The Spatial Distribution and Dynamics of CXCL13 in Lymphoid Tissues

Jason Cosgrove

PhD

University of York

Biology

September 2017

Abstract

Morphogens are soluble signalling molecules that regulate a broad spectrum of biological processes. However, the distances and scales over which this regulation occurs are unclear. To date, many studies have highlighted source-sink mechanisms for morphogen gradient formation but fail to take the role of the tissue microenvironment into account. Using a systems-based approach we show that the chemokine CXCL13 is regulated by the B-cell microenvironment on distinct but interconnected levels of biological organization.

CXCL13 is a key determinant of humoral immune responses, regulating the localisation of lymphocytes within lymphoid tissues. Due to a complex and dynamic interaction network occurring over broad spatiotemporal scales, mapping the spatial distribution of CXCL13 *in situ* is challenging. To address this we have mapped the 3-dimensional organisation of $CXCL13^{+}$ stromal cells *in situ* using a fluorescent reporter system, identifying three distinct but interconnected stromal subsets that are unique in their network properties. We quantify CXCL13 dynamics using high-speed narrowfield microscopy in collagen matrix and lymph node tissue sections with results suggesting that diffusion is highly constrained by local tissue microanatomy.

However, this data alone is insufficient to describe CXCL13 gradient formation. To consolidate this data we employ a quantitative modelling approach hybridising different techniques into a high fidelity *in silico* representation of the B-follicle, where immune cells can interact with stroma capable of creating and shaping complex physiological gradients. Simulation analyses and immunohistochemistry suggest that chemokine fields within the follicle are dynamic and non-uniform, with multiobjective optimization analysis suggesting that this spatial configuration is designed to promote scanning rates. Taken in concert, our data suggests that CXCL13 acts over short distances creating a complex landscape of expression. Importantly, this study provides a basis for understanding the spatial distribution of morphogens with complex binding behaviours.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank the many people who have contributed to the completion of this thesis. First and foremost I thank Professor Mark Coles and Professor Jon Timmis for their supervision, enthusiasm and expertise over the past 4 years. The work conducted in this thesis was possible thanks to funding from the Wellcome Trust for the Combating Infectious Disease: Computational Approaches in Translational Science (CIDCATS) doctoral training programme. I am grateful to the programme directors for granting me this opportunity and to my cohort for their support and friendship.

This PhD project was highly collaborative, with significant contributions from researchers spanning a number of scientific disciplines. From the Coles and Timmis research groups I would like to thank Anne Thuery and Emily Taylor for their help in the lab, as well as Kieran Alden and Simon Jarrett for their assistance with machine learning, model analysis and software engineering. From our collaborators I would like to thank Helen Miller for our collaboration on single molecule imaging; Natalia Pikor, Urs Mober, Lucas Onder and members of the Stein lab for providing imaging datasets; and to Mario Novkovic for sharing his expertise on network analysis. Contributions are noted on each relevant figure.

I am also very grateful to Professor Burkhard Ludewig, Professor Antal Rot, Professor Jens Stein, Professor Mark Leake, Dr. Peter O' Toole, Dr. Marlene Wolf and Dr. Andy Pomfret for sharing their expertise and providing feedback on the project. Lastly, I would like to thank my friends and family and particularly Manon Chaumely 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. Some of this work has already been presented, in: Cosgrove et al. (2015), Butler et al. (2016a), and Butler et al. (2016b). Data used from these works within the thesis body are cited accordingly. 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.

Chapter 1 Introduction

1. Overview of the Research Context

Through the formation of concentration gradients, morphogens such as chemokines, cytokines, and growth factors drive graded responses to extracellular signals, fine-tuning cell behaviours to regulate processes such as development, the maturation of adaptive immune responses, and tumorigenesis. Despite their fundamental importance, the precise mechanisms driving morphogen gradient formation in complex tissues have proven enigmatic and controversial, due in part to a dearth of experimental techniques capable of manipulating molecular gradients *in situ*

Recently, a diffusion-consumption scheme has been used to describe the spatial distribution of the cytokine IL-2 (Oyler-Yaniv et al., 2017). Consistent with the source-sink scheme of gradient formation, ACKR4 expressing Lymphatic Endothelial Cells (LECs) lining the ceiling of the subscapular sinus have been implicated in the formation of functional CCL21 gradients in the lymph node (Ulvmar et al., 2014). Interestingly, both molecules are known to dynamically interact with the extracellular matrix (ECM) through carbohydrate binding domains, a property common to many morphogens that may limit the capacity to undergo free diffusion, particularly in dense tissues (Barmore et al., 2016; Hasan et al., 1999; Wrenshall et al., 2003).

For many molecules, this ability to bind ECM components is a key determinant of functionality. *In vivo*, truncation of the highly charged C-terminus prevents its immobilization to high endothelial venules and affects lymphocyte homing to the lymph node (Stein et al., 2000). Mice carrying a mutated form of CXCL12 ($CXCL12^{gagtm}$) where interactions with the ECM are impaired show increased amounts of hematopoietic precursors, an impaired ability to support revascularization and impaired humoral immune responses. The structure of the germinal centre in $CXCL12^{gagtm}$ mice was impaired, as well as having fewer somatic mutations in immunoglobulin genes(Barinov et al., 2017). The role of the microenvironment in shaping gradients has also been described in the context of development. Fibroblastic Growth Factors (FGF) are critical for mesoderm induction as well as limb and lung development, modulating the proliferation and differentiation of a variety of cells of mesenchymal and neuroectodermal origin. Receptor dimerization is a prerequisite for FGF signalling and requires extracellular matrix components (Pellegrini et al., 2000). These experimental studies are supported by

mathematical analyses predicting that gradient formation is increased when chemokines are secreted in matrix-binding form as compared to a non matrix-interacting form (Fleury et al., 2006). In this thesis we focus on CXCL13, a chemokine that regulates the induction and maturation of antibody-mediated immune responses, seeking to address the following research question:

To what extent does the localised tissue microenvironment affect CXCL13 gradient formation, and how does the spatial distribution of CXCL13 affect the onset of humoral immune responses?

In the following sections we give a broad overview of antibody-mediated immune responses highlighting the key molecules, cell types, and time points of interest. Lastly we describe some emerging technologies used to study this pathway, focusing on imaging and systems biology approaches that are relevant to the work conducted in this thesis.

1.1 Overview of the Humoral Immune Response

The immune system is the collective term given to the cells, molecules, processes, tissues and organs that protect the health of an organism. It is responsible for mounting defences against pathogens and for clearing tumorous cells. Broadly speaking, the immune system can be divided into two main categories: innate and adaptive.

Innate immunity is of ancient origin with all animals and plants having some form of innate defence, such responses occur rapidly and in a non-specific manner upon infection. Adaptive responses take longer to develop but are capable of eliminating infections more efficiently through highly specific recognition of foreign substances, more commonly referred to as antigen (Ag). Specificity is provided through the immune repertoire; diversity in the repertoire is initially introduced via the process of VDJ recombination, where lymphocytes randomly assemble different gene segments – known as variable (V), diversity (D) and joining (J) genes – in order to generate unique receptors that can recognise a diverse array of molecules (Murphy, 2011).

In the following sections we focus on a specific component of adaptive immunity, the humoral immune response. Many of the microorganisms that cause infectious disease in humans multiply in the extracellular spaces of the body, and most intracellular pathogens spread by moving from cell to cell through the extracellular fluids (Murphy, 2011). The extracellular spaces are protected by the humoral immune response in which antibodies are produced by B-lymphocytes. Antibodies are dimeric proteins combined to protect against surface antigens expressed by bacteria, viruses and protozoa (**Figure 1.1**). This is achieved through three main mechanisms: neutralization, opsonisation and activation of the complement system (Murphy, 2011).

Given that bacteria and viruses can replicate rapidly, the immune system needs an efficient mechanism for creating large amounts of highly specific antibody-producing cells capable of dealing with high pathogen titres. During an immune response, the amount and affinity of immunoglobulin increases in a process known as affinity maturation (Victora, 2014). This process is tightly regulated, occurring in a specialized microenvironment known as the germinal centre (GC) where secondary diversification of the immune repertoire is achieved through three mechanisms: *(i) somatic hypermutation* (SHM), introducing point mutations into the variable regions to alter antigen affinity; *(ii)*

class switching which increases the functional diversity of the repertoire through replacement of the constant (C) region; and *(iii) gene conversion* which is similar to SHM but blocks of codons are donated, usually from an upstream site, instead of point mutations (Murphy, 2011).

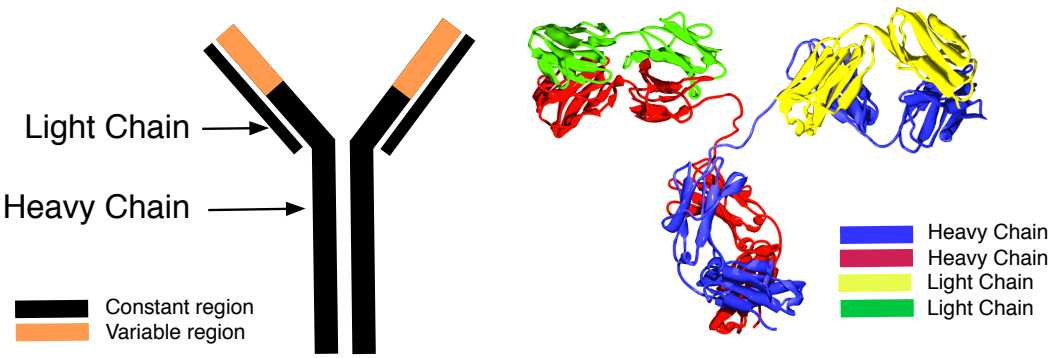

Figure 1.1: Immunoglobulin structure. On the right is the crystal structure of IgG2 for comparison with the basic scheme presented on the left.

Following diversification, B cells must compete for survival signals with those bearing the highest-affinity receptors surviving the selection process. B cells which pass this checkpoint may then contribute to immunological memory through two distinct populations of cells: *(i) long-lived plasma cells* (PCs) which secrete neutralizing antibodies long after Ag clearance and *(ii) memory B cells* that rapidly proliferate and differentiate into PCs following recurrent exposure to the initial immunizing Ag (Victora and Mesin, 2014).

It is important to note reports of somatic hypermutation occurring at extrafollicular sites in the context of T-independent antigens associated with salmonella infection (Di Niro et al., 2015). However, this has not been extensively studied and so the role of the localised tissue architecture in this context is poorly understood. For the remainder of this thesis we focus on the canonical antibody maturation pathway associated with T-cell dependent antigen.

1.2 The Induction of Humoral Immune Responses in Lymph Nodes

Dysregulation of affinity maturation can lead to inefficient or inappropriately directed immune responses, with most human B cell lymphomas originating from GC B cells (Basso and Dalla-Favera, 2015). To regulate the induction of affinity maturation, checkpoints must be reached before a B-cell can undergo activation and selection (**Figure 1.2**). Specifically, a B-cell must *(i) acquire and process antigen; (ii) present antigen to CD4$^+$ T helper cells via MHC-II; and (iii) organise into a germinal centre (Pereira et al., 2010)*. For this complex cascade of interactions to occur, a B-cell must navigate effectively through dense secondary lymphoid tissues. As few as one in 100,000 B and T lymphocytes are specific for a single Ag, yet these cells must come together if an effective antibody response is to occur (Murphy, 2011). As T cells and B cells mostly occupy two distinct zones in peripheral lymphoid tissues - the T-cell areas and the primary lymphoid follicles - there is a need for the precise regulation of cell migration. B-cell activation is thus highly dependent on the architecture of secondary lymphoid organs (Cremasco et al., 2014; Junt et al., 2008; Park et al., 2012).

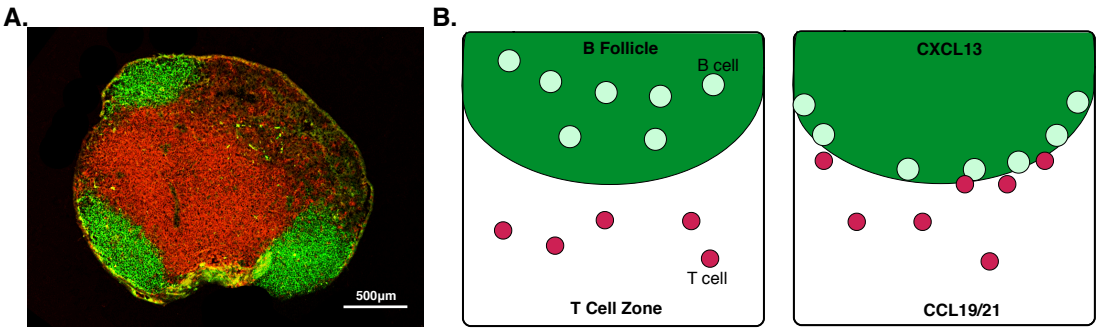

Figure 1.2 Structure of the B-cell microenvironment and the current model of B-cell activation: (a) Lymph nodes have a highly compartmentalised architecture with distinct B-follicles (green) and T-zones (red), scale bar 500µm. (b) Before becoming activated a B cell must interact with a specified pathogen. This is achieved through CXCL13 mediated scanning of the B-follicle. To undergo full activation, an antigen primed B-cell must subsequently respond to CCL19 and CCL21 to migrate to the T-zone to acquire an additional signal from antigen specific T-helper cells. Image (a) provided by Dr. A. Thuery

1.3 Regulation of B-Cell Activation by Lymph Node Architecture

Lymphoid tissues are responsible for the orchestration of functional immune responses. This is achieved through the development and maintenance of niches that support the retention, activation and proliferation of adaptive immune cells in response to antigenic stimulation. Lymph nodes (LNs) are strategically positioned collecting stations for Ags present in peripheral tissues of higher animals. A single inguinal murine LN recruits ~2% of the recirculating lymphocyte pool per day and displays a circadian rhythm; homing is highest at night, with more cells leaving the tissue during the day (Druzd et al., 2017).

During inflammation, the rates of lymph flow and lymphocyte accumulation are markedly increased, promoting antigen recognition and lymphocyte activation (Kumar et al., 2010). The dynamics of lymphatic flow patterns in lymph nodes are difficult to assess experimentally but have been quantified using a computational flow model (Jafarnejad et al., 2015a). This model predicts that about 90% of lymph takes a peripheral path via the subcapsular and medullary sinuses and that structural changes in the LN microenvironment, as well as changes in inflow/outflow conditions dramatically alter the distribution of lymph.

Naive lymphocytes search for cognate antigen during frequent visits to these local 'antigen libraries' entering via High Endothelial Venules (HEVs) in a chemokine and adhesion molecule regulated manner (Coelho et al., 2013). Upon entering the lymph node lymphocytes migrate to their respective niches (**Figure 1.3**) within the parenchyma to scan for antigen (Coelho et al., 2013; Gunn et al., 1998a; Pereira et al., 2010). The key niches within the lymph node are: *(i) the subcapsular sinus (SCS); (ii) the B-follicle; (iii) the T-zone and (iv) the medulla (Chang and Turley, 2015; Junt et al., 2008)*. The subcapsular sinuses are cavities that drain the afferent lymphatics and are found directly beneath the external fibrous capsule of the LN. Macrophages guard the entry of large antigen while antigen less than 70kDa can filter into the parenchyma through a conduit network of microchannels (Rantakari et al., 2015). The B cell follicles are structures responsible for regulating humoral immune responses. The follicles surround a central T-cell zone, which in turn leads to the medulla. The medulla contains large blood vessels, sinuses and medullary cords that contain antibody-secreting plasma cells and macrophages.

Non-hematopoietic stromal cells, which form the connective tissue required to support the function of lymphocytes, are key for creating and maintaining these niches (Junt et al., 2008). This support is provided by direct cell contact (supported by adhesion molecules such as integrins) or through soluble factors such as cytokines and chemokines. In both humans and mice, aged LNs are associated with a decreased number and size of GCs, as well as altered stromal cell numbers and chemokine expression patterns (Turner and Mabbott, 2017).

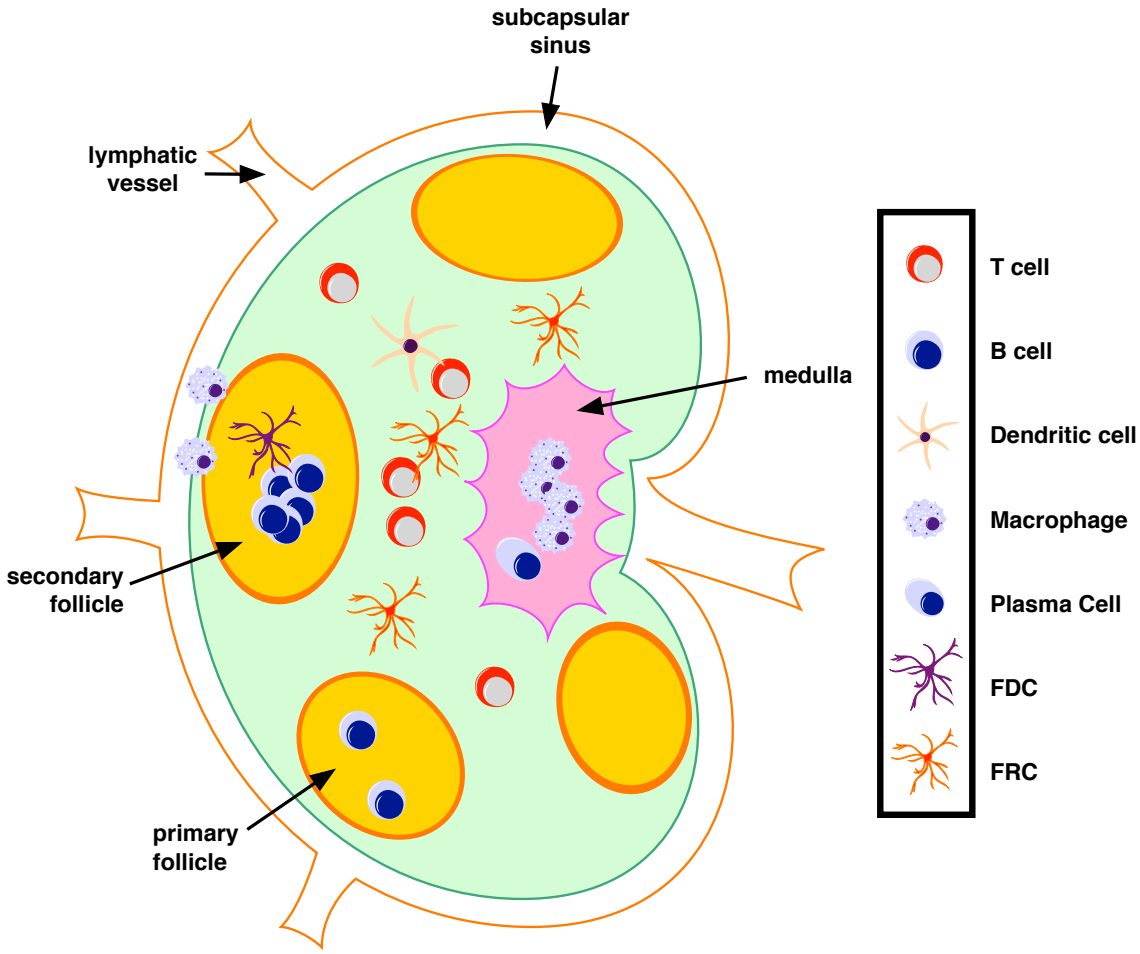

Figure 1.3: High level overview of lymph node structure and key cell types. Lymph flows into the lymph node via the afferent lymphatic vessels and drains in to the subscapular sinus. Within the lymphatic sinus are macrophages that guard the lymph node parenchyma from pathogens. Directly underneath the SCS are B-follicles. Primary follicles contain large numbers of B cells and an immature FDC network while secondary follicles contain germinal centers and mature FDC networks. T cells, dendritic cells and FRCs populate the T-cell zone, while the medulla contains large numbers of macrophages and plasma cells. In the context of the humoral immune response B cells enter via HEVs and migrate towards the follicles. Once in the follicles they scan for cognate antigen. If they fail to acquire antigen within 16-24 hours they leave the lymph node in an S1P dependent manner. If B cells receive the relevant signals they may go on to seed a germinal center reaction and undergo affinity maturation.

1.4 Regulation of Lymph Node Architecture by Stromal Cells

Lymphoid tissue architecture is organised by an interconnected network of stromal cells. Stromal-lymphocyte interactions are a core feature of lymphoid tissue formation and maintenance, and occur irrespective of the tissue type or anatomical location (Butler et al., 2016b). LNs start to develop at embryonic days 11 and 16 where pre-natal type 3 innate lymphoid cells known as lymphoid tissue inducer cells (LTis) are attracted to the chemokine CXCL13 produced by fibroblastic Lymphoid Tissue organizer cells (LTos)(Finke et al., 2002). LTα1β2 signals from LTi cells promotes LTβR–expressing LTos to up-regulate the chemokines CXCL13, CCL19 and CCL21, the cytokine Interleukin-7 (IL-7), as well as the adhesion molecules VCAM-1, ICAM-1 and MAdCAM-1, leading to an increased recruitment of cells and promoting the expansion and differentiation of the stromal network (Fütterer et al., 1998; van de Pavert and Mebius, 2010; Schmutz et al., 2009). Marginal Reticular Cells (MRCs) are among the first subset of stromal cells to appear, differentiating in response to lymphotoxin that is produced by LTi cells and subsequently localizing just below the lymph node SCS. Here, they produce CXCL13, facilitating LN colonisation by lymphocytes that drive the functional maturation of stromal cells leading to a rapid increase in LN size and to the formation of distinct T and B cell areas (Coles et al., 2006; Roozendaal and Mebius, 2011).

In the adult LN, distinct stromal subsets create and maintain the B-cell niche through the provision of key signals that drive lymphocyte migration, differentiation and activation (Gross et al., 2000; Gunn et al., 1998a; Moore et al., 1999; Rahman et al., 2003; Schneider et al., 1999; Wang et al., 2011a). Increased investigation of these cells has shown that they are heterogeneous with unique transcriptional profiles (**Table 1**) (Malhotra et al., 2012). This cross-talk is also a key determinant of B-cell homeostasis and ensures that each cell receives the appropriate signal at the appropriate time (Batista and Harwood, 2009; Pereira et al., 2010). A key mechanism by which this cross talk occurs is through soluble factors such as cytokines and chemokines.

Subset	Location	Function	Secretes	Markers
Fibroblastic Reticular Cells (FRC)	T-zone	Formation of conduit network Chemokine production APC adhesion Antigen presentation	CCL21 CCL19	CD45 - CD31 - Gp38 + ER-TR7+
Follicular Dendritic Cells (FDC)	Follicle	Antigen capture and presentation Chemokine production	CXCL13	CD45 - CD31 - Gp38 - CD21/35 + ER-TR7 -
Marginal Reticular Cells (MRC)	Subscapular zone	Primes innate cells Chemokine production	CXCL13	CD45 - CD31 - Gp38 + MadCam + ER-TR7 + RankL+
CXCL12^{+} Reticular Cells (CRC)	GC dark zone	Dark zone organisation	CXCL12	CD45 - Gp38+ CD21/35- ER-TR7-
Lymphatic Endothelial Cells (LEC)	Lymphatics	Transport of lymph, antigens and cells Chemokine production Formation of chemotactic gradients	S1P	CD45 - CD31 + Gp38 + ER-TR7+
Blood Endothelial Cells (BEC)	Cortex	Transport of blood Lymphocyte trafficking	CCL21	CD45 - CD31 + Gp38 - ER-TR7+
Pericytes	Vasculature	Regulate blood flow	CCL21 CCL19	CD45 – GP38- CD31- Integrin α7+

Table 1.1 Lymph node stromal cell subsets. The marker is CD45 is expressed by haematopoietic cells, and is a member of the protein tyrosine phosphatase family. It is an essential regulator of B and T cell receptor signalling. CD31 is an endothelial marker but is also found on the surface of platelets, monocytes, neutrophils and some types of T cells. It acts as an adhesion molecule but has been shown to have diverse roles in vascular biology(Woodfin et al., 2007). GP38 is a mucin type protein that contributes to the contractility of the lymph tissue reticulum (Astarita et al., 2015). Madcam is an adhesion molecule that contributes to lymphocyte entry into tissues. CD21/35 are complement receptors that process complement opsonised immune complexes. Rank ligand is a member of the tumour necrosis factor cytokine family, mice that overproduce RANKL in hair follicles display massive postnatal growth of skin-draining lymph nodes (Hess et al., 2012). Alpha-7 integrin mediates cell membrane interactions with the extracellular matrix.

1.5 Key Stromal Cell Subsets Within the Lymph Node

Within the developed follicle exists a dense network of Follicular Dendritic Cells (FDCs) whereby cell bodies are connected by an intricate mesh of cellular protrusions. Originally referred to as "antigen-retaining reticular cells" these cells are now understood to support B-cell homeostasis, activation and affinity maturation through the provision of survival, activation and migratory factors. Despite the name, these cells are of mesenchymal origin and are thought to arise from the clonal expansion and differentiation of MRCs, at least during an immune response (Jarjour et al., 2014) . Given their integral role in antibody responses it is unsurprising that this cell type has been implicated in many autoimmune disorders, contributing to the formation of ectopic germinal centres with sustained auto-antibody formation (Victoratos and Kollias, 2009).

In the primary follicle, FDCs are immature and exist in relatively low abundance but following immunisation or infection undergo maturation, upregulate CXCL13, complement and adhesion factors and accumulate in germinal centres (Aguzzi et al., 2014). During selective ablation of FDCs, follicles remodel into disorganised bands of B cells that retain CXCL13-expressing stromal cell populations, indicating that FDCs are not the only stromal-derived source of this chemokine (Wang et al., 2011b). Consequently, the role of FDCs in regulating migration in the primary follicle has proven enigmatic and controversial. The maturation of follicular stroma is dependent on B-cell derived TNFα and lymphotoxin signals that activate the NKkB pathway. Mice deficient in either of these signals fail to form both GCs and FDC networks (Wang et al., 2001). Components of the innate immune response are also implicated in FDC maturation. FDCs express toll-like receptors, with experiments in mice that lack TLR4 expression on stromal cells showing lower levels of SHM and high-affinity antibody production. Strikingly, immunization of mice with nanoparticles containing TLR4 and TLR7 ligands induced GCs that persisted for more than 1.5 years (Garin et al., 2010; Kasturi et al., 2011).

In the lymph node, the delivery of lymph-borne antigens to FDCs is determined by the size of antigen and whether it is opsonized with complement. C3d-coated immune complexes (ICs) larger than 70kDa (~5.5nm) are trafficked from $CR3^+$ SCS macrophages to $CR2^+$ B cells and finally to the FDC network while smaller antigen filter directly

through the conduit system (Batista and Harwood, 2009; Carrasco and Batista, 2007; Phan et al., 2007; Rantakari et al., 2015; Roozendaal et al., 2009). Mature FDCs express the Fc receptors CD16, CD23 and CD32, the complement receptors CR1 and CR2 (FDCs express a longer isoform of the molecule than B cells) and complement components, molecules that assist in IC capture and presentation (Aguzzi et al., 2014). ICs are expressed periodically on the cell surface; undergoing internalisation within a non-degradative cycling compartment to promote antigen presentation over extended periods (Heesters et al., 2013; Mandel et al., 1981). On the cell surface, ICs are periodically expressed along 20-50nm intervals, with live-cell imaging of the IC transfer process *in vitro* identifying a rapid dispersal upon contact with the naive B cell (Heesters et al., 2013; Mandel et al., 1981; Phan et al., 2009) However, it has not been assessed as to whether this spatial distribution of antigen aligns with the clusters of B-cell receptors on the surface on B cells to maximise antigen capture.

Marginal reticular cells (MRCs) are located immediately below the LN SCS (Katakai, 2012). Their expression patterns are distinct to those of FRCs and FDCs but are similar to lymphoid tissue organiser cells and during an immune response MRCs have the potential to differentiate into FDCs (Jarjour et al., 2014; Katakai, 2012). They secrete CXCL13 and express RANK-L and MAdCAM-1 and gp38 but not CR1/CR2 (Katakai, 2012). Given their location they may support natural killer cells and SCS macrophages but there is a dearth of direct experimental evidence available to confirm this.

Within the follicle are fibroblastic cells that are morphologically distinct from FDCs and MRCs and are responsible for the secretion of CXCL12 within the dark zone of the germinal centre (Rodda et al., 2015). They do not express the MRC markers RANK-L or MAdCAM-1 and have a distinct chemokine expression profile and are lineage marked by CD21 and Ccl19 fluorescent reporter mice (Rodda et al., 2015). CXCL12 reporter mice also identify reticular cells at the outer border of the primary (but not necessarily naive) follicle but it is not clear if these cells are distinct from dark zone associated $CXCL12^{+}$ stroma. Due to their anatomical location at the outer follicle, these cells may also contribute to EBI2 mediated migration as this is a key site of oxysterol metabolism (Cyster et al., 2014; Liu et al., 2011).

Fibroblastic Reticular Cells (FRCs) are specialised myofibroblasts of mesenchymal origin, characterised by their expression of podoplanin (PDPN) and platelet-derived growth factor receptor-α (PDGFRA), and their lack of CD45 and CD31 (Fletcher et al., 2015). They regulate lymphocyte homeostasis and activation through secretion of the survival factor IL-7, and the migratory factors CCL19/21. FRCs also have a key regulatory role, controlling the expansion of T cells through the regulated release of nitric oxide, and promoting peripheral tolerance through presentation of self-Ag (Fletcher et al., 2011a; Siegert et al., 2011).

FRCs are a key determinant of lymphoid tissue architecture, forming a complex 3D network with "small-world" properties, a topological feature that promotes network robustness (Novkovic et al., 2016). FRCs are important for maintaining follicular architecture while follicular reticular cells promote B-cell homeostasis through secretion of BAFF (Cremasco et al., 2014). In addition they stimulate the growth of High Endothelial Venules (HEVs) (through secretion of Vascular Endothelial Growth Factor (VEGF)), and generate and ensheath a network of collagen fibres known as the conduit system (Fletcher et al., 2015; Roozendaal et al., 2008). It is also important to note that the architecture of the FRC is highly dynamic; following immunisation and expansion, the FRC network expands rapidly in a process induced by the trapping of naïve lymphocytes or DCs (Yang et al., 2014). Under homeostatic conditions, the mucin type protein podoplanin is a key regulator of reticulum contractility; mature DCs expressing CLEC-2 attenuate this contractility relaxing the FRC network to enhance adaptive immune responses (Acton et al., 2012; Astarita et al., 2015).

Lymphatic endothelial cells line lymphatic vessels. They express adhesion molecules and chemokines, allowing the entry of hematopoietic cells into the lymphatic vessels. On the ceiling of the subcapsular sinus LECs express ACKR4, an atypical chemokine receptor that scavenges CCL21 facilitating the formation of functional gradients (Ulvmar et al., 2014). The reason that LECs on the SCS ceiling express ACKR4 has yet to be established but computational analysis of lymph flow suggests that there is increased shear stress on the ceiling of the SCS in comparison to the floor (Jafarnejad et al., 2015). These cells have also been reported to express peripheral tissue Ag, highlighting a role in maintaining peripheral tolerance (Cohen et al., 2010). Blood endothelial cells are specialized vascular endothelial cells that construct cortical blood vessels and capillaries, including high

endothelial venules (HEVs) specialized to attract naïve lymphocytes from the bloodstream. This process is further regulated by pericytes that wrap around blood vessels controlling the rate of vasoconstriction and vasodilation.

1.6 The Molecular Basis of B-Cell Migration

Cell migration in complex microenvironments comprises an array of interdependent biophysical and biochemical processes. The mode of migration is context-dependent; B-cells use strong forces at the rear to generate forward momentum and display flexible morphological adaptation to their environment rather than structurally changing matrix architecture as seen in the migration of stromal cells (Tschumperlin, 2013).

One of the main mechanisms by which stromal cells regulate B-cell homeostasis is through the secretion of signalling lipids and chemotactic cytokines known as chemokines (Pereira et al., 2010). Chemokines are small proteins (~10kDa), named after the number and spacing of conserved cysteine residues and transmit their signals through G-protein Coupled Receptors (GPCRs). Ligation of the chemokine to its respective GPCR leads to modification of the tertiary structure of the receptor, allowing the intracellular component of the receptor to bind and activate heterotrimeric G-proteins (Bennett et al., 2011). This facilitates downstream signalling cascades leading to the activation of four small GTPases: Rac, required for the actin polymerization at the pseudopod; Cdc42, required for the establishment of orientation machinery at the pseudopod and thus steering of a cell during chemotaxis; Rho, required for the retraction of the uropod; and Rap, required for integrin signalling (Rot and von Andrian, 2004). Activation of these molecules leads to adhesion via integrins or polarization of the actomyosin cytoskeleton causing migration along chemokine gradients (Rot et al., 2013). These factors also dynamically interact such that cdc42 inhibits Rho at the leading edge and Rho inhibits Rac at the uropod; thus maintaining polarity through the asymmetric distribution of signalling components (Maiuri et al., 2015).

The GPCRs that regulate migration are subject to a complex and dynamic layer of regulation (Bennett et al., 2011; Sh et al., 1982; Zigmond, 1981). While difficult to assess experimentally, mathematical modelling suggests that the dynamic modulation of

signalling is essential for effective migration within complex environments such as the germinal centre (Chan et al., 2013; Figge et al., 2008). Mechanisms of chemokine receptor silencing include: (i) the intrinsic GTPase activity of Gα subunits that acts as a built-in shut-off sequence; (ii) *desensitization*, caused by steric hindrance of G-protein activation due to receptor phosphorylation by G-protein-coupled receptor kinases (GRK); and (iii) *downregulation* caused by receptor internalization through clathrin-coated pits or the faster lipid raft caveolae pathway; receptors may then be degraded or recycled back to the cell surface (Bennett et al., 2011). Desensitization and internalization are regulated separately and require the phosphorylation of different residues in the C-terminal tail of chemokine receptors (Bennett et al., 2011). More recently atypical chemokine receptors have been described; such receptors do not induce the responses classically associated with GPCRs. However, they can impact chemokine availability and function in a wide array of molecular and cellular processes (Nibbs and Graham, 2013; Ulvmar et al., 2011, 2014).

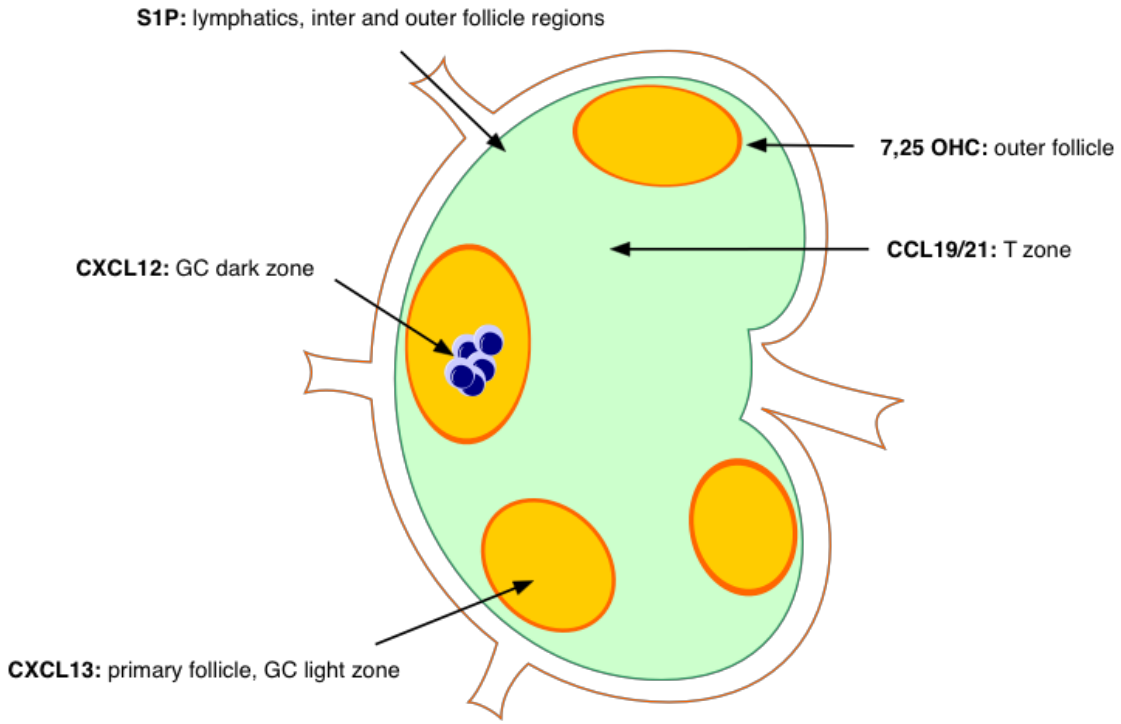

Figure 1.4 Chemotactic factors that regulate B cell migration and their sites of ligand production in lymph nodes. CXCL13 is found within lymph node follicles and germinal centre light zones and regulates lymphoid tissue structure and the induction and maturation of antibody-mediated responses. CCL19/21 is found in the T-cell zone and is implicated in B cells obtaining access to help from antigen specific CD4 T helper cells. 7α,25-Hydroxycholesterol is associated with scanning of the outer follicle regions, CXCL12 with the dark zone of the germinal centre while S1P is implicated in lymph node egress.

1.6.1 CXCL13

CXCL13 is a small globular protein with a molecular mass of 10.3 kDa that binds CXCR5, a GPCR (Förster et al., 1996a; Gunn et al., 1998a; Legler et al., 1998). CXCL13, like most chemokines, has a carbohydrate binding domain but it has yet to be determined if it acts principally in a surface bound way or as a soluble factor (Monneau et al., 2016). Expression of CXCL13 is regulated in a positive feedback loop involving CXCR5-mediated induction of $LT\alpha_1\beta_2$ expression by B cells which in turn contributes to maximal CXCL13 production through FDC maturation (Ansel et al., 2000). $CXCL13^{-/-}$ mice display poor follicular organization with B-cells observed in separate rings around T-zones but do not form polarized clusters but do have distinguishable Light Zones and Dark Zones. The GC FDC network still forms and expresses markers associated with maturation but it is not distal to the T zone as in a normal GC (Ansel et al., 2000). $CXCR5^{-/-}$ B cells injected into a WT host are known to have different migration characteristics within LN follicles in comparison to WT B cells (Coelho et al., 2013). However, CXCR5-mediated B-cell migration appears quite robust to the amount of chemokine ligand available (Coelho et al., 2013); this is consistent with studies that suggest a difference of 10 signalling receptors across a cell is sufficient to induce chemotaxis along a gradient (Herzmark et al., 2007; Zigmond, 1981). Western blotting of pooled lymph nodes suggests a concentration of 10-50nM (derivation presented in **Appendix 5**) (Luther et al., 2002) within the B follicle, a measure supported by *in vitro* analyses of B-cell migration in response to varying CXCL13 concentrations and by the dissociation constant for CXCR5-CXCL13 and other similar chemokine receptor and cognate ligand pairs (Barroso et al., 2012; Gunn et al., 1998a).

CXCR5 may also affect B-cell homeostasis and activation through indirect mechanisms. CXCR5 signalling has been shown shape B-cell receptor (BCR) triggered B-cell activation. CXCR5 regulated upregulation of LFA-1, an adhesion factor that binds ICAM-1, has been reported to lower the threshold for B cell activation through stable immune synapse formation (Carrasco et al., 2004; Sáez de Guinoa et al., 2011). It is important to note that naïve B-cells not only express CXCR5 but also CCR7, CXCR4 and EBI2 (Pereira et al., 2010). However, there are few studies looking at cross-talk between these receptors. It has been reported that EBI2 can dimerise with CXCR5 and lower its

affinity for ligand on B-cells but the functional consequences of this molecular interaction have yet to be investigated (Barroso et al., 2012).

1.6.2 CXCL12

C-X-C motif ligand 12 (CXCL12/SDF-1) is a chemokine and exists in 6 isoforms α, β, γ, δ, ε, and φ that differ in their affinities for CXCR4, CXCR7, ACKR3 and components of the extracellular matrix (Yu et al., 2006). The exact function of CXCL12 is difficult to discern given CXCL12 deficiency is embryonically lethal and it has different roles in different anatomical niches. In the context of the GC CXCL12 is essential for positioning $CXCR4^{+}$ B-cells in the dark zone. The cellular source of CXCL12 within lymph node follicles has been attributed to a sparse population of CXCL12 secreting $CD45^{-}$ $Gp38^{+}$ $CD21/35^{-}$ $ER\text{-}TR7^{-}$ fibroblasts (Rodda et al., 2015).

1.6.3 CCL19/CCL21

A third chemokine receptor system that influences B cell migration in the follicle is CCR7 and its ligands CCL21 and CCL19 (Gunn et al., 1998b; Yoshida et al., 1998). In resting LNs both CCL19 and CCL21 are secreted by T-Zone FRCs with bioavailability of CCL21 negatively regulated by LECs lining the ceiling of the subscapular sinus (Ulvmar et al., 2014). CCL21 differs to CCL19 by a highly charged 40 amino acid extension at its C terminus thus altering its ability to bind ECM components (Hirose et al., 2002; Patel et al., 2001; de Paz et al., 2007). *In vivo*, C-terminal truncation of CCL21 prevents its immobilization to high endothelial venules and consequently affects lymphocyte homing to the lymph node (Stein et al., 2000). This result is complemented by analyses of whole lymph node mRNA and extracellular protein on HEVs that reveal a daily oscillatory pattern of CCL21 expression that correlated with the kinetics of lymphocyte homing (Druzd et al., 2017). Mice, unlike humans, have two isoforms of CCL21 that differ in one amino acid; CCL21Ser is expressed by HEVs, whereas CCL21Leu is generated in the lymphatic endothelium (Rot and von Andrian, 2004). A mutant mouse strain known as *plt*/*plt* (paucity in lymph-node T cells) is deficient for CCL21Ser and CCL19. LNs of *plt*/*plt* mice and CCR7-deficient mice contain few naive T cells, but the B-cell compartment is less affected and there are marked numbers of memory T cells (Förster et al., 1999; Mori et al., 2001).

1.6.4 S1P

The lipid signalling molecule sphingosine 1-phosphate (S1P) is another key molecule governing the migration of B cells. S1P signals via GPCRs with 5 distinct receptors binding the ligand. In the context of B-cell migration receptors S1PR1-3 have begun to be characterised. S1PR1 is important for lymphocyte egress via the cortical lymph node sinuses as determined by studies using KO mice and pharmacological inhibition (Park et al., 2012). The lymphocyte activation marker CD69 inhibits S1PR1-mediated egress during an immune response; CD69 binds S1PR1 and the complex is internalised from the cell surface (Cyster, 2005). Sphingosine-1-phosphate receptor 2 (S1PR2) binds S1P but acts as an inhibitor of migration, promoting recruitment and retention of GC B cells and T-Follicular helper cells to the centre of the follicle through signalling via the G-protein Gα13. Loss of Gα13 mediated signalling has been shown to cause a lack of confinement of GC B cells, allowing egress into the circulation; this pathway is commonly disrupted in GC B cell derived lymphomas (Muppidi et al., 2014). In the absence of Gα13 mediated inhibition of migration, S1PR3 was shown to dominate the S1P response and allow cells that have entered the follicle to be guided into subcapsular or cortical sinuses for LN egress (Muppidi et al., 2015).

1.6.5 7α,25 Hydroxycholesterol

Epstein-Barr virus-induced G-protein coupled receptor 2 (EBI2, GPR183) is a GPCR that binds oxysterols (Liu et al., 2011). EBI2 guides B cell movement along the B-T boundary as well as the interfollicular and outer follicular regions with expression increasing during B cell activation but decreasing during germinal cell development due to the transcription factor BCL6 (Gatto and Brink, 2013; Gatto et al., 2009; Kelly et al., 2011; Liu et al., 2011; Pereira et al., 2009). $EBI2^{-/-}$ B cells prematurely accumulate at the centre of the follicle and associated early antibody responses are diminished (Pereira et al., 2009). The oxysterol 7α,25-dihydroxycholesterol (7α,25 OHC) is a potent binder to this receptor and is synthesized from cholesterol by the two enzymes (CH25H, CYP7B1) and is metabolized by another enzyme (HSD3B7) (Yi et al., 2012). The cellular sources of these enzymes are unknown but transcriptomic analyses[1] suggest that LECs have high expression of HSD3B7 indicating that this cell type might also act as a sink for oxysterols

[1] Data available on the Immgen database

as well as for CCL21. Collectively, these enzymes control EBI2 ligand concentration in lymphoid tissues (Cyster et al., 2014).

Gene Perturbation	Phenotype	Reference
CXCR5	Impaired homing to follicles	(Ansel et al., 2000; Förster et al., 1996b)
CXCR4	Impaired GC formation; lack of light zone and dark zone segregation	(Allen et al., 2004)
CCR7	Aberrant follicular organisation	(Förster et al., 1999)
CXCL12*	Impaired GC formation; lack of light zone and dark zone segregation	(Barinov et al., 2017)
CXCL13	Aberrant follicular organisation; B cells form a ring around the T-cell zone	(Ansel et al., 2000)
S1PR2	Increased numbers of GC B cells, perturbed LN architecture	(Green et al., 2011)
EBI2	Abberant positioning of activated B cells; perturbed IgG1 antibody responses	(Pereira et al., 2009)

Table 1.2. Summary of migratory factor KO experiments in mice and their effect on humoral immune responses. * CXCL12 deficiency is embryonic lethal so we refer to a study where the authors investigate the effects of a mutated form of the protein that cant bind extracellular matrix components.

1.7 Emerging Technologies to Study the Immune System

The molecular and cellular processes governing B-cell migration and activation described in the previous sections have been studied using a combination of gene knockouts, fluorescent reporter mice, gene expression analysis, immunohistochemistry and flow cytometry. These experimental studies have unearthed a complex and dynamic regulatory network acting over broad spatial and temporal scales. Consequently, the mechanisms by which chemotactic gradients form in lymphoid tissues and how cells respond to these gradients have proven difficult to assess. Recent technical advances have led to the development of new technologies with the capacity to further our understanding of this complex biological pathway. In the following sections we discuss emerging technologies from two key disciplines relevant to the work undertaken in this thesis: microscopy and systems biology.

1.8 Imaging Approaches

The cell types and molecules described in the previous section have been visualised using a suite of imaging techniques, with increased biological insight obtained through the integration of mathematical (image analysis), physical (multi-photon and super-resolution imaging), chemical (bio-compatible probes) and biological (fluorescent reporter animals) approaches (Butler et al., 2016a) (**Table 3.1**). In the following section we discuss established and emerging imaging technologies. We illustrate their utility with select references and discuss how they can be used to quantify the anatomy and dynamics of the lymph node follicle.

Confocal[2] and more recently, light sheet microscopy[3] have been used to study the spatial organisation of the lymph node, highlighting the importance of the immune microenvironment in regulating the induction and resolution of immune responses (Junt et al., 2008; Qi et al., 2014). Multi-photon confocal microscopy has been used extensively to image cell dynamics within the germinal centre and to highlight the role of chemokines in naïve B-cell trafficking (Allen et al., 2007; Coelho et al., 2013; Schwickert et al., 2007;

[2] In this approach a pinhole is placed at the confocal plane of the lens to eliminate out-of-focus light and increase the spatial resolution of the image

[3] The specimen is illuminated only in a single plane reducing phototoxic effects. By moving the sample through the light sheet, 3D images of a specimen can be recorded.

Shulman et al., 2013). This technique employs long-wavelength infrared lasers that allow observation deep into tissues, while their small excitation volume decreases photobleaching and phototoxicity. In addition, this technique can be used in the generation of second harmonics, which highlight the collagen fibres and the extracellular matrix that form the conduit network, the capsule, and structural support in lymph nodes.

Relative to tissue and cellular levels of organisation, molecular processes occur on much finer spatiotemporal scales (on the order of nanometres and milliseconds) and often require a distinct set of imaging modalities. Transmission electron microscopy and scanning electron microscopy use beams of electrons to create an image; because the wavelength of an electron is much shorter than light it provides increased resolution. These techniques have been used to image the morphological changes induced by antigen recognition and to show that antigen is dotted evenly on FDCs (El Shikh et al., 2010). Unlike electron microscopy, the spatial resolution of conventional fluorescence microscopy is limited by the diffraction of light at ~200nm. In this imaging process, light rays from each point on the object converge to a single point at the image plane. However, the diffraction of light prevents exact convergence of the rays, causing a sharp point on the object to blur. The three-dimensional (3D) intensity distribution of the image of a point object is called the point-spread function (PSF). The size of the PSF determines the resolution of the microscope: Two points closer than the full width at half-maximum (FWHM) of the PSF will be difficult to resolve because their images overlap substantially. This has hindered the visualization of molecular level processes where proteins may be just a few nanometres in diameter. In recent years, a number of "super-resolution" fluorescence microscopy techniques have been invented to circumvent the diffraction barrier, including techniques that can selectively deactivate fluorophores to minimise the area of illumination, such as Stimulated Emission Depletion (STED) microscopy (Hell and Wichmann, 1994; Schneider et al., 2015)), as well as techniques that are based on the localization of individual fluorescent molecules, such as stochastic optical reconstruction microscopy (STORM) (Rust et al., 2006). These methods have yielded an order of magnitude improvement in spatial resolution over conventional light microscopy facilitating measurements of single molecules. This increased resolution has yielded significant insights into lymphocyte biology showing that GC B cells recognize antigen through a specialized immune synapse (Nowosad et al., 2016). Fluorescence Energy Resonance Transfer (FRET) has been used to look at ligand-induced

conformational changes in antigen receptors and early signalling events in lymphocytes. Using this technique Tolar *et al* (2005) showed that the BCR exists as a monomer in resting cells but that antigen binding leads to BCR clustering and conformational changes to facilitate downstream signalling cascades.

Technique	Application	Exemplar Study
	Whole body	
MRI	Confirming accurate vaccine delivery	(de Vries et al., 2005)
PET	Imaging the infiltration of immune cells into tumours	(Rashidian et al., 2015)
	Tissue	
Confocal	Mapping of the FRC and conduit networks	(Novkovic et al., 2016)
Light Sheet	Mapping the volume of the pLN follicle	(Irla et al., 2013)
	Cellular	
Confocal	Mapping the expression of CCRL1 on LECs	(Ulvmar et al., 2014)
Multiphoton	B-cell interactions with T cells within the GC	(Shulman et al., 2014)
	Molecular	
FRET	Studying conformational changes of the B-cell receptor	(Tolar et al., 2005)
TEM	Mapping of antigen on FDCs	(El Shikh et al., 2010)

Table 1.3 Exemplar applications of imaging modalities

1.9 Systems Biology

Throughout the thesis we will refer to grouped biological entities as a complex system. In the context of this work, a complex system is defined as having high-level behaviours that cannot be deduced as a simple linear combination of lower level behaviours (P.Andrews et al., 2010). Systems-based approaches afford a quantitative perspective from which to consolidate large multivariate datasets, inform experimental design, and to assess the veracity of competing theories. In the following section, we provide a broad overview of systems biology approaches, comparing and contrasting key methodologies through published case studies.

