulation is based is an abstraction of the real system, and this must be considered when analysing what the result means in terms of the system being explored. In this study, the assumptions made in the creation of the simulation are documented in Tables 2.1-2.3. As these assumptions include the methods by which chemokines and adhesion factors are expressed, the results of statistical analyses that suggest these are the two influential factors in development may be subject to a degree of scepticism. In the case of the model presented in this thesis, the domain and platform models have been verified by two experimental immunologists with expertise in lymphoid organ development. Thus the assumptions made to model chemokine and adhesion factor expression could potentially impact the conclusions that are drawn, but these have been justified by those with domain specific expertise.

However scepticism in the assumptions that have been made should not necessarily be seen as a negative, but the start of a conversation that could improve the simulation further in the next iteration. Full transparency in the design of the simulation allows those in the field to assess the model and contribute their opinions and findings, potentially informing their own investigations as well as informing the development of the simulator. In the case of the Peyer's Patch simulation, the full simulation design (Domain and Platform Models), assumption tables, and the implemented simulator are available online at http://www.cs.york.ac.uk/immunesims/frontiers for this reason, opening this work to full scientific scrutiny.

Prior to any further iteration that seeks to add complexity to the current model and simulation and increase confidence in results generated, an evaluation should be per-

formed that details whether such an addition would provide any extra insight. Whereas biologists traditionally take a reductionist approach that focuses on the role of each factor in the system, the focus of models such as the one developed in this thesis is on the observable high-level behaviour, and how this emerges through interactions between individual biological factors (Germain *et al.*, 2011). Although the interaction may trigger a number of events that are both observable and implicit (such as intercellular pathways), the observable result is the focus, and thus implicit pathways can be assumed and abstracted from the model, as their inclusion would have no impact on the result. This needs to be taken into consideration when additions are suggested. The implementation of chemokine expression and response in this model is a good example, where three chemokines and two receptors have been captured as one. An obvious extension could be to add an additional chemokine and determine the effect this has on simulation result. However, previous laboratory studies have suggested that one chemokine, CXCL13, has a dominant role in the recruitment of LTi cells, and blocking CXCL13 response has a significant impact on patch formation (Luther *et al.*, 2003). The authors note that PP do form where CXCL13 response is inhibited, but these lack the structural characteristics of wild-type PP. Thus, it would be difficult to justify adding further chemokine complexity when the end result cannot be classified as a PP.

The implementation of the simulation using an agent-based approach introduces uncertainty that could also have an affect on the meaning of results generated (Helton, 2008), and thus these predictions. Agent-based simulations are inherently stochastic, and different results can be generated for the same parameter conditions (Read *et al.*, 2012). Through use of the consistency analysis technique developed by Read et al (2012), it has been assured that predictions made have been developed from simulation results that are representative of the condition on which the simulation has been run. This increases confidence that the results presented are robust, and the effect of inherent stochasticity is mitigated. With the exclusion of this study and Read's simulation of EAE (Read, 2011), evidence that such an analysis has been performed for other agent-based simulations in the literature is not apparent. This is not to say that a similar analysis technique has not been performed, but if this is the case results from such analyses have not been published. The examples in this thesis, Read's study (2011), and the release of the *spartan* toolkit constructed in the course of this study, could encourage more simulation developers to perform this analysis and ensure that results being presented are a robust representation of the exploration that was performed.

6.5 Simulating Peyer's Patch Formation Could Provide Insight on Lymphoid Organ Development

The accepted biological model of Peyer's Patch development has much in common with the formation of other lymphoid organs (Randall *et al.*, 2008). Thus it could be suggested that the predictions generated for Peyer's Patch development could also be applicable to the development of other secondary lymphoid organs such as lymph nodes and the spleen, and tertiary lymphoid organs that form during chronic infection (van de Pavert and Mebius, 2010).

Although the formation of lymph nodes does differ as development is encapsulated within the lymphatic epithelium (Randall *et al.*, 2008), primitive lymph nodes are aggregations of LTi cells around VCAM-1$^+$ LTo cells (Yoshida *et al.*, 2002) analogous to Peyer's Patch formation. Similarly to the PP model, ligation of LTβR on the LTo cell by LT$\alpha\beta$ on the LTi cell leads to the upregulation of chemokines CXCL13, CCL19, and CCL21 and adhesion factors VCAM-1, ICAM-1 and MAdCAM (Mebius, 2003). A clustering of LTi cells around VCAM-1 cells has also been detected in the developing spleen from E13, in the same lymphotoxin dependent manner (Withers *et al.*, 2007). With this similarity, it could be suggested that, if expression patterns were identical, the predictions generated through LTi and LTo cell interaction in the PP model could hold for lymphoid organ development in general, including the prediction that formation could be biphasic.

Such a hypothesis does not take the difference in physical environments into account. Yet a future iteration of the simulation could examine the effect the environment has on lymphoid tissue development. The simulation that has been implemented examines system behaviour at an interaction focused, higher-level, therefore certain pathways can be abstracted. Although there are differences between signalling pathways involved in PP, lymph node, and spleen development (the activation of LT$\alpha\beta$ by IL-7 in PP and TRANCE in lymph nodes being one (Cupedo *et al.*, 2004; Yoshida *et al.*, 2002)), such differences are lower level and are thus abstracted, leaving the behaviour that is similar between the three biological models of development. Predictions on the environmental effect on the formation of different secondary lymphoid organs could thus be drawn by altering just the simulation environment.

6.6 Novel use of Statistical Analysis Tools

The development of the *spartan* package, and the contribution this makes to the field of computer modelling and simulation, was noted in a previous section of this chapter. This section notes that this thesis has presented two novel applications of techniques within the *spartan* package.