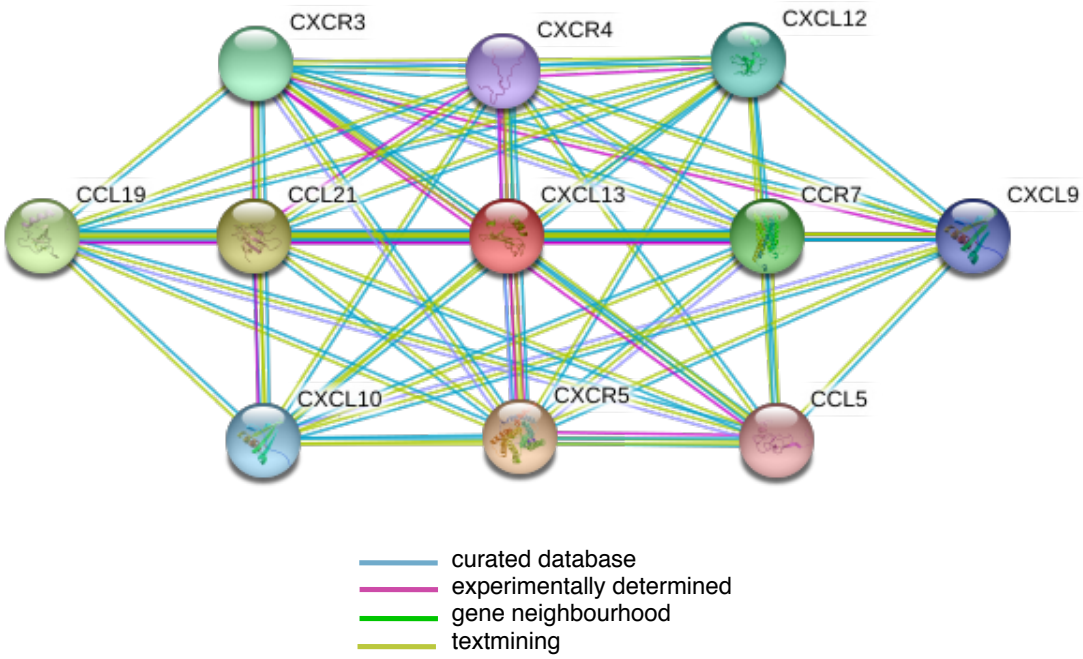

Figure 1.5 Interaction network for B-cell associated migratory factors and evidence available to support their interactions. This network was generated using the String database with protein-protein interactions and the type of evidence supporting the interaction highlighted (Szklarczyk et al., 2011). The String database assigns a confidence score from 0 to 1 based on the nature of the evidence. All interactions in this network have a confidence score greater than 0.9

Systems biology is an emerging discipline whose focus is to study the design, regulation, structure and dynamics of grouped biological entities (Kitano, 2002). Mathematical, statistical and computational analysis and modelling techniques are commonly employed to study these four key components. Broadly speaking, systems biology approaches can be split into two categories: top-down approaches which aim to discover patterns in complex multivariate datasets, and bottom up approaches which consolidate mechanistic information to identify knowledge gaps and predict system behaviours. Top down approaches are particularly suited to extracting information from datasets generated by modern high throughput genomic techniques such as DNA microarrays and RNA-Seq. Complex bioinformatics pipelines have been applied successfully to stratify genes into functional modules (Segal et al., 2003), to reconstruct interaction networks (Costanzo et al., 2016; Szklarczyk et al., 2011) (**Figure 1.5**) and to study the evolution of microorganisms (Brosch et al., 2001).

However, top-down approaches are susceptible to identifying correlations without understanding the underlying mechanistic processes. As such, it can be difficult to rigorously assess how a system is configured or predicting how it will respond to perturbations. Bottom-up modelling techniques are particularly suited to these tasks as they allow for biological processes to be represented as a set of mathematical expressions such that they can be quantitatively examined (Kitano, 2002). In this context a model is an abstraction of a complex system, formally expressed using mathematical concepts and language. The process of assembling the model unveils gaps in our knowledge and points out new directions for experimental studies that may not have been apparent *a priori* (Butler et al., 2016b). Simulation in the context of this thesis refers to the process of imitating the behaviour of a system through a suitably analogous piece of software, in order to gain insights into how the system is organized and how it functions. Thus a simulator can be considered as an implementation of a model, allowing for extrapolation of our current understanding as a means to testing the validity of competing hypotheses, examining gaps in our understanding and to obtain insights where experimentation is not possible.

1.9.1 Bottom-Up Modelling Approaches

To date, a number of mathematical modelling techniques have been used to model biological processes. In particular we discuss (i) ordinary differential equations (ODEs) that predict changes to components over one independent variable (e.g. time) on a continuous scale. **(ii)** partial differential equations (PDEs) are a technique that predicts changes to components over more than one independent variable (e.g. time and space). (iii) agent-based models (ABM) a model composed of individual entities, known as agents (**Figure 1.6**). The state of each agent is determined by the agent's attributes and location at a specific point in time with state transitions governed by a predetermined rule-set (iv) multiscale models which capture phenomena occurring across different spatiotemporal scales and (v) hybridised models that integrate different techniques to facilitate parsimony and efficiency (Cosgrove et al., 2015; Guo and Tay, 2008).

Differential equations (DEs) afford many advantages as they can concisely describe large amounts of information, can quickly obtain numerical solutions and there exists a number of established analysis techniques to evaluate these models (Mirsky et al., 2011). ODEs, which abstract interactions occurring in physical space through the use of contact frequency terms, are a commonly employed method to model the immune system, typically describing how grouped biological entities change over time. This approach was used to assess the dynamics of CXCR4 and CXCR5 within the GC (Chan et al., 2013). Analysis of the model shows that chemokine receptor expression is oscillatory and highlighted the importance of receptor down-regulation as a cell-intrinsic mechanism to govern migration patterns. This approach is also commonly employed in studies of viral dynamics and associated immune responses (Althaus and Boer, 2008; Elemans et al., 2011; Perelson and Ribeiro, 2013). Unlike ODEs, PDEs do not assume that space is one continuous compartment; consequently, they have been used to model the spatial distribution of interferons within the lymph node (Bocharov et al., 2011) and to model antigen independent proliferation of $CD8^+$ T cells (Antia et al., 2003).

An agent based modelling approach differs to equation-based systems, as they are composed of individual entities, known as agents; each agent exists in a well-defined state, determined by the agent's attributes and location at a specific point in time (Macal

and North, 2010). A transition into an alternative state is governed by a rule-set designed to capture the agent's interactions with other agents or the environment. Using these rules, agents calculate how to respond to features and stimuli within their local environment. The aggregate effects of these individual decisions lead to the emergence of system-wide patterns and behaviours that are not explicitly programmed or intuitively understood from the rules alone (Macal and North, 2010).

The key advantage of an agent-based approach is the ability to incorporate space, heterogeneity and stochasticity within a single model (Bauer et al., 2009; Cosgrove et al., 2015; Mirsky et al., 2011). Based on constraints obtained from the literature, Meyer-Hermann et al. (2006) exploited the emergent property of ABMs to test the veracity of different hypotheses regarding B-cell selection in the GC, ruling out models that failed to reproduce experimentally observed kinetics. Cellular Potts models (CPMs) are considered a specialised type of ABM that have also been used to model immune cell migration (Meyer-Hermann and Maini, 2005). In this scheme, agents are depicted as a set of contiguous lattice sites that can rearrange to produce cell motion and shape changes on the basis of a Hamiltonian (or cost function) that determines the probability of lattice updates. This can facilitate the incorporation of complex biophysical interactions but may be too computationally intensive to model large numbers of cells without the use of a parallelised implementation (Chen et al., 2007; Mirsky et al., 2011).

ABMs are designed to directly capture attributes and states of individual components; population counts and rates of change emerge during simulation execution rather than being explicitly specified. Accordingly, ABMs can be difficult to describe succinctly, may require more data to develop, and significant software engineering expertise to implement (Guo and Tay, 2008; Macal and North, 2010). Consequently, much research focus is being placed on methodologies to describe, design, and implement ABMs (discussed further in **Chapters 2 and 3**).

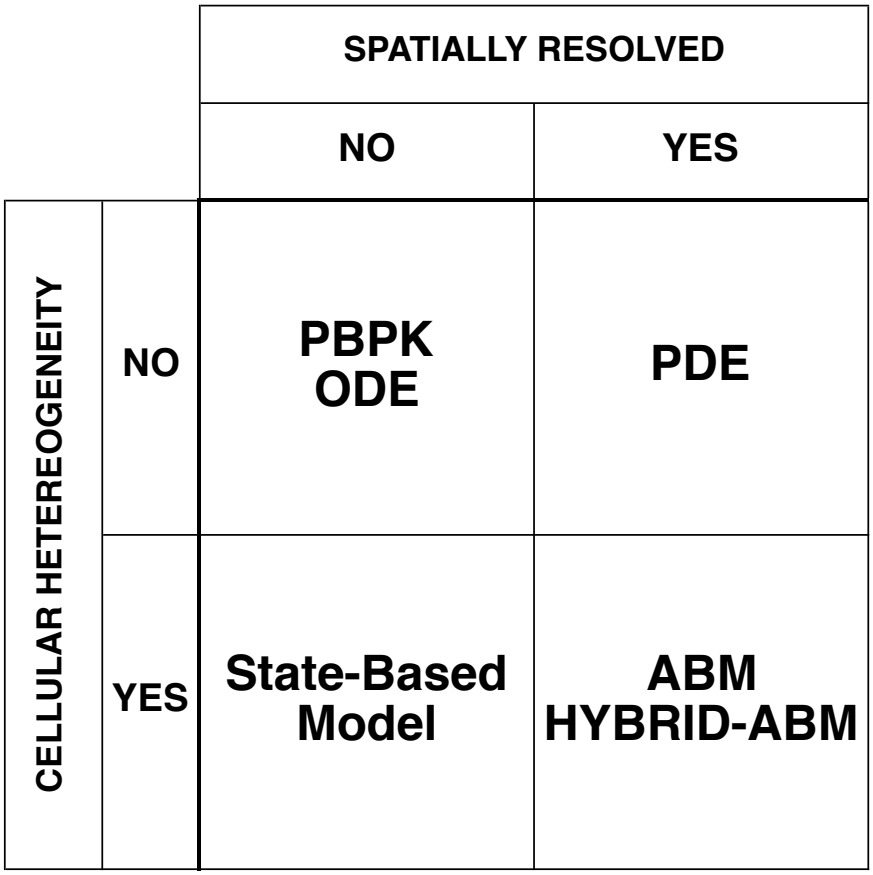

Figure 1.6 The capacity for various types of model to capture spatial resolution and cellular heterogeneity: When determining the appropriate modelling technique to employ it is important to consider the spatiotemporal scales relevant to the system and the heterogeneity of the entities of interest. Ordinary Differential Equations (ODEs) and Physiologically Based Pharmacokinetic (PBPK) models cannot capture systems with explicit spatial resolution (although compartmentalized systems are possible), relying on the abstract notion of well-mixed space. Partial Differential Equations (PDEs), and thereby, coupled systems of ODEs, are capable of spatial resolution, but to capture heterogeneous cellular phenotypes is often intractable. State-based modelling approaches enable heterogeneous phenotypes among cell populations but cannot in themselves capture spatial resolution (although they can model multiple, spatially disconnected compartments). ABMs incorporate state-based systems in spatial environments; as such, ABMs can capture both heterogeneous cell populations with an explicit notion of space and time. Figure taken from Cosgrove et al (2015).

1.9.2 Hybrid and Multiscale Modelling

In the previous sections we have detailed experimental efforts to dissect the complexity of B-cell activation with particular focus on imaging approaches. However, these techniques provide a narrow window of insight into a process that occurs over a timescale of days and weeks making it challenging to interpret results in the context of the wider literature. Modelling biological processes on diverse time and length scales is known as multiscale modelling, an approach well suited to bridging diverse imaging datasets within a single modelling framework. A key example of a multiscale model from the existing literature is an agent based model to examine the effects of a molecular switch, controlled via epidermal growth factor receptor signalling, on tumour spatial dynamics in the brain (Athale and Deisboeck, 2006). The model predicted that this switch could affect tumour expansion, leading to the development of novel hypotheses on the posttranslational regulation of protein expression.

Relative to cells, molecules are smaller, more abundant, have less complex behaviours and are well characterized by physical laws and equation based systems. Fick's laws of diffusion (discussed in **Chapter 4**) and the law of mass action[4] can describe molecular diffusion and interactions respectively; eliminating the need to model molecules individually (Guo et al., 2008b). However, these simplifying assumptions are not applicable to B-cell migration as they do not move like molecules, and display a wide range of roles and nonlinearly interacting behaviours (Mirsky et al., 2011). ABMs are capable of incorporating the stochastic, heterogeneous and spatial considerations of migration within a single system, and are thus better equipped to model cell migration.

The differences in scale, relative population sizes and complexity between molecules and cells thus warrants the integration of distinct but complementary modelling strategies in a process known as hybridisation (Guo and Tay, 2008). Through the hybridization of an ABM with other techniques it is possible to quantitatively examine how processes such as molecular diffusion, occurring over a timescale of milliseconds, can affect cellular motility occurring over minutes and hours (Guo et al., 2008b).

[4] The rate of a chemical reaction is proportional to the concentrations of the reacting substances.

We propose that a hybridised multiscale approach is best suited to describe B-cell migration within complex tissues. Furthermore, a hybrid approach allows us to build upon previous modelling work in the field that describes the processes of immune cell migration (Lin and Butcher, 2008), lymph node architecture (Kislitsyn et al., 2015; Novkovic et al., 2016), and physiological gradient formation (Wang and Irvine, 2013). However, consolidating these models is a challenging goal, given the dearth of quantitative measurements on the spatial organisation of the lymph node and the diffusivity of migratory factors within lymphoid tissues; key parameters required to accurately model immune responses in lymphoid tissues. Closer integration of theoretical and experimental approaches is thus required to consolidate these distinct modelling efforts into a high-fidelity representation of lymph node architecture, where immune cells can interact with stroma capable of creating and shaping complex physiological gradients (Butler et al., 2016a).

1.9.3 Emulation of Complex Systems Models

The drive to better capture biological complexity leads to more sophisticated simulators, it becomes challenging to perform statistical analyses that help translate predictions into increased understanding. These analyses may require repeated executions and extensive sampling of high-dimensional parameter spaces: analyses that may become intractable due to time and resource limitations. Although the composition of stochastic models may not be complex or computationally intensive, it can lead to a diverse set outputs for a fixed parameter input. This introduces aleatory uncertainty around those outputs that requires mitigation with replicate runs (Helton, 2008; Marino et al., 2008). Heterogeneity can be captured through modifying the simulation parameters to represent a range of individual responses from within a population. As it is common to simulate biological systems for which our understanding remains incomplete, values of a range of those simulation parameters may remain unknown or poorly constrained, leading to parametric uncertainty (Gutenkunst et al., 2007). However, for a model with a high-dimensional parameter space, systematic exploration of parameter values requires significant computational infrastructure, a particular problem for time-intensive non-deterministic models. Consequently, a surrogate tool, or emulator, that is capable of rapidly and accurately predicting simulation responses is an attractive option for reducing resource requirements associated with model evaluation (emulation is performed in **Chapter 5**)

(Kennedy and O'Hagan, 2001). In saving resources, emulation can serve as a useful adjunct to the original simulator, providing insights where complex analyses were previously intractable. In the context of systems biology, emulation has primarily been achieved through a Bayesian approach where a statistical model, usually a Gaussian process, is used to estimate simulator outputs. Such emulators have been applied to aid parameter estimation in a stochastic model of mitochondrial DNA population dynamics (Henderson et al., 2009), an epidemic model of influenza (Farah et al., 2014), and two models of hormonal crosstalk in Arabidopsis root development (Vernon et al., 2016).

1.10 Rationale for a Model-Driven Approach to Experimentation

Systematically integrating knowledge into a mechanistically coherent output is an important driver for rational experimental design. The benefits of integrating experimental and theoretical approaches, herein referred to as model-driven experimentation (MDE) are as follows: (i) simulations can capture and consolidate large quantities of data from disparate sources, identifying areas where biological knowledge is lacking, and assessing the veracity of competing theories (ii) simulations are more amenable to designing, conducting and collecting data from experimentation providing data at high spatiotemporal resolution (iii) simulation can help in directing wet-lab experimentation, permitting preliminary investigations that identify key time points and entities of the system under study (Butler et al., 2016b).

To illustrate the MDE approach we discuss two key case studies. Peyers' patches (PP) are specialised secondary lymphoid tissues of the intestine that develop during a fixed window in foetal development and have an essential role in maintaining intestinal immunity. Using several different gene knockouts, the key molecular regulators of PP formation were identified experimentally (Randall et al., 2008). However, a more nuanced view of the system in which the relative contributions of each molecule is understood, was lacking. This was addressed using an MDE approach that yielded the hypothesis upregulation of VCAM-1 on stromal cells was the key triggering event that determined the site of PP formation on the mid-gut while chemokine mediated mechanisms became more important at later timepoints (Alden et al., 2012a). Utilising this prediction, an *in vitro* assay was designed whereby foetal mid-gut explants were incubated in the presence or absence of anti-VCAM-1 antibodies. This assay verified that early upregulation of VCAM-1 was the triggering event that was essential for the initiation of LTi/LTin cell clustering (Patel et al., 2012). The MDE approach has also been demonstrated by Walker *et al*. who showed that an *in vitro* model of calcium-dependent wound closure could be quantitatively reproduced *in silico* using simple rule-based dynamics. Differences between *in silico* and *in vitro* models led to predictions for a role in wound-induced signalling events in urothelial cell cultures, prompting further experimental work (Walker et al., 2004). Experimental studies can thus be complemented with mechanistic models designed to consolidate existing information, identify key knowledge gaps and drive further experimentation.

1.11 Summary, Aims and Objectives

Morphogens are signalling molecules that facilitate cell-to-cell communication without the need for direct contact. Despite their fundamental importance, the mechanisms driving morphogen gradient formation in complex tissues are unclear, limiting understanding of how soluble factors are distributed *in vivo*. In this thesis we focus on CXCL13, a chemokine that regulates B-cell responses. Our key hypothesis is that lymphoid tissues are essential for the formation of CXCL13 gradients, and perturbation of the spatial distribution of CXCL13 can affect the onset of humoral immune responses. This hypothesis yields the following research questions:

1. How is CXCL13 distributed within primary lymph node follicles?
2. What are the key factors that regulate the spatial distribution of CXCL13 within primary lymph node follicles?
3. How do the dynamics of CXCL13-mediated cross talk between stromal cells and B-cells affect the induction of humoral immune responses?

These aims will be addressed with the following objectives:

1. Use visual notations to scope key components of the biological system and quantify these components using an ensemble of imaging and cytometry approaches.
2. Consolidate this data through development of a 3D hybrid multiscale simulator, CXCL13Sim.
3. Use CXCL13Sim to simulate CXCL13 gradient formation and associated B-cell responses.

1.12 Thesis Structure

The thesis addresses these aims and objectives in seven chapters, organised as follows:
Chapter 2 describes the materials and methods used to quantify key components of the biological system and to engineer a multiscale simulator, CXCL13Sim.

Chapter 3 describes a methodology designed to address challenges associated with multiscale model development. It also details the development of a conceptual model of CXCL13 mediated cross talk between follicular stromal cells and B cells identifying key limitations in our current understanding of the pathway.

Chapter 4 describes experimental work to address knowledge gaps identified in chapter 3. Specifically, we quantify the volume and density of the lymph node follicle, the 3-dimensional organisation of $CXCL13^{+}$ follicular stroma, and the CXCL13 diffusion constant. Using these measures in conjunction with data from the existing literature, we derive a quantitative description of CXCL13 within the primary LN follicle.

Chapter 5 details the consolidation of the results obtained in chapters 3 and 4 into a multiscale simulator, CXCL13Sim. Once implemented as an executable piece of software in Java we quantify the influence of stochastic and parametric uncertainty on simulation behaviours to facilitate interpretation of CXCL13Sim-derived results. We also discuss the use of machine learning approaches to enrich the evaluation of multiscale models.

Chapter 6 details the use of simulation analysis to assess the robustness of B-cell migration to perturbations to the spatial distribution of CXCL13 and to the dynamics of CXCR5 expression within the follicle. In addition, we enrich our simulation analyses through the use of multi-objective evolutionary algorithms and verify key predictions using immunohistochemistry.

Chapter 7 provides a critical review of the work that has been conducted in relation to the project aims identified above. Specifically, we highlight contributions to the fields of immunology and systems biology and conclude with a perspective on future trends emerging in systems biomedicine.

Chapter 2 Materials and Methods

2.1 Experimental Materials and Methods

2.1.1 Mice

6-8 week old wild type mice (C57BL/6) were housed in BSF at the University of York. All mice were housed in microisolator cages with a 12hrs dark/light cycle, and fed on a standard diet with autoclaved water. Mice were schedule 1 killed using increasing concentrations of CO_2 and then cervically dislocated. All experiments conformed to the ethical principles and guidelines approved by the University of York Institutional and Animal Care Use Committee in accordance with the European Union regulations and performed under a United-Kingdom Home Office License.

2.1.2 Sample Collection and Storage

Popliteal LNs were excised and excess fat or connective tissue removed with forceps. For flow cytometry they were placed in Fluorescence Activated Cell Sorting (FACS) wash (PBS containing 0.5% Bovine Serum Albumin (BSA) (Sigma Aldrich)). For immunohistochemistry, samples were transferred to optimal cutting temperature medium (OCT, Tissue-Tek, Sakura Finetek), snap frozen on dry ice and stored at -80°C. Optimal cutting temperature medium embedded tissue was sectioned into 20 μm sections using a cryostat and collected onto poly-L-Lysine coated microscope slides and stored at -20°C.

2.1.3 Enzymatic Digestion of Lymph Nodes

Digestion of lymph nodes was performed using an adaptation of a previously published protocol (Fletcher et al., 2011b). LNs were pierced with a fine forceps and incubated in enzymatic mix comprising 0.2mg/ml Collagenase P (Roche), 0.8mg/ml Dispase (Roche) and 0.1mg/ml DNaseI (Roche) in RPMI-1640 media (Life Technologies). Samples were incubated at 37°C in a thermomixer in 1.5ml tubes with gentle mixing. After 20 mins LNs were pipette mixed, the enzymatic mix was removed and added to 3ml of cold FACS wash or cell culture media (αMEM (Life Technologies), 10% Foetal Calf Serum (FCS) and 5% L-glutamine) and centrifuged at 300g for 5 mins at 4°C. Larger fragments that had not been digested that remained in the original digestion tube settle were pipette-mixed with enzymatic mix and incubated for 10 minutes at 37°C in a thermomixer with gentle mixing. After the incubation, cells were mixed vigorously, after the fragments were again allowed to settle, the supernatant was removed and added to the previously spun cell pellet and centrifuged again. Cells were filtered through 70μm cell strainers and then

used for flow cytometry.

2.1.4 Surface Marker Antibody Staining for Flow Cytometry

Flow cytometry staining was performed in a V-bottom 96 well plate. Cells were first resuspended in 100µl Fc block mix made up in FACS wash and incubated for 10 mins on ice. The Fc block mix included TruStain fcX (a rat anti-mouse CD16/32 antibody) used at a dilution of 1:200 (BioLegend) and rat IgG, 10µg/ml (sigma). Following the incubation the plate was spun for 5 min, 4°C, 300g and supernantant discarded. Cells were then resuspended in 100µl of antibody staining master mix containing fluorescent labelled monoclonal antibodies (mAbs) in FACS wash and incubated for 25-30 mins on ice (or at 37°C) in the dark. Single stains and unstained controls were also prepared. Cells were washed with FACS buffer, resuspended in 200µl of FACS wash and transferred into microtubes (Titertube micro test tubes, Bio-Rad) on ice. Samples were run on a Fortessa (BD Biosciences) flow cytometer and data was analysed using the FlowJo (Treestar) software package.

2.1.5 Live Cell / Dead Cell Discrimination

Following antibody staining, cells were resuspended in 100µl of PBS containing DAPI (1:5000 dilution), and then incubated for 5 min at room temperature in the dark. Cells were washed with FACS buffer (3 x 5 mins), resuspended in 200µl FACS wash, transferred into microtubes and kept on ice.

2.1.6 Quantifying Lymph Node Cellularity Using Flow Cytometry.

Accucheck counting (Invitrogen) beads were used to calculate total cellularity of samples. Following antibody staining, pellets were resuspended in 100µl FACs wash. 100µl of counting beads were mixed for 1 min to ensure they were evenly resuspended before running on the flow cytometer. **Figure 2.1** shows the counting beads on a flow cytometry plot. To ensure accuracy the beads are made up of two types of beads that differ in their fluorescent intensity, for accurate readings the two populations should be present at approximately 50:50 ratio. To calculate absolute cell number, the beads and cells of interest were gated on using the gating strategy shown in **Figure 2.1**. The following calculation was then made:

$$\text{Number of cells per } \mu l = \frac{\text{number of events(beads)}}{\text{number of events}} \text{ x number of beads per } \mu l \quad (2.1)$$

Where the number of beads per µl was provided by the supplier and varied with each batch. The total cellularity of the lymph node could be calculated using the values of cells per µl and final volume of FACS wash that contained the cells.

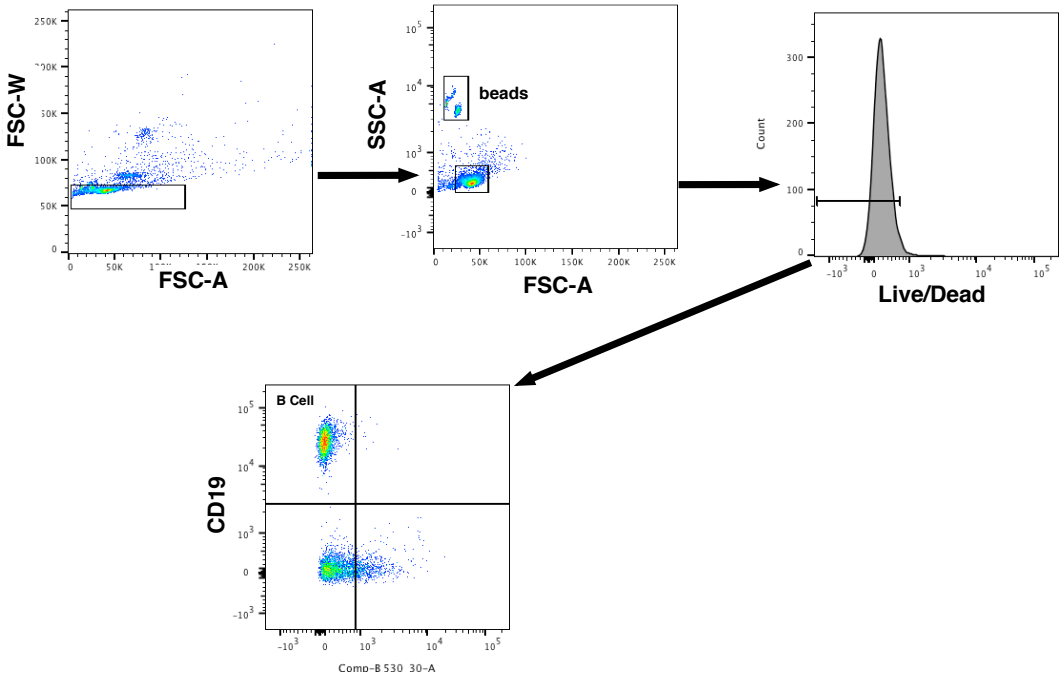

Figure 2.1 Gating strategy to quantify B-cell numbers. Representative flow cytometry plots showing the gating strategy used to quantify B cells. Singlet discrimination was achieved by using the forward scatter width. AccuCheck counting beads have a high side scatter (SSC) and a low forward scatter (FSC) enabling a distinction between beads and cells. The live-dead stain with DAPI only stains the dead cells as they do not have intact membranes. After gating on live cells, B-cells were identified using the marker CD19; a cell surface protein that modulates BCR mediated signals.

2.1.7 Immunohistochemistry

2.1.7.1 Immunofluorescent Staining of Frozen Tissue

Frozen lymph node sections on polylysine slides were incubated at room temperature for 30 mins. A circle was drawn around each section using a wax ImmEdge pen (Vector Laboratories), samples were fixed in 4% paraformaldehyde for 10 minutes at room temperature then hydrated and washed in PBS for 15 mins in total with changes of PBS every 5 mins. To prevent non-specific antibody staining, sections were incubated in a blocking buffer of PBS, 0.1% Tween-20 (Sigma-Aldrich), 0.1% Triton X-100 (Sigma-Aldrich) and 5% serum (the serum of the host the secondary antibody was raised in) at room temperature for 1 hour at room temperature. After blocking, sections were incubated in the primary antibody mix, made up in blocking buffer for 1 hour at RT. The slides were then washed and secondary antibody incubation was performed (if necessary). The slides were then washed. A drop of Prolong gold (Invitrogen) was added to each section, and then a No 1.5 glass coverslip (Fisher) mounted on top. The slides were incubated overnight at 4°C the next day slides were sealed using nail varnish. Stained slides were stored at 4°C.

2.1.7.2 Confocal Microscopy

Immunofluorescent stained sections were imaged using the Zeiss LSM 880 confocal microscope. Samples were excited with 405,488,561 and 633 nm lasers. Image acquisition was performed using the 63X oil objective. Tile scans and Z stacks were performed to image large tissue sections at high-resolution.

For imaging of chemokine gradients we used the Airyscan module to increase spatial resolution beyond the diffraction limit of light. On a standard confocal microscope out of focus emission light is excluded through use of a pinhole at the focal plane. Image resolution is increased by decreasing the pinhole, but with the caveat of losing light energy. Airyscan deals with this trade-off between resolution and light efficiency by using a hexagonal array of 32 detectors. Each detector acts as a small pinhole with positional information, this facilitates increased signal to noise ratio and spatial resolution in comparison to a standard confocal microscope (Huff, 2015).

Marker	Fluorophore	Supplier	Clone
Anti-mouse CD19	Alexa 488	eBioscience	eBio1D3
Anti-Human CD19	Alexa 488	Biolegend	HIB19
Anti-Human CXCL13	unconjugated	R&D Biosystems	53610
Anti-Human CD35	unconjugated	Santa Cruz	Ber-MAC-DRC

Table 2.1 List of antibodies used for flow cytometry and immunohistochemistry.

2.1.7.3 Quantification of LN topology.

Images of lymph nodes from *Cxcl13*-Cre/Tdtomato *R26R*- (abbreviated as *Cxcl13*-EYFP) mice (Onder et al., 2017) were obtained by laser scanning confocal microscopy (data provided by Dr. N. Pikor and Dr. L. Onder). The topological mapping (performed by J.C and Dr. Mario Novkovic) of follicular stromal cell network structure was created as an undirected, unweighted graph by defining nodes as the $EYFP^{+}$ centres of mass and edges as physical connections between neighbouring nodes. In this analysis images were manually annotated using Imaris (Bitplane) such that a straight line is draw between two stromal cell bodies that are connected by a protrusion or smaller branching process with no other cell body directly blocking this connection. An example graph is provided in **Chapter 4** (**Figure 4.3**). Analysis of topological parameters (described in **section 2.1.7.4**) was performed by Dr. Mario Novkovic using the iGraph[5] package and custom scripts in R.

[5] Available from https://cran.r-project.org/web/packages/igraph/index.html

2.1.7.4 Calculating Topological Parameters

In this study we focus on key metrics to characterise the topological properties of the follicular stromal cell network. Although a number of other metrics exist, these metrics are sufficient to assess whether the network has small-world properties, to facilitate a comparison of different stromal cell subsets and also provide quantitative data to inform the algorithmic reconstruction of an *in silico* stromal network. Specifically, we assess whether the network has small world properties (sigma and omega) (Humphries and Gurney, 2008; Watts and Strogatz, 1998) as has been reported for FRC networks (Novkovic et al., 2016) and assess the distances over which stromal cells interact how dense the interaction networks are.

Degree centrality: The number of edges e connected to a given node i..

Edge length: The length of an edge in microns. i.e. this represents the Euclidean distance between two nodes not the total length of interconnected cell protrusions.

Global clustering coefficient (C_{global}): The global clustering coefficient is given by the ratio of the triangles and the connected triples in the graph whereby a triplet is 3 connected nodes

$$c_{global} = \frac{number\ of\ closed\ triplets}{number\ of\ connected\ triplets\ of\ nodes} \qquad (2.2)$$

where δ_i represents the number of neighbors of node i. $\bar{C}$ represents the arithmetic mean of all nodes in the network.

Average local clustering coefficient ($\bar{C}$): The local clustering coefficient c_i of a node i is defined as:

$$c_i = \frac{2}{\delta_i(\delta_i - 1)} e_i \ , \ \ 0 \leq c_i \leq 1 \quad (2.3)$$

where δ_i represents the number of neighbors of node i. $\bar{C}$ represents the arithmetic mean of all nodes in the network.

Average shortest path length ($\bar{L}$): is defined as the arithmetic mean of all pairs of shortest distances between nodes i and j:

$$\bar{L} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} l_{ij} \ , \ 1 \leq \bar{L} < \infty \quad (2.4)$$

where l_{ij} is the length (number of edges) of the shortest path between nodes *i* and *j*, namely how many nodes one needs to pass in order to get from node *i* to node *j*. The maximum distance $\bar{L}_{max}$ is called the diameter of the network. In case of a complete network where all possible connections are present, all the node distances $l_{ij} = 1$, thus the sum $\sum_{i=1}^{n} \sum_{j=i+1}^{n} l_{ij} = \frac{n(n-1)}{2}$, which gives the minimal $\bar{L}_{min} = 1$.

Sigma factor (σ) : The small-world measure σ is defined as:

$$\sigma = \frac{\bar{C}/C_R}{\bar{L}/L_R} \ , \ 1 \leq \sigma < \infty \quad (2.5)$$

where C_R and L_R are the average clustering coefficient and average shortest path length of an equivalent random network averaged across 100 simulation runs. The equivalent random network has the same number of nodes and edges and is generated using the Erdos-Renyi model. A network has small world properties if $\bar{C} \gg C_R, \bar{L} \geq L_R$ and therefore $\sigma > 1$.

Omega factor (ω): The small-world measure ω is defined as:

$$\omega = \frac{L_R}{\bar{L}} - \frac{\bar{C}}{C_L} \ , \ -1 < \omega < 1 \quad (2.6)$$

C_L is the clustering coefficient of an equivalent lattice network and L_R is the average shortest path length of an equivalent Erdos-Renyi random network. A network will be classified as a small world network if: $-0.5 \leq \omega \leq 0.5$,

2.1.7.5 Quantification of Spatial Autocorrelation

To quantify the spatial autocorrelation of fluorescence intensity 2D confocal images were acquired on a Zeiss LSM 880 confocal microscope with the same laser settings and post processing for each sample. Processed .png files were then analysed in R using custom scripts. Briefly, this analysis involved discretising the image into 14.44μm^2 bins and calculating the spatial correlogram using the correlog function from the ncf[6] package. Spatial autocorrelation is quantified using Moran's I statistic with significance assessed through permutation testing (Bjørnstad et al., 1999; Moran, 1950).

[6] Available from https://cran.r-project.org/web/packages/ncf/index.html

2.1.8 Sub-Millisecond Single Molecule Imaging

2.1.8.1 Reagents

Human chemokines (CCL19 and CXCL13) labelled with the far-red fluorescent tag AF647 (Almac) were used to image molecular mobility in complex tissues **(Figure 2.2**).

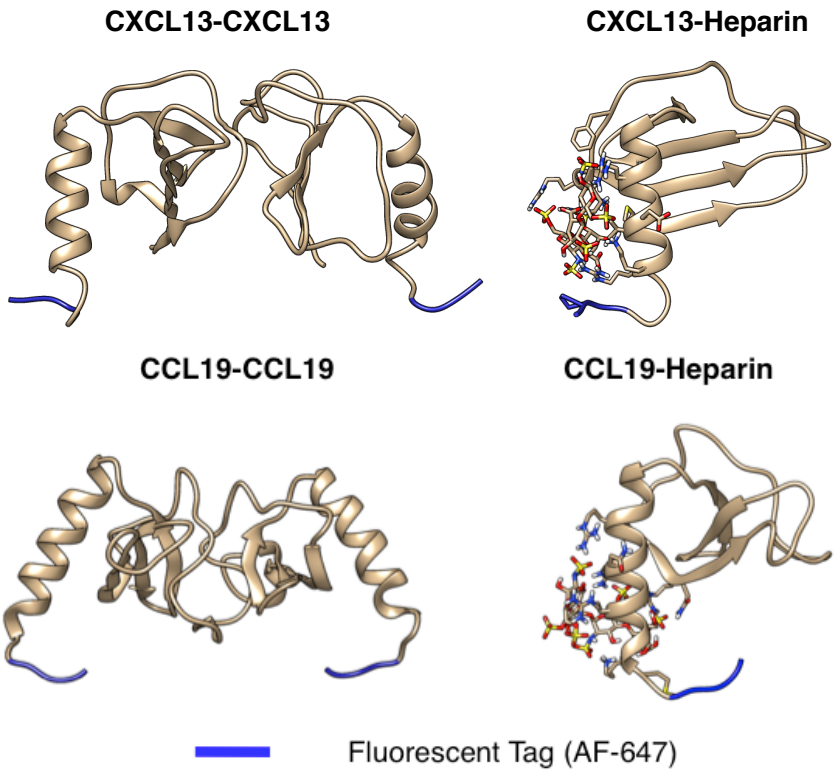

Figure 2.2. Fluorescently labelled chemokine molecules. The site where the fluorescent tag is added to the protein is highlighted in blue. The labelling site is distinct to putative interactions sites obtained from docking simulations performed on the ClusPro server (Kozakov et al., 2017).

2.1.8.2 Preparation of Collagen Matrices

Type I rat-tail collagen was obtained using the protocol described in (Barnes et al., 2016). Briefly, collagen was extracted from sterilized rat-tail tendons by dissolution in acetic acid and subsequent lyphophilization on ice. Lyophilized collagen I was dissolved in 0.02mol dm^{-3} acetic acid to provide a stock concentration of 6.6mg/ml.

Samples for fluorescence microscopy were prepared in tunnel slides formed by placing two parallel lines of double-sided tape on a standard microscopy slide approximately 5mm apart. A plasma-cleaned coverslip was placed on top to create a watertight tunnel and slides were cooled to 4˚C. Fluorescent chemokines were diluted to a working concentration of 111ng/ml and mixed with collagen type I rat-tail collagen diluted in PBS to a concentration of 3.3mg/ml and set to pH 7 with the addition of NaOH. Tunnel slides were then incubated at 15˚C for 30 minutes, followed by an additional 30-minute incubation at 37˚C. Formation of the collagen matrix was verified by second harmonic imaging (SHIM) performed on a Zeiss 780 Multiphoton microscope with a plan-apochromat 63x/1.4 oil objective lens with a shortpass 485 filter (**Figure 2.3**).

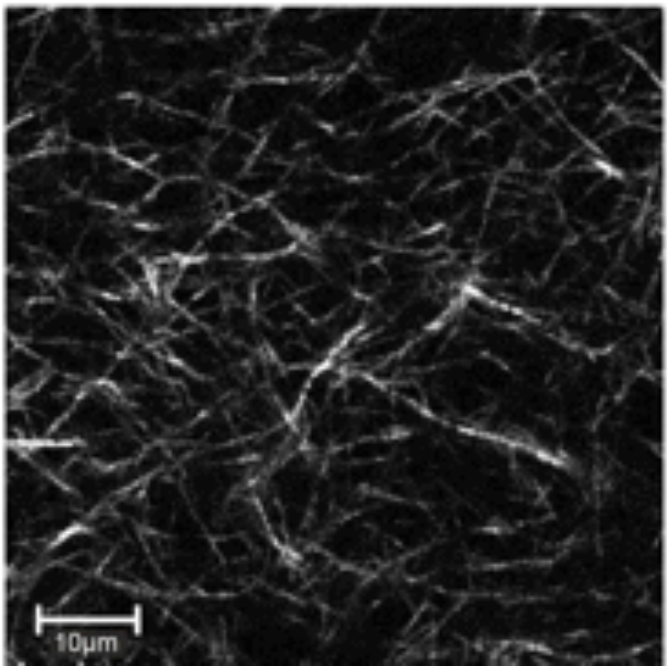

Figure 2.3 2D SHIM image of collagen network. Scale bar 10µm

2.1.8.3 Preparation of Heparan Sulphate Coated Tunnel Slides

Immobilised chemokine samples were prepared as follows: a plasma-cleaned coverslip (plasma-cleaning removed background autofluorescence) was incubated in Heparan Sulphate (50mg/ml) (Sigma) in PBS for 30 minutes. Slides were washed with PBS and left to dry at RT for 30 minutes before being placed onto a tunnel slide. 10nM fluorescently labelled chemokine solution in PBS was introduced to the tunnel slide and incubated in a humidity chamber for 15 minutes at 20˚C. Excess unbound chemokine was removed with a PBS wash.

2.1.8.4 Preparation of Lymph Node Tissue Sections for Super-Resolution Imaging

LNs were collected and stored as per **Section 2.1.2**. Frozen lymph node sections on polylysine slides were incubated at room temperature for 30 minutes. Samples were hydrated in PBS for 5 mins then left to dry and circles were drawn around each section with a wax ImmEdge pen (Vector Laboratories). Sections were incubated in a blocking buffer of PBS 5% goat serum (Sigma) at room temperature for 1 hour. After blocking, sections were incubated in primary antibody mix (anti-B220 FITC, eBioscience) made up in 1:200 blocking buffer for 1 hour at RT. Samples were washed with PBS for 3 x 5 minutes and 1μM of CXCL13-AF-647 added to the slides. Slides were left to incubate overnight at 4˚C after which slides were washed for 30 seconds in PBS and a No. 1.5 glass coverslip (Fisher) mounted on top. Slides were then sealed and imaged.

2.1.8.5 Narrowfield Fluorescence Microscopy

Bespoke fluorescence microscopy was performed on an inverted microscope body (Nikon Eclipse Ti-S) with a 100x NA 1.49 Nikon oil immersion lens and illumination from a supercontinuum laser (Fianium SC-400-6, Fianium Ltd.), controlled with an acousto-optical tuneable filter (AOTF) to produce a narrowfield excitation light centred on 619nm. The use of narrowfield imaging permits fluorescent excitation at distance of a few hundred nanometres above the coverslip thus mitigating some of the boundary effects which may be encountered using Total Internal Fluorescence (TIRF) microscopy where only a thin section directly above the coverslip is excited. A 633nm dichroic mirror and 647nm long-pass emission filter were used to filter the appropriate wavelengths of light emitted from the fluorescence images. Images were recorded on an emCCD camera (860 iXon^{+}, Andor Technology Ltd) cooled to -80˚C. 128x128 pixel images were acquired for

1000 frames with 1.98ms exposure times and 128 x 29 pixel image strips were acquired with 0.59ms exposure times. When imaging in tissue, sections were stained with an anti-B220 antibody conjugated to FITC. Samples were imaged at low (1.2μm/pixel) magnification with green illumination (470nm) to determine the location of the B cell follicles, before switching to high (120nm/pixel) magnification and red illumination to image chemokines in these areas.

Analysis of narrowfield microscopy data was performed by H.M. and involved the following steps. Single-molecules were identified and processed using ADEMS code, a bespoke software package in Matlab (Miller et al., 2015a). The micro-diffusion coefficient was calculated for each tracked particle from the gradient of a linear relation fitted to the first four steps in a track. The microdiffusion coefficient distributions comprised two distinct fractions: a low mobility fraction that had non-specifically adhered to the plasma-cleaned coverslip and a diffusive fraction. Consequently, the probability distribution of diffusion coefficients was modelled by a two gamma distribution (Qian and Elson, 1990):

$$F(x, D, N) = \frac{\left(\frac{N}{D}\right)^{N} x^{N-1} e^{-\frac{Nx}{D}}}{(N-1)!} \qquad (2.6)$$

where N is the number of independent steps in a track and D is the true diffusion coefficient. In fitting this model to the data the solution obtained is sensitive to the initial fitting conditions. To optimise these parameters simulated experimental data, with known input diffusion coefficient, was used to test that the diffusion constant derived from the fitting routine was consistent with this known input value. Specifically, to refine model-fitting parameters experimental data was simulated with and without noise, tracked, and the distribution of diffusion coefficients was fitted with the same constraints as the actual data. This process was repeated until the simulation represented the experimental data and the fit to the simulation data converged to the diffusion coefficient values simulated. Errors for measures of the diffusion coefficients were found by bootstrapping, an approach where a random 80% of the data is sampled and fitted as per the entire dataset. The standard deviation obtained from 10 repeats of this sampling and fitting was taken as the error for the measures obtained for 100% of the data.

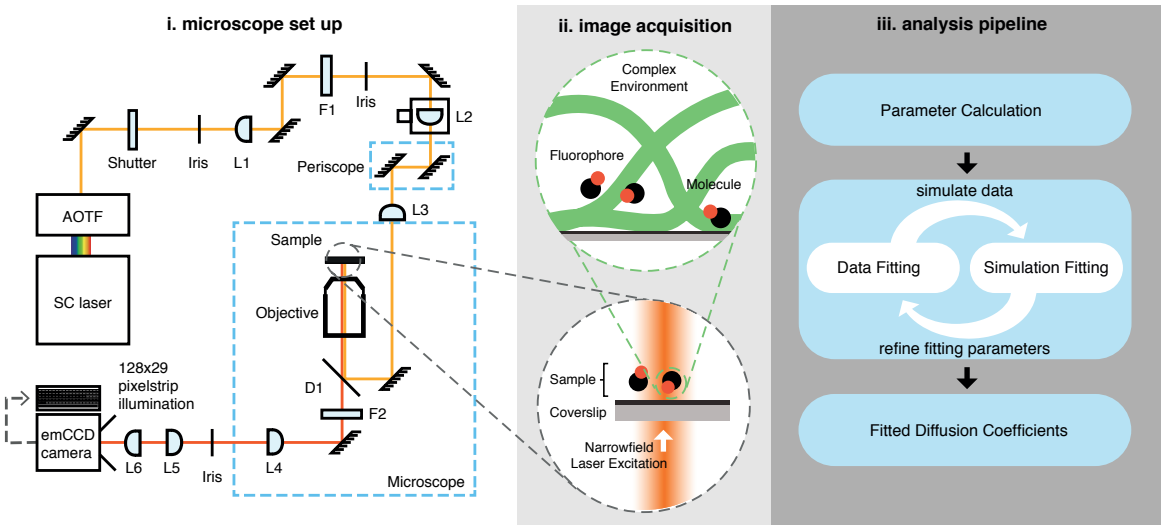

Figure 2.4 Overview of the super-resolution imaging platform. Single molecule images were obtained on an epifluorescent microscope, where illuminated and emitted light travels through the same objective lens. A continuous spectra of light was generated using a super continuum laser, this light then passes through an acousto-optic tuneable filter (AOTF), an electro-optical device that modulates the intensity and wavelength of multiple laser lines. Devices of this type rely on a specialized crystal whose optical properties vary upon interaction with an acoustic wave, enabling very rapid wavelength tuning. Using this system a narrowfield excitation light centred on 619nm is produced with a 633nm dichroic mirror and 647nm long-pass emission filter used to filter the appropriate wavelengths of light emitted from the sample. Images were recorded on an emCCD camera (860 iXon^{+}, Andor Technology Ltd) cooled to -80˚C. The cooling was necessary to increase the signal to noise ratio as at high temperatures you can get thermal fluctuations leading to noisier signals. 128x128 pixel images were acquired for 1000 frames with 1.98ms exposure times and 128 x 29 pixel image strips were acquired with 0.59ms exposure times.

2.1.9 Fluorescence Recovery After Photobleaching (FRAP)

Fluorescence Recovery After Photobleaching (FRAP) images were performed on a Zeiss LSM 880 microscope with an APO 110z/1.4 oil immersion lens. Fluorescently labelled chemokine diluted in PBS to a working concentration of 100nM were prepared in glass bottom petri dishes (1.5 coverglass, MatTek corporation). In these analyses an area of 4.9±0.1μm square was bleached with a 633nm laser. The bleached area was measured with an immobile sample of CXCL13-AF647. Thirty recovery traces (intensity(I) vs time (t)) were acquired for BSA-AF647. Analysis was performed by Helen Miller (H.M.) using the Zeiss Zen software to fit the data with the single exponential equation:

$$I = I_0 - I_1 e^{\left(-\frac{t}{T_1}\right)} \qquad (2.7)$$

where the initial intensity I_0 and drop in intensity, I_1 were fitted along with the decay constant T_1, from which the diffusion coefficient, D, is calculated via:

$$D = \frac{\omega^2}{8T_1} \qquad (2.8)$$

where ω is the full width at half maximum fluorescence intensity value of the bleached area.

2.1.10 Fluorescence Correlation Spectroscopy

Fluorescence correlation spectroscopy (FCS) was performed on a Zeiss LSM 880 microscope. Samples were prepared in glass bottom petri dishes (1.5 coverglass, MatTek corporation) and illuminated with a 633nm laser. The confocal volume was measured using a calibration sample with a known diffusion constant. To perform this calibration we measured the diffusion of BSA-AF647 in water and constraining the value to be 59$\mu m^2 s^{-1}$ (Putnam, 1975a) this allowed the structural parameter (s) to be fixed at 6.6. Three repeats of ten experiments were conducted and traces indicating the presence of multimeric clumps or proximity to the surface were excluded. Data analysis was performed by H.M. and required fitting autocorrelation traces, G(τ), to account for transient dark states[7] and diffusion using the following equation:

$$G(\tau) = 1 + \left(1 + \frac{Te^{-\left(\frac{\tau}{\tau_T}\right)}}{1-T}\right)\left(\frac{1}{V_{eff}\langle C\rangle}\frac{1}{\left(1+\frac{\tau}{\tau_D}\right)}\frac{1}{\sqrt{1+\left(\frac{1}{s}\right)^2\frac{\tau}{\tau_D}}}\right) \qquad (2.9)$$

where T is the fraction of molecules in dark state, τ_T is the time molecules spend in the dark state, τ_D is the time constant of diffusion across the confocal volume, V_{eff}, and <C> is the average concentration. D was calculated from T_D via the equation:

$$D = \frac{w^2}{4T_D} \qquad (2.10)$$

where w is the width of the focal spot (0.322μm).

[7] The fluorescence properties of a fluorophore can change over the course of an acquisition; the "flickering" in fluorescence intensity is the transition of the dye to the first excited triplet state. In this state, the chromophore needs more time to return to the ground state, during which time it cannot emit any fluorescence.

2.2 Computational Methods

2.2.1 Principled Design Framework

To ensure a principled and transparent design process we employ the CoSMoS (Complex System Modelling and Simulation) process, a framework to guide the modelling and analysis of complex systems (**Figure 2.5**) (P.Andrews J. Timmis et al., 2010). As suggested by the CoSMoS process, initial stages of model design involve the development of a domain model, a non-executable, conceptual model focusing purely on current biological understanding, disregarding any consideration of how to implement and simulate the conceptual model. The domain model specifies the states, relationships and methods of interaction (the rule-set) for the biological entities being captured. The platform model is akin to a software specification, as is standard in software engineering, and details how the biological processes specified in the domain model are to be implemented. The simulation platform is an executable piece of software, which implements the underlying conceptual model. The results model provides a structure to interpret data obtained from the Simulation Platform. 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 assess this result. Within the York Computational Immunology Lab a series of case studies have led to the development of an instantiation of the process, particularly suited to modelling the immune system. This is discussed further in **Chapter 3**.

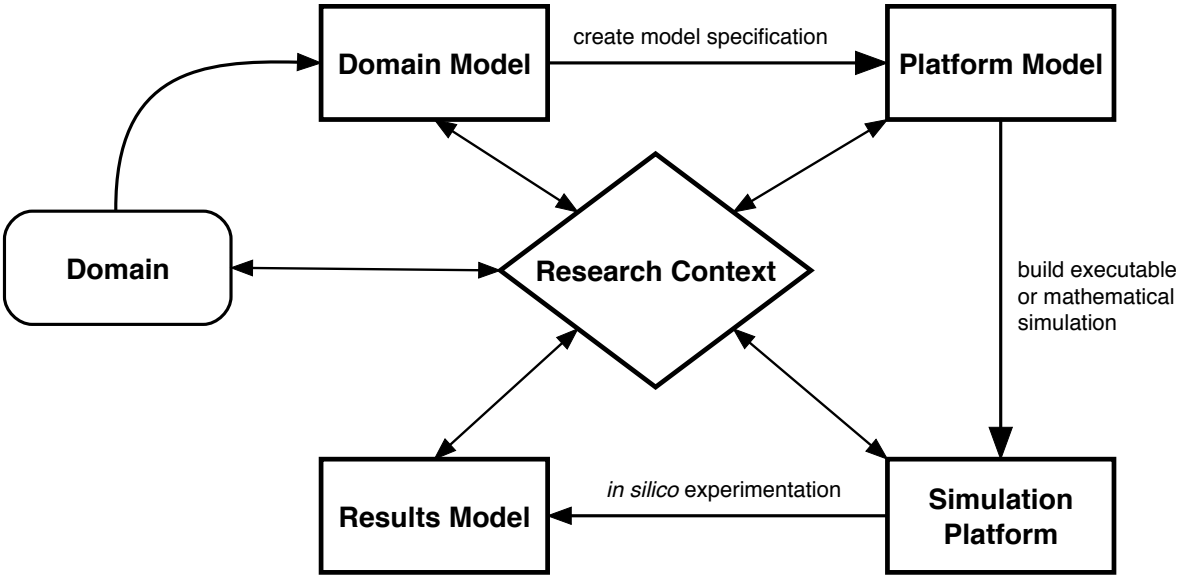

Figure 2.5. Overview of the CoSMoS process. In this approach the biological system of interest is referred to as the domain. Initial design phases lead to the development of a domain model, a non-executable specification focusing on current understanding with respect to the research context. The platform model represents a software blueprint while the simulation platform is an executable piece of software that implements the conceptual model. The results model summarizes the understanding generated from experimentation conducted using the simulator.

2.2.2 Modelling Biological Behaviours

The Unified Modelling Language (UML) is a general-purpose visual notation used to model the design of a system. While commonly used in software engineering, we use an adaptation of the UML as described by Read et al. (2014) to specify low-level behaviours of biological systems.

Expected Behaviours Diagram: In this diagram we present the research context within which the model is developed. This diagram specifies the research question and the observed emergent properties of the system, as well as the biological entities and mechanisms hypothesised to give rise to these properties. This approach helps to scope the research context, highlighting key model entities and time points of interest.

State Machine Diagram: For each model entity, states (a set of attributes and behaviours associated with a model entity at a specific moment in time) that the entity can exist in and the interactions that much take place for the state to change are examined and documented. In **Figure 2.6** we see that the cell enters the system in State A and if conditions P and Q are met the entity transitions into State B.

Activity Diagram: This diagram is used to specify a sequence of activities associated with model entities. For each entity, it details the workflow from an initial state to a finish point, detailing the decision paths and interactions with other entities that occur (**Figure 2.7**).

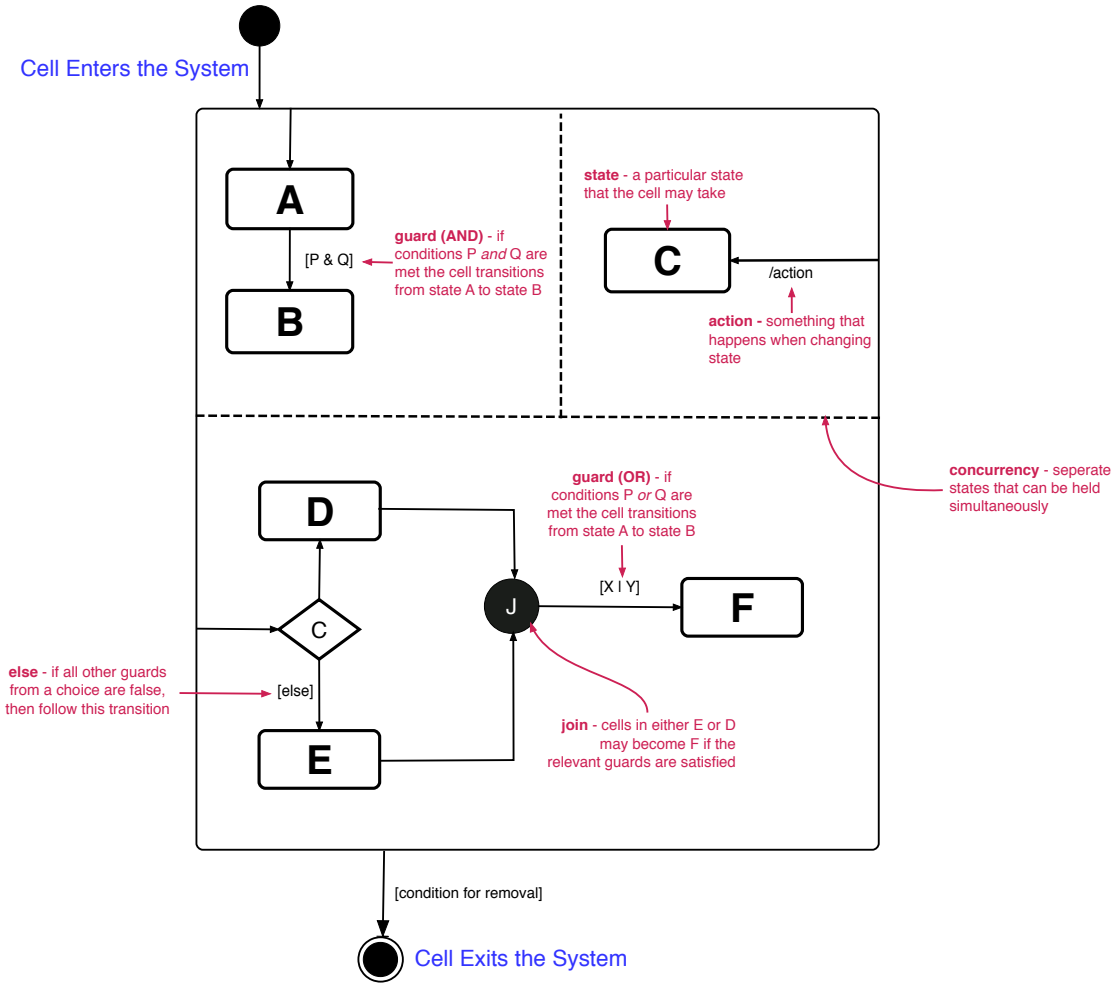

Figure 2.6. Key for Adapted Unified Modelling Language: State Machine. This is a key for the syntax used to describe biological processes as state machine diagrams in both the domain and platform models. *Rectangles* represent actions and agent states; *diamonds* represent decisions; *arrows* represent the order in which state transitions/activities may occur while a *circle* represents entry/exit into the system.

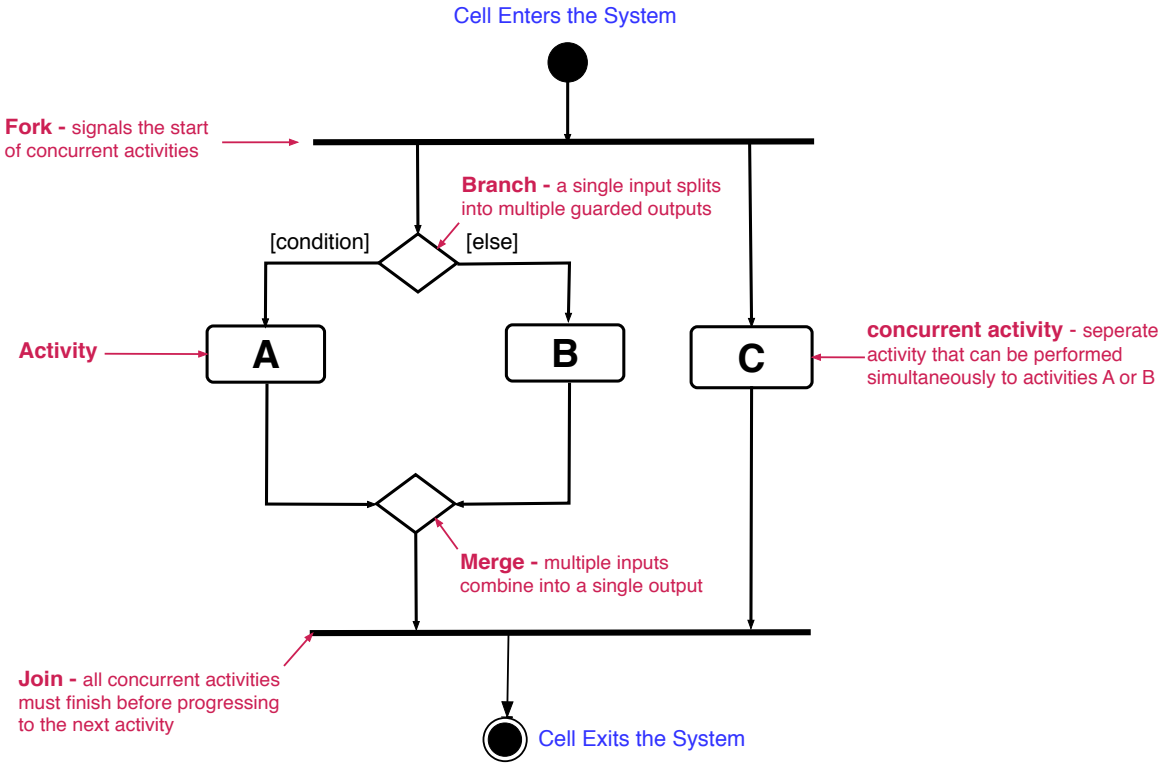

Figure 2.7. Key for Adapted Unified Modelling Language: Activity Diagram. This describes the syntax used to describe biological processes as activity diagrams in both the domain and platform models. *Rectangles* represent *actions or activities*; *diamonds* represent *decisions*; *horizontal bars* represent the start (*split*) or end (*join*) of concurrent activities; a *black circle* represents the start (*initial state*) of the workflow while an *encircled black circle* represents the end of the activity (*final state*).

2.2.3 Arguing that the Simulation is a Fit for Purpose Representation of the Biological System

The design and implementation decisions made when constructing a simulator are influenced by the overarching scientific objectives of the work, with simulation results interpreted in this context (Alden et al., 2015; Cosgrove et al., 2015). To argue that the simulator fulfils its remit, acceptance tests, key design decisions, and information used to inform the design, development and validation of the model and simulation are presented as arguments over evidence using a visual notation derived from goal structuring notation (**Figure 2.8**) and can be opened using the ARTOO tool[8] (Alden et al., 2015). This diagrammatic tool facilitates transparency of model design and analysis, capturing the reasoning behind the inclusion or exclusion of each biological feature and recording assumptions, as well as pointing to evidence supporting model-derived conclusions.

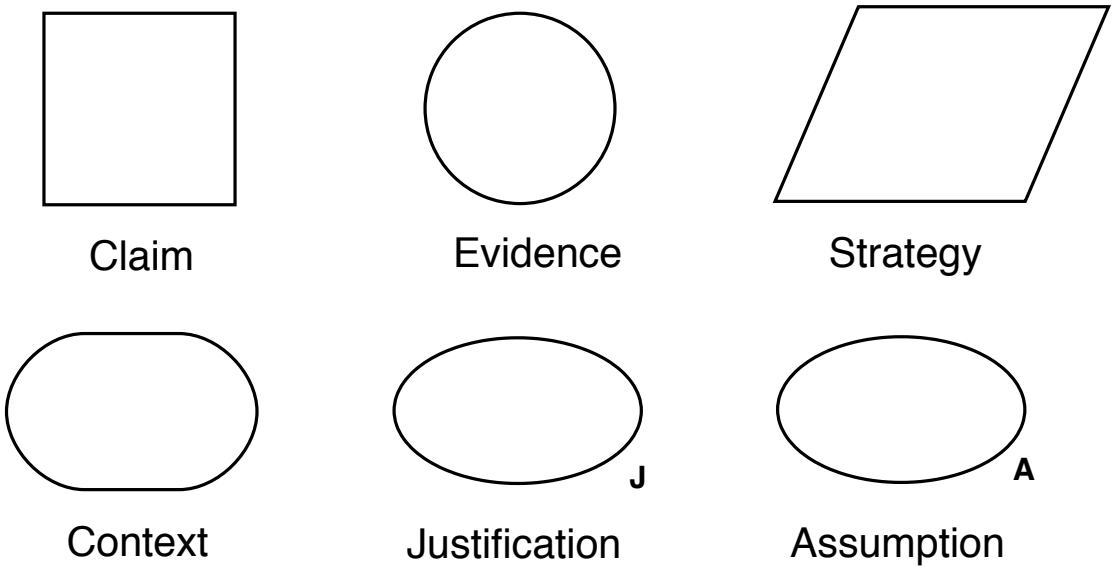

Figure 2.8. Key for Argumentation Notation: (i) claim the purpose of the argument we are seeking to support; (ii) Evidence that supports the argument made in the attached claim; (iii) strategy the steps that will be taken to argue that the claim is supported; (iv) context defines the purpose of the argument and key terms; (v) justification provides a reasoning behind the selection of a strategy or claim and (vi) assumption provides an explicit statement of any assumptions made in place of biological understanding. Figure reproduced from (Alden et al., 2015).

[8] Available from: https://www.york.ac.uk/computational-immunology/software/artoo/

2.2.4 Software Development and Computer Infrastructure

Within the simulation platform each module was developed using Java and the MASON[9] ABM library version 19 (Luke et al., 2005) in an iterative process of implementation, validation and refactoring using Acceptance Test-Driven Development (ATDD) (Sommerville, 2010). Tests are continually assessed and refined as the project progresses and are incorporated into an automated regression framework using the java library JUnit[10] (available from http://junit.org/junit4/) to ensure that new code does not disrupt existing functionality, expediting the development process. Test coverage was quantified using the eclipse plugin eclEmma[11]. Simulations were executed on a Linux Cluster made up of 64 CPU 256GB memory nodes and running Fedora 22 and an Oracle grid engine.

2.2.5 Calibration to Establish Baseline Simulation Behaviours

Model calibration is the process by which parameter values are tuned to reproduce experimentally measured behaviours. To constrain each parameter, lower and upper bounds were set on the basis of direct experimental measures or derived from indirect evidence from the wider literature (described further in **Chapter 4**). Parameters were systematically changed and compared to experimental datasets using the non-parametric Mann-Whitney U-Test given outputs were not normally distributed, as determined by a Shapiro-Wilk test for normality (Royston, 1992). To assess the robustness of our baseline calibrated parameter values, outputs from best-fit parameter sets were compared to gene knockout experiments. Parameter sets that failed to reproduce statistically comparable results to both wild type and gene deficient mice were omitted.

2.2.6 Quantification of Model Uncertainty

To quantify sources of uncertainty in the our simulator we used the R software package SPARTAN (Alden et al., 2013). This package contains a suite of statistical techniques (described in more detail in the following sections) specifically designed to help understand the relationship between the simulator and the physical system it describes.

[9] Available from http://cs.gmu.edu/~eclab/projects/mason
[10] Available from http://junit.org/junit4/
[11] Available from http://www.eclemma.org/download.html

2.2.7 Quantification of Aleatory Uncertainty

CXCL13Sim is non-deterministic and therefore can produce different outputs under the same parameter inputs. To determine how many runs are required to give a representative output for a given parameter set we perform an aleatory analysis (**Figure 2.9**). In this approach, distributions of simulation outputs generated using a fixed parameter set are compared. By varying the number of samples comprising the distributions, the analysis determines the minimum number of runs required to obtain statistically consistent distributions. Larger sample sizes produce increasingly similar distributions, thereby mitigating the effect of simulation stochasticity on results.

To understand aleatory uncertainty, 20 distributions were generated and contrasted for each sample size. A distribution of median responses for each simulation run is generated for each of the 20 subsets. Distributions 2–20 are contrasted with the distribution from the 1st set using the Vargha-Delaney A-Test (Vargha and Delaney, 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. Values of 0.5 indicate that the medians are the same while values of 1 and 0 mean that there is no overlap. In our analyses we set thresholds for small (0.56), medium (0.66) and large (0.71) effect sizes based on values suggested by Vargha and Delaney (2000) and define a significant behavioural alteration as one where the A-test statistic exceeds the medium threshold.

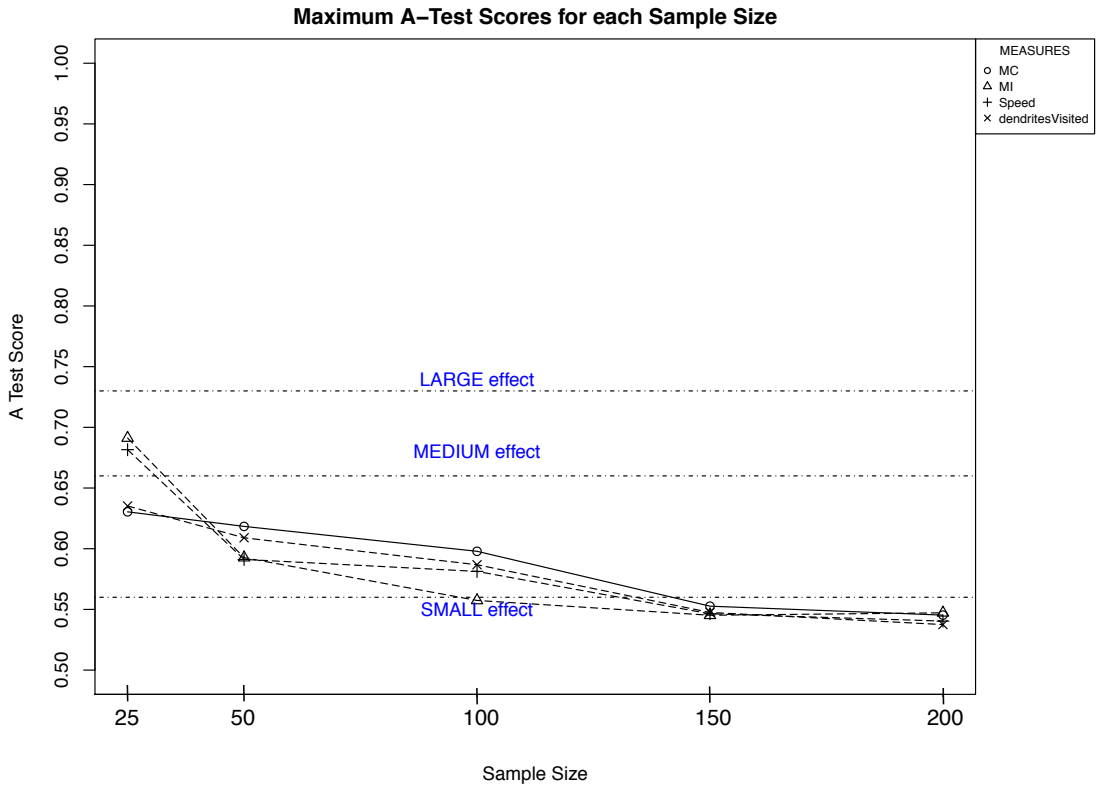

Figure 2.9. An illustratory example of output from an aleatory analysis using the R package SPARTAN. On the plot the A-score (Y-axis) is shown for each simulation output for a given sample size (X-axis). The A-score is the probability that a randomly selected sample from one of the 20 distributions will be larger than a randomly selected sample from another. As the number of replicate runs increases the value tends to a value of 0.5 suggesting that there is no difference between the distributions. Vargha & Delaney provide suggested thresholds for interpreting the effect size results above 0.71 or below 0.29 indicate a significant difference between the populations.

2.2.8 Quantification of Parameter Uncertainty

Prior to using a simulation for *in silico* experimentation, it is important to appreciate how sensitive outputs are to perturbations in parameters. Such an analysis can be performed through application of sensitivity analysis (SA) techniques. Through a systematic exploration of the parameter space, simulation inputs that have an influential effect on simulation behaviour are identified, aiding the biological interpretation of simulation results. SA techniques are split into two categories: local and global. Local analysis techniques examine how robust the simulation is to a perturbation of a single parameter value (herein referred to as an OAT analysis) (**Figure 2.10**). However, local SA techniques cannot reveal compound effects where one parameter's influence is dependent on the value of another. Such effects may be elucidated using global analysis techniques that perturb multiple parameters simultaneously.

For OAT analysis each parameter is adjusted, with all other parameters remaining at their calibrated value. The Vargha-Delaney A-Test described previously implemented in SPARTAN is employed to determine if changing the parameter value has led to a significant behavioural alteration in contrast to the baseline simulation. We define a significant behavioural alteration as one that leads to a medium effect size. Threshold values for determining small medium and large effect sizes are discussed in **section 2.2.7**

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. In **Figure 2.10a** we see an example output from an OAT analysis, low values of the parameter lead to a significant increase in all 3 output metrics. Parameter value 6 represents the baseline value (hence no statistically significant difference) and values greater than this all model outputs are reduced (shown by a high values of A test). In **Figure 2.10b** there is a large amount of uncertainty around this parameter as changing the parameters does not lead to a large difference in any of the three model outputs.

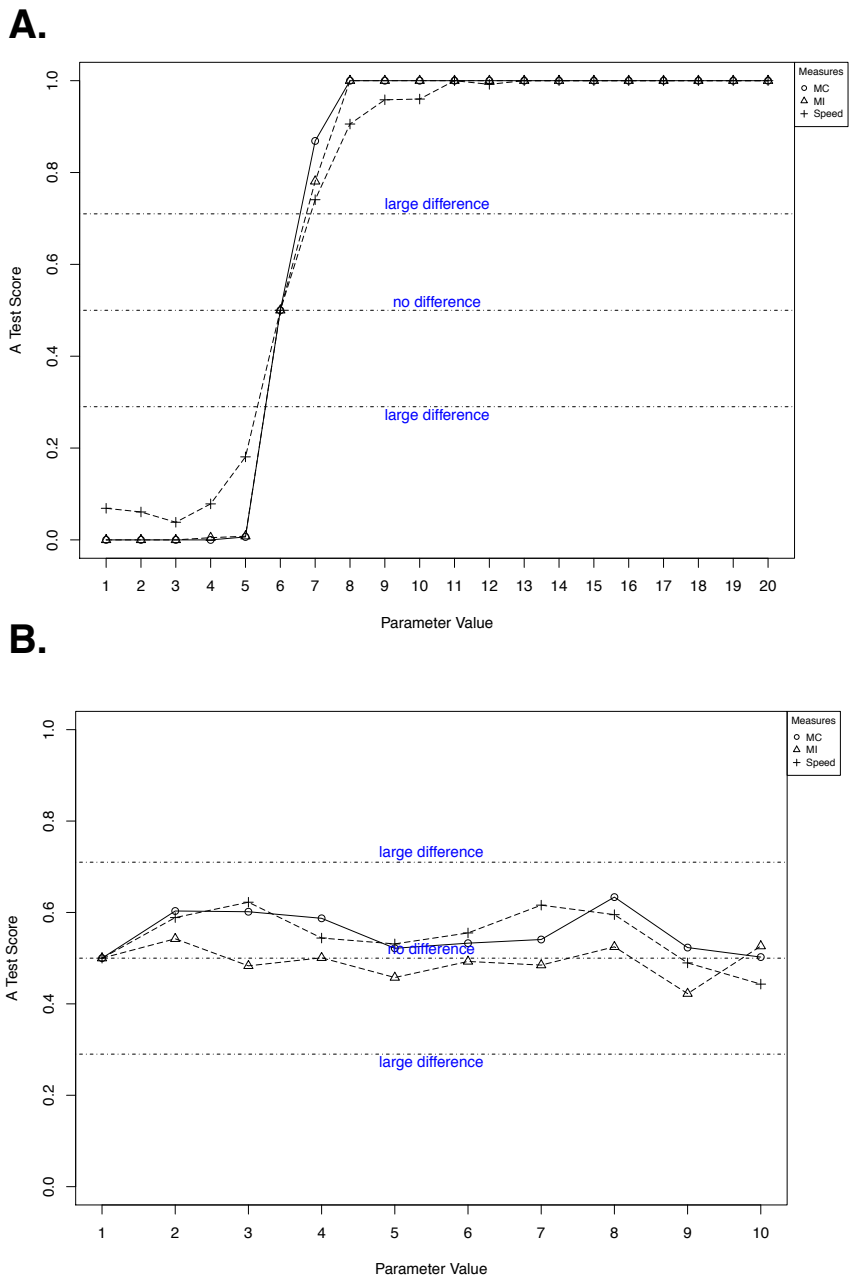

Figure 2.10. Example A-test scores when OAT adjusting parameters. These example plots show an illustratory example of the results obtained from an OAT analysis performed in SPARTAN. 3 Parameters were incrementally changed within their likelihood distributions with significant alteration in simulation outputs from baseline behaviours determined using the Vargha-Delaney A-Test. (A) an example output from an OAT analysis where low values of the parameter lead to a significant increase in all 3 output metrics. Parameter value 6 represents the baseline value (hence no statistically significant difference) and values greater than this all model outputs are reduced (shown by a high values of A test). (B) an example output showing a that changing the parameter does not lead to a large difference in any of the three model outputs. meandering index (MI) Δ; motility coefficient (MC) ¢ ; speed ✚.

2.2.9 Global Sensitivity Analysis

To perform a global sensitivity analysis we use two parameter sampling techniques, LHC (Latin-Hypercube) **(Figure 2.11)** and eFAST (Extended Fourier Amplitude Sampling Test). Through latin hypercube sampling, values for each parameter are selected with the aim of ensuring efficient coverage of the parameter space. Parameters that have significant impact on simulation behaviours are identified 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).

To calculate the PRCC the data are rank-transformed [12], and for each parameter, two linear regression models are found, the first representing the input parameter in terms of the other parameters and the second represents the output measures in terms of the other parameters. A Pearson correlation coefficient for the residuals from those two regression models gives the PRCC value for that specific parameter. Thus PRCCs characterise a linear relationship between input x and output y after the linear effects of the other inputs on y have been discounted. A significance test is performed to assess if a PRCC is significantly different from zero (Marino et al., 2008). Each PRCC (y) generates a value T, calculated as follows:

$$T = \gamma \sqrt{\frac{(N - 2 - p)}{1 - \gamma^2}} \qquad (2.11)$$

where T follows a student's t distribution with (N-2-p) degrees of freedom. N is the sample size and p is the number of parameters whose effects are discounted. In **Figure 2.12** we see PRCC values from an example LHC analysis, the parameter decay constant is significantly positively correlated with the output metric while Ka is negatively correlated. The parameter Rf is not significantly correlated with the output metric however.

[12] Inputs and outputs are sorted by magnitude

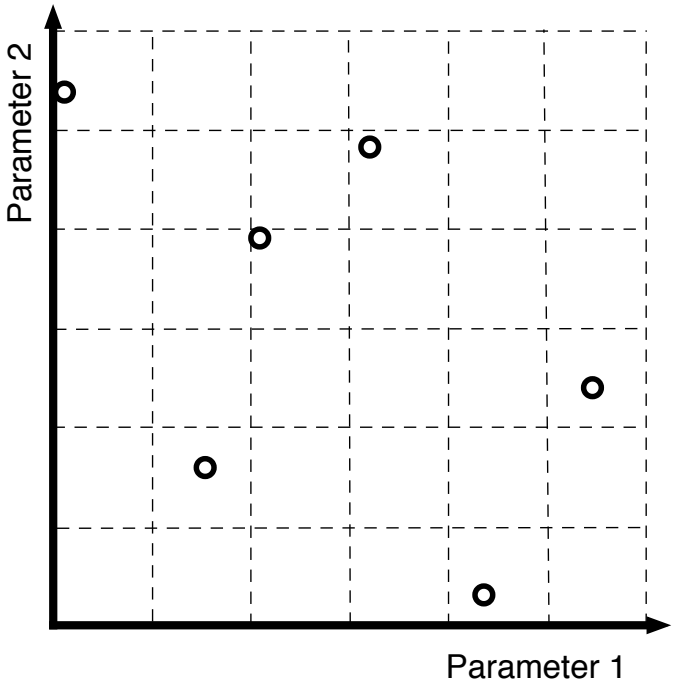

Figure 2.11. Overview of the latin hypercube sampling approach for 2 parameters. Latin Hypercube Sampling (LHS) is a type of stratified sampling. It works by controlling the way that random samples are generated for a parameter likelihood distribution. In this approach the parameter space is divided into subdomains (dotted lines). In this example 6 parameter combinations are taken, with one coming from each subdomain to ensure extensive exploration of the parameter space. Figure adapted from (Read, 2011).

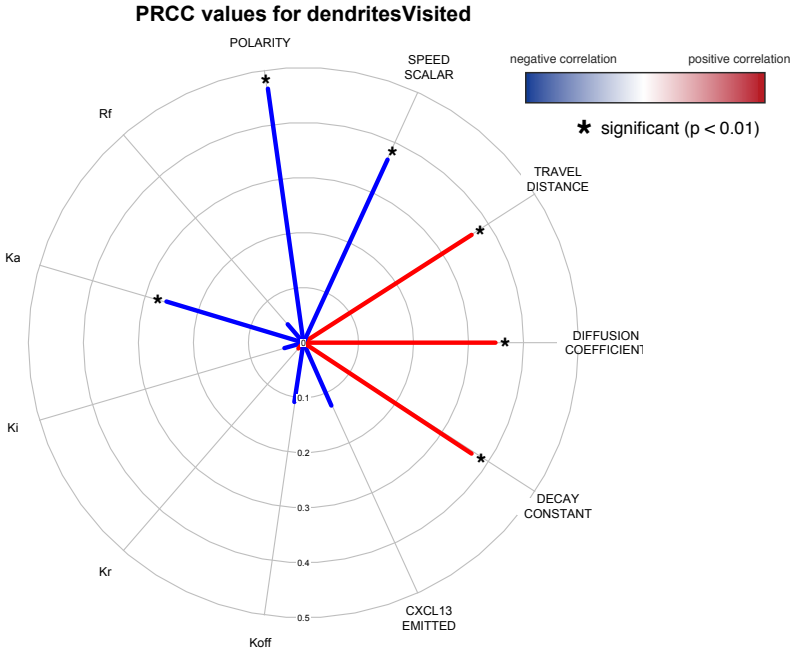

Figure 2.12. An example polar plot to visualize PRCCs determined following LHC sampling of the parameter space. In this plot the PRCC value is indicated with a line on the polar plot; red lines display positive correlations with respect to a given output while blue lines display negative correlations with respect to a simulation output.

In the previous global analysis approach we use Latin Hypercube Sampling (LHS) and Partial Rank Correlation Coefficients (PRCC) to identify parameters that have a significant effect on model outputs. This approach facilitates an understanding of what parameters should be targeted to achieve a desired response but does not indicate which parameter uncertainties have the greatest impact on output variability. Understanding this parametric uncertainty can allow for model refinement through directing experimental measurements. The extended Fourier Amplitude Sampling Test (eFAST) is a variance decomposition method (**Figure 2.15**) (analogous to ANOVA) that can be used to address this issue (Saltelli and Bolado, 1998; Saltelli et al., 2008). In this approach input parameters are varied, causing variation in model output. The algorithm then partitions the output variance, determining what fraction of the variance can be explained by variation in each input parameter.

In this scheme, parameters values are sampled using a sinusoidal function of a particular frequency. Each parameter is taken in turn and sampled at a frequency that is much larger than the other parameters (**Figure 2.13a**). Due to the symmetry of a sinusoidal function it is possible to choose the same parameter set more than once, therefore a re-sampling scheme in which a phase shift is introduced at each frequency is encouraged (Marino et al., 2008). 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. This process is repeated for an extra parameter, the 'dummy' parameter that 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.

To quantify the influence of each parameter, two sensitivity indexes are calculated for each parameter-response pairing: a first-order (Si) and total order sensitivity (STi) index (**Figure 2.13b**). The first indicates the fraction of output variance in that response that can be explained by the value assigned to the parameter. The latter indicates the variance in that response caused by higher-order non-linear effects between the parameter and the others under investigation. To determine the significance of these metrics, indexes are compared to those obtained for the 'dummy' parameter using a two-sample t-test.

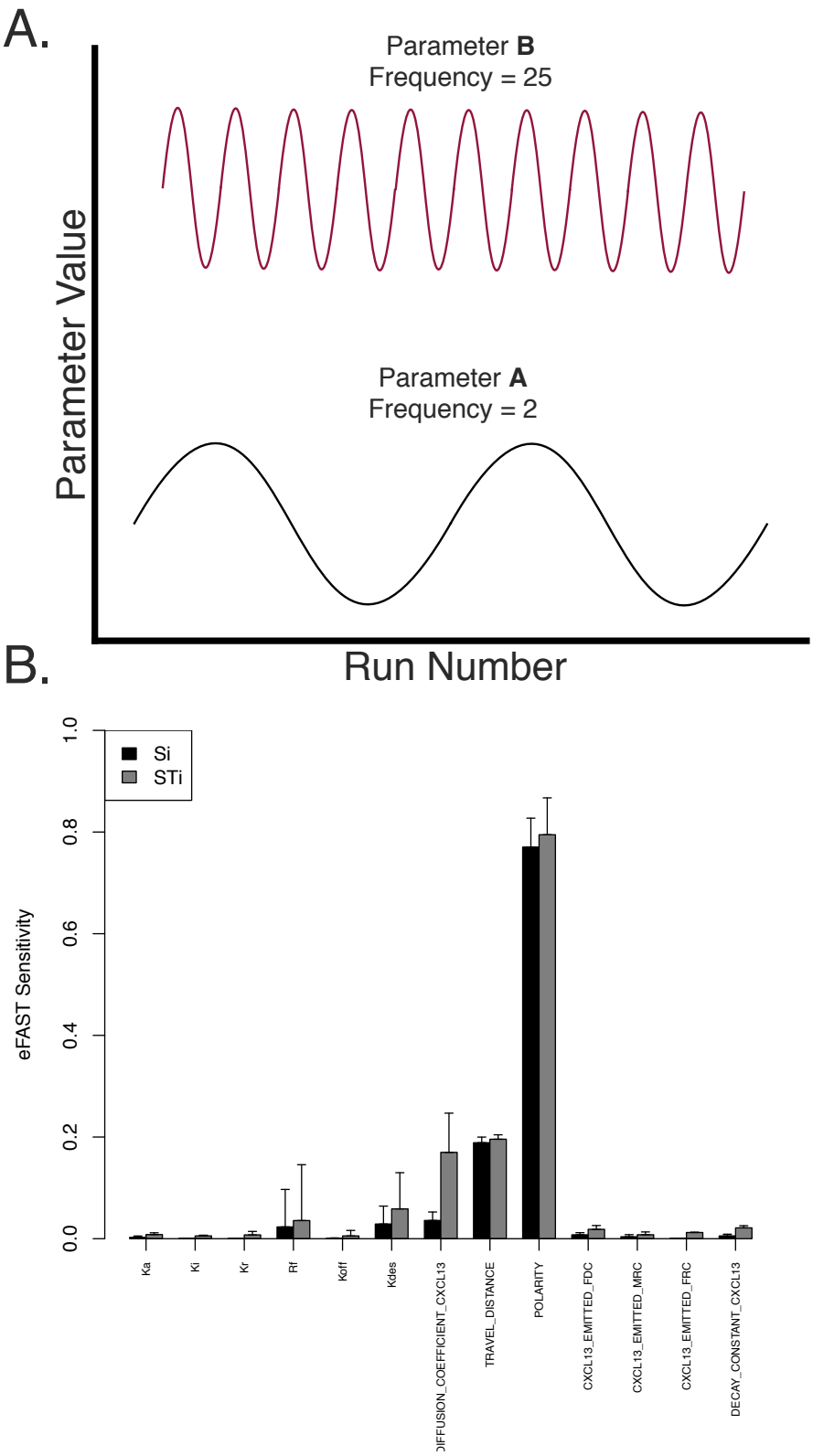

Figure 2.13. Overview of the extended Fourier Amplitude Sampling Test (eFAST). (A) In this scheme parameters are varied according to a sinusoidal function based on run number. The frequency is greatly increased for the parameter of interest. Figure adapted from Marino et al (2008). (B) Example SPARTAN output for an eFAST analysis. The Si (black) represents the fraction of output variance that is attributed to a parameter value. The STi (grey) represents the variance caused by higher order non-linear effects between that parameter and the others explored. The error bars represent the standard error over three resample curves.

2.2.10 Simulation Emulation with Machine Learning

The emulation pipeline is described in full in **Chapter 5.5** but briefly, involves the following: the training dataset for emulator development was obtained using latin hypercube sampling, with 1000 parameter sets. Each set was executed 100 times to mitigate aleatory uncertainty, and median responses calculated to summarize simulator performance under those conditions. The data set was partitioned into training (75%), testing (15%) and validation (10%) datasets.

The supervised learning approach used to generate CXCL13emulator was an artificial neural network, an approach inspired by the neuronal circuits in the brain, with computations structured in terms of an interconnected group of artificial neurons. During the learning phase, the weighting of connections between neurons are adjusted in such a way that the network can convert a set of inputs (simulation parameters) into a set of desired outputs (simulation responses). The ANN-based emulator was developed in the R package SPARTAN with supervised learning of the data achieved through backpropagation of errors. To determine optimal hyperparameters of the network we performed ten-fold cross validation on a selection of structures with thirteen inputs (the parameters) and four outputs (speed, meandering index, motility coefficient and checkpoints reached), with one to four hidden layers. The accuracy of each fold was determined to be the mean squared error between the predicted cell behaviour responses and those obtained by the simulator, and the accuracy of the network structure determined to be the average of these ten fold root mean squared errors.

2.2.11 Multiobjective Optimisation

Multiobjective optimisation (detailed further in **Chapter 6.1.2**) of the CXCL13emulator was performed using the non-dominated sorting genetic algorithm II (NSGA-II), a multiobjective genetic algorithm (Deb et al., 2002) (**Figure 2.14**). This analysis was performed in R using the package mco v15.0 [13]. The four objectives to be assessed by the algorithm were to: minimize the root mean squared error between emulator and simulator responses for cell speed, meandering index and motility coefficient; and maximize scanning rates. Values for generation number, mutation and crossover probabilities (**Table 2.2**) were determined by a global sensitivity analysis whereby values for mutation and crossover rates were sampled between 0.1 and 1.0 (intervals of 0.1) and values for the number of generations was sampled between 200 and 500 (intervals of 100). We chose parameters that performed well on all three objectives and maximized the variance of the parameter inputs.

Parameter	Value
Mutation rate	0.4
Crossover rate	0.9
Population Size	300
Generations	200

Table 2.2 Parameters used for NSGA-II

[13] Available from https://CRAN.R-project.org/package=mco

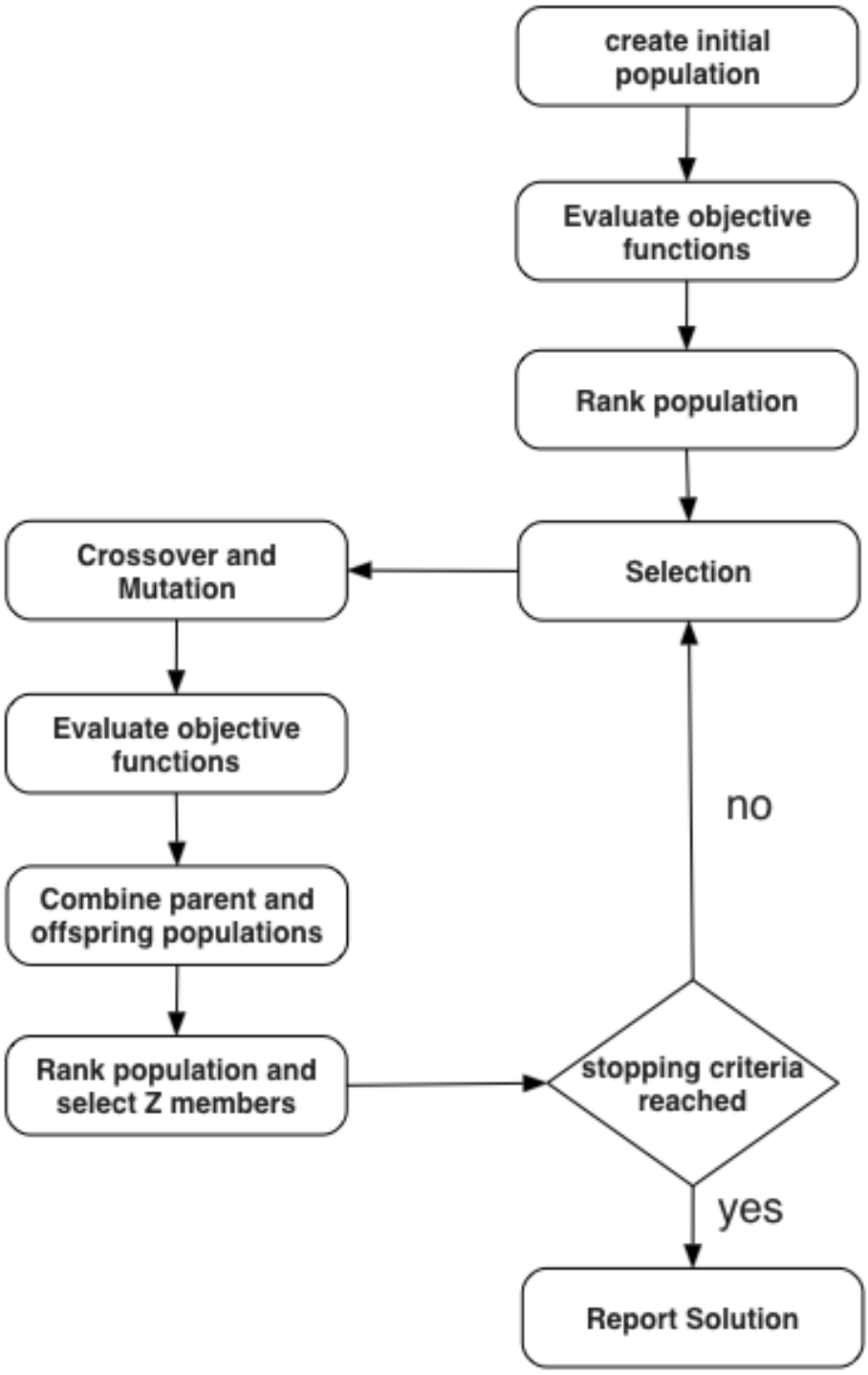

Figure 2.14. NSGA2 workflow. Step 1: Create a random population of size Z**.** Step 2: rank the random parent population based on non-domination. Step 3: For each non-dominated solution, assign a fitness score. Step 4: Create an offspring population using tournament selection, recombination, and diversity operators. Step 5: Create the next generation by combining parent and offspring population and sorting this population, of size 2Z, to identify all non-dominated fronts. The new parent population is then determined by taking the highest ranked (determined using a crowded comparison operator) non-dominated solutions. When the total non-dominated solutions exceed the population size Z, reject some of the lower ranked non-dominated solutions. Then perform the selection, crossover and mutation operations on the newly generated parent population, to create the new offspring population, of size Z. Step 6: Step 5 is repeated until the maximum number of iterations is reached.

Chapter 3 Agile Development of Multiscale Models

3.1 Introduction

3.1.1 Multiscale Modelling

The B-cell activation pathway discussed in **Chapter 1** is comprised of distinct layers of organisation, spanning broad time and length scales (**Figure 3.1**). Thus a key challenge in the field is to understand, in quantifiable terms, how the aggregate effects of molecular-level interactions manifest at higher levels of organisation. To address this challenge, the systems biology community has begun to develop theoretical models that describe cellular, tissue, organ and organism level processes concurrently; an approach more commonly referred to as multiscale modelling (Vicini, 2010). In this chapter we discuss the utility of multiscale models, how they are developed and what the limitations of this approach are. Following this, we present a framework to guide the development of multiscale models and use this framework to develop a Domain Model of CXCL13-mediated regulation of B-cell activation in lymph nodes.

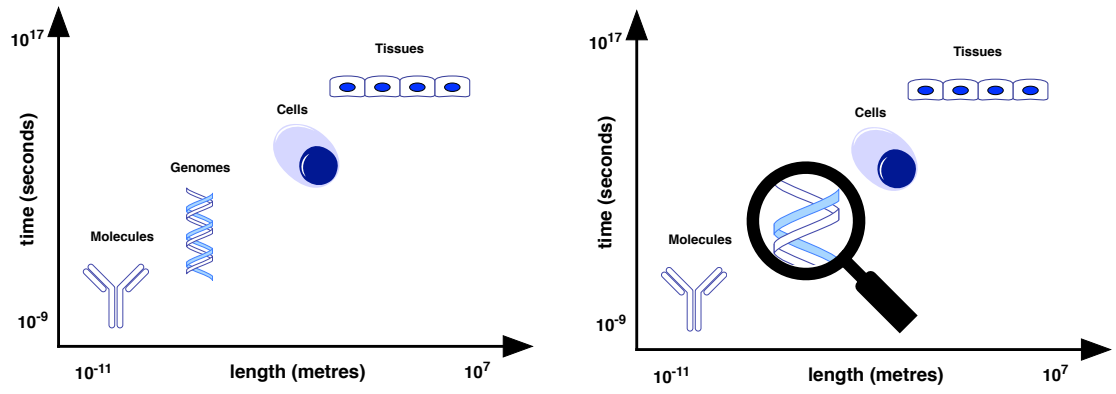

Figure 3.1. Biological processes occur across a broad range time and length scales. However, experimental methods, particularly imaging modalities, are optimised to measure specific time and length scales. This can make interpretation of a single dataset in the context of the wider literature challenging.

3.1.2 Multiscale Models: Advantages

Recent technological advances in microscopy have led to the unprecedented availability of data on the dynamics of B-cell migration and selection. However, as imaging techniques are optimised for a given time and length scale, they become increasingly limited in their ability to link molecular, cellular and tissue level processes (**Figure 3.1**). To address this issue, modelling approaches have been used to test the validity of different hypotheses for mechanisms controlling B-cell migration derived from multiphoton imaging data and to study the dynamics of G-Protein Coupled Receptors (GPCRs) responsible for B-cell migration within the Germinal Centre (Chan et al., 2013; Meyer-Hermann, 2006; Meyer-Hermann et al., 2012).

Multiscale modelling has also been applied to oncology, where clinical trials have the highest failure rate in comparison to other therapeutic areas (Begley and Ellis, 2012). In complex processes such as tumour formation, probing targetable mechanisms can be difficult owing to heterogeneity arising on multiple scales; cancerous cells adapt at genetic and molecular scales to survive in dynamic environments, altering cellular phenotypes and therefore treatment efficacy. In this context a hybrid agent-based approach incorporating a pharmacokinetic / pharmacodynamics (PK/PD) model was developed to explore the dynamics of tumour growth, as well as the penetration of the bio-reductive drug tirapazamine (TPZ) (Kazmi et al., 2012). This case study exploited the ability of multiscale models to consolidate information across different disciplines, affording a quantitative perspective from which to combine insights from cancer biology with pharmacological understanding of the therapeutic agent. In this study, a PK/PD approach was applied to model factors such as hydrogen ion production, nutrient distribution, and drug concentration, while an agent-based approach was used to model each individual cell over space and time, capturing interactions between cells within the tumour microenvironment. Combining both approaches showed that the drug was incapable of reaching the edge of the hypoxic region of the tumour, due to consumption of the drug as it diffused into the tumour.

3.1.3 Limitations of Multiscale Models: Focus on Uncertainty

The aforementioned case studies demonstrate the efficacy of multiscale models in bridging understanding across spatiotemporal scales and scientific disciplines. However, multiscale models require significant time and resource investments to engineer and evaluate, a core limitation restricting their wider use (Kirschner et al., 2014; Vicini, 2010). This issue is exacerbated by the lack of standardized practices and a well-defined validation and certification framework, without which it is difficult for regulation bodies to assess the validity of simulation-derived results.

Consequently, sources of uncertainty arise at all stages of the model development process. Datasets and published findings used to inform model development may come from many different experimental systems and organisms. This is particularly pertinent to multiscale models that require large amounts of data to develop. However, studies conducted by Amgen (Thousand Oaks, CA), with a view to confirming published results important to their R&D efforts, could only reproduce similar findings in 11% of cases (Begley and Ellis, 2012).

As the complexity of scientific software increases, greater emphasis must be placed on designing system architecture and testing strategies to avoid implementation errors and a code base that is difficult to maintain and repurpose. This is particularly pertinent to academia where there is typically: *(i) a high turnover of workers; (ii) a wide range of scientific skills and coding proficiency; (iii) the need for a maintainable and extensible code base that will outlive particular research grants and (iv) changing requirements that are driven by new scientific discoveries (Pitt-Francis et al., 2009).* Given these challenges we note the following features of academic software*: (i) they are not generic; (ii) they have not been completely tested and validated; (iii) they are not freely available to the scientific community and (iv) they do not include state-of-the-art techniques to improve performance* (Pitt-Francis et al., 2008).