The first of these is the application of the Extended Fourier Amplitude Sampling

Test (eFAST) (Saltelli, 2004; Saltelli and Bollardo, 1998) to results generated from an agent-based model. There are a number of examples in the literature of the eFAST technique, or its predecessor FAST (Cukier, 1973; McRae *et al.*, 1982) upon which eFAST is based, providing insight into results generated through simulation (King and Perera, 2007; Lu and Mohanty, 2001; Marino *et al.*, 2008). However none of the published examples applied the technique alongside an agent-based implementation. As was noted in the description of eFAST in section 2.4.4, this technique can be computationally expensive, especially for simulations with a large number of parameters (Ratto *et al.*, 2007; Tarantola *et al.*, 2006). In the case of the cell behaviour analysis in Chapter 3, eFAST produced 1,365 parameter value sets on which simulations needed to be run. As it was determined that each simulation needed to be run 500 times to mitigate aleatory uncertainty (section 3.3), the analysis required 682,500 simulation runs. Fortunately, simulation work conducted in this study could be run on a computer cluster, but if this was not available this analysis may not have been viable. Thus, the application of the technique is novel, but does come with a heavy resource cost.

The techniques in the *spartan* package have also been used to analyse simulated cell tracking data from a number of different time-points (Chapter 5), to understand whether the influence of certain parameters changes over the course of simulation time. The only example of such an analysis being performed previously is by Ray et al (2009), who examined how a calculated Partial Rank Correlation Coefficient changes over the course of simulation time in their model of TNF in controlling tuberculosis in a granuloma. The application of both the one-at-a-time robustness technique (Read *et al.*, 2012) and eFAST (Saltelli, 2004; Saltelli and Bollardo, 1998) over simulation time in this thesis is however novel. Functionality has been provided within the *spartan* toolkit to analyse simulation results captured at different time-points, thus the adoption of the package by the simulation developers may make such time-lapse analyses more common.

6.7   Future Directions

6.7.1   Investigating Cellular Mechanisms

In this study a variety of techniques, such as argument-based validation, calibration, and sensitivity analysis techniques, have been utilised to build confidence in the simulation as a tool representative of the biological domain that it captures. In their respective studies on confidence in simulation, both Bauer et al (2009) and Read et al (2011) note that the best practice in building confidence in predictions generated through simulation is to verify that prediction in the laboratory. Such an approach is also encouraged in the CoSMoS Framework (Andrews *et al.*, 2010) that has been adopted in this study. For the predictions that have been generated in the preceding

chapters, the loop is yet to be closed, as these have yet to be verified experimentally.

The prediction that the 72 hour period of PP formation may be biphasic has been generated by examining simulated cell behaviour at different time-points in the development period. The first time-point examined, 12 hours, is an *in silico* replication of an *ex vivo* investigation (Patel *et al.*, 2012). A similar culture system could be used to examine cell behaviour at a number of points from 12 hours onwards, producing responses that can be compared to those generated from the simulator. A statistical similarity between the simulation predictions and *ex vivo* data from a number of time-points would suggest the simulation parameters have been calibrated such that the emerging cell behaviour is correctly captured. A statistical difference may lead to some concern in the viability of the predictions that have been generated, yet it is important to recall that these predictions were formed based on the data available upon simulation construction. Data from numerous time-points could inform a new iteration of the modelling process, and if replicated by the simulator lead to stronger predictions in the long run. Although the prediction may not have proven correct, the simulation result has been used to inform biological exploration.

The above prediction could also be examined through performing experiments that block adhesion factor VCAM-1. Data presented in this thesis has suggested an important role for adhesion factor response and expression in the early stages, but this influence decreases as chemokines become influential between hours 24 and 36. It would be interesting to study the effect of inhibiting VCAM-1 at different time-points in the development period. To support the data presented in preceding chapters, one would expect to see no aggregations of hematopoietic cells if VCAM-1 is blocked in the early stages of development. However, if this factor was inhibited from hour 36 onwards, aggregations may still form, as the phase influenced by adhesion factors has elapsed. Such a result would strengthen both the hypothesis and confidence in the simulation as an adequate representation of the biological system.

Other predictions generated within this thesis, such as the role of LTin cells and the biological factors affecting the size of 2D cell aggregations that are formed, have a link to the number of hematopoietic (LTin/LTi) cells that are present in the gut. The current simulation bases cell number on estimates informed by flow cytometry results. A linear input rate is calculated such that the simulation creates the required number of cells with each time-step to meet this estimated number at E15.5. With no further data available, this rate has been continued through until E17.5, the end of the 72 hour development period. If flow cytometry could be performed on foetal intestines from more time-points, further estimates could be made such that the migration rate of LTin and LTi cells could be set more accurately. A more accurate number of cells would in turn lead to a more accurate number of cells in each aggregation, affecting predictions made on both number and 2D area of aggregations formed in simulation. It would then be possible to explore whether it is cell number that is a factor in limiting the number

of PP that are formed. The effect of altering LTin cell migration rate was investigated *in silico* in Chapter 4, by exploring two alternative migration rate functions, both set such that the required number of simulated cells were present at E15.5 (Figure 4.4a). Although this proved useful in suggesting that LTin migration rate could potentially have a role in limiting PP formation, this result is still based on the assumption that cell migration is constant throughout the period. A flow cytometry analysis would help reveal if this indeed is the case.