Lastly, as the dimensionality of the parameter space increases, or it becomes necessary to capture stochastic and heterogeneous behaviours, it becomes less tractable to robustly evaluate model behaviours within a time frame that can run parallel to laboratory or clinical studies. For example, agent-based simulations that capture both stochasticity and

heterogeneity can require hundreds of replicates to generate a representative output for a single parameter set (Alden et al., 2012b; Read et al., 2013a). In such settings, quantifying parametric uncertainty is time intensive, even with significant computational resources available. The extended Fourier amplitude sampling test (eFAST) analysis that partitions variance in simulation response between parameters of interest (Saltelli et al., 2008). However, the characteristics of this sampling technique (detailed in **Chapter 2.2.10**) give rise to a significant number of parameter sets, totalling 1,170 for a six-parameter study. Models with a long execution time may therefore limit the application of certain statistical analyses necessary to understand simulation behaviours and translate simulation-derived results back to the problem domain.

3.1.4 Limitations of Multiscale Models: Approaches to Address Uncertainty

Although there is currently no clear consensus on a standardized protocol for developing multiscale models for the biomedical sciences, some methodologies exist. Grimm et al. proposed a "three-block" standard protocol for the development of ABMs termed ODD (overview, design concepts, and details). The CoSMoS (Complex System Modelling and Simulation) process is a framework for the iterative development of complex systems models, in which model development is divided into 5 key phases (detailed in **Chapter 2**) (Alden et al., 2012b; Bown et al., 2012; Read, 2011). Irrespective of the approach-specific semantics, three core components remain consistent across most methodologies: *(i) designing a model, (ii) implementing a model, and (iii) validating a model*. To address uncertainty associated with designing a model, increased focus has been placed on closer integration of experimental and theoretical approaches, while argumentation structures have been applied to document all data used to inform a model as well as key assumptions made (Alden et al., 2015). In terms of model implementation, hybridisation has enabled researchers to tune model granularity at distinct scales to facilitate parsimony and boost performance (Guo and Tay, 2008). Lastly, at the model validation phase, extensive research focus has led to the development of a suite of statistical techniques to better relate simulator outputs to the research domain (Alden et al., 2013; Read et al., 2012). Despite this progress, there is a dearth of studies that look at how uncertainty can arise when the developing the computer code associated with a multiscale simulator or reducing the time taken to evaluate simulator behaviours.

3.1.5 Summary and Aims of the Chapter

Multiscale models afford a quantitative perspective from which to consolidate information across broad spatiotemporal scales and scientific disciplines. However, development of multiscale models is technically challenging, requiring significant time and computational resources. In this chapter, we aim to adapt the current instantiation of the CoSMoS process, to account for the challenges associated with developing multiscale models (**Chapter 3.1.4**). Using this updated instantiation seek we develop a conceptual model (herein referred to as a Domain model, as detailed in the CoSMoS process (**Chapter 2**)) of the pathway described in **section 3.4.1** using visual notations. In brief our aims were:

1. Update the current instantiation of the CoSMoS process to address issues associated with developing multiscale models (**section 3.2-3.3**).

2. Summarise current understanding of CXCL13-mediated regulation of B-cells in a Domain Model through visual notations (**section 3.4**).

3. Use the Domain Model to identify key entities and time points of interest as well as knowledge gaps (**section 3.4.3**).

3.2 Incorporating Agile Techniques into CoSMoS

Within the York Computational Immunology Lab, researchers have employed the CoSMoS process to design, develop and validate complex systems models (Alden et al., 2012b; Bown et al., 2012; Read et al., 2013a). While providing a generalised framework to build models, application of CoSMoS in a given research context should be tailored to suit the criticality and intended impact of the underlying research. Through the development of several case studies on the immune system, a novel instantiation of the process has evolved (Alden et al., 2012b; Read et al., 2013a). A core feature of this instantiation is the incorporation of engineering-derived approaches: *(i) the use of visual notations to describe models succinctly (Alden et al., 2015; Read et al., 2014) (ii) advanced statistical methodologies to deal with issues of domain, parameter and aleatory uncertainty (Alden et al., 2013; Read et al., 2012; Williams et al., 2016) and (iii) improving model parsimony or computational efficiency through model hybridisation (Guo et al., 2008b) and tuneable resolution (Kirschner et al., 2014).*

To extend this instantiation, we incorporate agile development approaches to address issues associated with developing software implementations of multiscale models. Core terminology relating to this updated instantiation is presented in **Table 3.1**. Although employed sparingly, agile approaches are well suited to the needs of small academic teams working within a highly dynamic problem domain (Pitt-Francis et al., 2009). A key case-study informing the incorporation of agile approaches is Chaste (cancer, heart and soft-tissue environment), a software library and a set of test suites for computational simulations of cancer, cardiac physiology and soft-tissue mechanics (Pitt-Francis et al., 2008). This software platform was developed with contributions from researchers across several disciplines with as much code modularity and reuse as possible. An independent review of the underlying source code from an industrial collaborator suggested that the framework performs very well in terms of code quality, readability, software architecture and general software engineering. Another case study for the use of modular design is a whole cell computational model of the human pathogen *Mycoplasma genitalium (Karr et al., 2012a)*. In this study the authors took a hybrid modelling approach, encapsulating functionality into distinct modules to expedite the development process.

Entity	An independent element of the model, such as a cell or protein.
Module	A self-contained model subunit that can be used to construct a more complex model
Modular	Model structure consisting of several standalone subunits that can function independently, but may be linked together to form a more complex representation of the system.
Cohesion	The degree to which elements of module are directed towards a single focused task
Coupling	The degree of interdependence between modules
Model	A non-executable description of a system, which may be described in an abstract manner, or for a platform-specific implementation as a simulation.
Simulation	An executable implementation of a model; a software platform capable of imitating the behaviour of a system to study its form and function.
Hybridisation	Using a combination of modelling techniques concurrently, to capture aspects of the system at different scales in a tractable manner and to overcome the limitations associated with using each technique in isolation.
Multiscale	A model combining processes occurring across multiple time and length scales.
ATDD	An agile approach to software-development in which testing and code development are interleaved. In this approach the functional requirements of the system, are specified as acceptance tests. Code is subsequently developed and tested incrementally, not proceeding to the next increment until all code passes associated tests.
Acceptance Test	A test conducted to determine if the requirements of a contract or specification are met
Unit Test	A test to verify a specific stand-alone function of a software method or class
Integration Test	A test to verify that modules are compatible, interact correctly and transfer the correct data across their interfaces at the correct time with respect to a given acceptance test

Table 3.1. Key definitions pertaining to the agile development workflow

Modularisation is a software design technique that emphasizes separating the functionality of a program into independent subunits (**Table 3.1**), with each distinct subunit responsible for only one aspect of the desired functionality (Baldwin and Clark, 2000). Modular design is not itself an agile approach but a fundamental engineering principle that can be applied in an agile way. Hybrid schemes are prone to having highly-coupled code that can be difficult to validate and to diagnose errors. It is also difficult to extend the functionality of highly coupled software; architectural decisions made in the past become difficult to change. Applying a modular design approach promotes looser coupling and higher cohesion between software subunits (**Table 3.1**), making code more reusable and facilitating model hybridisation (Baldwin and Clark, 2000).

Through the development of distinct modules, development can proceed in an incremental and iterative fashion where each component is designed and tested in isolation. With well-defined interfaces for how these modules interact they can be successfully integrated and separated at later points in time. To increase confidence in the function of each distinct module and the integrative system, test-first approaches are warranted.

Test-driven approaches to software development are characterized by the close integration of testing into code development. In this approach, tests are written before the code. Code is then developed and tested incrementally, not proceeding to the next increment until all tests are passed. Broadly speaking, testing can be split into two categories: validation and verification tests. Validation tests check that the specification captures the functional requirements of the software, while verification ensures that the software meets the specification. An adaptation of TDD, known as acceptance test-driven development (ATDD), incorporates both types of test to assess both the functional requirements of the software and the underlying code base (**Table 3.1**). In a multiscale modelling context, the ATDD approach affords the following advantages: *(i) ATDD can assess both engineering and biological aspects of the simulator (ii) ATDD provides a testable definition of fit for purpose (iii) argumentation structures can be built around testing data (iv) acceptance tests can inform the choice of modelling technique.*

Within our ATDD approach are 3 distinct levels of tests *(i) unit (ii) module and (iii) system/acceptance level* **(Figure 3.2a)**. Unit tests assess a specific stand-alone function of a module such as a method or a class with well-defined and simple functionality. Modules

involve a set of core data structures that are manipulated by a sequence of methods. Modules are tested by inputting a range of values and ensuring that outputs are consistent with expected values (examples provided in **Appendix 3**). System level or acceptance tests are designed to ensure that the simulator is a fit for purpose representation of the problem domain (**Figure 3.2a**). This yields a hierarchical testing strategy (**Figure 3.2b**) designed to produce a robust but malleable code base capable of responding effectively to changing requirements.

3.3 Development Workflow to Design and Implement CXCL13Sim

Initial stages of model design involve identifying and specifying the research focus through an expected behaviours diagram (Read, 2011) and performing a functional requirements analysis to identify core functionality that would provide a fit for purpose representation of the problem domain. Following these initial scoping exercises, the domain and platform models were developed and specified using an adaptation of the Unified Modelling Language (UML). Using this approach, we were able to effectively describe and communicate low-level behaviours in complex systems (Read et al., 2014).

The Simulation Platform begins with development of a hierarchical testing scheme in which unit and module (verification) tests are grouped under an associated acceptance test[14] (**Figure 3.2**). Tests are continually assessed and refined as the project progresses and are incorporated into an automated regression framework to ensure that new code does not disrupt existing functionality, expediting the development process.

To argue that the simulator fulfils its remit, acceptance tests, key design decisions, and information used to inform the design, development and validation of the model and simulation are presented as arguments over evidence using a visual notation derived from goal structuring notation (GSN) and can be opened using the GSN visualisation tool ARTOO (detailed further in **Chapter 2**) (Alden et al., 2015). In the following section, we use this development workflow to detail a Domain Model of CXCL13-mediated B-cell migration.

[14] Acceptance tests were informed by the functional requirements analysis

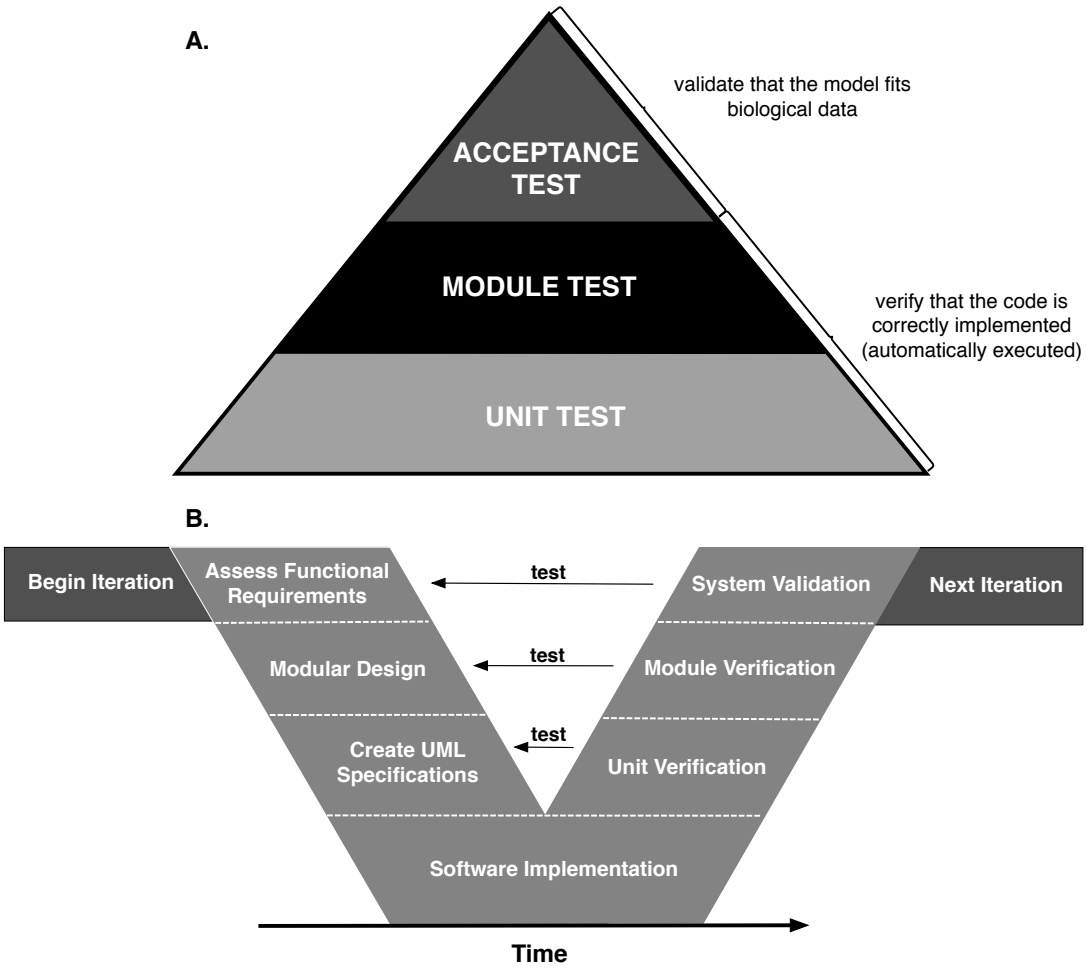

Figure 3.2. Agile Development Workflow. (A) In this instantiation each acceptance test is coupled to a set of verification tests at the unit and component level as part of a hierarchical testing framework. Unit tests are designed to test the core functionality of a stand-alone module while component tests are designed to test if modules are compatible, interact correctly and transfer the correct data across their interfaces at the correct time. (B) The workflow for each iteration is as follows: The functional requirements of the software are assessed and agreed upon with the Domain Expert. Subsequently, The system is divided into core modules, with system architecture mapped through an adaptation of the UML. At the Simulation Platform stage of development functional requirements were formalized as a set of acceptance test cases designed to validate the system. For each acceptance test, coupled sets of module and unit tests are specified. Code is then developed and tested incrementally, not proceeding to the next increment until all tests are passed.

3.4 Domain Model Development

3.4.1 Timeline of B-Cell Activation

Following immunization and infection, complement activation and deposition onto microbial antigens contributes to their efficient transport to the LN in a soluble form or through active transport by Dendritic Cells (DCs). Within the lymph node, antigen smaller than 70 kDa with a molecular radius less than ~4 nm filters into the parenchyma through reticular conduits or small pores in the subcapsular sinus (Roozendaal et al., 2008). The conduit system is guarded by PLVAP, a protein that forms a molecular sieve designed to regulate the parenchymal entry of lymphocytes and soluble antigens (Rantakari et al., 2015). However, not all lymph-borne molecules have free access to the lymphocyte compartment; larger antigen is retained at the subcapsular sinus, where macrophages and DCs sample the lymph and remove microorganisms, larger molecules and debris (Roozendaal et al., 2009).

Within the B-zone a dense web-like network of non-haematopoietic stromal cells are found. This network comprises 3 distinct but interconnected $CXCL13^{+}$ cell types: Marginal Reticular Cells (MRCs), B-zone Reticular Cells (BRCs) and Follicular Dendritic Cells (FDCs). FDCs not only express CXCL13 but also act as an important reservoir for trapped antigen that can stimulate naïve B cells (El Shikh et al., 2010).

Concurrently, B cells migrate across the endothelium into the cortex through the HEV at a rate dependent on the local vasculature. The arrest of blood-borne B cells in HEVs requires binding of the chemokine receptors CXCR4, CCR7 and to a minor extent CXCR5, leading to activation of the adhesion molecule LFA-1 that binds to ICAM-1 and ICAM-2 (Coelho et al., 2013). Once inside the lymph node, access to the follicle requires the G-protein-coupled receptor (GPCR) CXCR5 and is promoted by ICAM-1-expressing fibroblastic reticular cells (FRC) of the T-cell area which act as guidance structures (Coelho et al., 2013).

As B cells scan the follicle they respond to ligands for the receptors CXCR5 and EBI2; this promotes contact with FDCs and cells located around the follicle perimeter, including sinus associated macrophages, MRCs and DCs (Batista and Harwood, 2009; Carrasco and Batista, 2007; Pereira et al., 2010; Phan et al., 2007). Non-cognate interactions with

antigen via complement receptors CR1 and CR2 facilitate the mass transport of opsonized antigens from the exposed follicle perimeter to the protected centre for long-term display on FDCs, where they are trapped in the form of iccosomes by complement receptors present on the dense network of FDC processes (El Shikh et al., 2010).

If lymphocytes fail to recognize cognate antigen within a few hours to days, they return to the circulation in a sphingosine-1 phosphate receptor 1 (S1PR1) dependent manner through efferent lymph vessels and the thoracic duct (Grigorova et al., 2010; Matloubian et al., 2004). Notably, naïve B cells are resident in the LN longer than either $CD4^+$ or $CD8^+$ T cells (Tomura et al., 2008). However, if a B-cell does meet its cognate antigen it can internalise the antigen through its B-cell receptor. It can then degrade pathogen proteins into peptides for display on Major Histocompatibility Complex Class II (MHC-II) molecules on the B-cell surface. Following antigen priming, a B cell upregulates CD86 and CD80, proteins that provide co-stimulatory signals for T-cell activation while upregulation of the lymphocyte activation antigen CD69 inhibits the egress activity of S1PR1, leading to retention in the lymph node (Pereira et al., 2010). To maximise encounters with T cells, antigen-primed B cells exhibit a reduction in migration velocity, upregulate CCR7, (while CXCR5 expression remains unchanged) distribute themselves along the B/T border as a result of the balanced chemoattraction of CCL19/CCL21, CXCL13 and 7α,25 hydroxycholesterol (Pereira et al., 2010).

Helper T cells, primed by dendritic cells earlier in the infection, migrate around the T-zone in a CCR7 dependent manner. A subset of activated $CD4^+$ T cells upregulate CXCR5 and reduce CCR7 expression allowing them to migrate towards the edges of follicles to provide help to B cells. This activation occurs via CD40 ligation subsequently driving the resting B-cell into the cell-cycle and upregulating the transcription factor bcl6, therefore reducing the propensity to undergo apoptosis (Kitano et al., 2011). Activated B cells migrate to interfollicular and outer follicle regions after receiving T cell help, where they can undergo proliferation for 1-2 days before returning to the centre of the follicle in an S1PR2 dependent process to initiate GC clustering and acquire a GC phenotype associated with, amongst other molecules, the upregulation of the chemokine receptor CXCR4 and the glycan moiety GL7 (Allen et al., 2007; Chan and Brink, 2012). If BCs fail to acquire this TC help they revert back to a naïve phenotype, downregulating CCR7 and CD86 (Turner et al., 2017).

3.4.2 Characterising Current Understanding Through Visual Notations

To consolidate current understanding of this pathway we develop a Domain Model. The aim of this model is to summarise current understanding of the pathway with respect to the following question: *How does CXCL13 regulate the positioning of cognate B cells within primary lymph node follicles?* We define the research context using an expected behaviours diagram (**Figure 3.3**) (Read, 2011). This figure summarises the research scope, highlighting key entities and interactions hypothesized to give rise to the observant phenomena described in **section 3.4.1**. Using this diagram as a high-level reference, we further detail the behaviours of model entities using state-machine and activity diagrams (detailed in **Chapter 2.2**) **(Figures 3.4-3.8)** (Read et al., 2014). Key design decisions, including abstractions made and data used to inform the model, are presented as arguments over evidence. A subset of the argumentation structure for the Domain Model is presented in **Figure 3.9**, with the entire structure available online[15].

Through specification of the Domain Model we identify a number of key limitations to our understanding of the B-cell activation pathway. Specifically, the source of CXCL13 within the primary follicle is unclear. FDCs are implicated in the secretion of CXCL13 within the GC but their role within the primary follicle is less clear. In addition, the spatial distribution of CXCL13 is poorly understood. Mice deficient in the CXCL13 gene display poor follicular organisation but this system does not facilitate more a more nuanced understanding of the molecule. Lastly, a quantitative description of lymph node architecture is lacking with no quantification of follicular structure, size or cellularity.

3.4.3 Functional Requirements Analysis and Modularisation of the Pathway

Given the research scope defined in the expected behaviours diagram (**Figure 3.3**) the functional requirements of the system, key objectives required to provide a fit for purpose representation of the problem domain, are: *(i) to have a high fidelity representation of murine popliteal lymph node follicle size and B-cell density; (iii) to have a high fidelity representation of CXCL13 secreting stromal cells within the follicle and (iii) median values of B-cell migration metrics should display no statistically significant difference to those measured in vivo for wild-type or CXCR5$^{-/-}$ mice (Coelho et al., 2013).* The

[15] https://www.york.ac.uk/computational-immunology/software/cxcl13sim/

functional requirements analysis facilitated the identification of three core modules: *(i) chemokine, (ii) B cells, and (iii) stromal cells*.

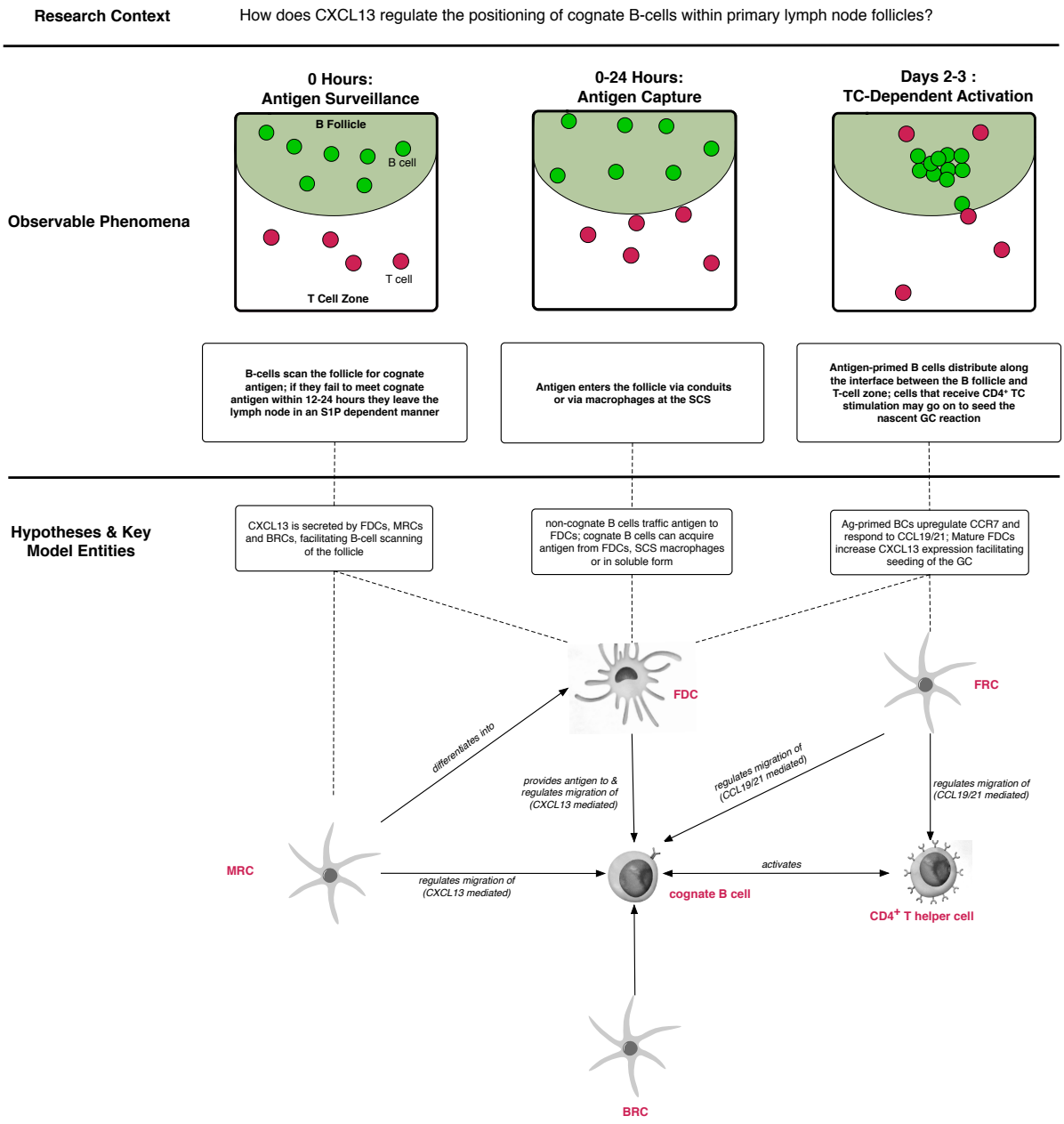

Figure 3.3. Expected Behaviours Diagram for CXCL13Sim. In this diagram we present the research context within which the model is developed. This diagram specifies the key emergent properties of the system as well as the low level mechanisms reported to give rise to these properties. Lastly, it defines key entities in the model and the relationships between them.

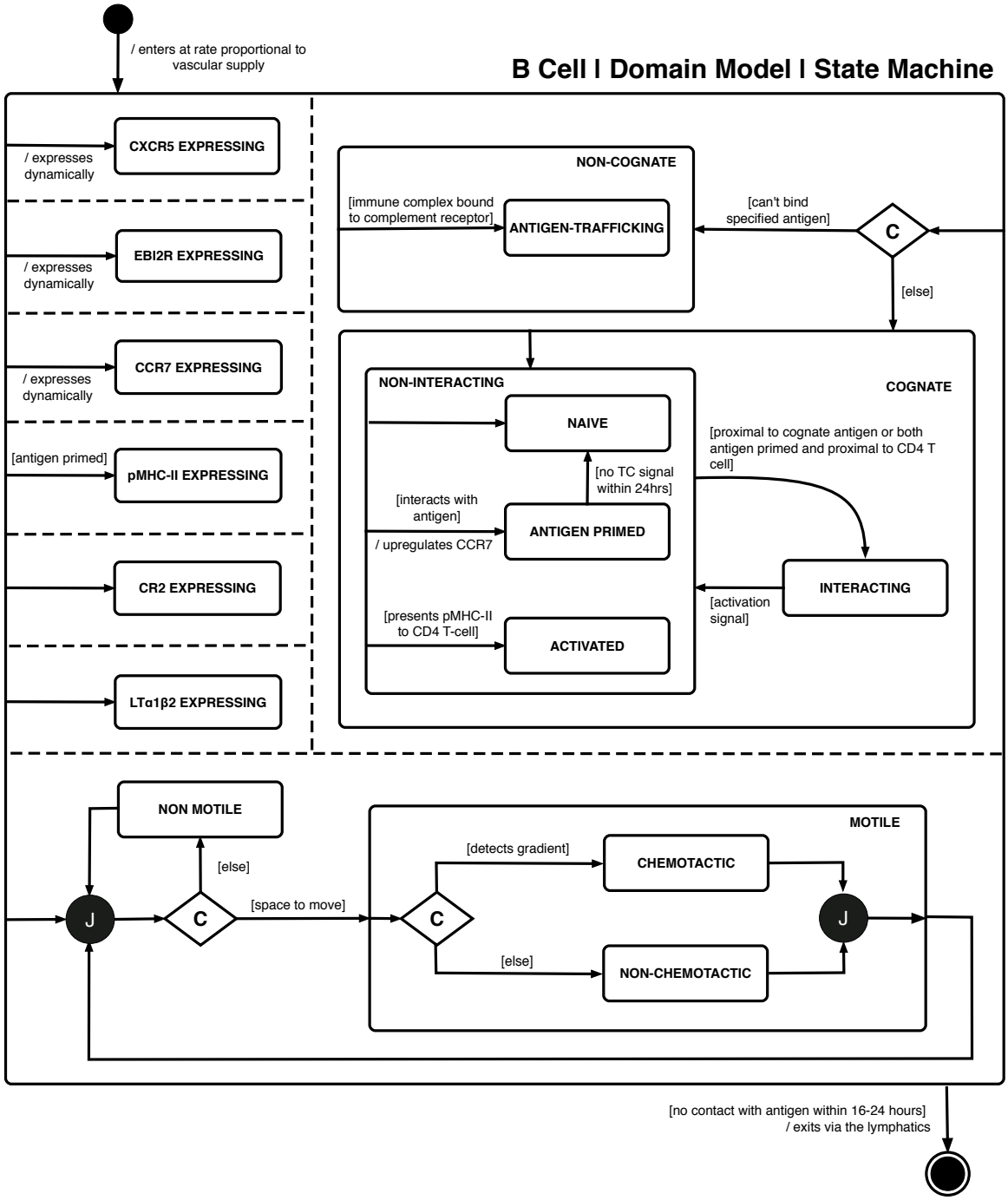

Figure 3.4. State Machine Diagram for a B cell: A B cell enters the LN at a fixed rate dependent on vascular supply, once in the LN a cognate B cell can become MHC expressing if it encounters antigen. Each B cell dynamically expresses CXCR5 and EBI2R and through detection of chemokine gradients can decide to move chemotactically or randomly if there is sufficient space to move. B-cells continue to express LTα1β2 at a fixed level and if 12-24 hours have passed and no interactions have occurred, the B cell exits the LN. If however, a BC encounters antigen it becomes antigen-primed, upregulating pMHC-II and CCR7 to facilitate interactions with T cells.

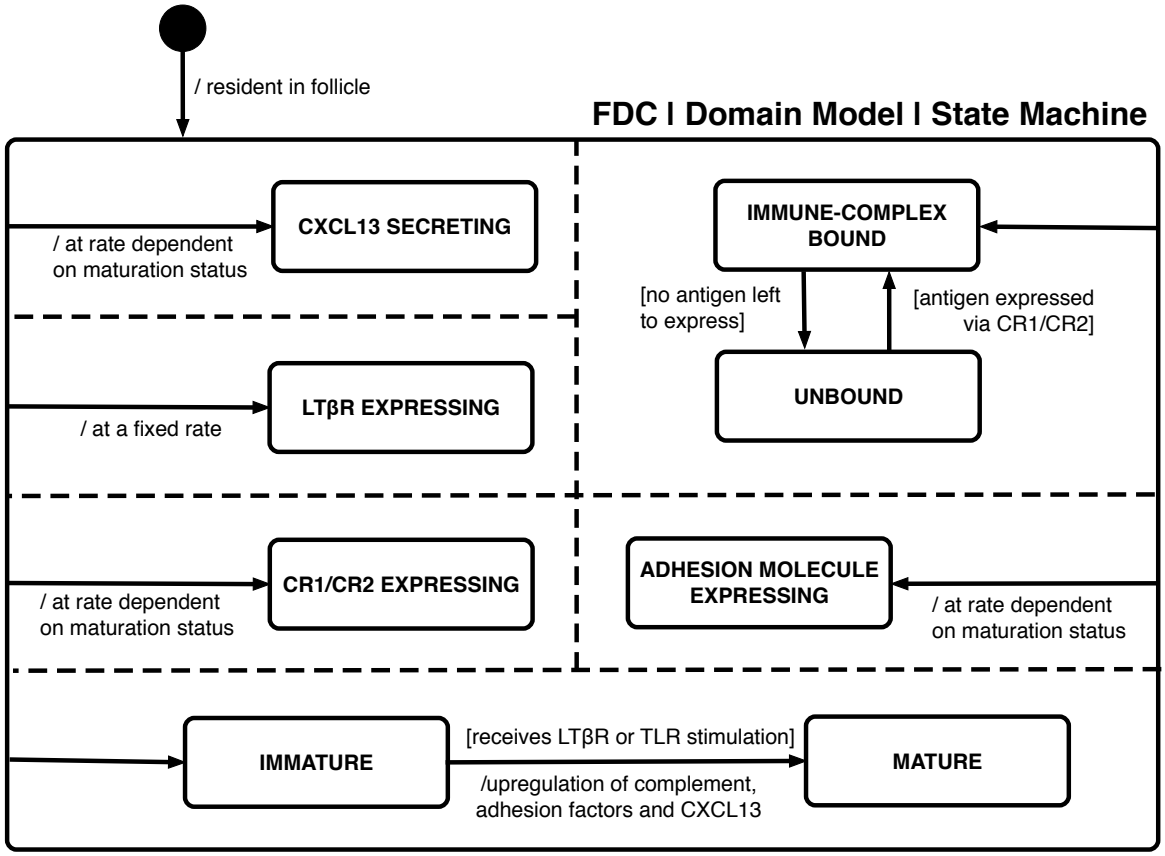

Figure 3.5. State Machine Diagram for a FDC: FDCs are resident in the system and are antigen presenting at time zero. FDCs secrete chemokine and express LTβR, complement receptors and adhesion molecules at a rate dependent on maturation status. FDCs also retain and present antigen unless cognate B cells capture all antigen

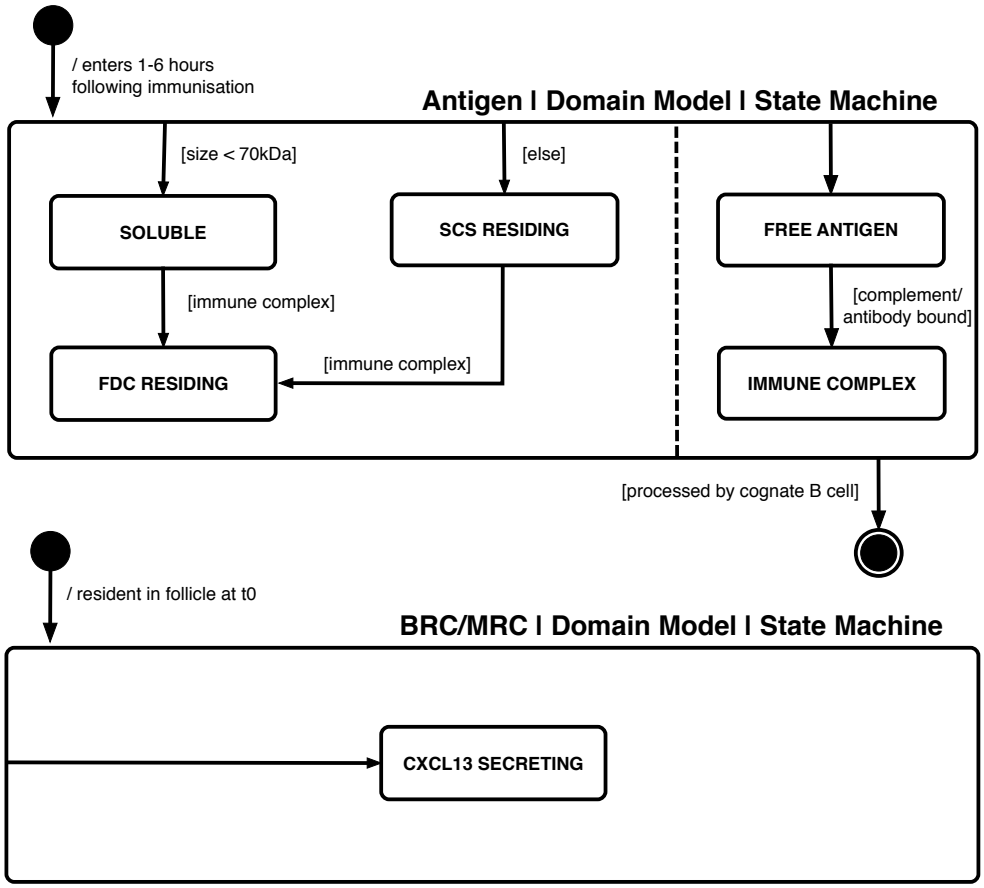

Figure 3.6. State machine diagrams for antigen, BRCs and MRCs: Antigen enters the follicle 1-6 hours following immunisation. It enters in a soluble or SCS residing form but may be transferred to FDCs if part of an immune complex. BRCs/MRCs are resident in the follicle at time zero and secrete CXCL13 at a fixed rate. SCS macrophages are in the system at time zero and can present antigen.

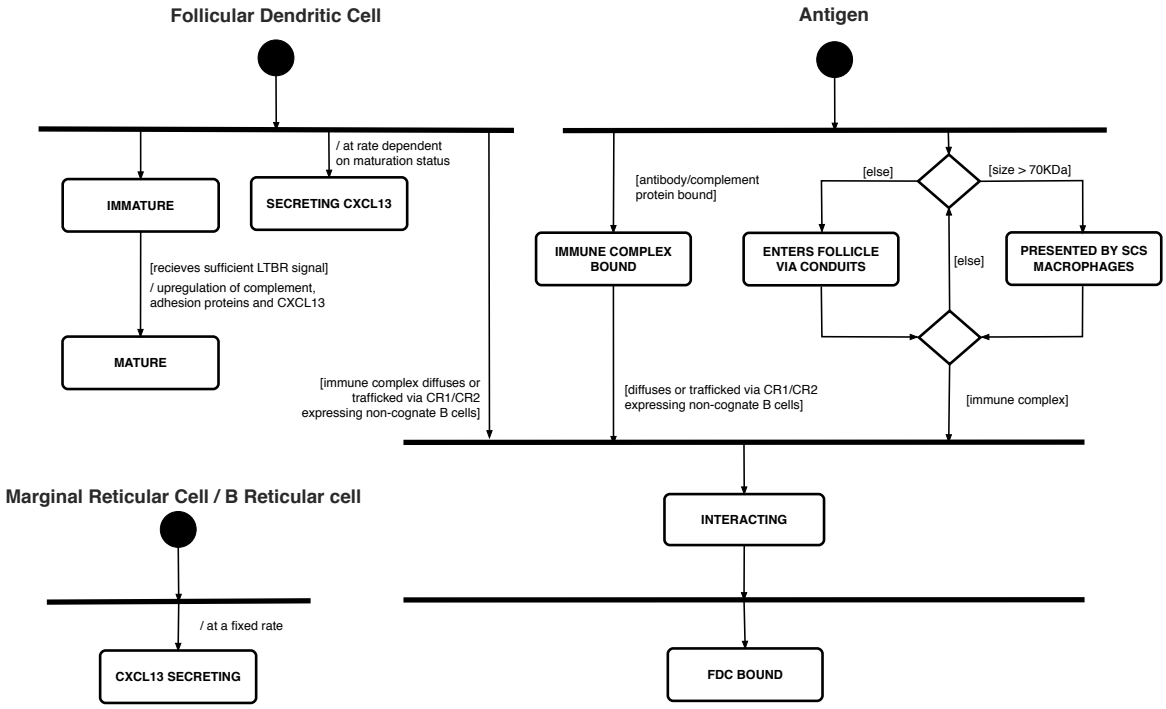

Figure 3.7. Activity Diagram 1: These diagrams detail the activities performed by each model entity and the conditions required for a change of activity to occur. At the beginning of the activity FDCs exist in an immature state but mature if sufficient LTβR signal is received. CXCL13 is secreted at a rate dependent on maturation status. Antigen enters the system via the conduits or may be presented by SCS macrophages if greater than 70kDa. If antigen is part of an immune complex it may become FDC bound. BRCs and MRCs secrete CXCL13 at a fixed rate.

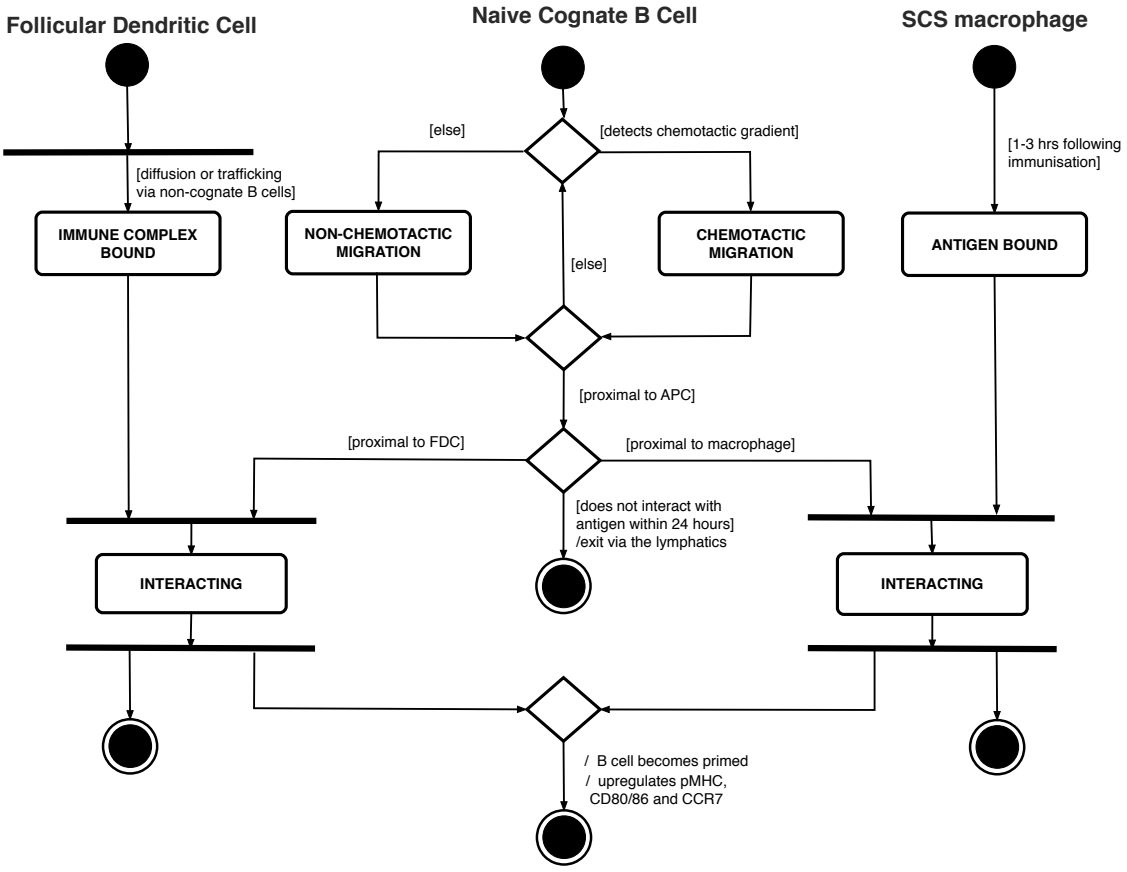

Figure 3.8 Activity Diagram 2: Activity diagrams detail the activities performed by each model entity and the conditions required for a change of activity to occur. At the beginning of the activity FDCs express antigen unless cognate B cells capture all antigen. Under homeostatic conditions FDCs secrete CXCL13 at a fixed rate. If a B cell expresses CXCR5 then it can orientate itself along a CXCL13 gradient, otherwise it randomly orientates itself and if there is free space available can migrate. Once proximal to an FDC/SCS macrophage a cognate B cell may interact with antigen via its B-cell receptor and express it via MHC-II.

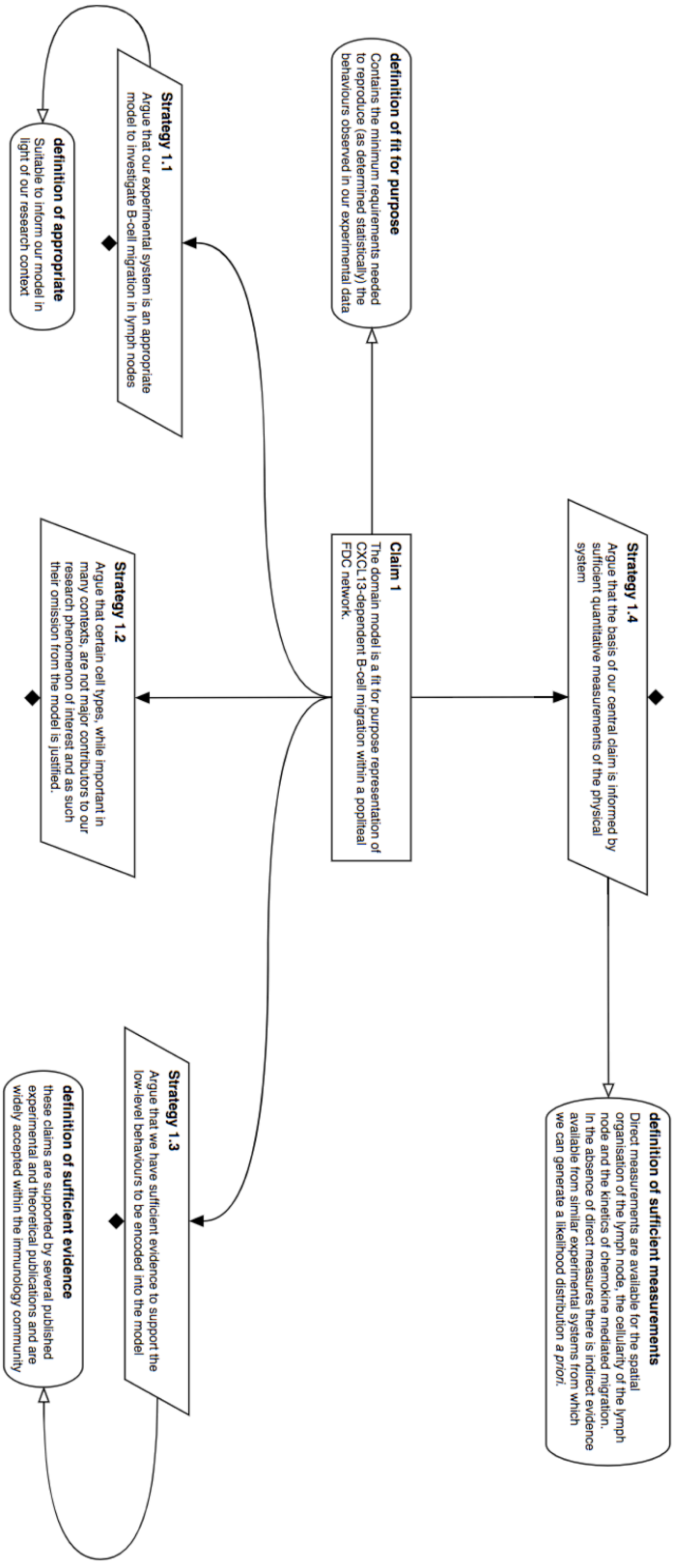

Figure 3.9 A subset of the argumentation structure for the Domain Model. The argumentation structure contains a central claim arguing that the Domain Model provides a fit for purpose representation of the problem domain. To provide arguments over evidence for our central claim we have 4 key strategies that have an associated set of claims and evidence nodes (not shown). The fully expanded argument structure is available from https://www.york.ac.uk/computational-immunology/software/cxcl13sim/.

3.5 Discussion

In this chapter we have discussed some of the key advantages afforded by a multiscale modelling approach. However, the need for significant time and resource requirements to design, implement and validate multiscale models limits their wider application. A key bottleneck in the development pipeline occurs during the implementation of multiscale models as executable software platforms. Given the highly dynamic nature of biomedical research programmes, multiscale software must be sufficiently lightweight and flexible to cope with constantly evolving requirements. Software that is difficult to test, redesign and refactor may limit the use of a simulator as a decision-making platform to complement an on-going laboratory or clinical study.

To address this issue we have extended an existing instantiation of the CoSMoS process through the incorporation of modular design and ATDD. These agile techniques are particularly suited to the development of hybridised models in that distinct models can be coupled together while promoting loose coupling and high cohesion between software subunits. This allows specific elements of the model to be validated in isolation such they can be more readily refactored and repurposed. Coupling of module and unit tests to acceptance test as part of a hierarchical testing strategy ensures that both engineering and biological aspects of the simulator are tested. An additional benefit of the test-driven approach is that it provides a means of testing whether the simulator is a fit for purpose representation of the problem domain. This feature may be useful in safety-critical contexts such as clinical trial design, where clear endpoints and objectives are agreed upon before the trial begins and the simulator can only be deemed fit for purpose if it can pass tests associated with pre-specified requirements.

Through development of a Domain Model of the pathway we have systematically scanned the existing literature. This led to the development of a conceptual model of CXCL13 mediated regulation of B-cell migration expressed through an adaptation of the UML, an expected behaviours diagram and an argumentation structure. The use of visual notations facilitated the communication of large amounts of data in a clear and concise manner that is accessible to interdisciplinary teams.

This process identified 3 core components of the pathway that we have specified as modules and has identified a number of key knowledge gaps that limit our understanding. Despite the extensive body of literature on this pathway, there are relatively few quantitative analyses of how these components interact, particularly in the context of the homeostatic follicle. We have found limited measures of follicular architecture and cellularity with no direct measures of the cellular sources or kinetics of CXCL13 within the follicle. The lack of such measures ultimately limits our ability to predict how CXCL13 is implicated in more complex situations such as autoimmune disease or following immunization and infection. In the following chapter we design experiments to address these knowledge gaps and further our understanding of the pathway.

Chapter 4 Quantifying Key Components of CXCL13 Bioavailability

4.1 Introduction

In the previous chapter we have identified key limitations to our understanding of CXCL13-mediated regulation of B-cell activation. In this chapter we focus on imaging and cytometry approaches to enumerate key components of the pathway. As imaging techniques are adapted and improved over time, they become increasingly specialised for a given task**;** physiological context must often be sacrificed to increase resolution or *vice versa*. One must consider the strengths and limitations of each approach with no single technology capable of characterising an entire pathway. To address this issue, imaging studies can be enriched through use of an ensemble of techniques, with data consolidated through advanced mathematical analyses. In the following section, we discuss key approaches to study the knowledge gaps identified in the previous chapter.

4.1.1 Imaging and Graph Theory Based Approaches to Study the Architecture of Stromal Networks

A key challenge in contemporary biology is to understand how grouped biological entities interact to control the behaviours of cells, tissues and organisms. Network theory has yielded the insight that despite the diversity of networks that occur in natural systems, their form and function is governed by a few simple principles (Barabási and Oltvai, 2004). Random networks have no underlying organization with edges and nodes randomly assigned to one another with degree centralities, the number of edges per node, following a Poisson distribution (Barabási and Oltvai, 2004). Random networks can be generated using the Erdos-Renyi model but typically do not occur in natural systems (Erdős and Rényi, 1960). Scale-free networks[16], are networks whose degree distributions follow a power law, containing an abundance of high degree nodes or hubs, this infers robustness to the network where random node deletion is unlikely to disrupt the entire network (Barabási and Oltvai, 2004). Examples of scale-free networks are the internet and protein interaction networks (Albert, 2005) or theoretical networks generated using the Barabasi-Albert model (Barabási and Oltvai, 2004). Small world networks are another class of networks and are characterized by low shortest path lengths through the network and a high clustering coefficient. This leads to localized cliques, and rapid information

[16] Called scale-free because power laws have the same functional form at all scales. I.e. zooming in on any part of the distribution doesn't change its shape: there are many nodes of low degree and a small, but significant number of hubs at each level of magnification.

transfer through the network; consequently a viral infection can spread more quickly through a network with small-world properties (Watts and Strogatz, 1998). Examples of small world networks include social networks (Telesford et al., 2011), airline routes (Telesford et al., 2011), fibroblastic reticular cell networks (Novkovic et al., 2016) and theoretical networks generated using the Watts-Strogatz model (Watts and Strogatz, 1998).

A preliminary study of FRC architecture exploited network theory to understand the migration of T cells. In this study the FRC network was modelled as a graph and *in silico* cell migration was confined to the edges (Donovan and Lythe, 2012). This model suggests that cell-turning angles are an emergent property of migrating along FRC processes. However a key limitation of this model is the lack of quantitative data informing network properties and the assumption that T cells are confined to the network. Recently, Novkovic *et al* (2016) used confocal microscopy in combination with fluorescent reporter mice and network theory to understand the spatial organisation of the lymph node FRC network. Through topological mapping of the confocal microscopy data the authors were able to perform *in silico* analyses of network robustness to viral attack, with simulation predictions that lymph nodes are able to tolerate a loss of approximately 50% of total FRCs verified experimentally.

4.1.2 Theoretical and Imaging Approaches to Study Diffusion

Studies of diffusion in the physical sciences have led to mathematical descriptions of the process in the form of Fick's laws, developed by Adolf Fick in the 19th century. The first law states that the molar flux due to diffusion is proportional to the concentration gradient. Mathematically, we can formalize this as:

$$N_i = -D_i \nabla c_i \quad (4.1)$$

Where for species i, N_i is the molar flux and c_i is the concentration; the negative sign signifies that diffusion occurs towards decreasing concentrations. The second law states that the rate of change (first derivative of concentration with respect to time) in concentration at a point in space with respect to time is proportional to gradient (the second derivative of concentration with respect to space).

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} \quad (4.2)$$

The value D that relates the first and second derivatives is known as the diffusion constant. Using the Einstein stokes relation we can predict the value of the diffusion constant as follows:

$$D = \frac{kT}{6\pi\eta r} \quad (4.3)$$

Where *k* is the Boltzmann constant, *T* is the temperature, η is the viscosity of the medium in which the molecule is travelling, and r is the Stokes radius of the molecule. Once this constant is determined we can then predict the mean squared displacement of the molecule over time using the following expression:

$$\langle x^2 \rangle = 2dDt \quad (4.4)$$

Where x^2 is the distance squared, d is the number of dimensions. However, to apply this model we must assume that the viscosity of the medium in which the molecule is travelling is constant and that the molecule is undergoing free diffusion, and therefore no transient biochemical interactions. Given the density of the follicle microenvironment and the biochemical properties of chemokines this model is not appropriate to describe the diffusion of CXCL13 *in vivo*.

Currently, there are a number of experimental approaches to assess molecular dynamics at both the population and single-molecule level (**Figure 4.1**). Fluorescence Recovery After Photobleaching (FRAP) involves rapidly bleaching a sample with a high-intensity laser. The movement of unbleached molecules from neighbouring areas into the bleached area is then recorded by time-lapse microscopy (Phair and Misteli, 2001a). This method is minimally invasive but as an ensemble method lacks the sensitivity associated with single molecule approaches.

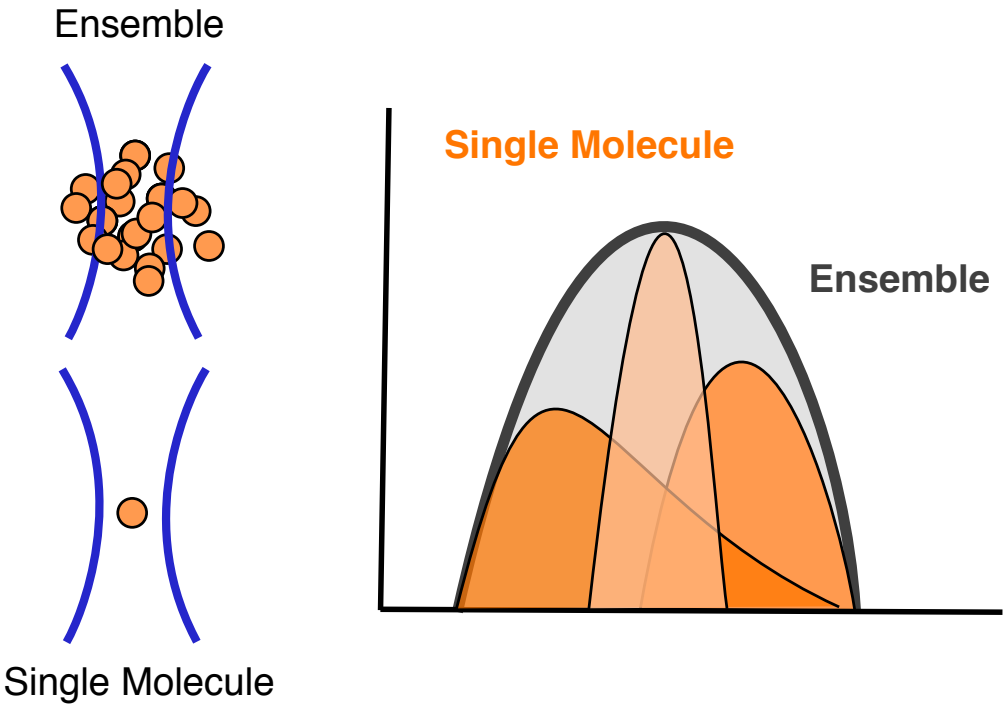

Figure 4.1 Comparison of single molecule and ensemble approaches in detecting heterogeneous behaviours. Ensemble approaches that measure the average properties of large concentrations of molecules have used extensively to analyse molecular mobility. However, single molecule approaches can identify a spectrum of diffusive behaviours within the same sample and thus provide a unique insight into molecular heterogeneity**.**

One approach to assess single molecules is Fluorescence correlation spectroscopy (FCS) (Dong et al., 2012), a technique characterised by the statistical analysis of the fluctuations of fluorescence in a system in order to decipher dynamic molecular events, such as diffusion or conformational fluctuations of biomolecules (Gryczynski, 2008; Phair and Misteli, 2001a). FCS measures fluctuations of fluorescence signal intensity in a small detection volume as molecules diffuse through via Brownian motion. This generates a fluorescence intensity profile from which we can determine an autocorrelation function. The autocorrelation function represents the correlation coefficient between the intensity at time *t* and all later time points. As the shift in time progresses, the correlation coefficient (R) decreases. R as a function of time then represents the probability that the correlation decays. This technique is useful as it is a non-invasive with minimal photobleaching, it also requires less fluorophores per field of view and thus physiological concentrations of proteins can be used. Using dual-colour fluorescence cross-correlation spectroscopy, Lilliemeier *et al* showed that identified distinct T-cell receptor (TCR) and linker for activation of T cells (LAT) domains which appear in distinct clusters on naïve cells come together upon TCR activation (Lillemeier et al., 2010). Key limitations for applications of FCS are the lack of spatial resolution and the need for robust and fast data analysis. This approach is typically used for homogeneous samples; however many molecules comprise multiple and diverse species resulting from protein interactions (association/aggregation) which increases the complexity of data analysis.

More recently, increased emphasis is placed on single molecule imaging approaches with high temporal resolution, given the high speed at which molecular interactions can occur. Despite technical advances, the direct visualisation of dynamic molecular interactions at the single molecule level is still a challenging imaging frontier. STORM has integration times of ~tens of milliseconds for individual image frames with full reconstructions taking several seconds. Scanning fluorescence methods such as STED are limited to ~1 Hz frame rates with faster imaging up to ~1,000 Hz possible by trading image quality, while widefield approaches such providing super-resolution information in living cells at ~ millisecond timescales(Plank et al., 2009; Reyes-Lamothe et al., 2010).

4.1.3 Summary and Aims of the Chapter

Development of a Domain Model (**Chapter 3**) identified a lack of quantitative measures of follicular architecture, cellularity and CXCL13 dynamics. We address this using an ensemble of imaging and cytometry approaches in collaboration with specialists in the fields of super-resolution imaging (Leake lab, University of York, UK), cell migration (Stein lab, University of Bern, Switzerland) and stromal immunology (Ludewig lab, Kantonsspital St. Gallen, Switzerland). Consolidating this work, we derive a quantitative description of the primary lymph node follicle. Our key aims are summarized as follows:

1. Quantify the network architecture of $CXCL13^{+}$ follicular stromal cells using network theory based analysis approaches (**section 4.2**)
2. Quantify the diffusion constant of CXCL13 in collagen matrix and lymph node using single molecule imaging approaches (**section 4.3**)
3. Derive a quantitative description of the canonical primary lymph node follicle (**section 4.4**).

4.2 The Microanatomy of CXCL13 Production in the Primary Follicle

To characterize the anatomy of CXCL13 secretion, we annotate 3D confocal images (typically 450µm x 450µm x 35µm) of lymph node follicles (provided by the Ludewig lab) taken from *Cxcl13*-Cre/Tdtomato *R26R*-EYFP (abbreviated as *Cxcl13*-EYFP) mice (Onder et al., 2017) (**Figure 4.2**). Briefly, this *in vivo* reporter strain of mice works as follows: CXCL13 promoter activity causes expression of the red fluorescent protein TdTomato, this in turn causes activation of a cre recombinase enzyme which removes a stop signal ubiquitously expressed on the rosa26 locus inducing expression of the fluorescent eYFP protein (**Figure 4.2**). EYFP thus acts as a lineage marker, being constantly expressed in cells that originate from a CXCL13 producing precursor while TdTomato expression is confined to cells with current CXCL13 promoter activity.

In this approach, (detailed further in **Chapter 2.1.7.3**) images are manually curated in the IMARIS image analysis software[17] package by marking the coordinates of each cell body and marking connections with lines if a physical connection in the form of a cell process exists between 2 cells bodies and no other cell body blocks the path (**Figure 4.3**). Mapping the stromal network in this way generates a topological mapping which can be further assessed using custom scripts in R and the iGraph package (Csardi and Nepusz, 2006) (detailed further in **Chapter 2.1.7.4**).

[17] Software available from http://bitplane.com

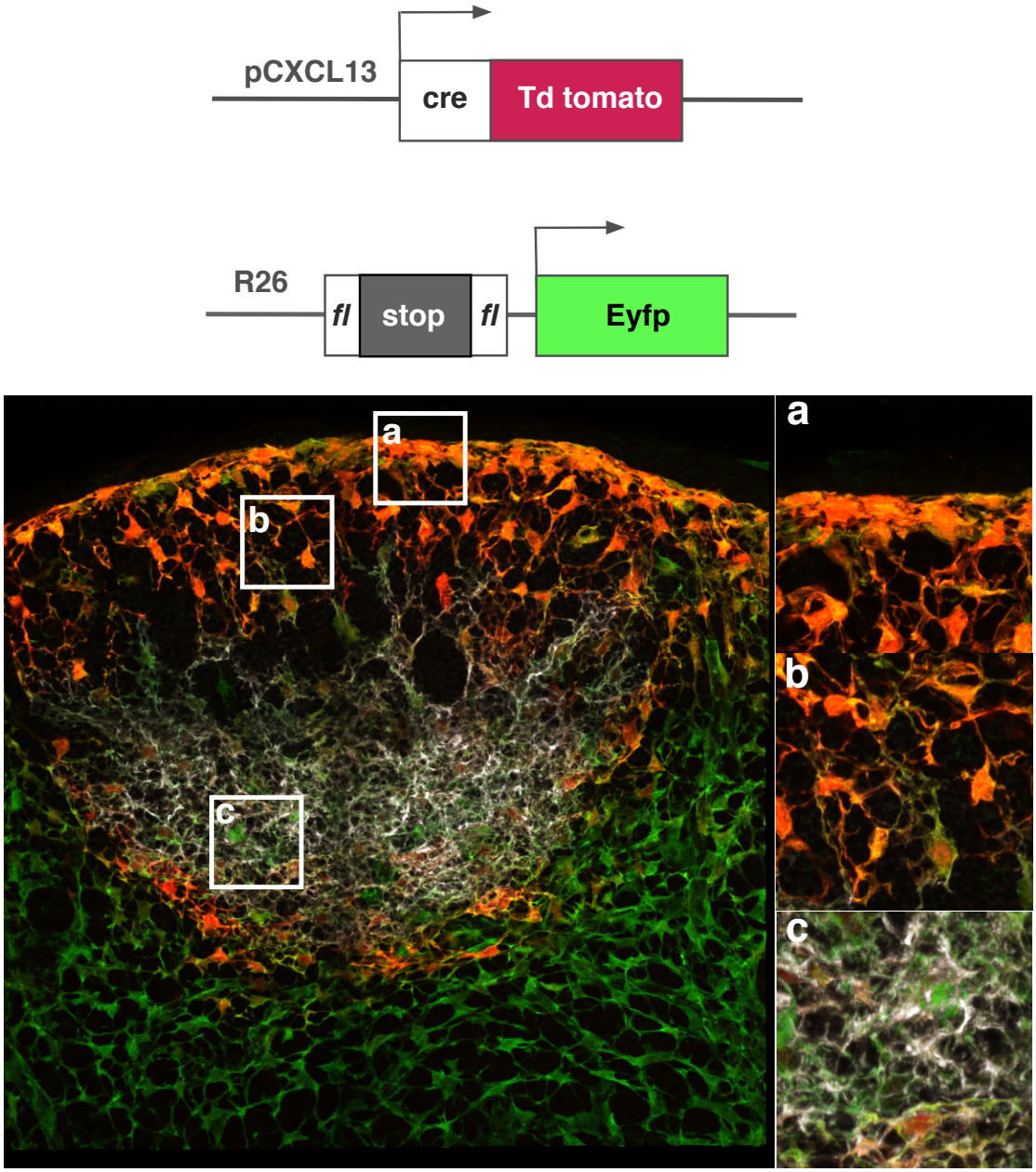

Figure 4.2 The morphology of lymph node follicular stroma. The *Cxcl13*-EYFP mouse is a reporter strain that facilitates visualisation of cells currently expressing CXCL13 (red) and cells that originate from a CXCL13 expressing precursor (green). In this system cxcl13 promoter activity leads to TdTomato expression and also activates the cre enzyme that removes the stop sequence in the promoter of the ubiquitously expressed rosa26 locus giving constant expression of eYFP. Within the follicle we focus on 3 distinct stromal subsets (a) MRCs (b) BRCs and (c) FDCs. TdTomato (red), eYFP (green) CD21/35 (white). Data for figure provided by Dr. Natalia Pikor.

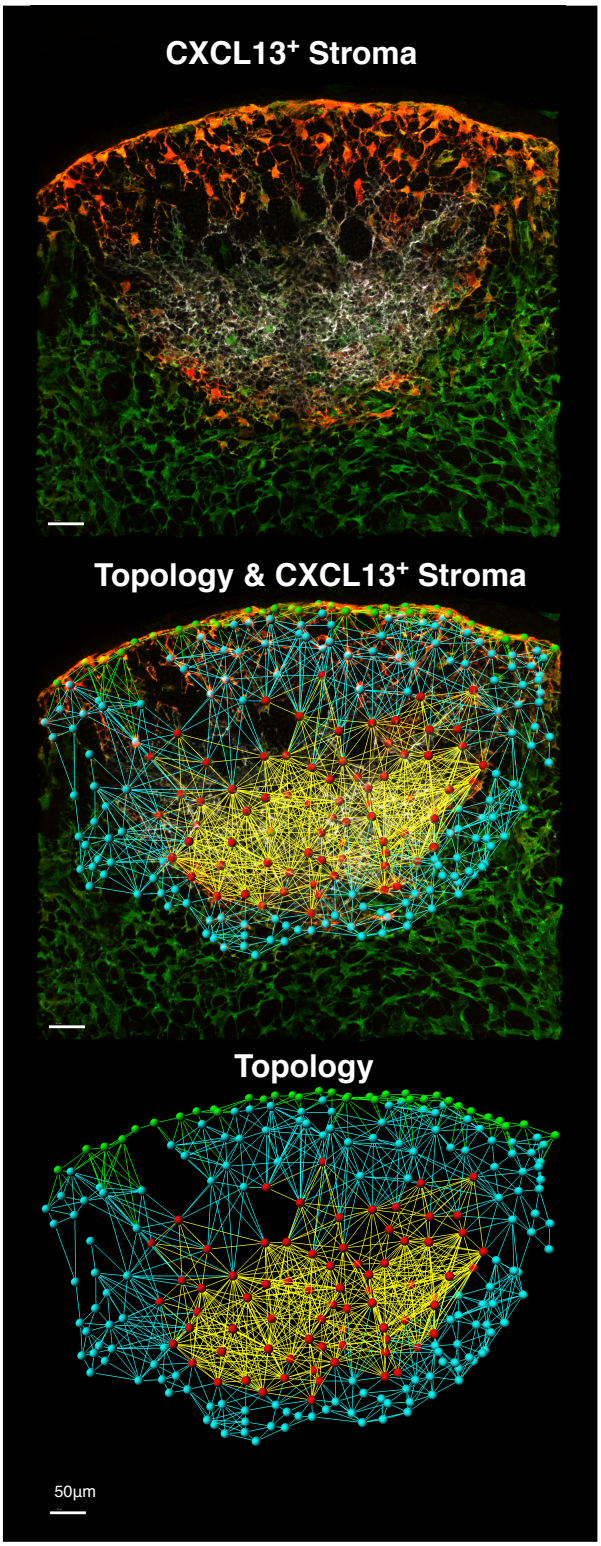

Figure 4.3 Topological mapping of follicular stroma. in this analysis we treat cell bodies as nodes and cellular protrusions between cell bodies as edges. FDC subnetwork highlighted in red, BRC subnetwork in blue and MRC subnetwork in green. Experiments performed by N.P and L.O, analysis performed by J.C and M.N Scale bar = 50µm.

From the images we map a network of 256 nodes and 1764 edges (**Table 4.1**). The median edge length through the network was 34.49μm with highest values observed in FDCs (**Figure 4.4a).** The median degree centrality was 9.5 edges for the whole network with similar median values and distributions observed for both BRCs and MRCs while FDCs were observed to be significantly different (**Figure 4.4b, Table 4.1**). The high density of FDC connections could be attributed to their web-like morphology, which served to maximize network connectivity in comparison to both BRCs, and MRCs. Stromal subsets were also heterogeneous in their fluorescent reporter expression for CXCL13, with highest values recorded for MRCs suggestive of a CXCL13 gradient towards the SCS, a major site of antigen entry into the follicle (**Figure 4.4e**). From this analysis we conclude that FDCs are the most topologically distinct cell type within the follicle (**Figure 4.4**).

The global and local clustering coefficients (0.52 and 0.44 respectively) are higher than those expected for a random network with the same number of nodes and edges (0.054 and 0.0541). The mean shortest path length through the follicle network is 3.58 while for an equivalent random network we obtain a value of 2.39. The high clustering coefficient and low shortest path length suggested that the network may have small world properties, as has been reported for fibroblastic reticular networks in the T-cell zone (**Figure 4.5**). This was confirmed by calculating the small world network metrics σ and ω (detailed in **Chapter 2.7.1.4**) (**Table 4.1**). σ is calculated by comparing clustering and path length of a given network to an equivalent random network with same degree on average (Humphries and Gurney, 2008). σ values greater than 1 are indicative of small-worldness; in the follicle, σ is 6.4 (**Table 4.1**). ω is another method for quantifying network small-worldness and compares the clustering of a given network to an equivalent lattice network and its path length to an equivalent random network (Telesford et al., 2011). The metric is constrained between -1 and 1 with values close to zero indicative of small-world properties; in the follicle ω is 0.075 (**Table 4.1**).

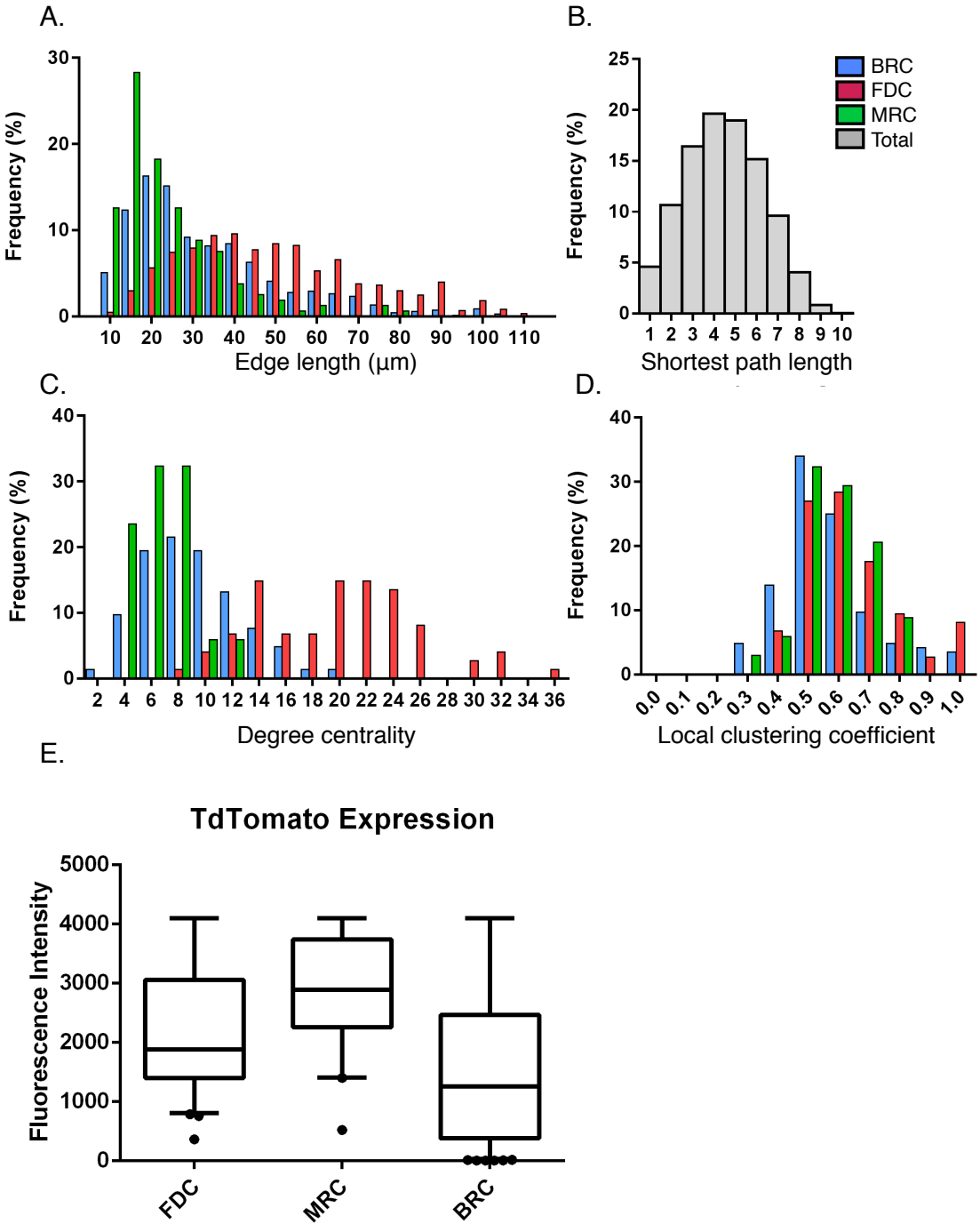

Figure 4.4 Network properties of follicular stromal cells. (A) Distributions of edge lengths with distinct subsets highlighted. (B) Distribution of shortest path lengths through the entire follicular network. (C) Distributions of degree centralities within the network. (D) Distributions of local clustering coefficients FDCs shown in red, BRCs in blue and MRCs in green. (E) TdTomato fluorescent reporter expression for CXCL13. Experiments performed by N.P and L.O, analysis performed by J.C and M.N

Metric	Value
Number of Nodes	256 nodes
Number of Edges	1764 edges
Median Degree Centrality	9.5 edges
Median Edge Length	34.49 µm
Local Clustering Coefficient	0.52
Global Clustering Coefficient	0.44
Shortest Path Length	3.58 nodes
Sigma	6.4
Omega	-0.075

Table 4.1 Summary of CXCL13^{+} stromal cell network properties

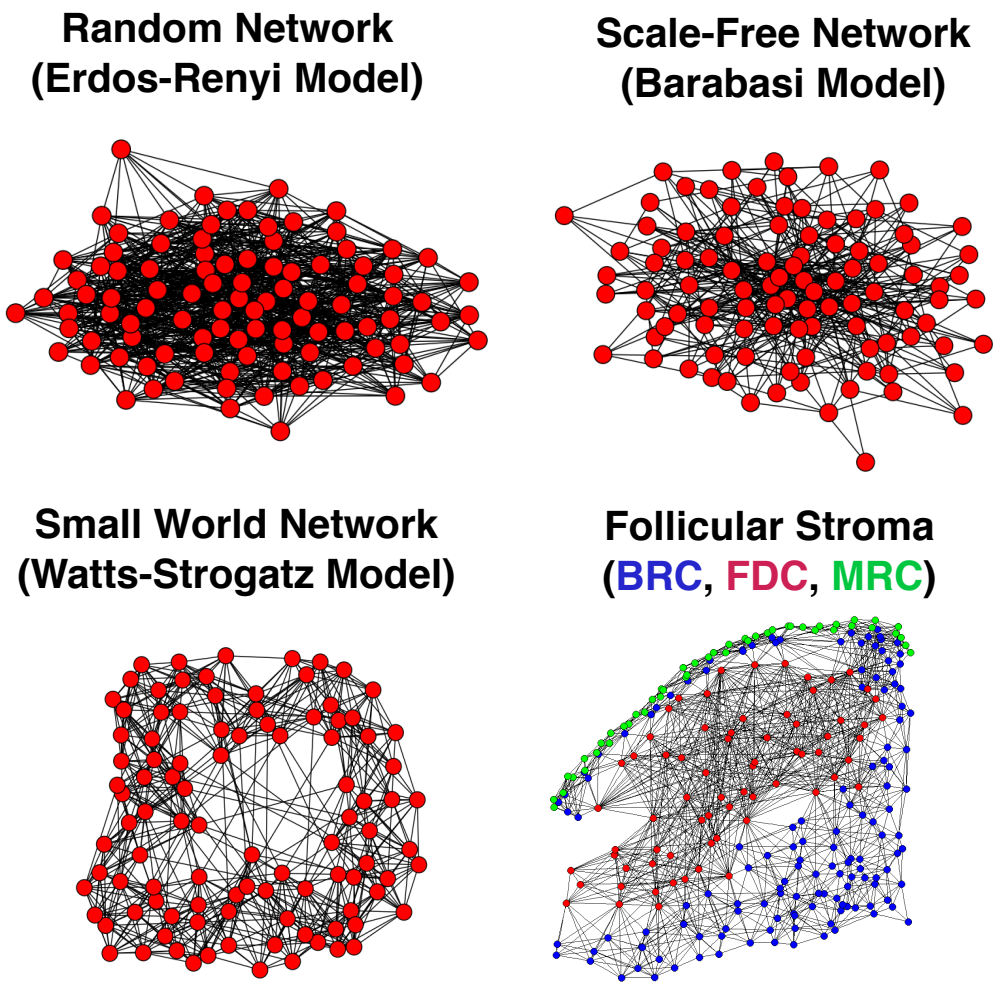

Figure 4.5 Exemplar illustration of random, scale-free, small world and follicular stromal networks. Random networks have no underlying organization with edges and nodes randomly assigned to one another, Scale-free networks, for example protein interaction networks have an abundance of high degree nodes or hubs, this infers robustness to the network where random node deletion is unlikely to disrupt the entire network. Low shortest path lengths through the network and a high clustering coefficient promoting fast information transfer characterize small-world networks. These characteristics are also prevalent in follicular stromal cell networks as verified through calculation of the small world network metrics σ and ω**.**

4.3 Determination of the CXCL13 Diffusion Coefficient

To measure the diffusion constant of CXCL13 we first assessed 3 distinct imaging modalities in collaboration with the Leake group: FRAP, FCS and a bespoke single-molecule narrowfield imaging approach. The latter is detailed further in the **Chapter 2** but briefly involved adapting a standard inverted epifluorescence microscope to facilitate minimally perturbative sub-millisecond super-resolution localization of rapidly diffusing fluorescently labelled biomolecules. A key feature of the system is the use of an ultrasensitive back-illuminated EMCCD detector (iXon^{+}, Andor) which could be sub-arrayed to 128 x 29 pixels to enable rapid frame rates up to ~2,270 Hz. Imaging data was subsequently analysed in bespoke software (Miller et al., 2015b) written in MATLAB (Mathworks), which enabled precise quantification of the underlying molecular stoichiometry[18] and mobility[19] of tracked particles with ~40nm spatial precision and sub-millisecond sampling.

To assess how the different techniques compared we measured the diffusion characteristics of the previously-characterised Bovine Serum Albumin (BSA) protein (Putnam, 1975b) that was conjugated with the far-red fluorophore Alexa-647 (**Figure 4.5**). Due to the small size of BSA (66kDa) it was not possible to track diffusion over consecutive imaging frames in PBS buffer at the single molecule level, consistent with theoretical estimates for the diffusion constant D (> 100 $\mu m^2\ s^{-1}$). To address this, we assessed diffusion in 10% Ficoll solution, which increased the fluid viscosity and facilitated single particle tracking. Using this experimental assay we then compared our empirically derived values for D, with theoretical expectations based on hydrodynamic modelling of BSA as a Stokes sphere of radius 3.48nm incorporating Faxens law to account for friction encountered in a viscous environment (**Table 4.3**) (Axelsson, 1978).

Using FRAP and FCS we measure values of D = 7.1 ± 0.3 $\mu m^2 s^{-1}$ and D = 18.8 ± 0.3 $\mu m^2 s^{-1}$ respectively for BSA-AF647 in Ficoll (**Table 4.2**). The result for FRAP is lower than the theoretical value, even considering boundary effects and temperature fluctuation whilst the result from FCS is higher (**Table 4.2**). The FRAP and FCS results differ by a

[18] By comparing the integrated brightness with that measured for a single fluorophore

[19] By determining the microscopic diffusion coefficient *D* from the measured mean square displacement (Robson et al., 2013)

factor of 2.6, consistent with other published results (Calizo and Scarlata, 2013; Erika M. Adkins et al., 2007; Guo et al., 2008a). Differences between the techniques have been attributed to the different spatial scales of the two measurements or the high number of assumptions required in fitting FRAP data (Macháň et al., 2016). Single molecule imaging of BSA-AF647 yielded two distinct populations of molecules. Low mobility tracks where molecules interact with the coverslip and a freely diffusing mobile population with D_{mobile}= 9.3 ± 0.4 $\mu m^2 s^{-1}$ (**Table 4.2**). Thus in addition to the consistency between single molecule and theoretically expected values, the approach is capable of assessing the underlying heterogeneity of molecular mobility. Consequently, we conclude that a single molecule approach is best suited to characterise chemokine mobility.

Condition	Diffusion Constant ($\mu m^2 s^{-1}$)
Theoretical	12.3 ± 0.1
FCS	18.8 ± 0.3
FRAP	7.1 ± 0.3
Single mol.	9.3 ± 0.4

Table 4.2. Comparison of diffusion constant measures for BSA-AF647 obtained theoretically, or using FCS, FRAP and single molecule imaging. Data preparation and acquisition performed by J.C and H.M. Data analysis performed by H.M

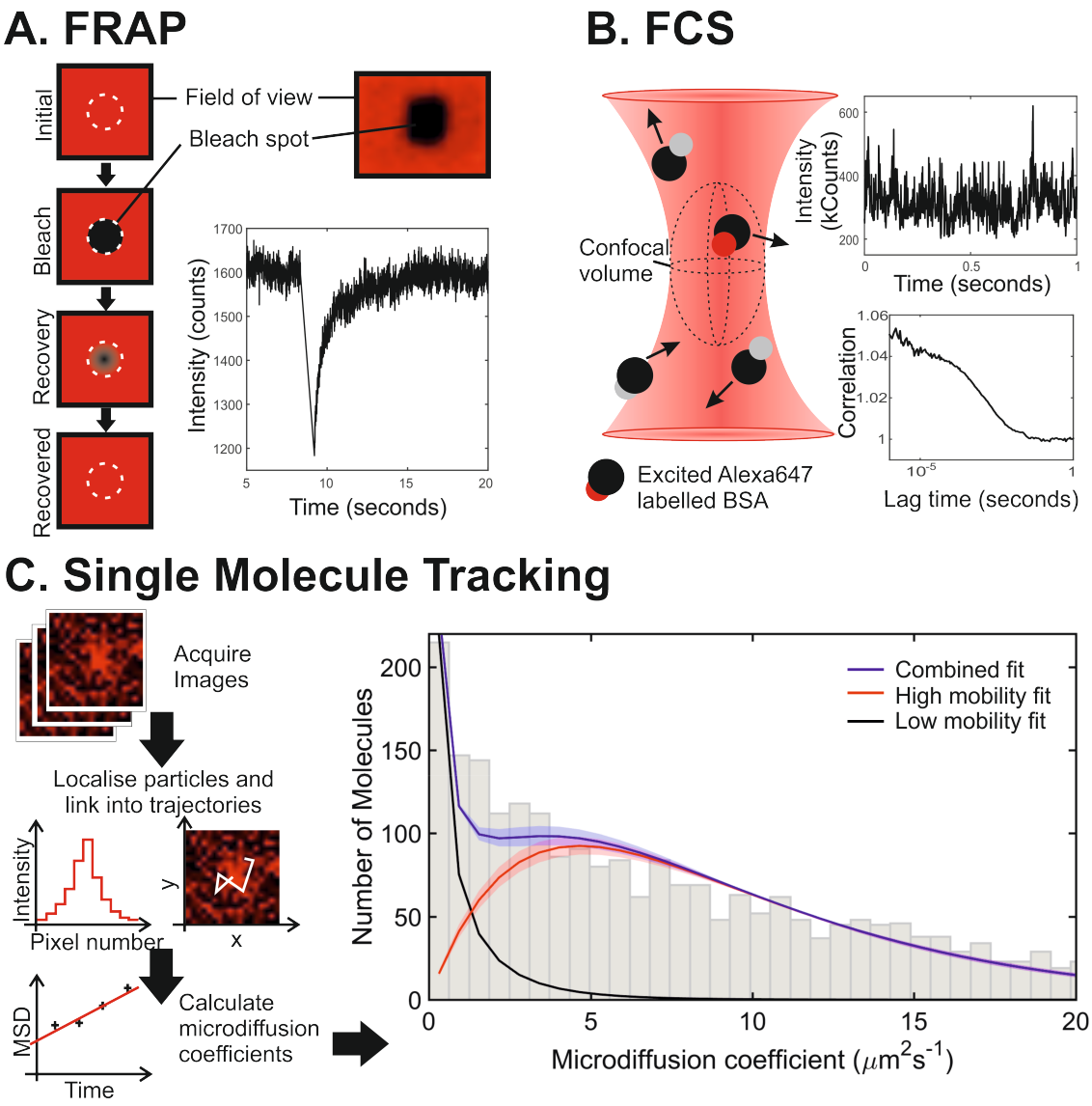

Figure 4.6 Measurement of the diffusion coefficient of BSA-AF647 in 10% ficoll 400 with three techniques. (A) FRAP: schematic of technique and fluorescence intensity recovery trace. (B) FCS: Schematic of the confocal volume, section of intensity fluctuation trace and example correlation curve. (C) Single-molecule tracking with high speed narrowfield microscopy: Simplified schematic of the stages in tracking and the resulting fit with error bounds of one standard deviation. J.C performed sample prep, J.C. and H.M. imaged the samples and H.M. performed image analysis

Using the single molecule assay we investigated Alexa Fluor 647 (AF647) tagged CXCL13, in addition to CCL19, with narrowfield excitation exploring a range of sample exposure times between 0.44-2.04 ms per frame, acquiring most data using 0.59 ms per frame as a compromise between sampling speed and accuracy of detection. We assessed mobility in 3 different extracellular environments: *(i) surface-immobilized heparan sulphate (ii) collagen gel matrix and (iii) lymph node tissue sections co-stained with anti-B220 to identify B-follicles*.

To assess whether different chemokines were heterogeneous in their diffusion characteristics, we measured diffusion in the controlled collagen matrix environment. Images of AF647 labelled chemokines were collected at high speed (**Figure 4.5a-b**), and these were confirmed to be single molecules (**Figures 4.5**). AF647-labelled chemokines have a high-affinity for glass, sticking non-specifically to the coverslip surface. This population was characterised by tracking chemokine immobilised to the surface of the coverslip with heparin sulphate. This produced a distribution of low mobility molecules that can be filtered from the mobile fraction, facilitating data filtering and analysis. Interestingly, the diffusion rates for CXCL13 and CCL19 differed (6.2 ± 0.3 $\mu m^2 s^{-1}$ and 8.4 ± 0.2 $\mu m^2 s^{-1}$ respectively, **Table 4.3**). This relative difference between the molecules was anticipated given the differences in the masses of the molecules (CXCL13 = 12.6 kDa; CCL19 = 10.9 kDa). Theoretically expected values derived for free diffusion in water (146 $\mu m^2 s^{-1}$ and 149 $\mu m^2 s^{-1}$) were 23.5 and 17.7 times faster than for those measured in collagen (**Table 4.3**). These values suggest that it would take ~ 45 seconds for CXCL13 and CCL19 to diffuse 200μm freely in water, while in collagen it would take 17.92 and 13.23 minutes for CXCL13 and CCL19 to diffuse the same distance (**Table 4.4**). Analysis of the single particle tracking data from CXCL13-AF647 and CCL19-AF647 showed broad probability distributions for *D* under all different conditions (**Figure 4.7**) indicative of underlying heterogeneity in molecular mobility.

Following this analysis, we performed single molecule imaging of CXCL13 in lymph node follicles. Within the samples, extracellular matrix components were highly autofluorescent and so it was not possible to track as many molecules. The resultant distribution of microdiffusion coefficients was broad yielding a bimodal distribution of high and low mobility diffusion rates (**Figure 4.8**). Superimposing these rates on the spatial distribution of CXCL13 within the tissue indicates that low mobility chemokines

are transiently interacting with extracellular matrix components, while high mobility molecules are freely diffusing in spaces between these components. Thus the microanatomy of the follicle constrains the rate of diffusion, with the density of the LN giving rise to diffusion rates of 7.6 ± 1.0 $\mu m^2 s^{-1}$ substantially lower than for free diffusion in water (**Table 4.3**). This suggests that it would take an average of 14.62 minutes to travel 200μm in tissues (**Table 4.4**).

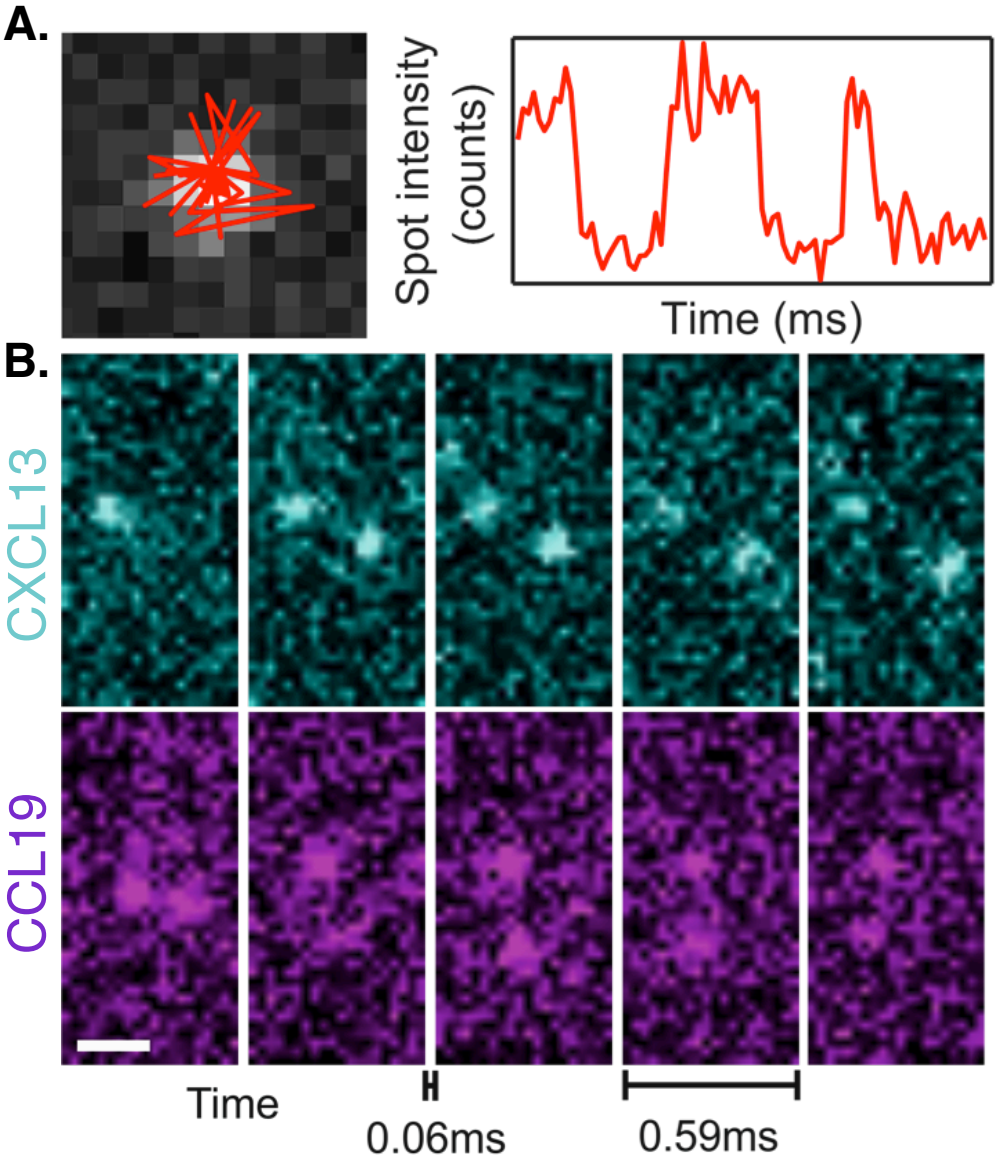

Figure 4.7 Tracking single molecules. (A) Localisation of a single particle with tracking indicated by the red line. RHS: a fluorescence intensity plot showing a blinking pattern characteristic of single molecules. (B) Raw images showing CXCL13 (cyan) and CCL19 (purple) molecules labelled with Alexa-647. Scale bar = 1μm

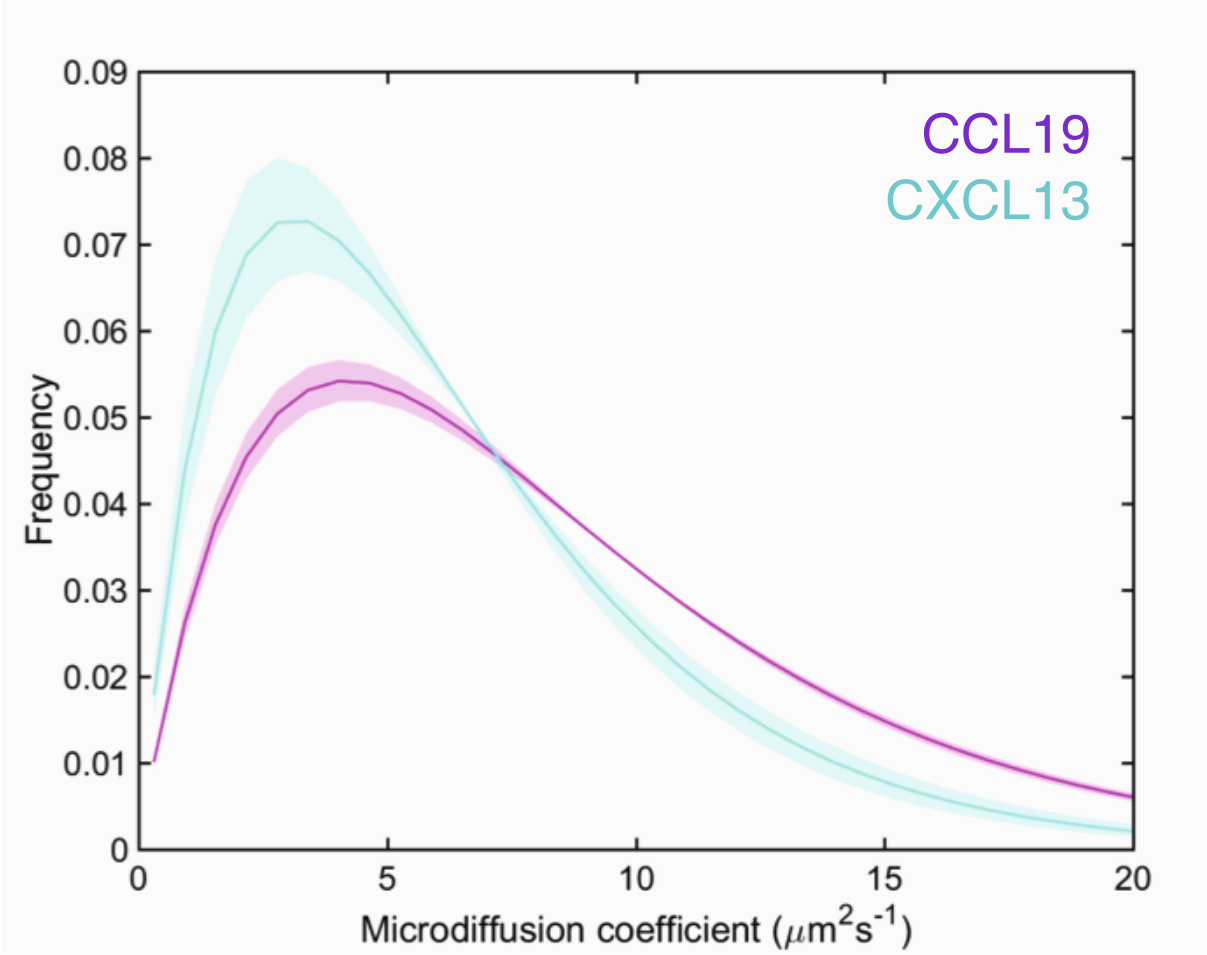

Figure 4.8. Comparison of CCL19 and CXCL13 microdiffusion coefficient distributions in collagen matrix. Distributions of CXCL13 microdiffusion coefficients are shown in cyan, with CCL19 shown in purple.

	D_{free} ($\mu m^2 s^{-1}$)	$D_{collagen}$ ($\mu m^2 s^{-1}$)	Error	Number of mobile tracks	R^2 of fit
CXCL13	146	6.2	0.3	1930	0.98
CCL19	149	8.4	0.2	4859	0.984

Table 4.3. Summary of diffusion constants obtained for CXCL13 and CCL19 using the Einstein-Stokes relation (D_{free}) and in collagen ($D_{collagen}$). Also reported are the error, number of mobile tracks and the R2 of model fitting to experimental data.

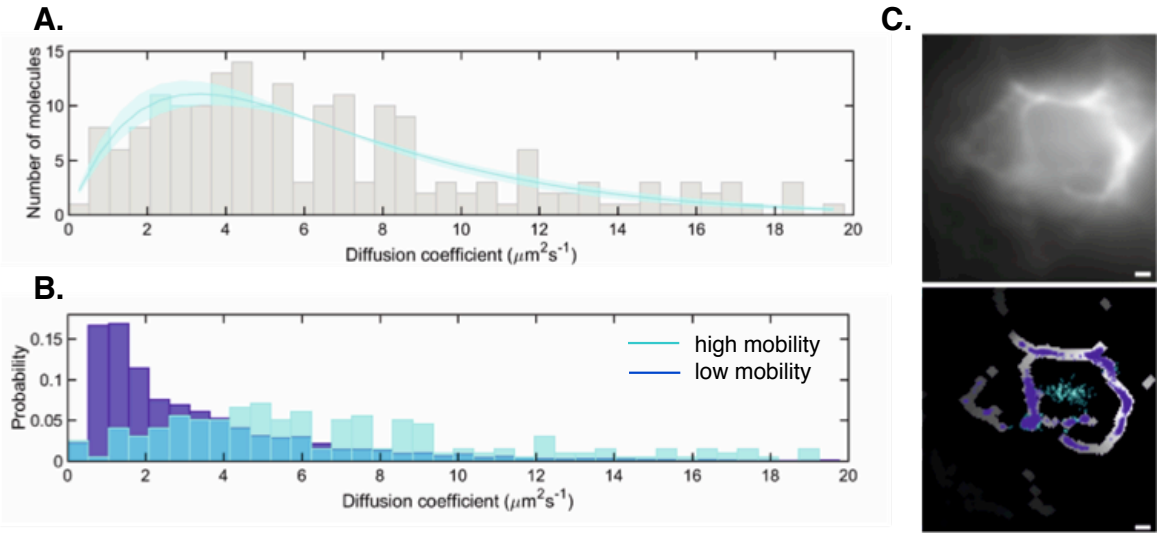

Figure 4.9 Single molecule analysis of chemokine diffusion in lymph nodes. (A) Distribution and fit of diffusion coefficients of CXC13-AF647 in tissue, shaded area indicates one standard deviation. (B) Comparison of diffusion coefficients for the immobile (blue) and mobile (cyan) populations when tracking CXCL13-AF647 in lymph node tissue shown in (C). (C) Sub-pixel super resolution localization precision of the immobile fraction tracked in tissue. All scale bars 1µm. J.C performed sample prep, J.C. and H.M. imaged the samples and H.M. performed image analysis

Condition	Time taken to diffuse 200µm (s)	Time taken to diffuse 200µm (min)
$CXCL13_{water}$	45.66	0.76
$CXCL13_{collagen}$	1075.27	17.92
$CXCL13_{tissue}$	877.19	14.62
$CCL19_{water}$	44.74	0.75
$CCL19_{collagen}$	793.64	13.23

Table 4.4 Time taken to diffuse 200µm for a given diffusion constant. Details of the calculation are provided in Appendix 1.

4.4 The Canonical Homeostatic Murine Popliteal Follicle

Development of a conceptual model prompted experimental work to quantify the diffusion of CXCL13 and the anatomy of $CXCL13^+$ stromal cells. However, we can also draw upon data from the wider literature to derive additional information that can be used to better understand our research context. In the following section, we summarise our findings to provide a quantitative description of the homeostatic follicle.

Quantification of pLN follicle size and cellularity: Follicle volume was obtained from two publications (Irla et al., 2013; Kumar et al., 2010). Kumar *et al* determined the mean volume of a popliteal LN follicle using OPT scanning on BABB-immersed pLNs in conjunction with software based 3D reconstruction and quantification. Their analysis shows that typical pLN volume is $1.25mm^3$, dividing that by 10% (the percentage of volume accounted for by follicles) gives a volume of 1.25×10^8 μm^3 for all of the B-zones in one LN and dividing that by the number of follicles (18 [7-35]) gives 6.94×10^6 μm^3. In addition the authors observed a strong correlation ($r^2 = 0.93$) between the size of the lymph node and the number of follicles. Values obtained by Irla *et al*. of inguinal lymph nodes suggest that total LN volume is $2.4mm^3$. Dividing this by 18% gives the % follicular volume. We then divide this value by the number of follicles (12.5) to get the volume of one follicle (3.46×10^7 μm^3). The discrepancy between the two sets of measurements can most likely be attributed to differences in lymph node morphology at different anatomical sites and different experimental procedures.