6.7.2   Statistical Analysis

The statistical analyses in the preceding chapters have examined the influence of simulation parameters for which a value is unknown (Table 3.1). These are parameters where the value cannot be directly translated from the biological system. For example, two of these are constants used in the sigmoid curve that captures chemokine expression (as detailed in Figure 2.8). The analyses have revealed how robust the simulation is to alterations in the value of these parameters, and determined the parameters in this subset that influence emergent behaviours. Future work could consider extending the application of sensitivity analysis techniques to a wider number of parameters in the simulation.

In his study of building confidence in a simulation of EAE, Read (2011) explores how robust the simulation is to an alteration in all parameter values, including those for which a biological value can be derived. Such an analysis could potentially reveal further insight into how robust PP formation is to changes in biological as well as the implementation specific parameters. Including parameters such as the LTin and LTi number estimates in a full global sensitivity analysis could give a statistical indication of how influential cell number is in both cell behaviour and patch characteristic simulation responses, in comparison to all other factors. However, with this simulation being agent-based, and thus requiring a large number of replicate result sets for each condition on which it is run, one would have to deduce whether such results are worthy of a time and resource heavy analysis. For example, altering the parameters that capture the range in which an LTin and LTi cell speed exists would potentially have an impact on the result, as this change would directly influence the cell displacement response. Performing thousands of simulation runs would verify that hypothesis yet not generate a result that offers any insight. It may be more beneficial to ascertain the level of uncertainty in the biological parameter first, and include this in any future analysis if deemed to be high.

6.7.3   Extending the Simulation

The simulation presented in this thesis captures an abstraction of the current understanding of PP development gathered both from the literature and through collabora-

tion with experimental immunologists. Through use of statistical techniques included in the *spartan* toolkit, the simulator has been used to generate predictions on the factors that influence cell behaviours and aggregation characteristics that emerge during the 72 hour period. There may however be scope to extend the simulations predictive capabilities further. One such example is detailed below.

Secondary lymphoid organogenesis occurs during foetal development, alongside the development of other biological systems. In this instance, PP formation may be occurring alongside the development of the enteric nervous system (ENS) within the gut, a process dependent on interactions between RET expressing neural crest cells and stromal cells on the intestine wall (Patel *et al.*, 2012). As both these neural crest cells and LTin cells express RET (Patel *et al.*, 2012; Veiga-Fernandes *et al.*, 2007), there could be potential for neural crest cells to influence the development of PP. Although there is little biological evidence to support such a hypothesis, a new iteration of the model could include the addition of additional non-LTin RET ligand expressing cells, to determine the effect the existence of such cells could potentially have on PP formation.

Glossary

antimesenteric Area of small intestine opposite the mesentery (that attaches the small intestine to the abdomen).

agent-based modelling A modelling approach where each individual entity, such as a cell, is represented explicitly, and can thus maintain its own attributes and cell state. Agent behaviour is specified as rules that determine the set of states an agent may exist within, and the event that must occur for an agent to change state.

domain The system of interest that is being modelled, for example the foetal development of a PP.

domain model The current scientific understanding of the dynamics of the system to be modelled.

ex vivo A procedure in which an organ, cells, or tissue are taken out of a living body for an experimental procedure. The experiment utilises the tissue rather than artificial medias, as performed *in vitro*.

Goal-Structuring Notation A visual notation for performing Argument-Based Validation, providing a method of structuring such an analysis to ensure each step in an implementation is validated, the reasoning behind the inclusion or exclusion of a feature or assumption is provided, and evidence given as to why this conclusion has been drawn.

in silico Method of performing an exploration on a computer via a simulation tool rather than in the laboratory.

in vitro An experimental procedure performed outside of a living organism, such as in a test tube or laboratory dish.

in vivo An experimental procedure performed within a living organism.

latin-hypercube Parameter sampling approach that selects values for each parameter from the value space, aiming to reduce any possible correlations while ensuring efficient coverage of the space over a minimal number of samples.

LTi   Lymphoid Tissue Inducer Cell. Hematopoietic cell that migrates from
                 the fetal liver into the small intestine. Believed to be responsible for
                 initiating PP development.

LTin   Lymphoid Tissue Initiator Cells. Hematopoietic cell that migrates from
                 the fetal liver into the small intestine. Can respond to chemokine
                 expression and thus form most of the aggregation of cells within a
                 forming PP.

LTo   Lymphoid Tissue Organiser Cells. Non-hematopoietic cell that is
                 expressed on the epithelium of the small intestine. Differentiates on
                 stable bind with an LTin and LTi cell

Mann-Whitney U-Test Non-parametric statistical test for comparing two
                 distributions. Null hypothesis is that the two distributions could have
                 been taken from the same sample.

organogenesis The process of organ development, in this case secondary lymphoid
                 organs

ODE   Ordinary Differential Equations. Used to explore the dynamics of a
                 population of biological factors (such as target and infected cells in
                 influenza) using a series of differential equations. An equation is
                 compiled that specifies the impact each factor has on the size of the
                 population of its complementary factors.

PP   Peyer's Patches. Secondary lymphoid organs located in the small
                 intestine that trigger adaptive immune responses to antigen.

platform model Model that describes how the information in the domain model will
                 be encoded as a computer simulation.

PRCC   Partial Rank Correlation Coefficient. A robust measure for quantifying
                 non-linear relationships between an input and output.

secondary lymphoid organ Organs located at drainage points in lymphatic vessels
                 that initiate protective immune responses to antigens from peripheral
                 tissues.

sensitivity analysis The application of statistical techniques to examine how a
                 system responds to an alteration in input parameter values.

Si   eFAST First-Order Sensitivity Index. Indicates the fraction of output
                 variance that can be explained by the value assigned to a parameter.

STi eFAST Total-Order Sensitivity Index. Indicates the variance caused by higher-order non-linear effects between that parameter and the others explored.