B-cell density was determined by quantifying $CD19^+$ cells using flow cytometry in conjunction with Accucheck counting beads (n = 5 mice, 3 separate experiments) and averaging total pLN counts by the mean number of follicles per lymph node (**Figure 4.10**). Stromal cell densities within a fixed follicular volume were determined from the number of nodes obtained from a 400μm x 400μm x 30μm image (**Table 2**) of a follicle obtained from a *Cxcl13*-EYFP reporter mouse.

From this data we define a popliteal lymph node as having a volume of $1.25\ mm^3$, 15% of which is defined as a B-zone split between 15 follicles. Within this lymph node each follicle is spheroidal with a total volume of $1.25 \times 10^7\ \mu m^3$ (~ 250 x 250 x 350 μm) and contains 4.8×10^4 $CD19^+$ B cells (**Table 5**).

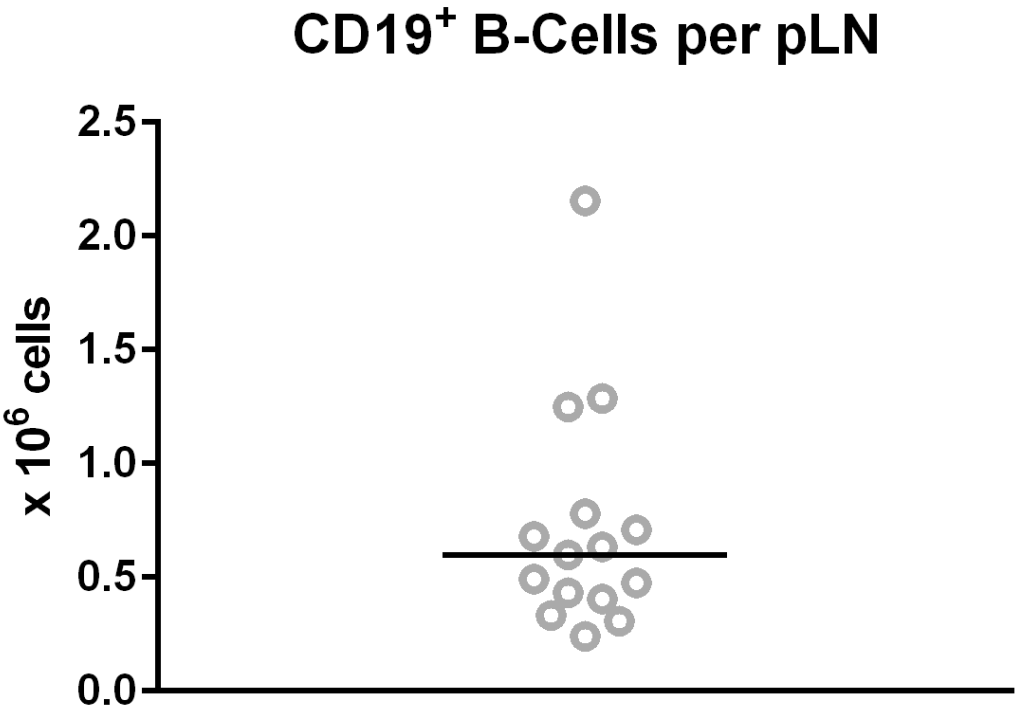

Figure 4.10. B cell numbers per popliteal pLN. B-cell density was determined by quantifying $CD19^+$ cells using flow cytometry in conjunction with Accucheck counting beads as described in section (Chapter 2) and averaging total pLN counts by the mean number of follicles per lymph nodes. Grey circles represent individual datapoints, black line represents the median value. N= 15 mice, data pooled from 3 separate experiments. Experiments and analysis performed by J.C and A.T

B-Cell Migration: The migration patterns of B cells were measured by injecting 5 x 10^6 purified B cells from C57BL/6 or CXCR5 deficient mice labelled with the fluorescent cell staining dyes Carboxyfluorescein succinimidyl ester and CellTracker Orange (5-(and-6)-(((4-chloromethyl)benzoyl)amino)tetramethylrhodamine) CMTMR into age- and sex-matched wild type mice and imaged with two-photon microscopy (Coelho et al., 2013). This data was provided by the Stein lab. WT B cells have a velocity (expressed as median[lower quartile – upper quartile]) of 8.0 [2.3 - 10.3] $\mu m^2 min^{-1}$, a meandering index[20] of 1.1 [0.1-2.4] and a motility coefficient[21] of 15.6 [0.1 – 96.7] $\mu m^2 min^{-1}$.

CXCR5 expression: The number of CXCR5 molecules per naïve B-cells in lymph nodes is unknown, however data from similar systems suggests that the amount of receptors is in the range of 10,000 – 100,000 receptors (Sklar et al., 1984; Tilo Beyer, 2007).

CXCL13 diffusion constant: A value of 7.6 ± 1.0 $\mu m^2 s^{-1}$ was obtained in LN sections using high-speed single-molecule imaging, as discussed in **section 4.3**. An upper bound for this value was determined using the Einstein stokes relation, assuming a Stokes radius of 3.48 nm and that the molecule undergoes free diffusion in water (146 $\mu m^2 s^{-1}$).

Affinity Constant (K_d) and CXCL13 secretion rate: The total amount of chemokine from pooled lymph nodes (Luther et al., 2002), *in vitro* migration assays (Gunn et al., 1998a), and ligand binding constants (Barroso et al., 2012) were used to derive upper and lower limits of a likelihood distribution for CXCL13 concentrations *in vivo*. Taken in concert these analyses suggest that the value lies in the range 1 - 50nM with a baseline follicle concentration assumed to be 10nM that we set as our binding affinity K_d. This is consistent with expected ranges of K_d for CCR7 and CCL19 where similar concentrations have been described to sufficiently trigger downstream signalling of CCR7 following binding of CCL19 in G-protein loading assays (Kohout et al., 2004), downstream signalling assays (Otero et al., 2006) and microfluidic migration assays (Schwarz, 2016).

K_i and K_r: The two rate constants K_i and K_r associated with receptor internalization and recycling were estimated from experimental data on receptor desensitization and

[20] A measure of the straightness or confinement of cell tracks

[21] Analogous to the diffusion constant

resensitization neutrophils and from mathematical modelling of B-cell migration (Sklar et al., 1984; Tilo Beyer, 2007).

CXCL13 Decay Rate: Chemokine's are further regulated *in vivo* by proteases providing rapid enzymatic modulation of bioactivity and availability. Systematic analyses of proteomic half lives suggests a broad range of possible values, which we constrain between 15 minutes to 48 hours (Eden et al., 2011; Schwanhäusser et al., 2011). Assuming a constant decay rate[22] this yields rates between 0.015 and 0.0002s^{-1} (Milo and Phillips, 2015)

Metric	Value
pLN volume	1.25 mm^3
% B-cell volume	15%
Number of follicles per LN	15 follicles
Follicle volume	1.25 x 10^7 μm^3 (~ 250 x 250 x 350µm)
Number of B cells per follicle	4.8 x 10^4 $CD19^+$ B cells
CXCR5 per B cell	10000-100000 receptors
CXCL13 baseline concentration	10nM
CXCL13 diffusion constant	7.6 $\mu m^2 s^{-1}$
Secreted CXCL13 half life	0.25 – 48 hours

Table 4.5. Quantitative description of CXCL13 bioavailability in primary lymph node follicles

[22] Half life and decay rate are related through the following expression: $t_{0.5} = \ln(2)/\text{decay rate}$

4.5 Discussion

In this chapter we have used an ensemble of imaging approaches to address knowledge gaps identified in **Chapter 3**. Specifically, we quantified the topological characteristics of $CXCL13^{+}$ stromal cells in the primary follicle using an adaptation of a previously published approach (Novkovic et al., 2016). The authors of this study were key collaborators, providing imaging datasets and sharing expertise in image annotation and network analysis. This analysis identified an underlying regulation to the spatial organization of CXCL13 at the cellular level, with secretion coming from 3 distinct cell types organized into a small-world network. Each cell type is unique in its topological properties, supporting the idea that the primary B-follicle is compartmentalized into distinct niches with distinct functionality as has been established for the germinal center [23] (Allen et al., 2004). FDCs were the most topologically distinct subset, a property that may promote their unique role in antigen presentation. The high connectivity of the FDC network creates a labyrinth of single-cell niches, maximizing contact with B cells that floss the network for antigen. Topological analysis of the global network identifies regions of high connectivity, with long-range connections between these cliques. These guidance structures are likely to promote trafficking both within and between the different niches of the B-cell microenvironment.

We also note that the highest levels of CXCL13 reporter activity occur in MRCs, which lie proximal to a major site of antigen entry. In the lymph node, the delivery of lymph-borne antigens to FDCs is determined by the size of antigen and whether it is opsonized with complement. Lymph-borne antigens larger than 70kDa (~5.5nm) are presented by SCS macrophages that bind C3d-coated immune complexes via CR3. Immune complexes are subsequently shuttled to the FDC network via CR2 expressing B cells (Batista and Harwood, 2009; Carrasco and Batista, 2007; Phan et al., 2007; Rantakari et al., 2015). Our data thus provides a unique insight into how the primary follicle is structurally organized to promote B-cell homeostasis and activation.

[23] the compartmentalisation of the GC was first described in 1930 following analysis of cat lymph nodes: Beitrag zur Cytologie der Keimzentren der Lymphknoten

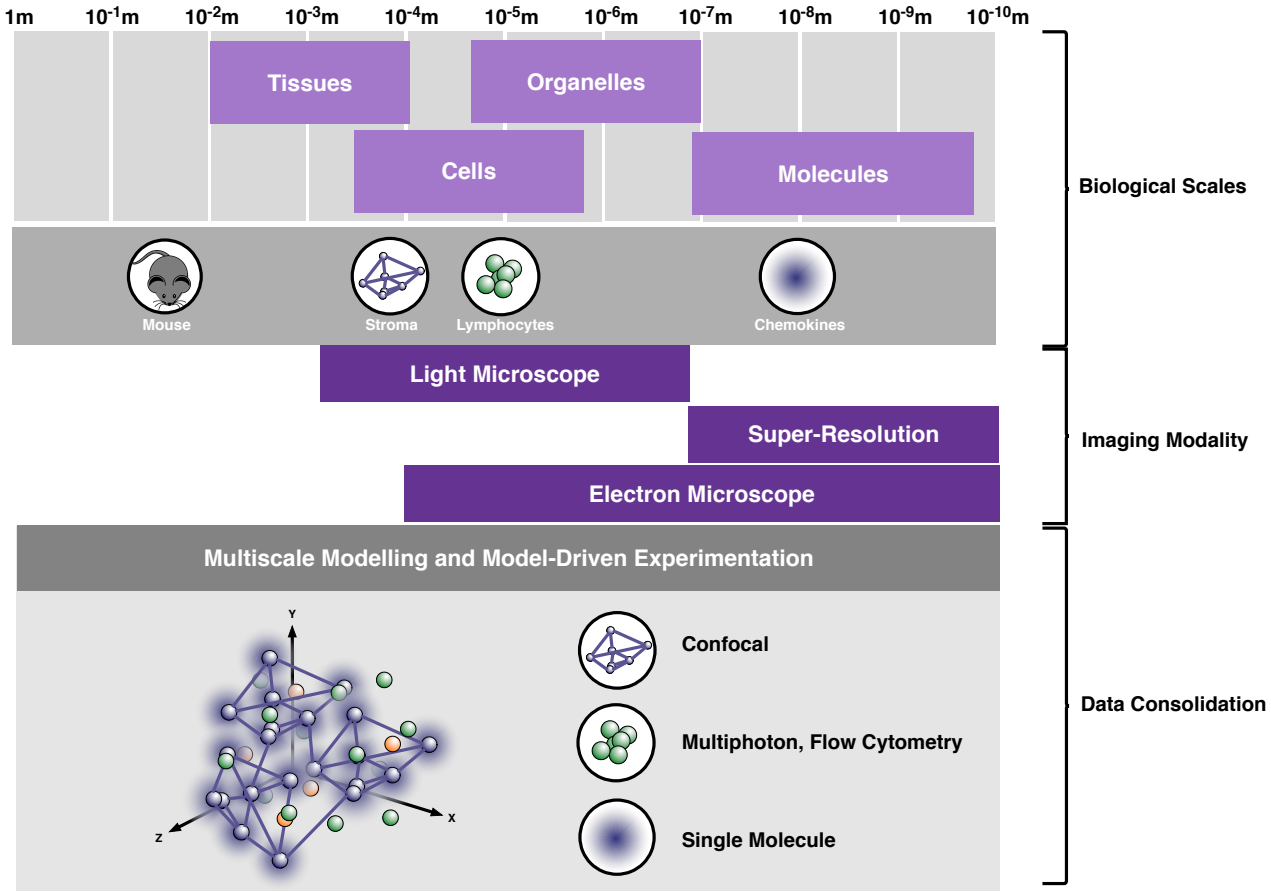

Figure 4.11. Consolidating imaging datasets with multiscale modelling. The research context we assess incorporates phenomena across broad spatiotemporal scales. Each distinct model entity: stroma, lymphocytes and chemokines, required a distinct imaging modality to visualise and quantify. Subsequently, multiscale modelling is employed to consolidate these disparate datasets. Through simulation analysis additional insights can be gleaned from the data prompting additional experimentation.

To measure CXCL13 mobility, we collaborated with members of the Leake lab to develop an assay capable of quantifying chemokine mobility in complex tissues. This was achieved by using sub-millisecond narrowfield microscopy, an approach shown to outperform both fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS) given its high spatiotemporal resolution and ability to dissect the heterogeneity of molecular mobility. The ability to assess the heterogeneity of molecular mobility in a complex environment is a key advantage as studies of the chemokine interactome suggest that chemokines undergo a dynamic array of molecular interactions (Hundelshausen et al., 2017). Our analyses show that chemokines display distinct mobility patterns, with CCL19 diffusing faster on average than CXCL13. This is consistent with theoretical predictions given the relative difference in the masses of the molecules. However, the assumptions required to employ the Einstein stokes relation make it inapplicable to predict the diffusion of chemokines through complex tissues. This is because it fails to account for the density of the local environment or the transient molecular interactions that occur *in vivo*. Further experimental analysis of chemokine mobility and biochemical interactions could facilitate adaptations to existing theoretical frameworks to better predict diffusivity.

We combine these datasets with flow cytometry measurements of B-cell density within a follicle, multiphoton imaging measurements and values derived from the wider literature to provide a more complete mapping of the pathway. However, despite this systematic analysis of the pathway a number of key questions remain, due in part to a dearth of experimental techniques capable of manipulating molecular gradients *in situ*. In addition, our datasets have been obtained using an ensemble of techniques, making them difficult to consolidate and analyse at a systems level. Theoretical models can facilitate better interpretation of this data and through simulation-based experimentation can generate insights that drive further experimental work (Butler et al., 2016a, 2016b). In the following chapters, we aim to extend our CXCL13 case study using an iterative, synergistic cycle of modelling and experimentation (**Figure 4.11**).

Chapter 5 Reconstruction of the Primary B-Follicle *In Silico*

5.1 Introduction

5.1.1 Implementing Multiscale Models: Key Technical Considerations

In this chapter we detail the implementation and validation of a multiscale simulator using the agile CoSMoS instantiation detailed in **Chapter 3**. In the following section, we highlight some of the key technical considerations associated with designing, implementing and validating software implementations of multiscale models.

Prior to implementation as computer code, biological processes are abstracted into mathematical constructs to move from the Domain Model towards a simulator (detailed further in **Chapter 2.2.1**). We refer to this step as the Platform Model, a process analogous to a software specification used in software engineering. This is a key step in the development process as inappropriate decisions and abstractions can render the simulation an inaccurate reflection of the Domain Model (Alden et al., 2012b; Cosgrove et al., 2015). At this phase of development, it is important that any explicit references to emergent system-level properties are omitted; such properties should emerge from the low-level behaviours, validating the simulator as a fit for purpose representation of the problem domain.

Designing the mathematical constructs and architecture underpinning multiscale platforms is particularly challenging, in certain research contexts a single modelling approach is incapable of accurately describing a biological system. A common approach to hybridise different techniques is to model molecules using population-level ordinary or partial differential equations, and then solve them on a per-agent basis within an ABM (Guo et al., 2008b). At the molecular level, the application of population-based as opposed to agent-based approaches can save considerable computational power reducing the expense incurred if the highest resolution timescale were to be used for each phenomenon. By selectively tuning the granularity at each scale, a single model can examine processes across diverse scales (Guo and Tay, 2008; Kirschner et al., 2014).

The Simulation Platform represents a computer code implementation of the Platform model. During this step, low-level behaviours are encoded using control ("if – then") statements. The Object Oriented programming (OOP) paradigm uses classes, templates for program components, to separate and encapsulate functionality. This feature lends itself well to the implementation of multiagent systems (Johnson et al., 2004). Agile

techniques, such as those incorporated into the novel CoSMoS instantiation, may also improve the quality of scientific software (discussed further in **Chapter 3.2**) (Pitt-Francis et al., 2008).

Once the code has been developed it is important to identify parameter values that align the simulation's behaviour with experimental data, a process referred to as calibration. Calibration establishes baseline model behaviours, a benchmark to which all subsequent perturbations are compared. This is a challenging step, as the parameters controlling mathematical constructs may not translate directly to the biological system, may simultaneously account for several biological factors, or the parameter value may be unknown. Studies of parameter sensitivities of systems biology models suggest that fitting parameters against multiple datasets will often leave many parameters poorly constrained while direct parameter measurements would require formidably precise and complete to usefully constrain many model predictions (Brown et al., 2004; Gutenkunst et al., 2007). Model calibration is thus an active research area, with recent focus on the use of Bayesian statistics (Coelho et al., 2011), heuristic optimisation approaches (Read et al., 2013b), and advanced sampling schemes (Gilks et al., 1995; McKay et al., 1979) to constrain parameter values and efficiently scan high dimensional parameter spaces where values for a single parameter can span several orders of magnitude (Read, 2011).

Incorporation of heterogeneity and/or stochasticity within a model can lead to different results from the same parameter inputs, leading to aleatory uncertainty (discussed in **Chapter 2.2.8**). Once calibrated, empirical methods exist to assess the effect of stochasticity on aggregated results (Read et al., 2012). By comparing simulation responses obtained under the same conditions, the influence of stochastic variation on simulation outputs can be quantified.

To fully evaluate simulation outputs it is also important to identify critical system parameters and quantify their influence on simulation predictions. In a sensitivity analysis (SA), the modeller performs a systematic exploration of the parameter space, recording the influence of parameter inputs on simulation outputs (discussed in **Chapter 2.2.9**) (Ray et al., 2009; Read et al., 2012). The benefits of SA are two-fold in that they can identify key pathways and entities as well as quantifying parameter sensitivities. A SA approach was employed by Ray et al to examine the effects of TNF-α on the formation of

granulomas[24], predicting that multiple TNF-α activities and macrophage activation are key contributors to the control of infection within a granuloma (Ray et al., 2009).

5.1.2 Emulating Multiscale Models with Supervised Learning Algorithms

In this thesis, we focus on the use of agile development techniques as a means of reducing the time taken to develop multiscale models. However significant bottlenecks may still occur at the Results Model phase of development. For non-deterministic models where input parameters are poorly constrained, certain analyses may become intractable. For example when quantifying parametric uncertainty it is possible to use an OAT analysis to look at local sensitivity while using latin hypercube sampling with PRCCs to quantify global sensitivity. However, these analyses alone may not be well suited to looking at non-monotonic relationships between parameter inputs and emergent outputs of the simulator. In addition, the use of PRCCs only looks at correlations between inputs and outputs and does not indicate the magnitude of the effect. Consequently, the sensitivity score for each parameter will highly depend on the analysis being performed. It is thus to use a number of different approaches to understand this parametric uncertainty. For example, the extended Fourier amplitude sampling test (eFAST) provides an alternative global sensitivity analysis that partitions variance in simulation response between parameters of interest (Saltelli et al., 2008). As a variance based approach it is particularly suited to looking at non-linear non-monotonic relationships between model inputs and outputs (Marino et al., 2008). In addition, it is well-suited to assessing parameter interactions through the STi metric (detailed further in Chapter 2.1.10) where as OAT analysis looks at each parameter in isolation and PRCCs do not explicitly indicate the extent to which the parameter interacts with other parameters. However, the characteristics of this sampling technique (**detailed in Chapter 2.1.10**) give rise to a significant number of parameter sets. For an ABM of lymphoid tissue organogenesis (detailed in **Chapter 1.10**) a single eFAST analysis required 585,000 executions (Alden et al., 2012b). Even with the availability of high-performance computing resources, such resource-intensive analyses become intractable for simulators with a long execution time. This is particularly problematic when one considers that model development is an

[24] Granulomas are macrophage-rich aggregates of immune cells that can determine host response to a number of infections, including *mycobacterium tuberculosis*

iterative process, with several rounds of analysis and refinement required to fully design, develop and validate a model.

Given these issues, we reason that emulation, that is the use of a surrogate tool to rapidly and accurately predict simulation responses, has the potential to permit enriched evaluation and refinement (introduced briefly in **Chapter 1.9.3** and detailed further in **Chapter 5.5**), easing translation of simulator insights into increased biological understanding. Although used sparingly in this context, machine learning approaches have also shown promise in identifying complex non-linear relationships within large multivariate datasets (Inza et al., 2010). Using supervised learning approaches, it is possible for a machine-learning algorithm to learn the emergent behaviours of a simulator and quickly and accurately predict the simulation response for parameter sets the algorithm has yet to observe (Pruett and Hester, 2016). This attribute makes machine learning algorithms well placed to emulate simulators of biological systems, as illustrated by the use of support vector machines to emulate models of haemorrhage and renal denervation, resulting in a 6-fold decrease in computation time (Pruett and Hester, 2016). Another key benefit of the approach is that a number of different machine learning approaches have been developed (Bishop, 1996; Breiman, 2001; Cortes and Vapnik, 1995). In particular we focus on Artificial Neural Networks (ANN), a machine learning technique inspired by the neuronal circuits in the brain with computations structured in terms of an interconnected group of artificial neurons. The weighting of connections between neurons are adjusted in such a way that the network can convert a set of inputs into a set of desired outputs. This attribute makes ANNs well suited to learn the behaviours of a multiscale simulator.

Through reducing the time taken to perform sophisticated statistical analyses, emulators can inform experiments to quantify sensitive parameters, and identify sections of the simulator that are highly influential and may require further refinement. From an engineering perspective, the application of emulation may expedite simulator development by permitting rapid prototyping and identification of errors in model design, parameterization, and software infrastructure. Due to the complexity of scientific software it can be challenging to locate errors before running time intensive statistical analyses. Errors located late in the development process incur significant time penalties, in contrast

with those that could be encountered early by testing an emulation of that simulator. In overcoming some of the resource limitations associated with multiscale simulators, emulation permits enriched evaluation and refinement, facilitating translation of simulator insights into increased biological understanding.

5.1.3 Summary and Aims

Developing multiscale software platforms is challenging, with many technical considerations that affect confidence in simulation-derived results (**Chapter 5.1.1**). In this chapter we develop a hybrid multiscale model designed to address our research hypothesis that lymphoid tissue architecture is a key determinant of CXCL13 patterning. Given our research context and the associated domain and platform models we require a high fidelity representation of follicular architecture to provide a fit for purpose representation of the problem domain. In addition, we use this multiscale model to showcase the use of agile development methods and statistical analysis techniques within the CoSMoS framework to increase confidence in simulation-derived results. Specifically, we aim to achieve the following:

1. Develop a modular Platform Model for CXCL13Sim (**section 5.2**)
2. Implement a Simulation Platform of CXCL13Sim in JAVA using ATDD (**section 5.3**)
3. Quantify parametric and aleatory uncertainty in CXCL13Sim using the SPARTAN package in R (**section 5.4**)
4. Develop and validate an emulator of CXCL13Sim using supervised learning approaches (**section 5.5**)

5.2 Platform Model

5.2.1 Overview of the Platform Model

The Domain Model detailed in **Chapter 3** describes a set of biological processes occurring on molecular, cellular and tissue levels of organization. To provide a fit for purpose representation of the domain model while also promoting model parsimony and efficiency we hybridise different modelling techniques into a multiscale platform, with selection of a given technique informed by a functional requirements analysis (**Chapter 3**).

An overview of the scheme (**Figure 5.1**) is as follows: *In silico* stromal cells (**Module 1**) are modelled as a series of nodes and edges (Kislitsyn et al., 2015) capable of interacting with lymphocytes and secreting CXCL13 (**Module 2**). CXCL13 diffusion is modelled using a discretized PDE (Grajdeanu, 2007). Individual lymphocytes are distinct entities displaying heterogeneity in receptor expression and activation status; as such we have modelled them as agents that adjust their behaviours with respect to vector and ordinary differential equation-based calculations (**Module 3**). This was achieved through adaptation of a published scheme which explicitly accounts for gradient detection and the dynamics of GPCR expression on the cell surface (Lin and Butcher, 2008; Wu and Lin, 2011). Within this system agents exist within a continuous environment with chemokine existing in discrete grid spaces. An analysis of an agent-based model of the number of effector T cells leaving the LN suggests that a 3D model is required to adequately capture the dynamics of T-cell output from the lymph node during infection. 2D models were shown to underestimate lymph node output because the distance between antigen-presenting cells is overestimated in 2D with respect to 3D (Gong et al., 2013). As such 3D was deemed most appropriate to model the efficacy of antigen-mediated encounters by B cells.

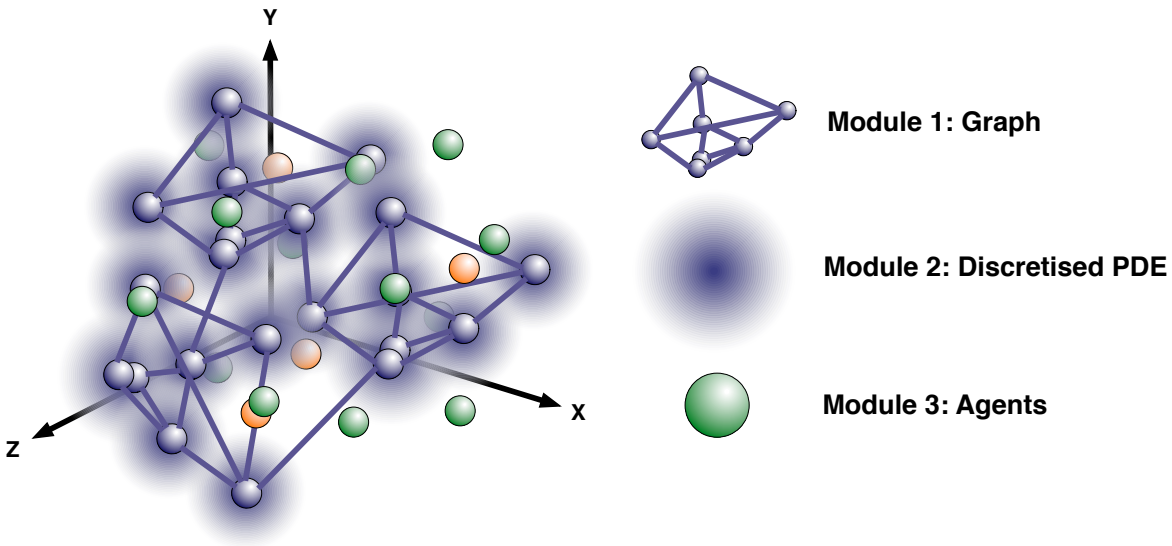

Figure 5.1. Hybrid 3D multiscale representation of a follicle. In this system stromal cells are modelled as a graph (Module 1), chemokine diffusion is modelled as a discretised partial differential equation (Module 2), while B cells are modelled as rich agents which can interact with their local environment through a set of coupled differential equations and vector based calculations (Module 3).

5.2.2. Module 1: Stroma

In silico stromal networks are generated using an adaptation of the algorithm developed by Kislitsyn *et al*. (2015). The algorithm stochastically builds a network from an initial node, picking the nearest non-expanded node and generates a set of vectors that will eventually become edges to new nodes. Each vector has a direction and length randomly chosen, but conforming to values derived from experimental data, and the directions are chosen such that the sum of all the vectors in the set is approximately zero. A new node is then created at the end of each vector, and an edge connects them. This process is repeated to generate a graph.

However, this algorithm has some limitations in that it can only describe one stromal subset yet in the follicle there are 3 distinct subsets with unique morphological features (**Chapter 4**). In addition, our datasets also show that a number of cell protrusions directly connect to one another, a property which can affect network topology. To account for this, branches between edges are added by creating a vector connecting the midpoints of each edge, subject to subtype-specific constraints on the maximum edge length and the local density such that the degree centrality matches our *in vivo* datasets. This approach is used to generate BRC and FDC networks while MRCs are stochastically seeded underneath the subscapular sinus by random sampling of X and Y values subject to density constraints to ensure consistency between *in vivo* and *in silico* edge lengths and degree centrailities. Edges between MRCs and other subtypes are created stochastically subject to distance and local density constraints. To represent the heterogeneity of network structures observed *in vivo* and to ensure no biases were introduced through a specific stromal architecture, the algorithm generates a unique network at the beginning of each simulation run. A suite of automated tests were developed using the pipeline described in **Chapter 3.3** to assess whether edge lengths and degree centralities are within expected bounds and that there are no overlapping nodes or edges.

Using this approach we generate a unique network at the start of each simulation run. Running 250 simulations we find that this approach yields networks with median sigma and omega values of **12.00** and **-0.097** respectively, confirming that the network has small world properties (**Figure 5.2**). The discrepancy between *in vivo* and *in silico* sigma values was anticipated as sigma scales with network size and the *in silico* follicle is

approximately 4 times the volume of the tissue section used to perform the topological mapping. Comparison of the median values for both edge lengths and degree centralities for *in silico* and *in vivo* networks shows no statistically significant differences (**Figure 5.3-5.4**).

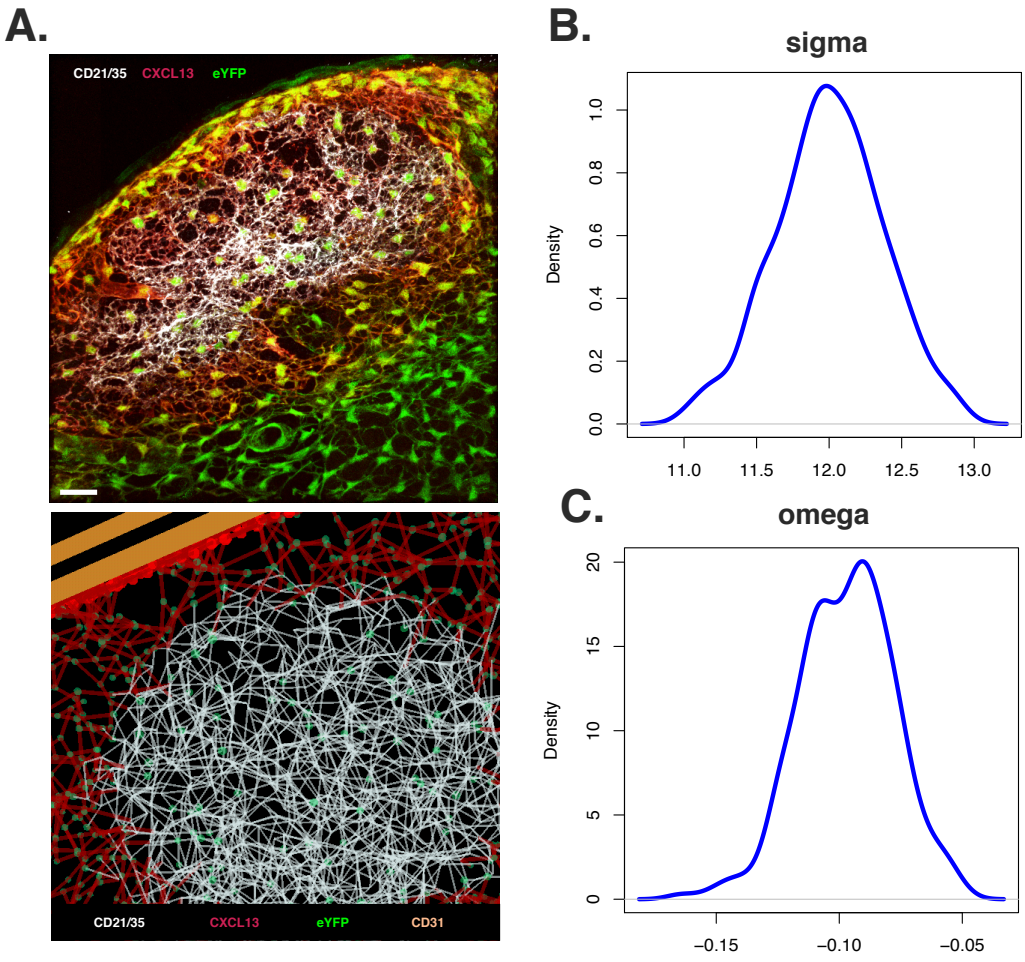

Figure 5.2. Development and validation of the Stroma module. (a) Top figure shows a lymph node follicle showing tdtomato (red) and eyfp (green) expression from CXCL13-EYFP mice. FDCs are marked with an antibody against CD21/35 (white). Bottom figure shows *in silico* stromal networks. (b) This figure shows the distribution of sigma values obtained under baseline parameter values from 250 simulation runs. (c) This figure shows the distribution of omega values obtained under baseline parameter values from 250 simulation runs. Scale bar = 50μm. Java code was developed by J.C and S.J.

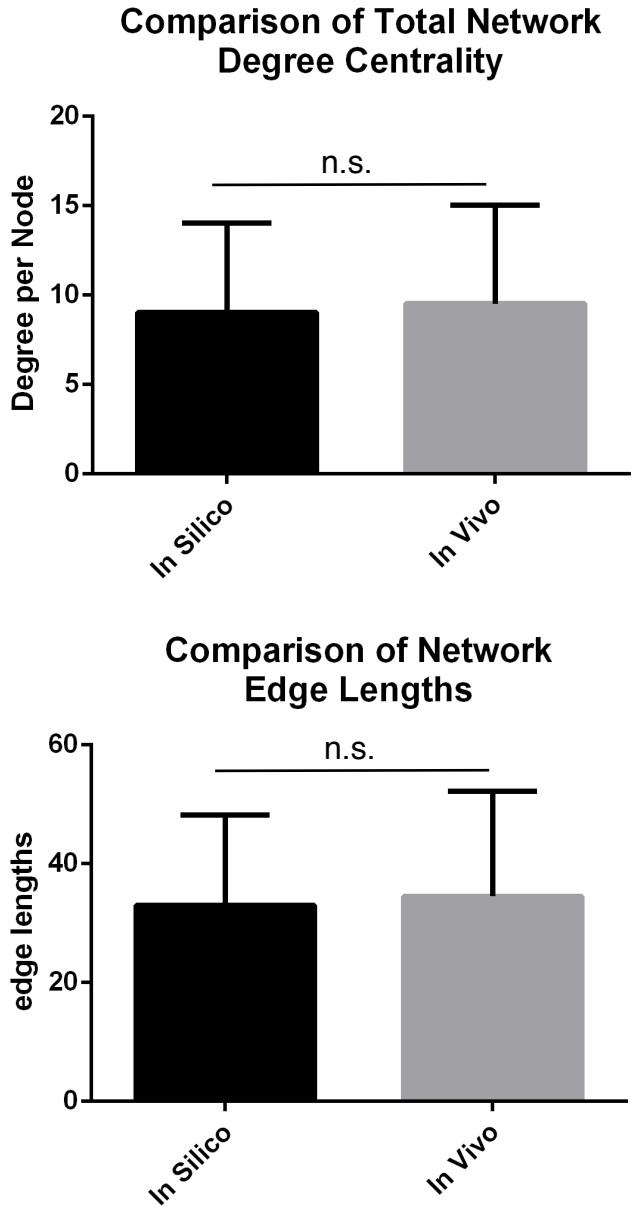

Figure 5.3 Comparison of *in vivo* and *in silico* edge lengths and degree centralities for the entire follicular network. No statistically significant differences were found between the median values for *in silico* and *in vivo* datasets as determined by a Mann-Whitney test with $p < 0.05$ representing a statistically significant result. Bars represent the median values and error bars represent the I.Q.R.

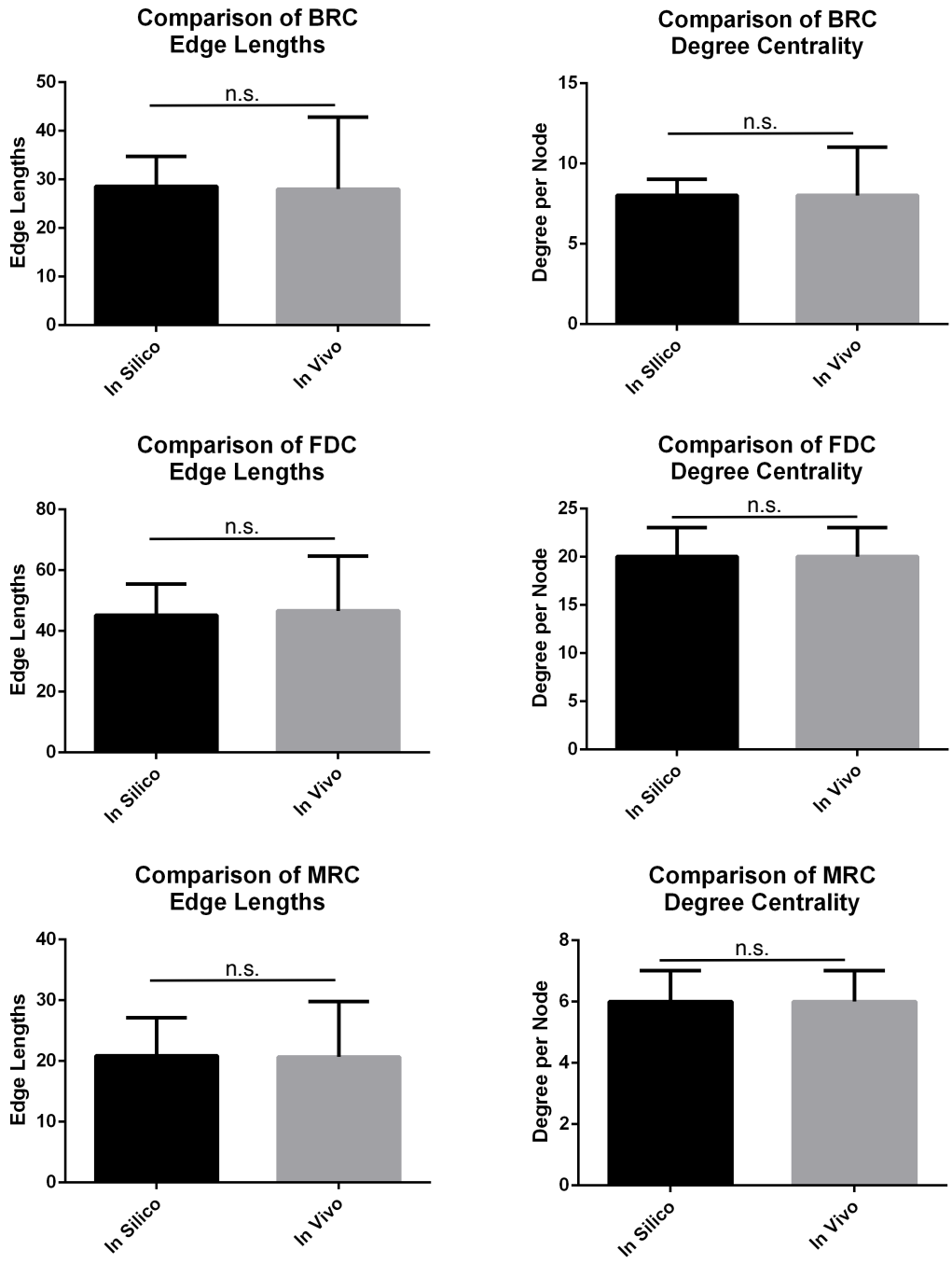

Figure 5.4 Comparison of *in vivo* and *in silico* edge lengths and degree centralities for the entire follicular network. No statistically significant differences were found between the median values for *in silico* and *in vivo* datasets as determined by a Mann-Whitney test with $p < 0.05$ representing a statistically significant result. Bars represent the median values and error bars represent the I.Q.R.

5.2.3. Module 2: Chemokine

Many different techniques exist to model molecules *in silico,* each with associated advantages and limitations. A common approach to model diffusion is through functions that relate molecular concentration to distance from a source (Alden et al., 2012b), or by PDEs (**Figure 5.5**) (Bocharov et al., 2011; Guo et al., 2008). As they describe molecular concentrations on a continuous scale at the population level, it can be difficult to incorporate complex behaviours such as localised binding effects into PDEs and distance-concentration functions. At the other extreme, it may also be intractable to model molecules (which exist in much larger numbers and move on faster scales than immune cells) using individual-based approaches. This limitation can make it difficult to simulate the dynamics of chemokine field formation where molecules simultaneously undergo production, diffusion, decay, binding and scavenging, key mechanisms required to shape functional chemotactic gradients (Rot and von Andrian, 2004). In cases where high model granularity is required at the molecular level, soluble factors can be represented as floating point values on discretised grids (**Figure 5.5**). The scheme we implement is a discretised form of the heat equation (Grajdeanu, 2007). This mathematical construct is capable of isotropic diffusion[25], can diffuse to an arbitrary number of neighbours and is applicable to linear, planar, spatial and n-dimensional implementations. These attributes make it well suited to studies of molecular components of the immune system (**Figure 5.5**).

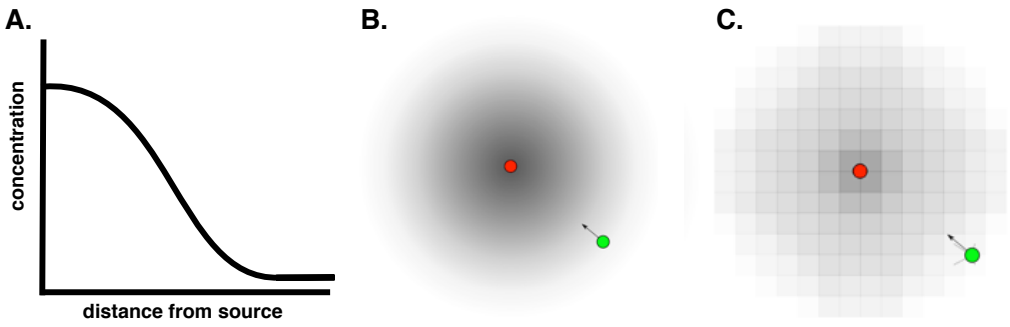

Figure 5.5. **Different schemes to model diffusion in theoretical models.** (A) Functions that relate concentration to distance from a source. (B) PDEs predict changes in concentration over time and space on a continuous scale. (C) Discrete PDEs in which the environment is binned into discrete gridspaces; within each gridspace concentration is homogeneous.

[25] Anisotropy is an implementation artifact where diffusion occurs faster in certain directions than others, making the diffusion neighborhood look square

In this scheme, chemokine molecules diffuses through a discrete 3D environment where the number of moles of chemokine molecules in each grid space *(x,y,z)* is denoted $\varphi(x,y,z)$. The change in the spatial distribution of molecules is then subject to the following simultaneously occurring processes *(i) production (ii) diffusion (iii) decay and (iv) consumption*. As we are using agents to model individual cells, terms (i) and (iv) emerge from the simulation. Chemokine is secreted by each stromal cell at a fixed rate and is removed from the grid at a rate that is proportional to its current value λ.

$$\varphi(\mathrm{x},\mathrm{y},\mathrm{z})_{t+1} = (1-\lambda)\varphi(\mathrm{x},\mathrm{y},\mathrm{z})_t \qquad (5.1)$$

where *t* represents the time step. At each discrete time step chemokine diffuses to the grid spaces adjacent and diagonally adjacent to each grid space. The coefficient for the amount of chemokine diffused to each grid space is:

$$A\big(\varphi(x) - \varphi(\mathrm{y})\big)e^{-\frac{d_y^2}{\mu}} \qquad (5.2)$$

$$\mu = 4Dt \qquad (5.3)$$

where chemokine in grid space *x* ($\varphi(x)$) is being diffused to grid space *y*, d_y^2 is the distance squared between *x* and *y*, *D* is the diffusion constant, *t* is the time step, and A is a normalizing constant that ensures the total amount being diffused is less than or equal to the amount that exists:

$$A\ \sum_{y=1}^{n} e^{-\frac{d_y^2}{\mu}} = 1 \qquad (5.4)$$

for an arbitrary grid space where n represents the number of surrounding grid spaces in the diffusion neighbourhood, which in CXCL13Sim is a 3D Moore neighbourhood with 26 neighbouring grid spaces. This approach mitigates an artefact known as anisotropy, where diffusion appears "square" because treating each grid space equally would favour diffusion diagonally due to the larger distance to the corner neighbours than the lateral ones (**Figure 5.6**). This implementation explicitly takes the distance (d_y^2) between gridspace *x* and its adjacent grid space *y* into account to avoid this effect without the need to implement a tessellated hexagonal grid, which can only be done in 2D. The borders of the grid follow Dirichlet boundary conditions.

When modelling diffusion in discrete space, the speed at which chemokine diffuses is limited by the time step and size of the diffusion neighbourhood. Increasing the size of the diffusion neighbourhood would allow greater diffusion coefficients without a smaller time step, but would significantly increase the time resources required to run the algorithm. To account for this, the time step we use is $\frac{\delta^2}{kD}$ where δ is the length of an edge of the discretized grid spaces, D is the diffusion coefficient, and k is a constant, empirically determined in order to match the measured diffusion speed with the mathematically derived value for mean-squared displacement for a given diffusion constant. The diffusion grid is updated every second while agents are updated once every minute and are assumed quasi-static with respect to diffusion. Automated tests were developed to ensure that chemokine is conserved when the decay constant is set to zero and no agents are in the simulation, that the diffusion coefficient input gives the expected mean-squared displacement output and that the rate of diffusion per time step cannot extend beyond the diffusion neighbourhood.

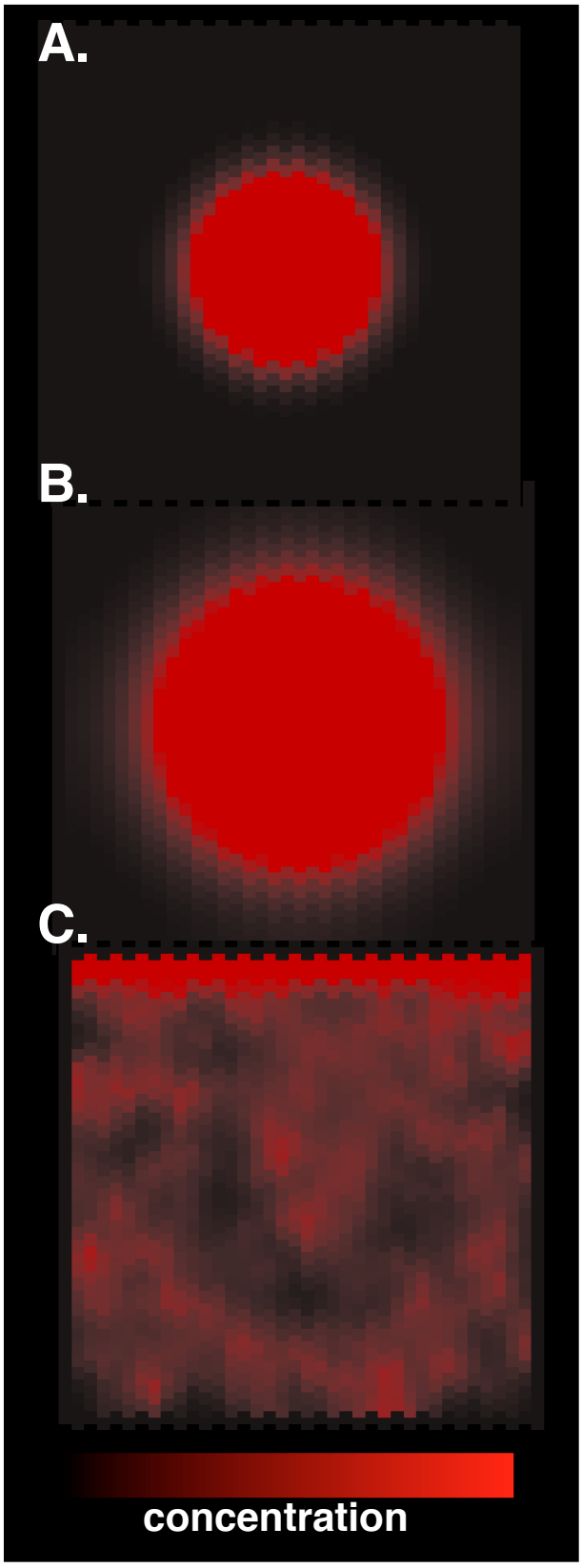

Figure 5.6. **Examples of gradients formed using the Gradjeanu scheme.** (A) a chemokine field with low diffusion rate and high decay rate leading to a small circle of CXCL13 expression (red) (B) a chemokine field with high diffusion rate and low decay rate leading to a larger CXCL13 field (C) shows a complex chemokine field generated by a stromal network generated by the algorithm described in section 5.2.2. Java code was developed by S.J and J.C.

5.2.4. Module 3: Lymphocytes

5.2.4.1. Lymphocyte Migration

Chemotaxis, and chemokine receptor internalisation and recycling are key mechanisms governing the fine-tuning of responses to chemokines *in vivo* (Bennett et al., 2011; Lin and Butcher, 2008; Rot and von Andrian, 2004). In addition, the follicle is a highly dense structure and so it is important to account for interactions between cells. To model these phenomena *in silico,* we adapt the scheme developed by Lin *et al* (Lin and Butcher, 2008; Wu and Lin, 2011). In this scheme an agent samples local chemokine concentrations using 6 sampling pseudopodia (**Figure 5.7**). At each pseudopod there are a population of receptors on the cell surface [R_f] that are free to bind ligand [L] at a rate K_{on}. Receptor dynamics are controlled by a set of ordinary differential equations solved on a per agent basis using a 4^{th} order Runge-Kutta scheme (Press et al., 2007). Ligation of the chemokine to its respective receptor [LR] leads to downstream signalling cascades and localisation of actin with ligand dissociating at a rate K_{off}. Following binding receptors are desensitized at a rate K_{des}, internalised at a rate K_i and are recycled at a rate K_r (**Figure 5.7**). The values for K_{on} and K_{off} are set so that the affinity constant (K_d), that is the ligand concentration at which half of the receptors are bound, is set to 10nM, the baseline concentration within the simulator.