UML Unified Modelling Language. A diagram notation widely used in software engineering, that has also found application in the specification of models of biological systems.

Vargha-Delaney A-Test An effect magnitude test used to examine and quantify the difference between two distributions.

Bibliography

Adachi, S., Yoshida, H., Kataoka, H., Nishikawa, S. (1997). Three distinctive steps in Peyer's patch formation of murine embryo. *International Immunology* **9**: 507–514.

Adachi, S., Yoshida, H., Maki, K., Saijo, K., Ikuta, K., Saito, T., Nishikawa, S.-i. (1998). Essential role of IL-7 receptor α in the formation of Peyer s patch anlage. *International Immunology* **10**: 1–6.

Alarcon, T., Byrne, H. M., Maini, P. K. (2005). A Multiple Scale Model for Tumor Growth. *Multiscale Modeling & Simulation* **3**: 440.

Alden, K., Read, M., Timmis, J., Andrews, P. S., Veiga-Fernandes, H., Coles, M. C. (2012a). Spartan: A Comprehensive Tool for Understanding Uncertainty in Simulations of Biological Systems. *In Press* .

Alden, K., Timmis, J., Andrews, P. S., Veiga-Fernandes, H., Coles, M. C. (2012b). Pairing experimentation and computational modelling to understand the role of tissue inducer cells in the development of lymphoid organs. *Frontiers in Immunology* **3**: 1–20.

An, G. (2001). Agent-Based Computer Simulation and SIRS: building a bridge between basic science and clinical trials. *Shock* **16**: 266–273.

An, G. (2006). Concepts for developing a collaborative in silico model of the acute inflammatory response using agent-based modeling. *Journal of critical care* **21**: 105–10; discussion 110–1.

Andrews, P. S., Polack, F., Sampson, A. T., Timmis, J., Scott, L., Coles, M. (2008). Simulating biology : towards understanding what the simulation shows. In Proceedings of the 2008 Workshop on Complex Systems Modelling and Simulation, 93–123.

Andrews, P. S., Polack, F. A. C., Sampson, A. T., Stepney, S., Timmis, J. (2010). The CoSMoS Process , Version 0.1 : A Process for the Modelling and Simulation of Complex Systems. *Technical Report YCS-2010-453, Department of Computer Science, University of York, 2010* 1–40.

Ansel, K. M., Ngo, V. N., Hyman, P. L., Luther, S. a., Förster, R., Sedgwick, J. D., Browning, J. L., Lipp, M., Cyster, J. G. (2000). A chemokine-driven positive feedback loop organizes lymphoid follicles. *Nature* **406**: 309–14.

Antia, R., Ganusov, V. V., Ahmed, R. (2005). The role of models in understanding CD8+ T-cell memory. *Nature reviews. Immunology* **5**: 101–11.

Baccam, P., Beauchemin, C., Macken, C. a., Hayden, F. G., Perelson, A. S. (2006). Kinetics of influenza A virus infection in humans. *Journal of virology* **80**: 7590–9.

Banks, T. A., Rouse, B. T., Kerley, M. K., Blair, P. J., Codfrey, V., Kuklin, N. A., Bouley, D. M., Thomas, J., Kanangat, S., Mucenskit, M. (1995). Lymphotoxin-alpha-Deficient Mice. Effects on secondary lymphoid organ development and humoral immune responsiveness. *Journal of immunology* **155**: 1685–1693.

Bauer, A. L., Beauchemin, C., Perelson, A. S. (2009). Agent-based modeling of host-pathogen systems: The successes and challenges. *Information Sciences* **179**: 1379–1389.

Beauchemin, C. (2006). Probing the effects of the well-mixed assumption on viral infection dynamics. *Journal of theoretical biology* **242**: 464–77.

Beauchemin, C., McSharry, J. J., Drusano, G. L., Nguyen, J. T., Went, G. T., Ribeiro, R. M., Perelson, A. S. (2008). Modeling amantadine treatment of influenza A virus in vitro. *Journal of theoretical biology* **254**: 439–51.

Beauchemin, C., Samuel, J., Tuszynski, J. (2005). A simple cellular automaton model for influenza A viral infections. *Journal of theoretical biology* **232**: 223–34.

Bersini, H., Carneiro, J. (2006). Immune System Modeling : The OO Way. In ICARIS 2006, LNCS 4163, 150 – 163.

Catron, D. M., Itano, A. a., Pape, K. a., Mueller, D. L., Jenkins, M. K. (2004). Visualizing the first 50 hr of the primary immune response to a soluble antigen. *Immunity* **21**: 341–7.

Celada, F., Seiden, P. (1998). Modeling immune cognition. *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218)* **4**: 3787–3792.

Celada, F., Seiden, P. E. (1996). Affinity maturation and hypermutation in a simulation of the humoral immune response. *European journal of immunology* **26**: 1350–8.

Celada, F., Selden, P. E. (1992). A computer model of cellular interactions in the immune system. *Immunology Today* **13**: 56–62.

Chakraborty, A. K., Das, J. (2010). Pairing computation with experimentation: a powerful coupling for understanding T cell signalling. *Nature reviews. Immunology* **10**: 59–71.

Chakraborty, A. K., Dustin, M. L., Shaw, A. S. (2003). In silico models for cellular and molecular immunology : successes , promises and challenges **4**: 933–936.

Chao, D. L., Davenport, M. P., Forrest, S., Perelson, A. S. (2003). Stochastic stage-structured modeling of the adaptive immune system. *Proceedings / IEEE Computer Society Bioinformatics Conference. IEEE Computer Society Bioinformatics Conference* **2**: 124–31.