To quantify the influence of these parameters on the ODE outputs we performed a global sensitivity analysis using latin hypercube sampling and PRCCs (**Chapter 2**) in ASPASIA, a toolkit for evaluating interventions on systems biology markup language (SBML) model behaviours (**Figure 5.9**) (Evans et al., 2017). It is important to note that by looking at the rank correlation gives an incomplete understanding of parametric uncertainty as it does not take the magnitude of the change of model outputs into account. This is addressed by the use of an eFAST analysis in **Chapter 5.6**. From calculating [LR] at each pseudopod, a gradient vector, $\overrightarrow{LR}$ is calculated across the cell along 3 axes with ρ representing each individual pseudopod. If $\overrightarrow{LR}$ exceeds a threshold then the cell will become chemotactic, with the overall orientation vector of the cell $\overrightarrow{LR_{total}}$ taken as a sum of $\overrightarrow{LR}$ with a leading edge vector $\overrightarrow{LR_m}$ that accounts for the orientation of the cell from the previous time step.

$$\overrightarrow{LR} = \left([LR_{\rho 1}] - [LR_{\rho 4}]\right) + \left([LR_{\rho 2}] - [LR_{\rho 5}]\right) + \left([LR_{\rho 3}] - [LR_{\rho 6}]\right) \quad (5.5)$$

$$\overrightarrow{LR_{total}} = \frac{\alpha\overrightarrow{LR} + \overrightarrow{LR_m}}{\left|\alpha\overrightarrow{LR} + \overrightarrow{LR_m}\right|} \quad (5.6)$$

$$\frac{d[\mathrm{LR}]}{dT} = K_{on}[L]\left[R_f\right] - K_{des}[LR] - K_{off}[LR] \quad (5.7)$$

$$\frac{d[R_f]}{dT} = K_r[R_i] - K_{on}[L]\left[R_f\right] + K_{off}[LR] \quad (5.8)$$

$$\frac{d[R_i]}{dT} = K_i[R_{des}] - K_r[R_i] \quad (5.9)$$

$$\frac{d[R_{des}]}{dT} = K_{des}[LR] - K_i[R_{des}] \quad (5.10)$$

In the presence of chemotactic gradients actin flows polarize at the leading edge of the cell thus the relative weighting between $\overrightarrow{LR}$ and $\overrightarrow{LR_m}$ is scaled by a constant α to represent the persistence of the cell; the value of α is dependent on the chemotactic state of the agent. As a universal coupling exists between actin flows and cell speed (Maiuri et al., 2015) and relate the increase in velocity v^* observed during chemokinesis to cell persistence α using the following expression:

$$v* = \frac{LN(\alpha)}{\gamma} \quad (5.11)$$

$$\alpha = \begin{cases} \alpha_1, \ if\ chemotactic \\ \alpha_2, \ if\ not\ chemotactic \end{cases} \quad (5.12)$$

The value of the scalar term γ was determined empirically, by fitting the model to WT migration data and verifying against CXCR5$^{-/-}$ migration patterns (Coelho et al., 2013). A number of automated tests were developed to ensure that total receptor values are conserved over time, and that agents move towards high concentrations of chemokines when expressing CXCR5.

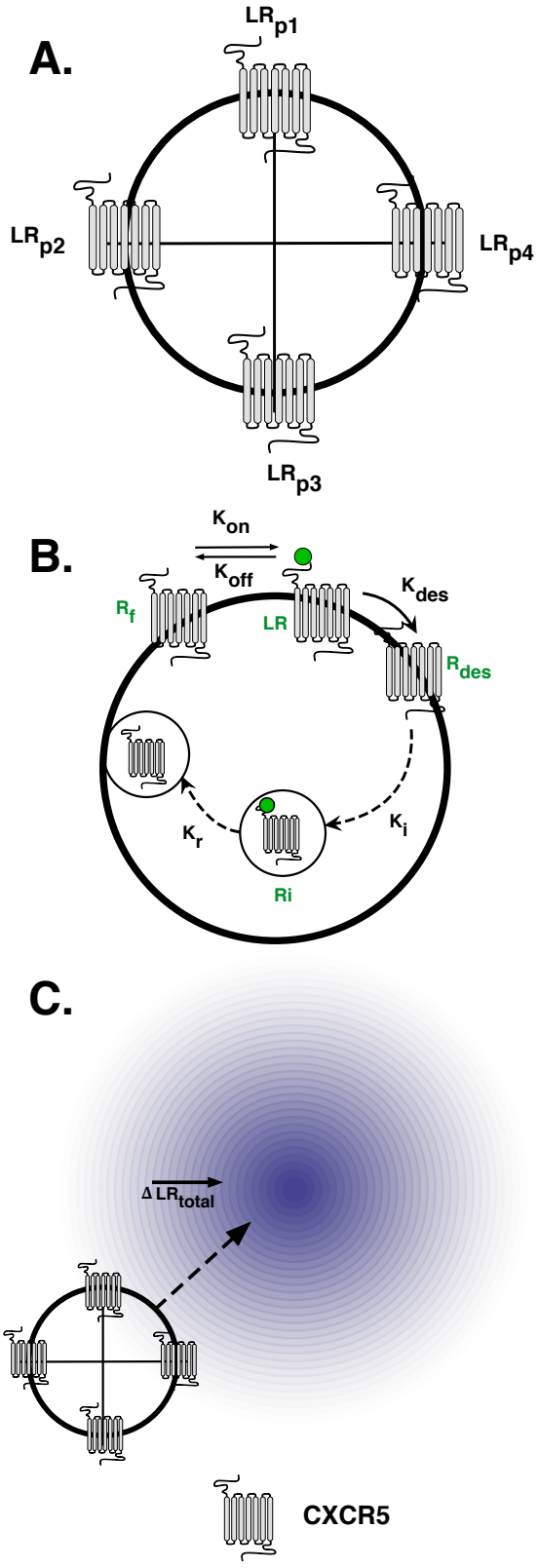

Figure 5.7 Overview of receptor and migration kinetics in CXCL13Sim. Each cell has 6 chemokine sampling pseudopods (4 are shown). At each pseudopod, signalling is a function of local chemokine concentrations and receptor expression. Receptor expression is dynamic and subject to ligand association/dissociation as well as receptor desensitization, internalization and recycling. From the amount of receptors signalling at each pseudopod, gradient vectors are calculated along 3 axes. The overall net movement vector LRtotal determined by summing these vectors with a polarity vector representing cell persistence in a given direction.

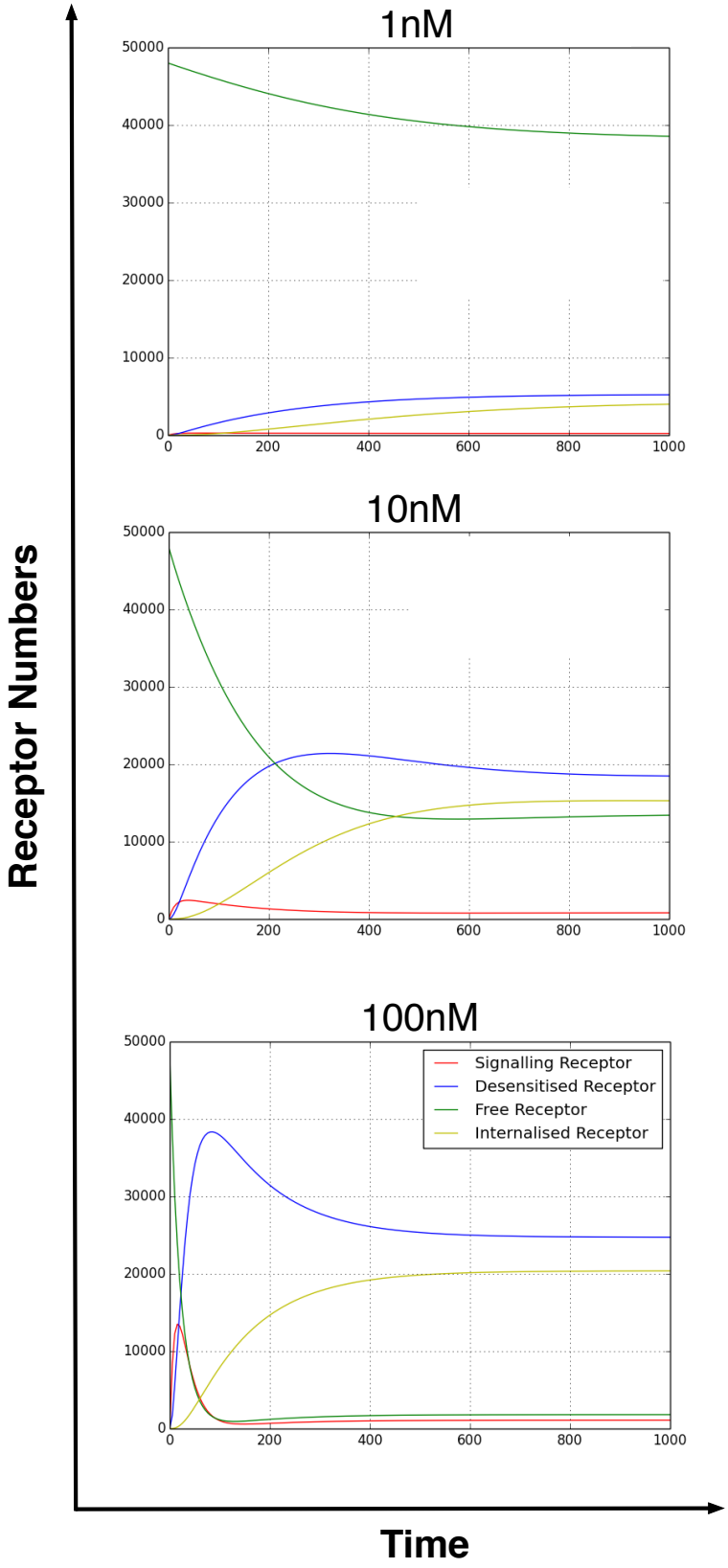

Figure 5.8. **CXCR5 Kinetics for 1nM, 10nM and 100nM CXCL13.** Receptor kinetics with calibrated parameter values in response to 1nM show large numbers of free receptors on the cell surface with relatively low levels of desensitised, internalised and signalling receptors. As the amount of chemokine increases, the amount of free receptors decreases and the number of desensitised receptors increases.

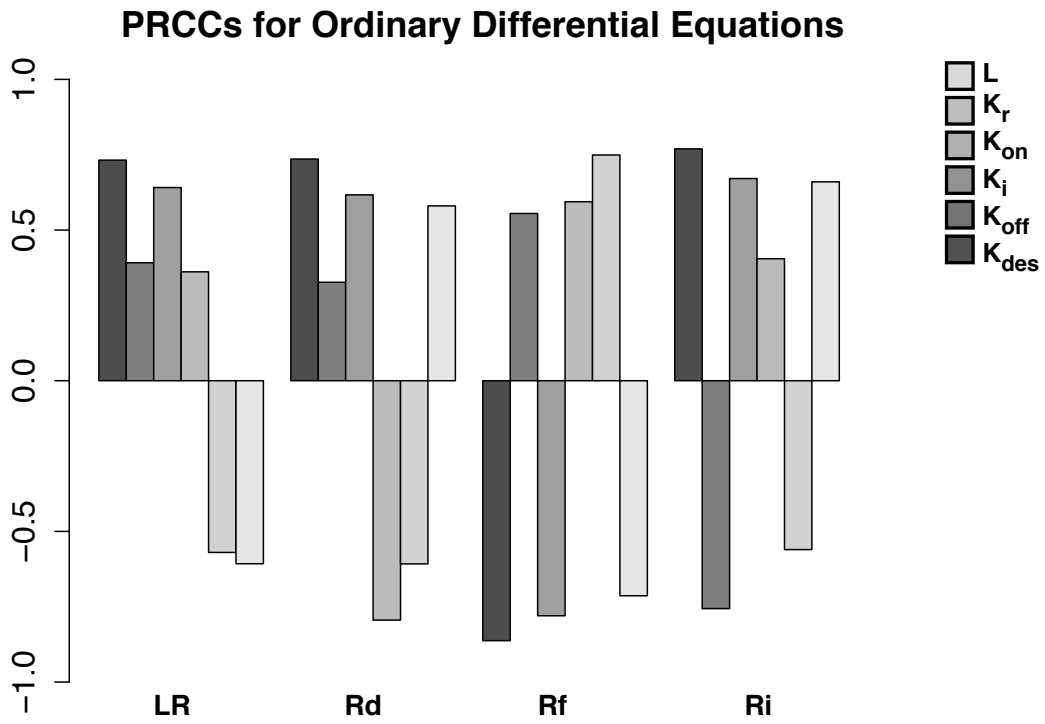

Figure 5.9. Parameter sensitivities for ODE parameters performed using ASPASIA (Evans et al., 2017). The influence of a parameter on model outputs is summarised using a partial rank correlation coefficient.

5.2.4.2 Lymphocyte Interactions

To account for dense lymph node environment lymphocyte migration must take into account interactions with other cell types. As time proceeds in fixed discrete intervals (each time step representing 1 minute in the physical system) we treat both the movement vector of lymphocytes and the edge of the stromal cell as lines (**Figure 5.10**). To determine if the two agents are interacting we calculate whether the closest point between the two lines is less than the sum of their diameters. To determine the closest point, we define the lines L1 and L2 as follows:

$$L1(s) = P1 + sd1, \text{where } d1 = Q1 - P1 \quad (5.13)$$
$$L2(t) = P2 + td2, \text{where } d2 = Q2 - P2 \quad (5.14)$$

Then for some pair of values for s and t, L1 (s) and L2 (t) correspond to the closest points on the lines, and v(s,t) describes a vector between them.

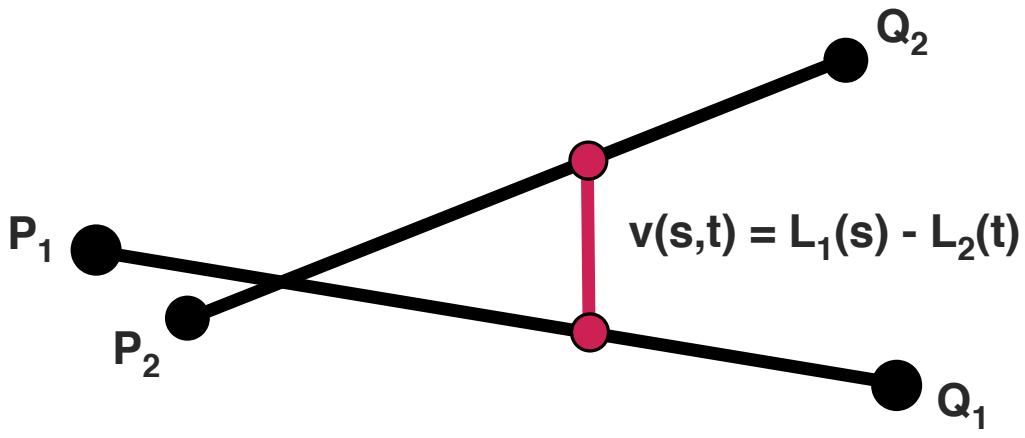

Figure 5.10 Modelling lymphocyte interactions with stroma. In this scheme the movement vector of lymphocytes and stromal cell processes are modelled as lines. Each line is defined by two points, Pi and Qi. To determine if the two agents are interacting we calculate whether the closest point between the two lines is less than the sum of their diameters. Then for some pair of values for s and t, L1 (s) and L2 (t) correspond to the closest points on the lines, and v(s,t) describes a vector between them.

The closest point between the two lines is obtained when the vector is perpendicular to both lines (**Figure 5.10**) i.e. when the dot product of the two vectors is equal to zero or

$$d1 * v(s,t) = 0 \quad (5.15)$$
$$d2 * v(s,t) = 0 \quad (5.16)$$

Using Cramer's rule we can then solve this system of equations to determine which values s and t where L1(s) and L2(t) correspond to the closest points on the lines. Additionally, cells may interact with each other, however cell structure is dynamic and cells are observed to slide over one another. To account for this, once an agent has determined the direction in which to move, the probability that the cell can move towards the target location is determined as $e^{-\delta}$, where δ is the number of cells in the target location. Automated tests were developed to ensure that lymphocytes completely caged within a tight network of stromal cell protrusions cannot pass through due to interacting with the network, and that the number of agents per gridspace does not exceed a threshold value.

5.2.5 Integration of Model Subunits

System architecture is modelled using an adaptation of the UML as per the domain model. This specification defines how model subunits interface and details the flow of information through the system (**Figure 5.11 – 5.15**). On the diagrams, the modules are specified using: *M1, M2* and *M3* to represent stroma, chemokine and lymphocytes respectively. Key decisions and abstractions made during the development process are presented as arguments over evidence[26] using an adaptation of goal structuring notation.

[26] Available from https://www.york.ac.uk/computational-immunology/software/cxcl13sim/

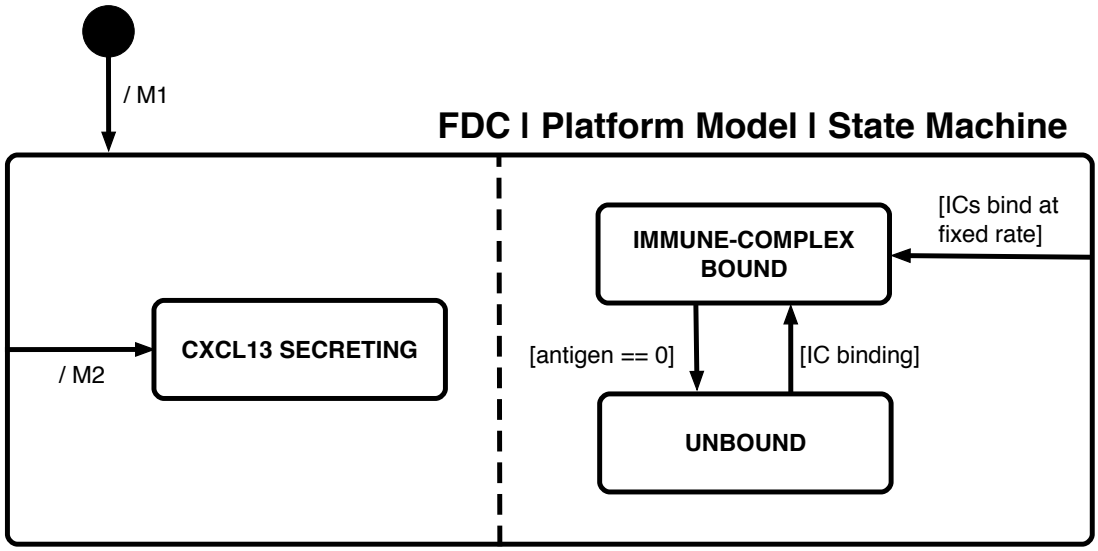

Figure 5.11. State machine diagram for a Follicular Dendritic Cell. FDCs are resident in the system at t0 and are generated using the algorithm described in M1. FDCs secrete chemokine at a fixed rate and once secreted, chemokine diffuses as described in M2. Antigen levels are expressed as integers and are decreased following interactions with cognate B cells. LTβR mediated stimulation of CXCL13 production is assumed constant in homeostatic conditions and is thus not explicitly referenced in the Platform Model.

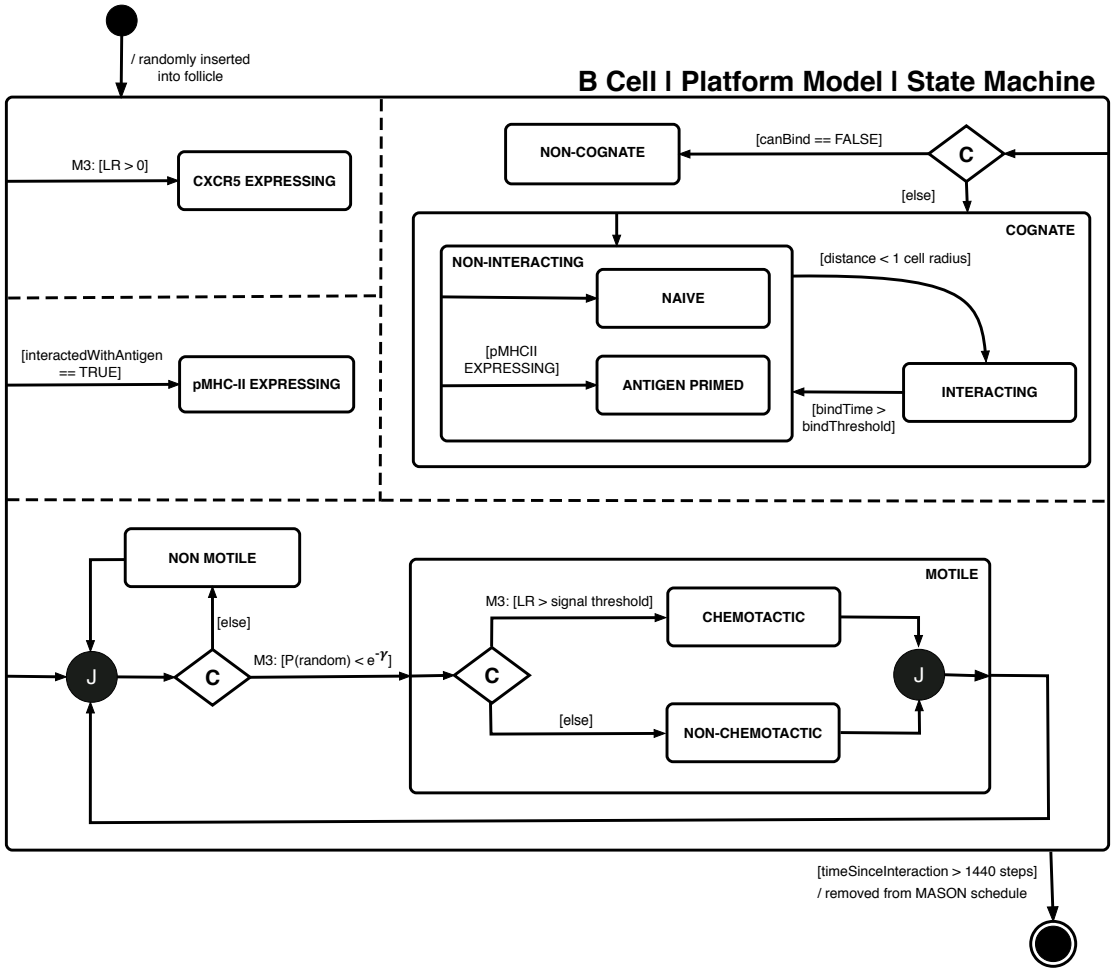

Figure 5.12. State machine diagram for a B cell. B cells are seeded randomly in the follicle at the start of the simulation in a cognate or non-cognate state on the basis of antigen specificity. For a cognate cell to become primed it must be within 1 cell radius of cognate antigen. With respect to M3, a B cell may change into a CXCL13 desensitised state if the magnitude of the cells orientation vector does not exceed a threshold value. To migrate, a B cell must determine if there is space to move and will then move either randomly or in the direction of a gradient on the basis of it sensitivity to ligand.

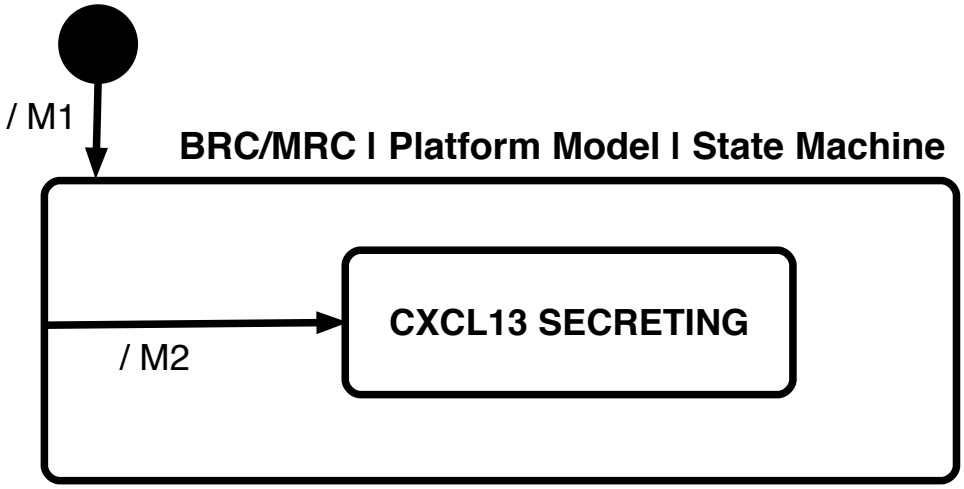

Figure 5.13. **State machine diagram for Marginal and B reticular cells.** MRCs/BRCs are resident in the system at the start of the simulation and are generated using the algorithm described in M1. They secrete chemokine at a fixed rate and once secreted, chemokine diffuses as described in M2

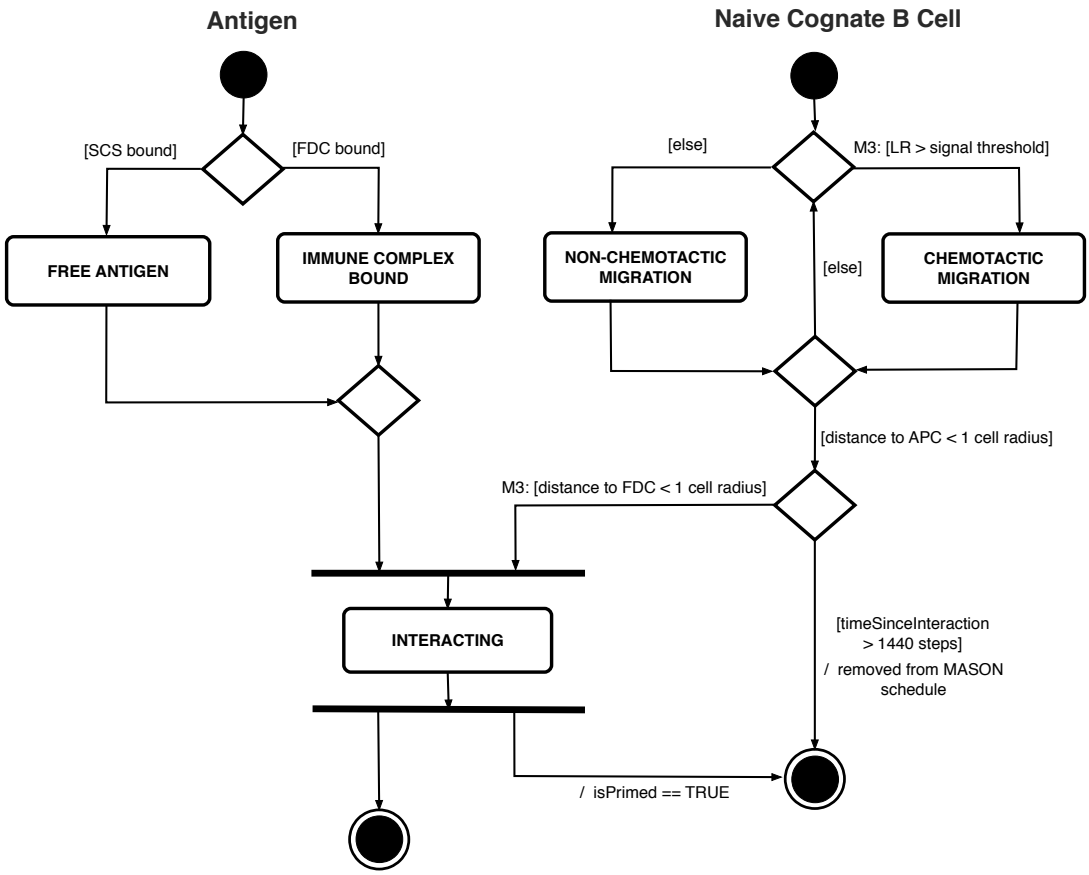

Figure 5.14. Activity diagram for B cells: At the beginning of the activity FDCs express a fixed number of antigen, specified by an integer value. CXCL13 is secreted by FDCs, diffuses and is removed from the grid as specified in M2. If the number of signalling CXCR5 molecules ([LR]) exceeds a threshold value then the B cell is capable of detecting a chemokine gradient. When the magnitude of the orientation vector LRtotal exceeds a signalling threshold then the cell will orientate towards the gradient otherwise it will orientate randomly. Before migrating a B cell must also determine if it has space available and to become primed the B cell must interact with antigen.

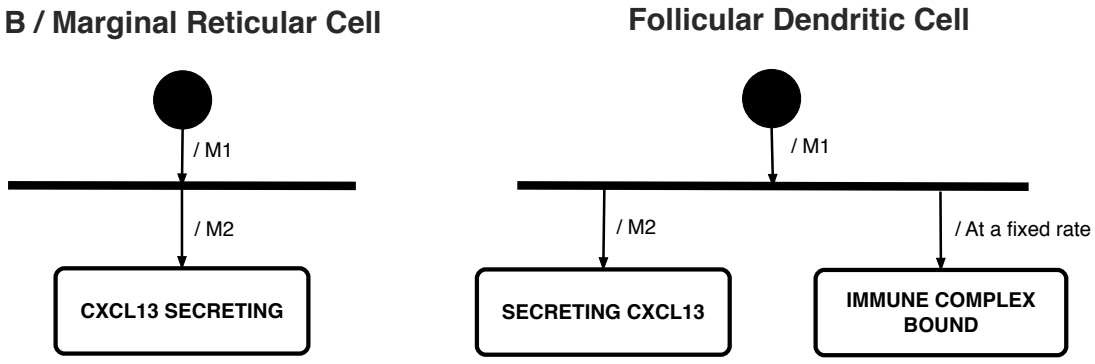

Figure 5.15. **Activity diagram for stromal cells:** At the beginning of the activity FDCs express a fixed number of antigen, specified by an integer value. CXCL13 is secreted by FDCs, BRCs and MRCs at a fixed rate.

5.2.6 Model Outputs

Once a simulation run is complete, the simulator writes a .csv file with the metrics detailed in **Table 5.1** for each cognate B-cell agent. These metrics facilitate comparison with experimental measures of migration and are used to assess the influence of parameter perturbations on the emergent cellular behaviours.

Measure	Description
Total Displacement	Record the steps taken by cells and calculate displacement over a fixed time period using vector addition.
Net Displacement	Euclidean distance between the first and last position of the cell
Cell Velocity	Total displacement / time
Motility Coefficient	Net displacement^2 / 6* time
Meandering Index	√Time * (net displacement / total displacement)
checkPointsReached	Number of unique gridspaces reached

Table 5.1 Summary of model outputs. Following each individual simulation run, the following metrics are calculated for each B-cell agent.

5.3 Simulation Platform

5.3.1 Model Calibration and Validation

Within the simulation platform each module was developed using Java and the multi agent simulation library MASON (Luke et al., 2005). CXCL13Sim passes all associated tests with 97% code coverage. A total of 138 tests are incorporated into an automated regression-testing pipeline using the eclEmma package (details of how to access the package are provided in Chapter 2.2.4). Example code is provided for a small subset of tests in **Appendix 3**.

To calibrate the simulator, the 13 free parameters were systematically changed and outputs were compared to *in vivo* multiphoton datasets. Fluorescently labelled wild-type and $CXCR5^{-/-}$ B cells were adoptively transferred into WT hosts and their distribution and migration patterns inside popliteal lymph nodes analysed using selective plane illumination microscopy, which preserves the three-dimensional organ structure. The median values for 250 simulation runs are 1.15 for the meandering index, 13.28 for motility coefficient and 8.11 for speed (**Figure 5.16**). Comparison of a single *In silico* run with *in vivo* datasets showed no significant differences between motility coefficients, meandering indices and velocities for (a) wild-type or (b) $CXCR5^{-/-}$ B cells; assessed using a Mann-Whitney test at a significance level of 5% (**Figure 5.17**). The additional 9 parameters were fixed at empirically determined or calibrated values representing key attributes such as cell size and the density of the stromal cell network that are not designed to change between simulation runs.

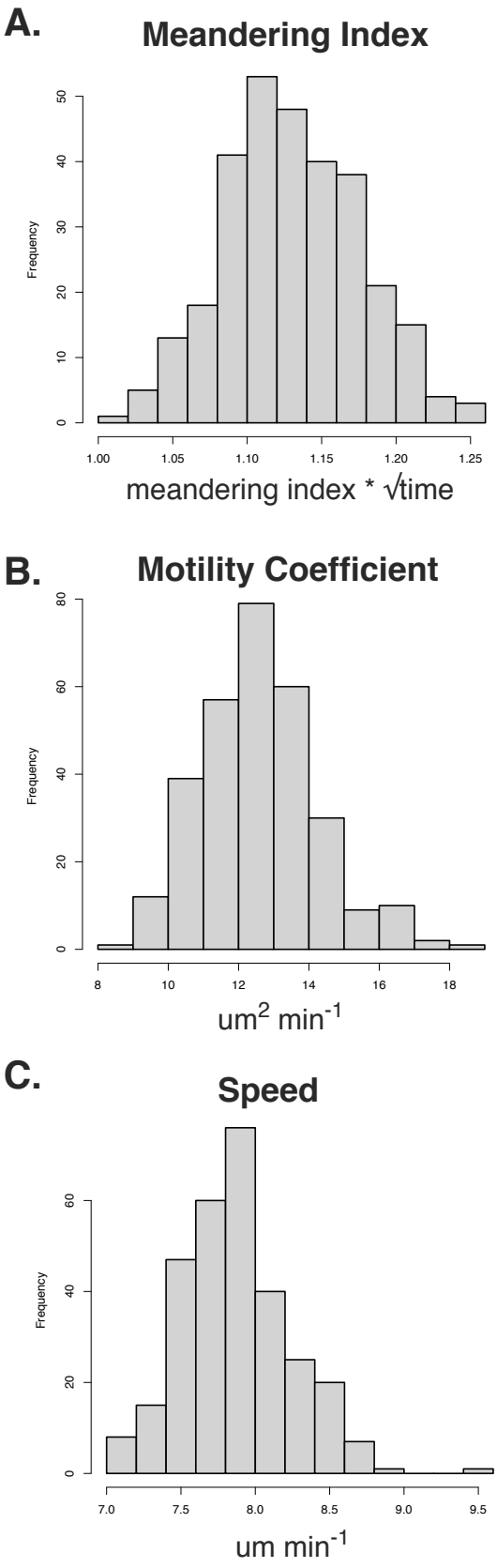

Figure 5.16. Distributions of emergent cell behaviours under baseline conditions. (A) Distribution of meandering index values for baseline parameter values. (B) Distribution of motility coefficient values for baseline parameter values. (C) Distribution of speed values for baseline parameter values.

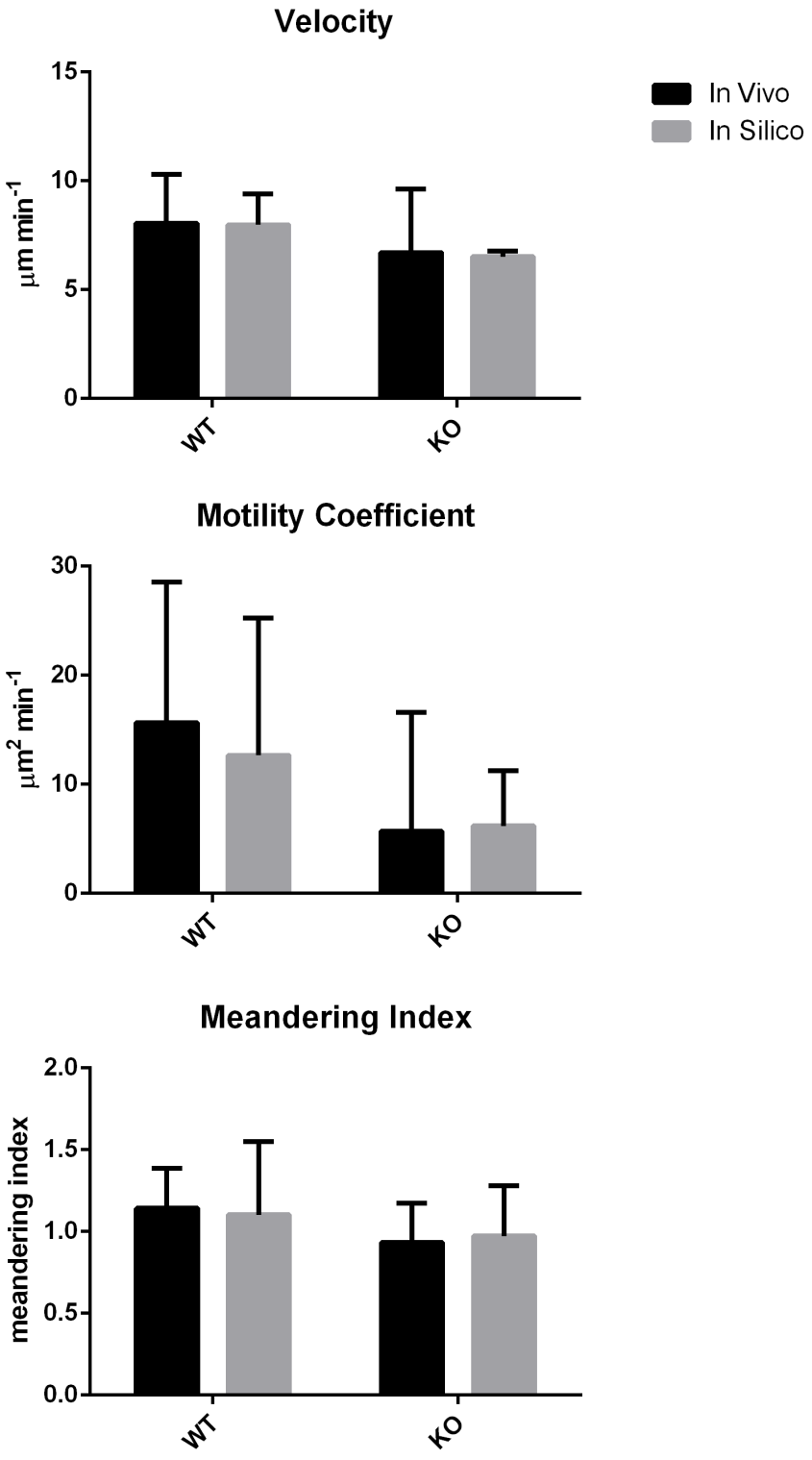

Figure 5.17 Calibration of emergent behaviours against experimental behaviours. No significant differences were observed between *in silico* and *in vivo* motility coefficients, meandering indices and velocities for wild type or CXCR5$^{-/-}$ B cells. All datasets were non-normal as determined by a Shapiro-Wilk test, subsequently; a Mann-Whitney test was used to determine if datasets were significantly different at a significance level of 5%. Bar charts represent the median value with error bars representing the I.Q.R.

Parameter	Value	Unit	Range	Reference
B Cell Size	7	µm	Constant	(Monroe and Cambier, 1983)
Total Number of B cells	6000	cells	Constant	Measured
Total Number of MRCs	100	cells	Constant	Measured
Total Number of FDCs	~200	cells	Constant	Measured
Total Number of BRCs	~450	cells	Constant	Measured
Proportion of Cognate Cells	5	%	Constant	-
Displacement constant	7.4	µm min^{-1}	[1-10]	Calibrated
Signal threshold	10	ΔLR	Constant	(Herzmark et al., 2007; Zigmond, 1981)
Maximum turn angle	180	Degrees	Constant	(Coelho et al., 2013)
Total receptor number	48,000	Receptors	[10,000-100,000]	(Sh et al., 1982)
K_{on}	4.8 x 10^5	M s^{-1}	[1x10^5-1x10^6]	(Barroso et al., 2012)
K_i	0.0033	s^{-1}	[0.001-0.01]	(Sh et al., 1982; Tilo Beyer, 2007)
K_{des}	0.075	s^{-1}	[0.01-0.1]	(Sh et al., 1982; Tilo Beyer, 2007)
K_r	0.004	s^{-1}	[0.001-0.01]	(Sh et al., 1982; Tilo Beyer, 2007)
K_{off}	0.0048	s^{-1}	[0.001-0.01]	(Sh et al., 1982; Tilo Beyer, 2007)
FDC secretion rate	0.18	fg min^{-1} $cell^{-1}$	[0.1-0.5]	(Gunn et al., 1998a; Luther et al., 2002)
MRC secretion rate	0.18	fg min^{-1} $cell^{-1}$	[0.1-0.5]	(Gunn et al., 1998a; Luther et al., 2002)
BRC secretion rate	0.18	fg min^{-1} $cell^{-1}$	[0.1-0.5]	(Gunn et al., 1998a; Luther et al., 2002)
CXCL13 decay rate	0.007	s^{-1}	[0.0002-0.05]	(Phair and Misteli, 2001b; Wang and Irvine, 2013)
CXCL13 diffusion rate	7.6	$µm^2$ s^{-1}	[0-146]	Measured
Polarity	0.475	-	0-1	Calibrated
Random Polarity	3.8	-	Constant	Calibrated

Table 5.2. Summary of parameter values. For each parameter the name, baseline value and range used for uncertainty and sensitivity analyses is provided. Parameter values were determined experimentally or in cases where no direct experimental value exists, upper and lower limits were derived from indirect evidence, baseline values were then determined by fitting the model to experimental datasets (calibration). The model was further validated against migration data from $CXCR5^{-/-}$ B cells and parameters were removed where possible. The values for stromal cells are averaged over 250 runs with individual values varying to a small extent between runs due to stochastic network formation.

5.4 Understanding Aleatory and Parameter Uncertainty

5.4.1 Mitigation of Aleatory Uncertainty

CXCL13Sim is non-deterministic and therefore, repeat experiments using the same parameter set can lead to differing results (**Figure 5.16**). This variation is termed aleatory uncertainty and because of this effect multiple simulation executions must be performed to obtain a representative result. To determine how many runs are required to give a representative output for a given parameter set we perform an aleatory analysis (detailed further in **Chapter 2.2.7**) (Alden et al., 2013; Read, 2011) (**Figure 5.18**). In this approach, distributions of simulation outputs generated using a fixed parameter set are compared. By varying the number of samples comprising the distributions, the analysis determines the number required to obtain statistically consistent distributions (as detailed in **Chapter 2.2.7**). This analysis shows that large differences between parameter sets can be detected with 50 replicate run for all outputs while smaller differences require 250 runs (threshold values to determine each effect size are detailed further in **Chapter 2.2.7**).

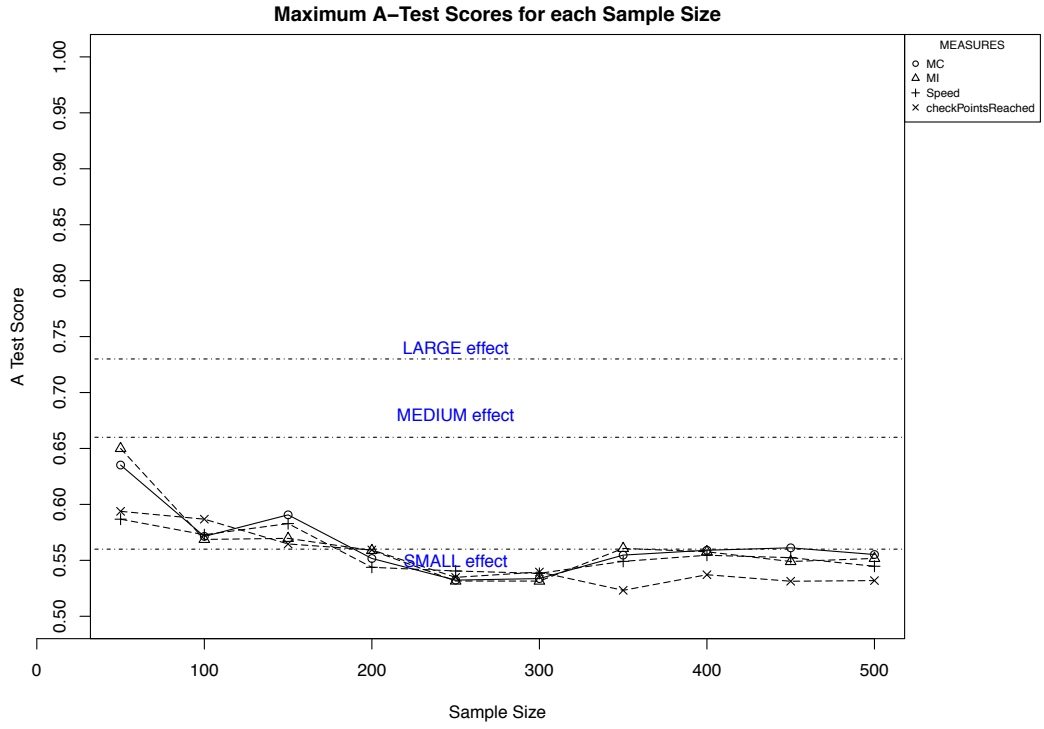

Figure 5.18. **Mitigating aleatory uncertainty.** An aleatory uncertainty analysis was performed using the SPARTAN package in R. This analysis shows that 250 runs are required to provide enough power to detect small effects in all model outputs. Based on this analysis parameter sensitivity analyses were run with 50 replicates while *in silico* experiments are run with 250 replicates. checkPointsReached (the amount of unique grid spaces reached within a simulation run) ✖; meandering index (MI) Δ; motility coefficient (MC) ○; speed ✚.

5.4.2 Local Sensitivity Analysis

To quantify parameter uncertainty in CXCL13Sim we first perform a local parameter SA using the SPARTAN statistical package as follows: each parameter is adjusted within the ranges specified in **Table 5.2**, with all other parameters remaining at their calibrated value with 100 replicates used for each parameter set to mitigate aleatory uncertainty. The Vargha-Delaney A-Test described previously (**section 2.2.8**) is employed to determine if changing the parameter value has led to significant difference in comparison with baseline behaviours (Vargha and Delaney, 2000).

In this analysis we divided parameters into 3 groups. Parameters controlling B cell migration (B-cell displacement constant and cell polarity) had a significant impact on both cell migration and scanning rates (**Figure 5.19**). Analysis of parameters controlling CXCR5 expression (K_{des}, K_{off}, K_r, K_i, K_{on} and R_f) (**Figure 5.20**) show a high level of uncertainty around parameters K_{off}, K_r, K_i with perturbations not leading to a large difference in cell migration and scanning rates. However, K_{des}, K_{on} and R_f were influential in regulating cell migration. Only perturbed rates of K_{des} altered the rate of scanning (**Figure 5.20**). Perturbations to parameters controlling CXCL13 expression (cell secretion rates, diffusion constant and decay rate) led to significant changes in both cell behaviours and scanning rates (**Figure 5.21**).

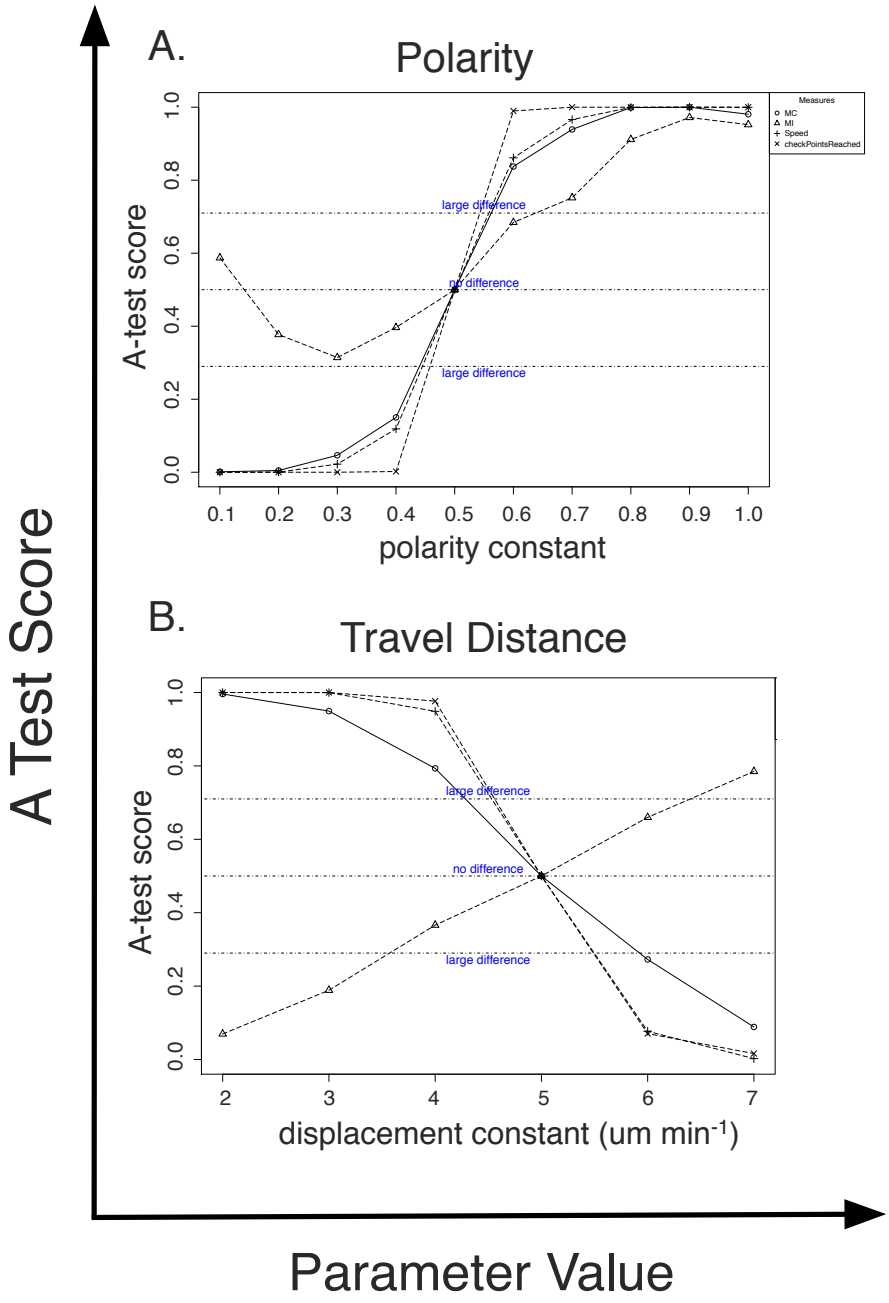

Figure 5.19. A-test scores when OAT adjusting parameters which relate to B-cell migration and numbers. Parameters *polarity (A) and travel distance (B)* were incrementally changed within their likelihood distributions using an OAT parameter robustness approach. A significant alteration in simulation outputs from baseline behaviours was determined using the Vargha-Delaney A-Test. checkPointsReached ×; meandering index (MI) Δ; motility coefficient (MC) **O**; speed **+**.