Cohen, I. R. (2007). Modeling immune behavior for experimentalists. *Immunological reviews* **216**: 232–6.

Cornes, J. S. (1965). Number, size, and distribution of Peyer's patches in the human small intestine: Part I The development of Peyer's patches. *Gut* **6**: 225–229.

Cukier, R. I. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics* **59**: 3873.

Cupedo, T., Lund, F. E., Ngo, V. N., Randall, T. D., Jansen, W., Greuter, M. J., de Waal-Malefyt, R., Kraal, G., Cyster, J. G., Mebius, R. E. (2004). Initiation of cellular organization in lymph nodes is regulated by non-B cell-derived signals and is not dependent on CXC chemokine ligand 13. *Journal of immunology (Baltimore, Md. : 1950)* **173**: 4889–96.

Cyster, J. G. (1999). Chemokines and cell migration in secondary lymphoid organs. *Science* **286**: 2098–102.

De Togni, P., Goellner, J., Ruddle, N., Streeter, P., Fick, A., Mariathasan, S., Smith, S., Carlson, R., Shornick, L., Strauss-Schoenberger, J. (1994). Abnormal development of peripheral lymphoid organs in mice deficient in lymphotoxin. *Science* 703–707.

Dejardin, E., Droin, N. M., Delhase, M., Haas, E., Cao, Y., Makris, C., Li, Z.-w., Karin, M., Ware, C. F., Green, D. R. (2002). The Lymphotoxin-Beta Receptor Induces Different Patterns of Gene Expression via Two NF- kB Pathways. *Immunity* **17**: 525–535.

Di Paulo, E. A., Noble, J., Bullock, S. (2000). Simulation Models as Opaque Thought Experiments. In Bedau, M. A., McCaskill, J. S., Packard, N., Rasmussen, S. (editors), The Seventh International Conference on Artificial Life, 497–506. MIT Press.

Eberl, G., Marmon, S., Sunshine, M.-J., Rennert, P. D., Choi, Y., Littman, D. R. (2004). An essential function for the nuclear receptor ROR gamma in the generation of fetal lymphoid tissue inducer cells. *Nature Immunity* **5**: 64–73.

Efroni, S., Harel, D., Cohen, I. R. (2003). Toward Rigorous Comprehension of Biological Complexity : Toward Rigorous Comprehension of Biological Complexity : Modeling , Execution , and Visualization of Thymic T-Cell Maturation. *Genome Research* 2485–2497.

Efroni, S., Harel, D., Cohen, I. R. (2005). Reactive Animation: Realistic Modeling of Complex Dynamic Systems. *Computer* **38**: 38–47.

Efroni, S., Harel, D., Cohen, I. R. (2007). Emergent dynamics of thymocyte development and lineage determination. *PLoS computational biology* **3**: e13.

Figge, M. T., Garin, A., Gunzer, M., Kosco-Vilbois, M., Toellner, K.-M., Meyer-Hermann, M. (2008). Deriving a germinal center lymphocyte migration model from two-photon data. *The Journal of experimental medicine* **205**: 3019–29.

Finke, D., Mattis, A., Lipp, M., Kraehenbuhl, J. P. (2002). CD4+CD3- Cells Induce Peyers Patch Development: Role of alpha4beta1 Integrin Activation by CXCR5. *Immunity* **17**: 363–373.

Forrest, S., Beauchemin, C. (2007). Computer Immunology. *Immunological Reviews* **216**: 176–197.

Fu, Y. X., Chaplin, D. D. (1999). Development and maturation of secondary lymphoid tissues. *Annual review of immunology* **17**: 399–433.

Fukuyama, S., Kiyono, H. (2007). Neuroregulator RET initiates Peyer's-patch tissue genesis. *Immunity* **26**: 393–5.

Futterer, A., Mink, K., Luz, A., Kosco-Vilbois, M., Pfeffer, K. (1998). The lymphotoxin beta receptor controls organogenesis and affinity maturation in peripheral lymphoid tissues. *Immunity* **9**: 59–70.

Garnett, P., Stepney, S., Leyser, O. (2008). Towards an Executable Model of Auxin Transport Canalisation. In Proceedings of the 2008 Workshop on Complex Systems Modelling and Simulation, 63–91.

Germain, R. N. (2001). The art of the probable: system control in the adaptive immune system. *Science (New York, N.Y.)* **293**: 240–5.

Germain, R. N., Meier-schellersheim, M., Nita-lazar, A., Fraser, I. D. C. (2011). Systems Biology in Immunology : A Computational Modeling Perspective. *Annual Review of Immunology* **29**: 527–585.

Ghetiu, T., Alexander, R. D., Andrews, P. S., Polack, F. A. C., Bown, J. (2009). Equivalence Arguments for Complex Systems Simulations A Case-Study. In 2nd CoSMoS Workshop. Luniver Press.

Ghetiu, T., Polack, F. A. C., Bown, J. (2010). Argument-driven Validation of Computer Simulations A Necessity Rather Than an Option. In VALID 2010: The Second International Conference on Advances in System Testing and Validation Lifecycle, 1–4.

Goodnow, C. C. (1997). Chance encounters and organized rendezvous. *Immunological reviews* **156**: 5–10.

Guo, Z., Tay, J. C. (2005). A Comparative Study on Modeling Strategies for Immune System Dynamics Under HIV-1 Infection. In Artificial Immune Systems, 4th International Conference, ICARIS 2005, LNCS 3627, 220 – 233. Springer.

Haeno, H., Gonen, M., Davis, M. B., Herman, J. M., Iacobuzio-Donahue, C. A., Michor, F. (2012). Computational Modeling of Pancreatic Cancer Reveals Kinetics of Metastasis Suggesting Optimum Treatment Strategies. *Cell* **148**: 362–375.

Hashi, H., Yoshida, H., Honda, K., Fraser, S., Kubo, H., Awane, M., Takabayashi, A., Nakano, H., Yamaoka, Y., Nishikawa, S.-I. (2001). Compartmentalization of Peyers Patch Anlagen Before Lymphocyte Entry. *Journal of immunology* **166**: 3702–3709.

Helton, J. C. (2008). Uncertainty and sensitivity analysis for models of complex systems. In Barth, T. J., Griebel, M., Keyes, D. E., Nieminen, R. M., Roose, D., Schlick, T. (editors), Computational Methods in Transport: Verification and Validation, 207–228. Springer.

Honda, K., Nakano, H., Yoshida, H., Nishikawa, S., Rennert, P., Ikuta, K., Tamechika, M., Yamaguchi, K., Fukumoto, T., Chiba, T., Nishikawa, S.-I. (2001). Molecular Basis for Hematopoietic/Mesenchymal Interaction during Initiation of Peyer's Patch Organogenesis. *Journal of Experimental Medicine* **193**: 621–630.

Hoorweg, K., Cupedo, T. (2008). Development of human lymph nodes and Peyer's patches. *Seminars in immunology* **20**: 164–70.

Hu, J., Sealfon, S. C., Hayot, F., Jayaprakash, C., Kumar, M., Pendleton, A. C., Ganee, A., Fernandez-Sesma, A., Moran, T. M., Wetmur, J. G. (2007). Chromosome-specific and noisy IFNB1 transcription in individual virus-infected human primary dendritic cells. *Nucleic acids research* **35**: 5232–41.

Jacob, C., Litorco, J., Lee, L. (2004). Immunity through Swarms : Agent-based Simulations of the Human Immune System. In Nicosia, G., Cutello, V., Bentley, P., Timmis, J. (editors), Artificial Immune Systems, International Conference, ICARIS 2004 Proceedings. Lecture Notes in Computer Science 3239, 400–412. Springer.

Jiang, Y., Pjesivac-Grbovic, J., Cantrell, C., Freyer, J. P. (2005). A multiscale model for avascular tumor growth. *Biophysical journal* **89**: 3884–94.

Jung, C., Hugot, J.-P., Barreau, F. (2010). Peyer's Patches: The Immune Sensors of the Intestine. *International journal of inflammation* 823710.

Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Assad-Garcia, N., Glass, J. I., Covert, M. W. (2012). A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell* **150**: 389–401.

Kelly, T. P. (1999). Arguing Safety A Systematic Approach to Managing Safety Cases. Ph.D. thesis, University of York.

King, D., Perera, B. J. C. (2007). Sensitivity Analysis for Evaluating Importance of Variables Used in an Urban Water Supply Planning Model. In Kulasiri, L. O., D. (editors), Proceedings of the International Congress on Modelling and Simulation (MODSIM '07), 2768–2774. Modelling and Simulation Society of Australia and New Zealand.

Kirschner, D. E., Linderman, J. J. (2009). Mathematical and computational approaches can complement experimental studies of host-pathogen interactions. *Cellular Microbiology* **11**: 531–539.

Klein, J. (2002). BREVE : a 3D Environment for the Simulation of Decentralized Systems and Artificial Life. In Proceedings of Artificial Life VIII, the 8th International Conference on the Simulation and Synthesis of Living Systems. MIT Press.

Kleinstein, S. H. (2008). Getting started in computational immunology. *PLoS computational biology* **4**: e1000128.

Kohler, B., Puzone, R., Seiden, P. E., Celada, F. (2000). A systematic approach to vaccine complexity using an automaton model of the cellular and humoral immune system. *Vaccine* **19**: 862–876.

Kumar, N., Hendriks, B. S., Janes, K. a., de Graaf, D., Lauffenburger, D. a. (2006). Applying computational modeling to drug discovery and development. *Drug discovery today* **11**: 806–11.

Lagreca, M. C., de Almeida, R. M. C., Zorzenon dos Santos, R. M. (2001). A dynamical model for the immune repertoire. *Physica A* **289**: 191–207.

Linderman, J. J., Rigg, T., Pande, M., Miller, M., Marino, S., Kirschner, D. E. (2011). Characterizing the dynamics of CD4+ T cell priming within a lymph node. *Journal of immunology* **184**: 2873–2885.

Lu, Y., Mohanty, S. (2001). Sensitivity analysis of a complex, proposed geologic waste disposal system using the Fourier Amplitude Sensitivity Test method. *Reliability Engineering and System Safety* **72**: 275–291.

Luke, S. (2005). MASON: A Multiagent Simulation Environment. *Simulation* **81**: 517–527.

Luther, S., Ansel, K. M., Cyster, J. G. (2003). Overlapping roles of CXCL13, interleukin 7 receptor alpha, and CCR7 ligands in lymph node development. *The Journal of experimental medicine* **197**: 1191–8.

Mahajan, S. D., Schwartz, S. a., Nair, M. P. (2003). Immunological assays for chemokine detection in in-vitro culture of CNS cells. *Biological procedures online* **5**: 90–102.

Marino, S., Hogue, I. B., Ray, C. J., Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology* **254**: 178–96.

Mata, J., Cohn, M. (2007). Cellular automata-based modeling program: synthetic immune system. *Immunological reviews* **216**: 198–212.

McKay, M. D., Beckman, R. J., Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Techometrics* **21**: 239–245.

McRae, G. J., Tilden, J. W., Seinfeld, J. H. (1982). Global Sensitivity Analysis - A Computational Implementation of the Fourier Amplitude Sampling Test (FAST). *Computers and Chemical Engineering* **6**: 15–25.

Mebius, R. E. (2003). Organogenesis of lymphoid tissues. *Nature reviews. Immunology* **3**: 292–303.

Mebius, R. E., Miyamoto, T., Christensen, J., Domen, J., Cupedo, T., Weissman, I. L., Akashi, K. (2001). The fetal liver counterpart of adult common lymphoid progenitors gives rise to all lymphoid lineages, CD45+CD4+CD3- cells, as well as macrophages. *Journal of immunology* **166**: 6593–601.

Meier, D., Bornmann, C., Chappaz, S., Schmutz, S., Otten, L. A., Ceredig, R., Acha-Orbea, H., Finke, D. (2007). Ectopic Lymphoid-Organ Development Occurs through Interleukin 7-Mediated Enhanced Survival of Lymphoid-Tissue-Inducer Cells. *Immunity* **26**: 643–654.

Meier-Schellersheim, M., Xu, X., Angermann, B., Kunkel, E. J., Jin, T., Germain, R. N. (2006). Key role of local regulation in chemosensing revealed by a new molecular interaction-based modeling method. *PLoS Computational Biology* **2**: 710–724.

Milanesi, L., Romano, P., Castellani, G., Remondini, D., Liò, P. (2009). Trends in modeling Biomedical Complex Systems. *BMC bioinformatics* **10 Suppl 1**: I1.

Motta, S., Castiglione, F., Lollini, P., Pappalardo, F. (2005). Modelling vaccination schedules for a cancer immunoprevention vaccine. *Immunome research* **1**: 5.

Mowat, A. M. (2003). Anatomical basis of tolerance and immunity to intestinal antigens. *Nature reviews. Immunology* **3**: 331–41.

Nishikawa, S.-I., Honda, K., Vieira, P., Yoshida, H. (2003). Organogenesis of peripheral lymphoid organs. *Immunological reviews* **195**: 72–80.

Ohl, L., Henning, G., Krautwald, S., Lipp, M., Hardtke, S., Bernhardt, G., Pabst, O., Förster, R. (2003). Cooperating mechanisms of CXCR5 and CCR7 in development and organization of secondary lymphoid organs. *The Journal of experimental medicine* **197**: 1199–204.

Owen, R., Bhalla, D. (1974). Epithelial cell specialisation within human Peyer's Patches: an ultrastructural study of intestinal lymphoid follicles. *Gastroenterology* **66**: 189–203.

Pasparakis, M., Alexopoulou, L., Grell, M., Pfizenmaier, K., Bluethmann, H., Peyer, K. G. (1997). Peyer s patch organogenesis - cytokines rule , OK ? *Gut* **41**: 707–709.

Patel, A., Harker, N., Moreira-Santos, L., Ferreira, M., Alden, K., Timmis, J., Foster, K. E., Garefalaki, A., Pachnis, P., Andrews, P. S., Enomoto, H., Milbrandt, J., Pachnis, V., Coles, M. C., Kioussis, D., Veiga-Fernandes, H. (2012). Differential RET responses orchestrate lymphoid and nervous enteric system development. *Science Signalling* **5**.

van de Pavert, S. a., Mebius, R. E. (2010). New insights into the development of lymphoid tissues. *Nature reviews. Immunology* 1–11.

Perelson, A. S. (2002). Modelling viral and immune system dynamics. *Nature reviews. Immunology* **2**: 28–36.

Perelson, A. S., Essunger, P., Cao, Y., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M., Ho, D. D. (1997). Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**: 188–191.

Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M., Ho., D. D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**: 1582–1586.

Polack, F. A. C., Andrews, P. S., Ghetiu, T., Read, M., Stepney, S., Timmis, J., Sampson, A. T. (2010). Reflections on the Simulation of Complex Systems for Science. In Kwiatkowska, R. C., Paige, R., Marta (editors), ICECCS 2010: 15th IEEE International Conference on Engineering of Complex Computer Systems, Oxford, UK, March 2010, 276–285.

Polack, F. A. C., Droop, A., Garnett, P., Ghetiu, T., Stepney, S. (2011). Simulation validation : exploring the suitability of a simulation of cell division and differentiation in the prostate. In CoSMoS 2011: Proceedings of the 2011 Workshop on Complex Systems Modelling and Simulation.

Polack, F. A. C., Hoverd, T., Sampson, A. T., Stepney, S., Timmis, J. (2008). Complex Systems Models : Engineering Simulations. In ALife XI, Winchester, UK, September 2008, 2002, 482–489.

Puzone, R., Kohler, B., Seiden, P., Celada, F. (2002). IMMSIM, a flexible model for in machina experiments on immune system responses. *Future Generation Computer Systems* **18**: 961–972.

Randall, T. D., Carragher, D. M., Rangel-Moreno, J. (2008). Development of secondary lymphoid organs. *Annual Review Immunology* **26**: 627–650.

Ratto, M., Pagano, A., Young, P. (2007). State Dependent Parameter metamodelling and sensitivity analysis. *Computer Physics Communications* **177**: 863–876.

Ray, J. C. J., Flynn, J. L., Kirschner, D. E. (2009). Synergy between individual TNF-dependent functions determines granuloma performance for controlling mycobacterium tuberculosis infection. *Journal of theoretical biology* **182**: 3706–3717.

Read, M. (2011). Statistical and Modelling Techniques to Build Confidence in the Investigation of Immunology through Agent-Based Simulation. Ph.D. thesis, University of York.

Read, M., Andrews, P. S., Timmis, J., Kumar, V. (2012). Techniques for Grounding Agent-Based Simulations in the Real Domain : a case study in Experimental Autoimmune Encephalomyelitis. *Mathematical and Computer Modelling of Dynamical Systems* **18**: 67–86.

Read, M., Timmis, J., Andrews, P. S., Kumar, V. (2009). A Domain Model of Experimental Autoimmune Encephalomyelitis. In 2nd Workshop on Complex Systems Modelling and Simulation, 9–44.

Reis, B. S., Mucida, D. (2012). The role of the intestinal context in the generation of tolerance and inflammation. *Clinical & developmental immunology* **2012**: 157948.

Rumbaugh, J., Jacobson, I., Booch, G. (2005). The Unified Modelling Language Reference Manual. Addison-Wesley, 2 edition.

Saltelli, A. (2004). Sensitivity Analysis in practice: A guide to assessing scientific models. Wiley.

Saltelli, A., Bollardo, R. (1998). An alternative way to compute Fourier amplitude sensitivity test (FAST). *Comput. Stat. Data Anal.* **26**: 445–460.

Saltelli, A., Chan, K., Scott, E. M. (2000). Sensitivity Analysis. Wiley series in probability and statistics Wiley.

Sidorenko, Y., Reichl, U. (2004). Structured model of influenza virus replication in MDCK Cells. *Biotechnol Bioend* **88**: 1–14.

Sieburg, H. B., McCutchan, J. A., Clay, O. K., Cabalerro, L., Ostlund, J. J. (1990). Simulation of HIV Infection in Artificial Immune Systems. *Physica D* **45**: 208–227.

Smith, A. M., Adler, F. R., McAuley, J. L., Gutenkunst, R. N., Ribeiro, R. M., McCullers, J. a., Perelson, A. S. (2011). Effect of 1918 PB1-F2 expression on influenza A virus infection kinetics. *PLoS computational biology* **7**: e1001081.

Smith, A. M., Perelson, A. S. (2012). Influenza A Virus Infection Kinetics: Quantitative Data and Models. *Wiley Interdiscip Rev Syst Biol Med.* **3**: 429–445.

Stafford, M. A., Corey, L., Cao, Y., Daar, E. S., Ho, D. D., Perelson, A. S. (2000). Modeling Plasma Virus Concentration during Primary HIV Infection. *Journal of The* **203**: 285–301.

Strain, M. C., Richmanz, D. D., Wongz, J. K., Levinew, H. (2002). Spatiotemporal Dynamics of HIV Propagation. *Journal of theoretical biology* **218**: 85–96.

Sun, Z. (2000). Requirement for RORgamma in Thymocyte Survival and Lymphoid Organ Development. *Science* **288**: 2369–2373.

Tarantola, S., Gatelli, D., Mara, T. (2006). Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering & System Safety* **91**: 717–727.

Vaidya, N. K., Rong, L., Marconi, V. C., Kuritzkes, D. R., Deeks, S. G., Perelson., A. S. (2010). Treatment-mediated alterations in HIV fitness preserve CD4+ T cell counts but have minimal effects on viral load. *PLoS Computational Biology* **6**: 1–14.

Vargha, A., Delaney, H. D. (2000). A critique and improvement of the CL Common Language Effect Size Statistics of McGraw and Wong. *Journal of Educational and Behavioural Statistics* **25**: 101–132.

Veiga-Fernandes, H., Coles, M. C., Foster, K. E., Patel, A., Williams, A., Natarajan, D., Barlow, A., Pachnis, V., Kioussis, D. (2007). Tyrosine kinase receptor RET is a key regulator of Peyer's patch organogenesis. *Nature* **446**: 547–51.

Watson, J., Crick, F. (1953). Molecular Structure of Nucleic Acids. *Nature* **171**: 737–738.

Withers, D. R., Kim, M.-Y., Bekiaris, V., Rossi, S. W., Jenkinson, W. E., Gaspal, F., McConnell, F., Caamano, J. H., Anderson, G., Lane, P. J. L. (2007). The role of lymphoid tissue inducer cells in splenic white pulp development. *European journal of immunology* **37**: 3240–5.

Yokota, Y., Mansouri, A., Mori, S., Sugawara, S., Adachi, S., Nishikawa, S.-I., Gruss, P. (1999). Development of peripheral lymphoid organs and natural killer cells depends on the helix-loop-helix inhibitor Id2. *Nature* **397**: 702–706.

Yoshida, H., Kawamoto, H., Santee, S., Hashi, H., Honda, K. (2001). Expression of a4b7 integrin defines a distinct pathway of lymphoid progenitors committed to T cells, fetal intestinal lymphoxin producer, NK, and dendritic cells. *Journal of immunology* **167**: 2511–2521.

Yoshida, H., Naito, J., Inoue, J. (2002). Different cytokines induce surface lymphotoxin-ab on IL-7 receptor-a cells that differentially engender lymph nodes and Peyer's Patches. *Immunity* **17**: 823–833.

Yoshida, H., Shinkura, R., Adachi, S., Nishikawa, S., Maki, K., Ikuta, K., Nishikawa, S.-i. (1999). IL-7 receptor alpha+ CD3 cells in the embryonic intestine induces the organizing center of Peyer s patches **11**: 643–655.

Zhao, J., Tiede, C. (2011). Using a variance-based sensitivity analysis for analyzing the relation between measurements and unknown parameters of a physical model. *Nonlinear Processes in Geophysics* **18**: 269–276.

Zheng, H., Jin, B., Henrickson, S. E., Perelson, A. S., von Andrian, U. H., Chakraborty, A. K. (2008). How antigen quantity and quality determine T-cell decisions in lymphoid tissue. *Molecular and cellular biology* **28**: 4040–51.