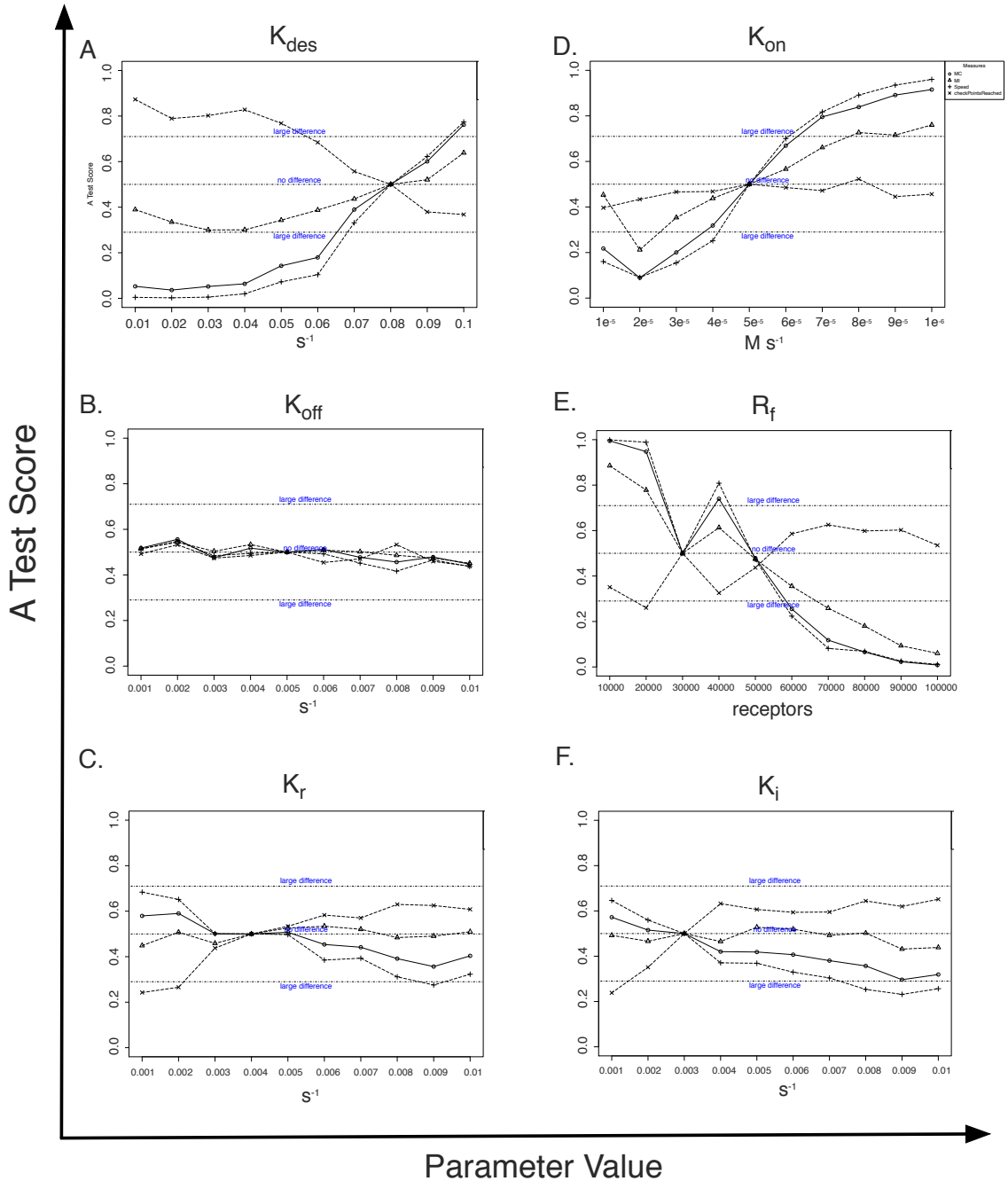

Figure 5.20. A-test scores when OAT adjusting parameters which relate to CXCR5 expression. Parameters K_{des} *(A),* K_{off} *(B),* K_r *(C),* K_{on} *(D),* Rf (E), and K_i *(F),* were incrementally changed within their likelihood distributions using an OAT parameter robustness approach. A significant alteration in simulation outputs from baseline behaviours was determined using the Vargha-Delaney A-Test. checkPointsReached ×; meandering index (MI) Δ; motility coefficient (MC) **O**; speed **+**.

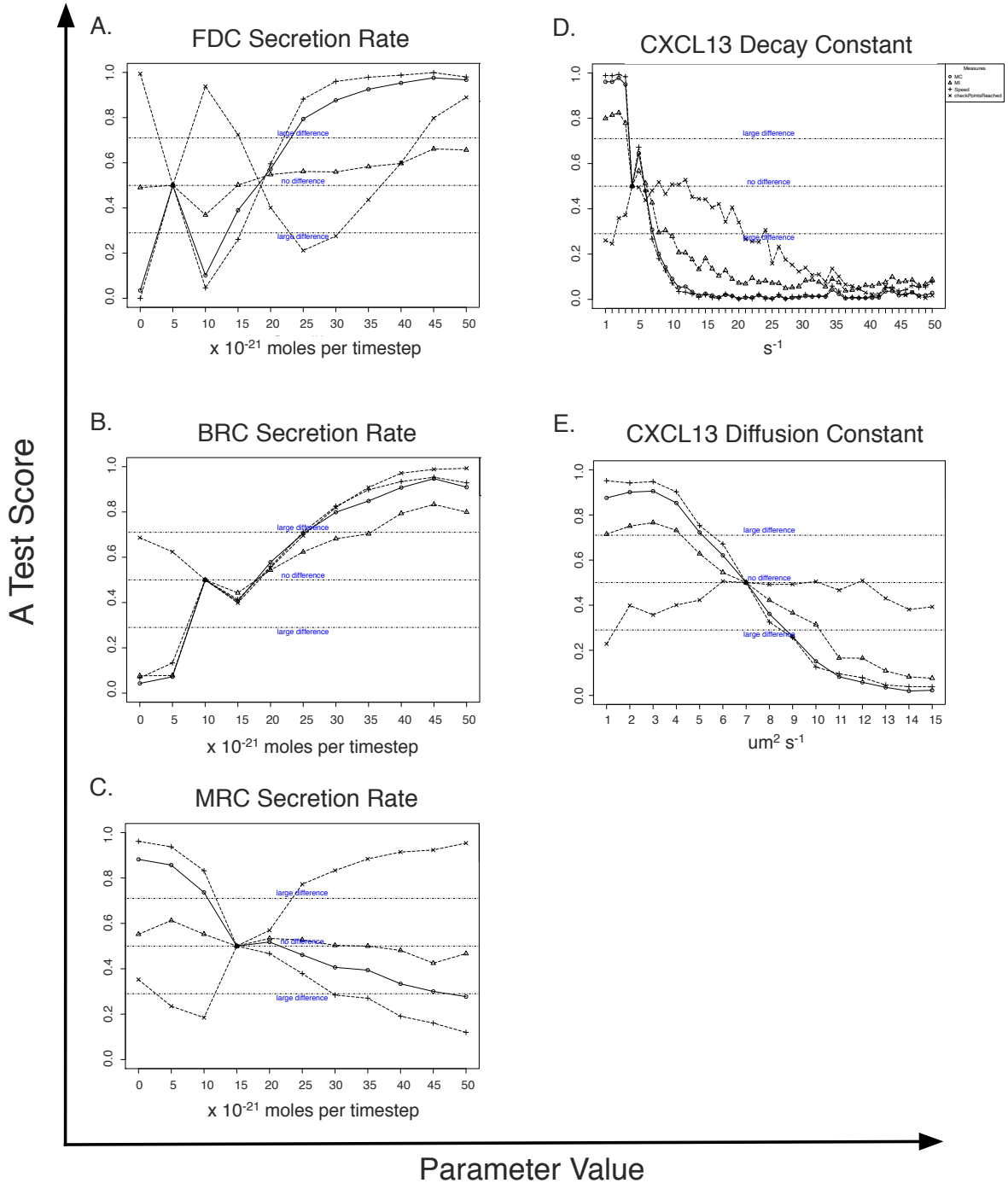

Figure 5.21. A-test scores when OAT adjusting parameters which relate to CXCL13 bioavailability. Parameters *cxcl13 emitted (A-C), decay constant (D) and diffusion coefficient (E)* were incrementally changed within their likelihood distributions using an OAT parameter robustness approach. A significant alteration in simulation outputs from baseline behaviours was determined using the Vargha-Delaney A-Test. checkPointsReached ×; meandering index (MI) Δ; motility coefficient (MC) **o**; speed **+**.

5.4.3 Global Sensitivity Analysis

To assess combinatorial effects of parameter perturbations we performed a global sensitivity analysis using latin hypercube sampling that partitions the distribution of each input parameter into intervals of equal probability, selecting one sample from each interval (**Figure 5.22-5.23**). LHC sampling generated 1000 parameter sets, 100 executions per parameter set were performed on a high-performance cluster and the influence of each parameter was quantified using a PRCC (detailed in **Chapter 2.2.9**). The parameters polarity, travel distance, and total number of CXCR5 receptors were key determinants of cell migration and scanning rates (**Figure 5.22-5.23**).

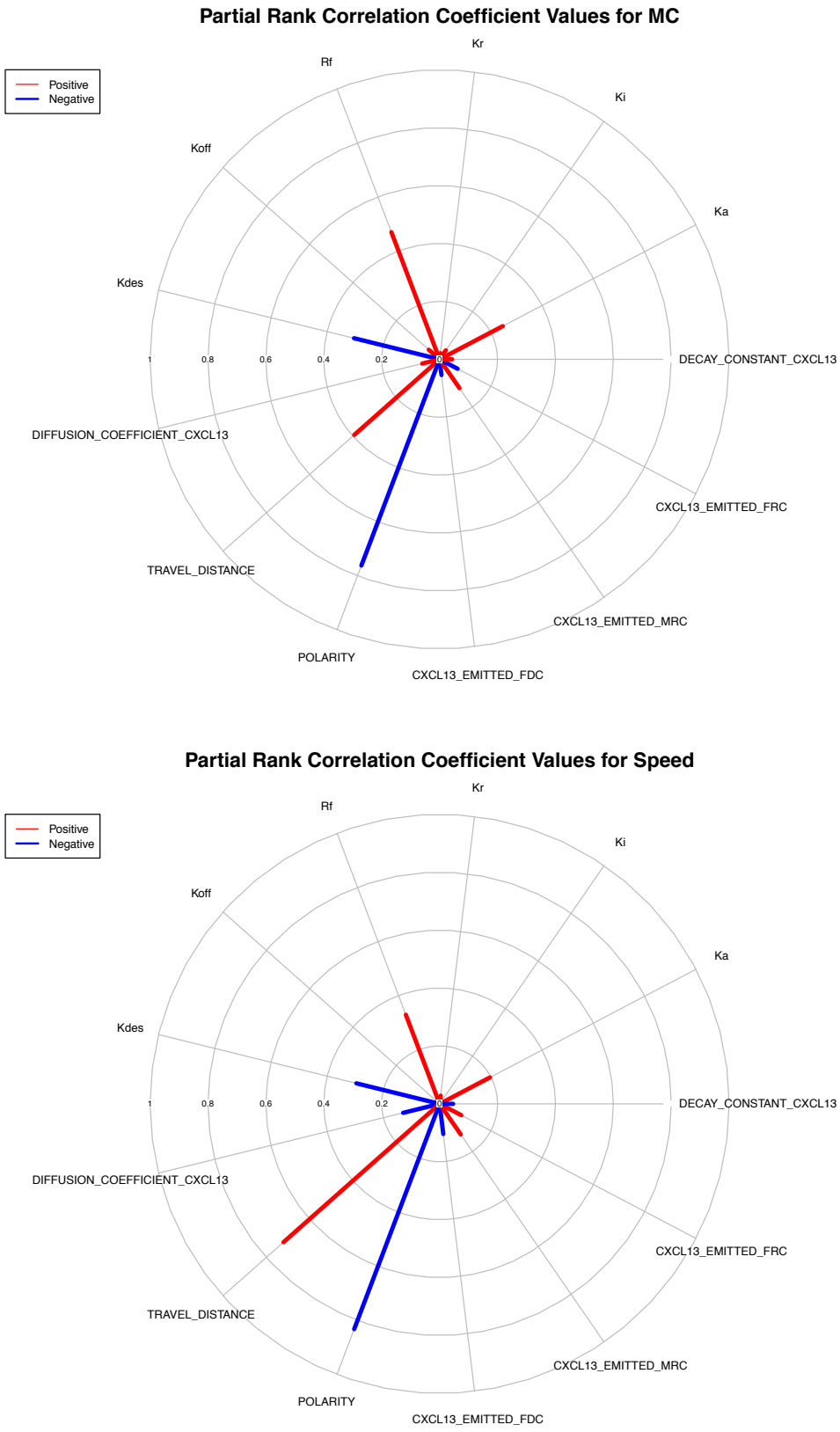

Figure 5.22. Partial rank correlation coefficients for Motility Coefficient (top) and Speed (bottom). Red lines show positive correlations, blue lines show negative correlations. The length of each line represents the magnitude of the correlation between a parameter and an output discounting the linear effects of other parameters

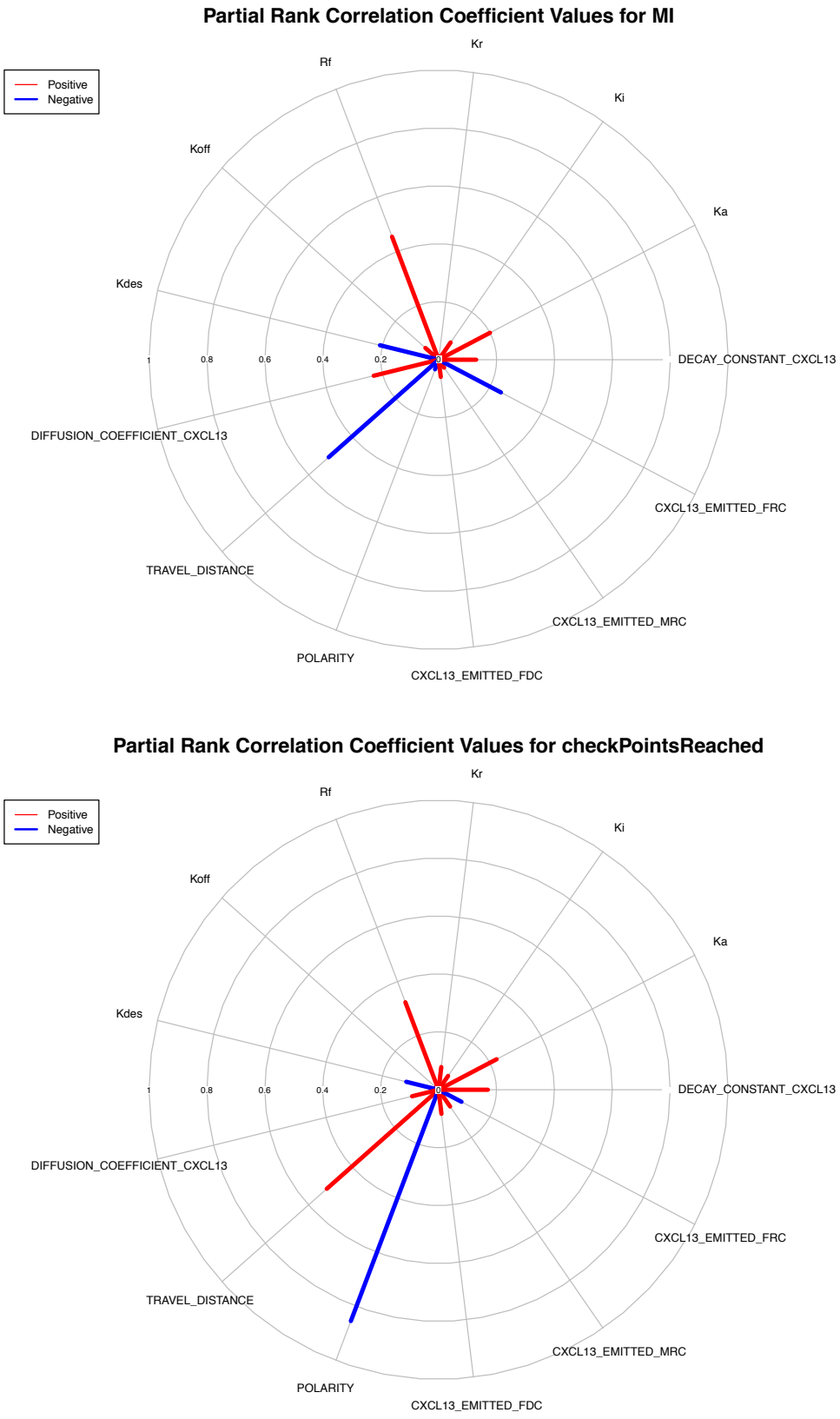

Figure 5.23. Partial rank correlation coefficients for Meandering Index (top) and checkPointsReached (bottom). Red lines show positive correlations, blue lines show negative correlations. The length of each line represents the magnitude of the correlation between a parameter and an output discounting the linear effects of other parameters

5.5 Emulator Development and Validation

CXCL13Sim is non-deterministic, capturing the stochastic nature of cell migration and the heterogeneity of $CXCL13^{+}$ stromal networks. Consequently, it required significant resources to quantify parametric and aleatory uncertainty. Specifically, it took ~123,000 simulation runs (each run takes 20 minutes on a macbook pro laptop with a 2.5 GHz Intel Core i7 processor and 8 cores [27]) to complete a local, global and aleatory analysis for one model iteration, with several iterations required. Using a high performance computer cluster (specifications are listed in **Chapter 2.2.4**) it was possible to perform each analysis on a timescale of days or weeks. The significant time resources required to evaluate the model warrant the use of an emulator to perform experiments where use of the simulator is intractable. Thus the emulator serves to complement the simulator as a tool to study CXCL13-mediated regulation of B cell migration (**Figure 5.24**).

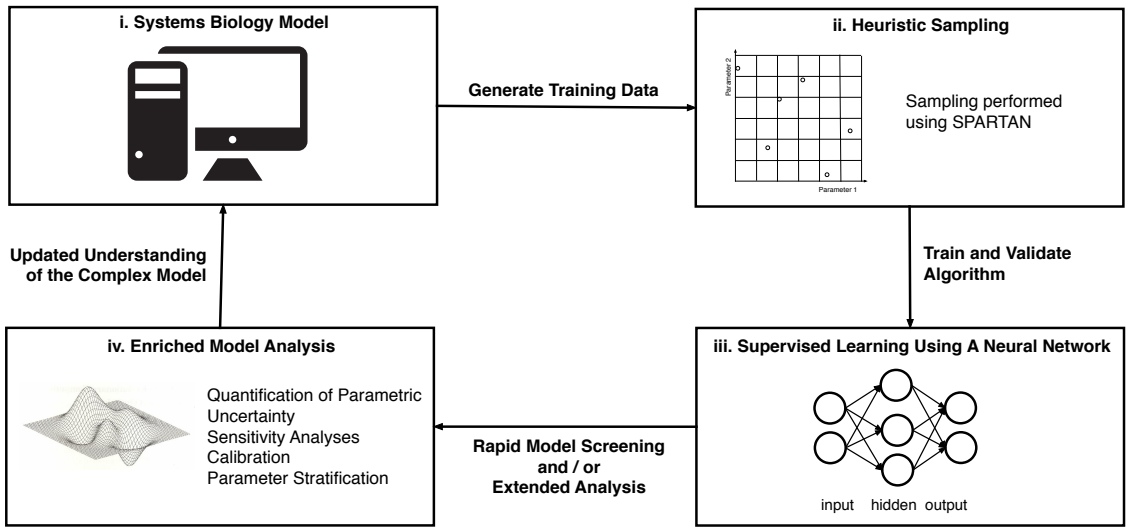

Figure 5.24 Using emulation to enrich understanding of complex systems models. Issues of time and resource limitations incurred in simulator analysis can be addressed by integrating machine-learning approaches within the process of simulator development, analysis, refinement, and translation.

[27] Based on baseline parameters. The time varies depending on the parameter inputs.

As an agent based model a number of high-level properties emerge from the simulator due to aggregated interactions between agents and their environment (Cosgrove et al., 2015; Macal and North, 2010). To learn the complex relationship between parameter inputs and emergent agent behaviours we employ a supervised machine learning approach. Supervised learning involves generating a dataset of inputs (x) and outputs (y) and then teaching an algorithm to approximate a mapping function between the two. With a sufficiently accurate mapping function it is then possible to predict Y for a set of unobserved values of x.

The initial dataset used to teach the algorithm how to map between inputs and outputs is a key determinant of performance, effectively acting as a teacher in the learning process. To generate this data we used the global sensitivity analysis, that is generating 1000 parameter sets and associated outputs by LHC sampling in SPARTAN. To map the complex relationship between parameter inputs and the emergent properties of the model we train an artificial neural network (ANN). ANNs are a technique inspired by the neuronal circuits in the brain, with computations structured in terms of an interconnected group of artificial neurons organised in layers (Bishop, 1996).

In this scheme parameter inputs are passed into the network and iteratively processed by a number of hidden layers. Within each hidden layer the sum of products of inputs and their corresponding weights are passed through a sigmoidal activation function[28] which is fed as inputs into the next layer. This process is repeated until the output layer is reached and we have a prediction for the output values. During the learning phase, the weighting of connections between neurons is adjusted in such a way that the network can convert a set of inputs (simulation parameters) into a set of desired outputs (simulation responses) (Bishop, 1996).

A key technical consideration when developing neural networks is how to evaluate predictive power. Testing predictive performance on the training data is not useful as it can lead to over fitting, whereby the network is poor at predicting previously unobserved data. To solve this problem, a proportion of the dataset is omitted from the training dataset and used to validate algorithm performance.

[28] The activation function confers non-linear properties to the network and allows it to learn complex relationships between inputs and outputs.

Paritioning the data incurs a cost however, as we reduce the amount of samples used for training the model. In addition, the data used to train the model, even if not used in the evaluation process, can have a significant impact on predictive performance. To address these issues we perform a procedure known as k-folds cross validation. In this scheme the data is partitioned into k-folds and the algorithm learns the mapping between inputs and outputs using k-1 folds as training data with validation performed on the remaining part of the data. This process is repeated until each fold is used as the test set with overall performance taking as the average for each fold. This approach can be computationally expensive, but does address the aforementioned issues.

To develop our ANN we generate a number of neural network structures with different number of hidden layers and nodes within each layer (so called hyperparameters) but fixed input and output layers (one node for each distinct input and output respectively. K-folds cross validation is used to determine which structure best minimises the root mean squared error between emulator and simulator responses. The performance of the best neural network structure on the validation dataset is shown in **figure 5.25**.

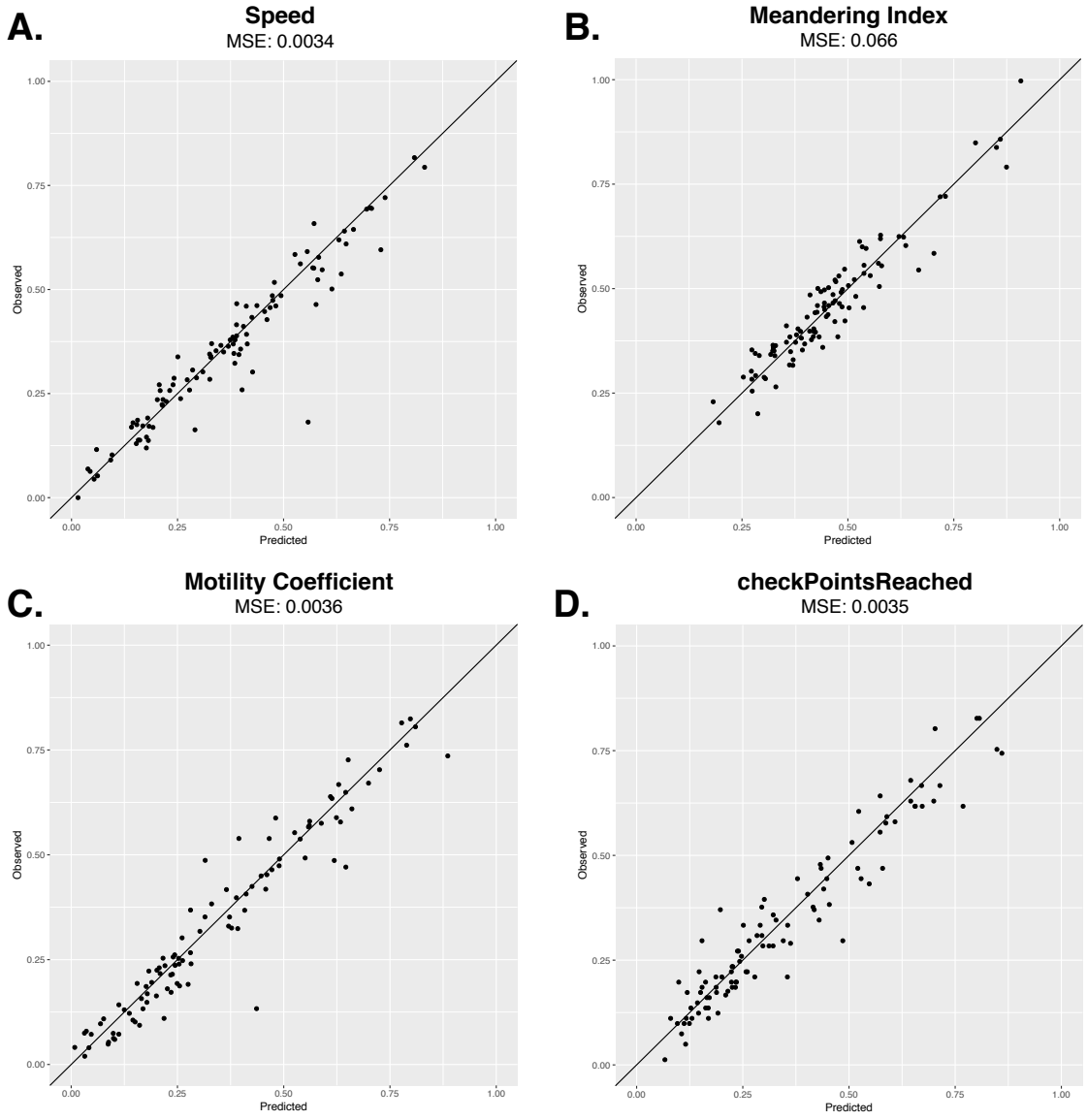

Figure 5.25 Emulator performance on the validation dataset. Each datapoint represents the predicted (emulator) vs observed (simulator) output value for a distinct set of parameter inputs. (A) Predictive performance for the output Speed. (B) Predictive performance for the output Meandering Index. (C) Predictive performance for the output Motility Coefficient. (D) Predictive performance for the output checkpointsReached.

5.6 eFAST Analysis using CXCL13emulator

In **section 5.4.3** we perform a global sensitivity analysis using latin-hypercube sampling and quantifying the non-linear relationship between parameters and responses, correcting for the effects of other parameters on the response. This facilitates an understanding of what parameters should be targeted to achieve a target response. However, it only provides a rank correlation between inputs and outputs and does not indicate the magnitude by which outputs change following a perturbation. Consequently a PRCC metric does not indicate which parameter uncertainties have the greatest impact on output variability. In addition this variance based approach allows investigation of both non-linear and non-monotonic relationships between model inputs and outputs (Marino et al., 2008)

To address this we employ the eFAST variance based approach (**Chapter 2.2.9**). Due to time constraints it was intractable to perform an eFAST analysis using the CXCL13simulator so we use the CXCL13emulator. In this analysis we find that polarity is a key determinant of each output metric with a significant difference between Si and STi values (detailed further in **Chapter 2.2.9**) in comparison with the dummy variable.

Interestingly, eFAST analysis highlights a number of parameters that differ substantially between their Si and STi values. From this analysis we see a number of parameters that do not contribute highly to the regulation of cell behaviours in isolation but do contribute to cell behaviours through non-linear relationships with other parameters, the CXCL13 diffusion constant being one example. This is consistent with analyses of previous model iterations where model configurations with less parameters (data not shown) were unable to produce emergent behaviours consistent with our experimental data.

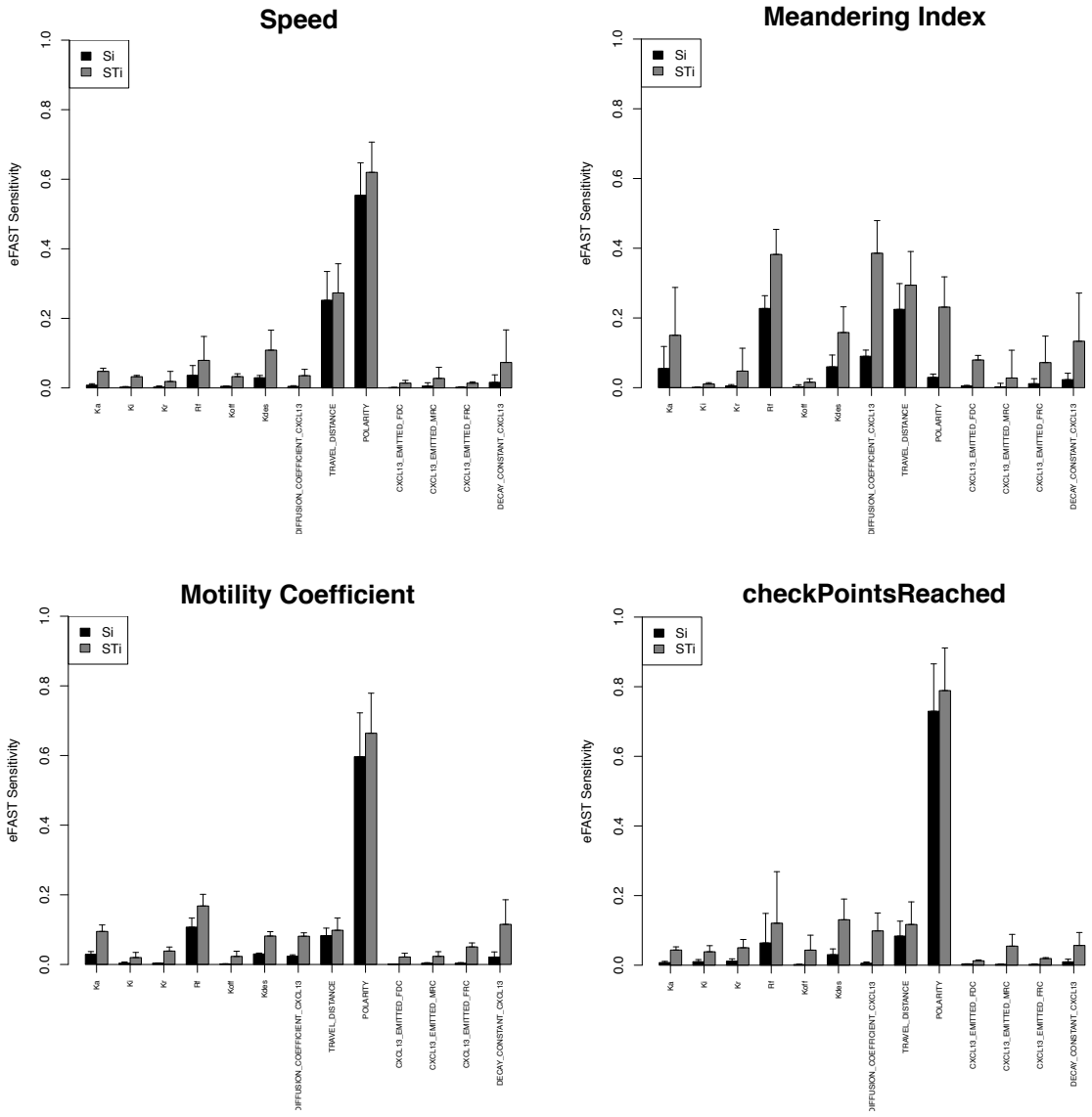

Figure 5.26 eFAST analysis using the CXCL13emulator. Si (black) represents the fraction of output variance that can be explained by the value assigned to that parameter. STi (grey) represents the variance caused by higher order non-linear effects between that parameter and others explored. Bars represent the mean value for either Si or STi, with error bars representing the standard error over three resample curves.

5.7 Discussion

This chapter detailed the development of a multiscale software tool, designed to assess the veracity of our research hypothesis that lymphoid tissues are essential for the formation of CXCL13 gradients, and perturbation of the spatial distribution of CXCL13 can affect the onset of humoral immune responses. Platform Model development was facilitated by modularisation of the domain into distinct subunits. Through modularisation it was possible to identify existing modelling techniques that could capture the key features of each model subunit at the desired granularity. This yielded an executable software platform where *in silico* B cells can dynamically detect chemotactic gradients within the context of the follicular microenvironment.

Interestingly, there are limited theoretical studies that combine all 3 components. Guo *et al*. (2008) developed a hybrid model of chemotaxis in a reductionist *in vitro* assay and did not take the lymph node environment or agent interactions into account (Guo et al., 2008b). Donovan *et al* (2016) looked at T-cell migration on an FRC network but had limited data to inform the properties of the FRC network and had no explicit reference to chemokines. The lack of studies looking at migration *in vivo* can be attributed in part to the lack of quantitative information available in the literature and the challenges associated with developing multiscale models. To address this we have incorporated the data from **Chapter 4** into a hybrid scheme to enrich model granularity without becoming prohibitively slow or difficult to interpret. This ultimately provides an executable software platform capable of addressing our research question and driving further experimentation (**Chapter 6**).

Through incorporation of agile techniques within the CoSMoS framework it was possible to create a robust but malleable code base that could cope with constantly evolving requirements. Automated tests at the verification level expedited the development process, ensuring that changes at later phases of development did not affect the functionality of the original code base, while validation tests facilitate translation of simulation-derived results back to the biological system. Importantly, the testing process identified a number of design and software errors that may not have been encountered until late into the development process where errors become more resource costly to

diagnose and address.

Following development, model parameters were calibrated against experimental data to establish a baseline set of behaviours to which all subsequent perturbations are compared. This required large amounts of disparate datasets and derivations from the wider literature to constrain parameter values and perform fitting and validation. The CXCR5 deficient multiphoton experiments were key to ensuring that the model was not overfitted and where possible the amount of parameters was reduced. Subsequently, we quantified the effect of stochasticity on baseline simulation behaviours, with a view to determining how many replicate runs are required to mitigate this aleatory uncertainty. Parametric uncertainty was quantified using both local and global approaches, identifying the migration displacement constant, the numbers of CXCR5 receptors on the cell surface and cell polarity as key regulators of simulation outputs.

Key attributes of the simulator include the high speed of data acquisition relative to laboratory experiments and the high granularity at which we can observe mechanistic processes. Each simulation run takes on average 15-20 minutes to run, and represents 24 hours. With high performance computing, hundreds of experiments can be run in parallel dramatically outcompeting laboratory experiments in terms of the time resources required to obtain data. Incorporating equation-based systems within agent-based models facilitates single-cell analyses of receptor expression and can facilitate analyses of how the local environment can alter molecular phenotypes with high temporal resolution over long timescales. Thus the simulation platform may provide a useful adjunct to imaging approaches to understand the dynamics and microanatomy of CXCL13, affording single cell tracking precision within a complex tissue over a timescale of several hours.

The analyses performed in this chapter are key to understanding simulation behaviours and to fully evaluate simulation-derived results. To facilitate an enriched suite of analyses within the project time constraints we employed supervised learning algorithms to develop an emulator tool capable of rapidly and accurately reproducing emergent behaviours of the multiscale simulator. In reducing the time taken to evaluate parameter sets, the emulator was able to rapidly perform an eFAST analysis (on the order of 10 seconds) providing a better understanding of parametric uncertainty. In the following chapter, we use CXCL13Sim and the CXCL13emulator in combination with experimental approaches, to map, perturb and optimise CXCL13 gradients within the primary follicle.

Chapter 6 Mapping CXCL13 Gradients *In Situ*

6.1 Introduction

6.1.1 Current Model of Morphogen Bioavailability *In Situ*

CXCL13 has emerged as a key regulator of follicular architecture and B-cell responses. Despite this progress a more nuanced understanding of how the molecule functions is lacking. The precise spatial distribution of the molecule is poorly understood and it is unclear if the molecule acts principally in a soluble or an immobilised form. Consequently, we have a limited understanding of how CXCL13 gradients form within complex lymphoid tissues and how B cells respond to them.

To better understand CXCL13 bioavailability, it is beneficial to discuss the behaviours of a broader class of molecules known as morphogens. The molecular mechanisms governing gradient formation are highly nuanced and context specific with a number of prevalent views existing. One such mechanism was first described by Alan Turing, who demonstrated that two diffuse morphogens that dynamically influence one another can give rise to symmetry breaking and thus the emergence of complex patterns (Turing, 1952). This work was later extended by Meinhardt and Gierer who produced a number of naturally occurring patterns through numerical simulation (Meinhardt and Gierer, 2000) and by experimental studies of development (Kondo and Miura, 2010).

In 1969, Lewis Wolpert suggested a scheme of gradient formation where smoothly declining gradients were formed by the diffusion of morphogens from a source to a sink. The role of the morphogen is then dependent on the concentration which the cell perceives – once the localised concentration exceeds a certain threshold genes can be switched on or off (Wolpert, 1969). Recently, a secretion-consumption scheme has been used to describe the spatial distribution of the cytokine IL-2 (Oyler-Yaniv et al., 2017) . Consistent with the source-sink scheme of gradient formation CCRL1 expressing LECs lining the ceiling of the subscapular sinus have been implicated in the formation of functional CCL21 gradients in the lymph node (Ulvmar et al., 2014).

However, CCL21 is also regulated by the microenvironment. CCL21 is secreted by a complex network of T-Zone fibroblastic reticular cells (FRCs) and undergoes interactions with ECM components (Hirose et al., 2002; Patel et al., 2001; de Paz et al., 2007). *In vivo*, truncation of the highly charged C-terminus prevents its immobilization to high

endothelial venules and consequently affects lymphocyte homing to the lymph node (Stein et al., 2000). The ability to bind the ECM is not a CCL21 specific characteristic; *In vivo* studies of CXCL12 using a mouse strain carrying a mutated form of CXCL12 where interactions with the extracellular matrix are impaired ($CXCL12^{gagtm}$ mice) show increased amounts of hematopoietic precursors, an impaired ability to support revascularisation and impaired humoral immune responses. In these mice the structure of the GC was impaired, as well as having fewer somatic mutations in Ig genes (Barinov et al., 2017).

The role of the microenvironment in shaping gradients has also been described in the context of development. Fibroblastic Growth Factors (FGF) are critical for mesoderm induction as well as limb and lung development, modulating the proliferation and differentiation of a variety of cells of mesenchymal and neuroectodermal origin. Receptor dimerization is a prerequisite for FGF signalling and requires extracellular matrix components (Pellegrini et al., 2000). Previous studies have identified the secreted serine protease HtrA1 as a key mediator of ECM binding, promoting the mobilization of FGF/proteoglycan complexes and long-range FGF signalling during mesoderm induction and posteriorization in Xenopus embryos (Hou et al., 2007). These experimental studies are supported by mathematical analyses predicting that gradient formation is increased when morphogens are secreted in matrix-binding form as compared to a non matrix-interacting form (Fleury et al., 2006)

Given that many soluble factors display complex binding behaviours, interacting with components of the extracellular matrix (Hasan et al., 1999; Patel et al., 2001) and other molecules (Dyer et al., 2016; Hundelshausen et al., 2017) it is unclear if a source-sink or reaction diffusion model can provide a general-purpose description of morphogen bioavailability. In terms of a source-sink scheme for CXCL13 we note that the molecule is negatively regulated by the internalisation of typical and atypical chemokine receptors and by protease-mediated enzymatic degradation (Nibbs and Graham, 2013; Ulvmar et al., 2014; Zabel et al., 2006). CXCL13 binds the atypical chemokine receptors ACKR1 and ACKR4 (ACKR4 binds human but not murine CXCL13) but to date, their role in regulating CXCL13 is unknown (Nibbs and Graham, 2013). Due to the density of the B follicle CXCR5-mediated scavenging could also contribute significantly to gradient formation but the extent to which it contributes is also unclear.

Measurements of CXCL13 diffusion show that the ECM constrains mobility (**Chapter 4.3**) suggesting that CXCL13 bioavailability is likely to be regulated, at least in part, by the localised microenvironment. This result was unsurprising given that chemokines are highly basic proteins (the isoelectric point of CXCL13 is 10.5), having a net positive charge at physiological pHs. This promotes interactions with components of the extracellular matrix such as glycosaminoglycans (GAG), highly negatively charged molecules (Handel et al., 2005). All chemokines bind GAGs, but are heterogeneous in their binding affinities (Patel et al., 2001; de Paz et al., 2007). An additional layer of complexity is added by the heterogeneous distribution of ECM proteins within the follicle. An immunohistochemistry study of the murine lymph node shows signal for fibronectin and collagen type IV. Positive signal was also observed for the proteoglycan perlecan with positive signal highly concentrated around HEVs and just under the subcapsular sinus. ER-TR7 was present in the follicle but staining was substantially reduced in comparison to the T-zone (Ma et al., 2007). Another factor that may contribute to gradient formation is the architecture of CXCL13 stromal cell networks. Data from **Chapter 4.2** suggests that the small-world network topology is configured to promote B-cell trafficking but it has yet to be determined how this architecture influences the spatial distribution of CXCL13.

6.1.2 Multiobjective Optimisation of Multiscale Simulators

CXCL13 bioavailability is subject to the following simultaneously occurring processes: secretion, diffusion, decay and scavenging. Given the highly dynamic nature of this regulatory network, it is unclear what spatial configuration of CXCL13 is best suited to promote scanning within the primary follicle. This question requires systematic tuning of each process, thus it is intractable to address it experimentally.

However, the mathematical constructs that model these processes can be easily tuned *in silico* through parameter perturbation, with the added benefit that they are much faster to evaluate than in the laboratory. As such, CXCL13Sim is well suited to addressing this research question. Despite these advantages, this analysis is technically challenging due to a: (i) highly complex search space, and (ii) multiple, conflicting objectives.

Due to the high dimensional parameter space and the need for replicate runs to mitigate aleatory uncertainty it is intractable to evaluate every possible parameter set. Consequently it is beneficial to use a search heuristic such as an evolutionary algorithm (EA). EAs mimic evolution using biological mechanisms of reproduction, mutation, recombination, and selection to iteratively generate, execute and evaluate solutions to a complex problem. This allows the algorithm to efficiently scan complex multidimensional parameter spaces to find solutions to a given problem.

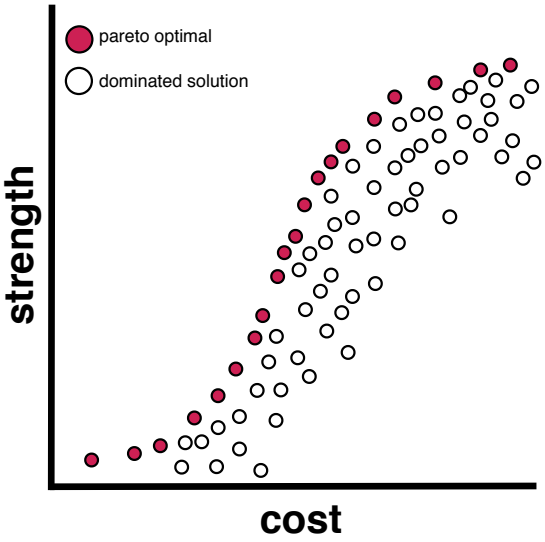

Figure 6.1. Example Pareto front for the bridge problem. This plot shows an example Pareto front, populated by solutions where improvement in one objective cannot be obtained without compromising performance for another objective. Pareto optimal solutions are highlighted in red while dominated solutions are shown in white. A solution is dominated if there exists another with at least equal performance on all objectives and superior performance on at least one.

A certain class of mathematical problems have multiple objectives to optimise. For example, the key objectives when building a bridge might be to maximise the strength while minimising cost. However, improvements gained in strength may compromise cost efficiency and *vice versa*. For such problems a specialised class of EAs known as multiobjective optimisation evolutionary algorithms (MOEAs) are used (Deb, 2001; Deb et al., 2002; Zitzler and Thiele, 1998). This class of algorithms uses a heuristic search approach to obtain a population of solutions (a Pareto front) where improvement in one objective cannot be obtained without compromising performance for another objective (**Figure 6.1**). For the bridge problem the Pareto front represents a set of solutions that are best trade-off between cost and strength. Some solutions will have a high cost but also high strength; some solutions will have a low cost but a lower strength and so forth.

6.1.3 Summary and Aims

The aim of this chapter is to use the *in silico* tool developed in **Chapter 5,** in conjunction with experimental approaches to examine chemokine field formation and associated B-cell responses in lymph node follicles. A key advantage of this system is the ability to link molecular, cellular and tissue level components to understand how gradients form within complex tissues. The key questions we wish to address include: (i) what is the spatial distribution of CXCL13 in the primary follicle (ii) how sensitive are scanning rates to perturbations to CXCR5 and CXCL13 and (iii) is there an optimal configuration of CXCL13 in the follicle with respect to network scanning. The key aims of the chapter are summarised as follows:

1. Combine theoretical (modelling and simulation) and experimental (immunohistochemistry) approaches to map the spatial distribution of CXCL13 within the primary follicle (**section 6.2-6.3**)
2. Determine the Pareto optimal[29] configuration of CXCL13 *in silico* using a MOEA approach in conjunction with the CXCL13emulator (**section 6.4**)
3. Assess the sensitivity of B cell scanning to perturbations in CXCR5 *in silico* (**section 6.5**) and perform single cell tracking to determine spatial and temporal patterns of expression within the context of a complex tissue (**section 6.7**).
4. Use an MOEA approach in conjunction with the CXCL13emulator to determine the Pareto optimal[30] configuration of CXCR5 signalling with respect to scanning *in silico* (**section 6.8**).

[29] the optimisation problem is defined in section **2.2.11**
[30] the optimisation problem is defined in section **2.2.11**

6.2. Assessing Cell-Specific Contributions to CXCL13 Secretion

Confocal images of lymph node follicles taken from *Cxcl13*-EYFP reporter mice implicate 3 key stromal cell subtypes in the secretion of CXCL13 within the primary follicle; FDCs, MRCs, BRCs (detailed further in **Chapter 4**). However, the relative contribution of each cell type to the regulation of naïve B cells within the primary follicle is poorly understood. To address this we systematically perturbed the secretion rate of each cell type. Maximal scanning rates were observed for intermediate rates of secretion, with perturbations to FDC secretion yielding the biggest changes in agent behaviours (**Figure 6.2a-c**). For both MRCs and BRCs, lower rates of secretion gave improved scanning rates. We further assessed cell-specific contributions to CXCL13 bioavailability by stopping secretion in each subset (**Figure 6.2d**). In this analysis, loss of CXCL13 from MRCs or BRCs did not significantly alter scanning rates while loss of CXCL13 from FDCs did.

In vivo quantification of fluorescent intensity of each cell type suggests that secretion may be heterogeneous, with MRCs having the highest levels of reporter expression, followed by FDCs and then BRCs **(Figure 6.3a).** To assess what effect this spatial configuration might have on scanning rates we used TdTomato expression from CXCL13-EYFP mice as inputs into our simulator and quantified scanning rates. Interestingly, this configuration led to a small but significant increase in the scanning of the entire follicle but did lead to a more dramatic increase in scanning at the SCS, a major site of antigen entry into the primary follicle **(Figure 6.3b-c).** This result suggests that fine-tuning of CXCL13 secretion can generate gradients capable of directing B cells to different follicular niches. Taken in concert, these analyses identify a complex non-linear relationship between CXCL13 concentration and B-cell scanning.

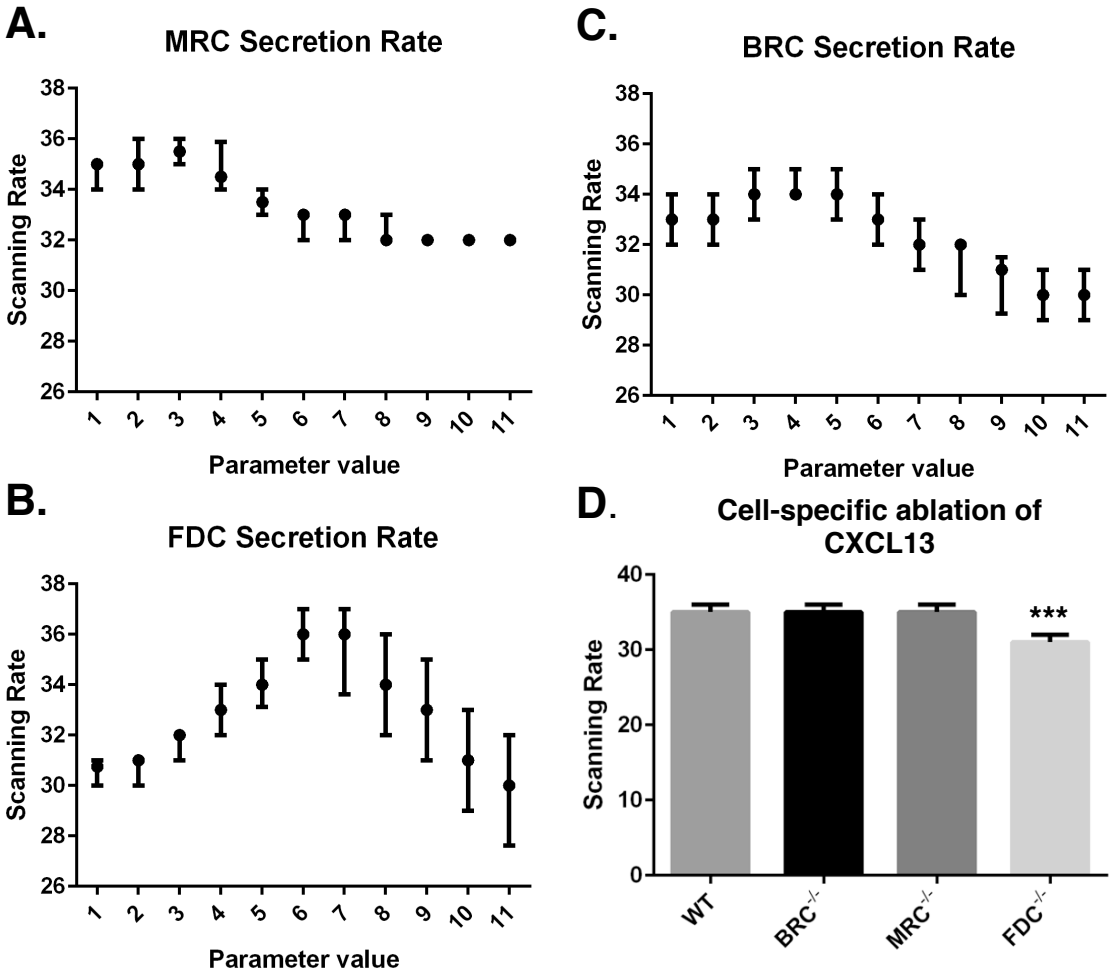

Figure 6.2 Assessing cell-specific contributions to CXCL13 bioavailability. (A-C) OAT perturbations of cell-specific secretion rates. Dots represent the median values with error bars displaying the I.Q.R (D) In this experiment baseline behaviours (WT), were compared to those for follicles where BRCs (BRCs$^{-/-}$), MRCs (MRCs$^{-/-}$) or FDCs (FDCs$^{-/-}$) don't secrete CXCL13. In this plot median values are shown with error bars representing the I.Q.R. Significance assessed using the Mann-Whitney test.

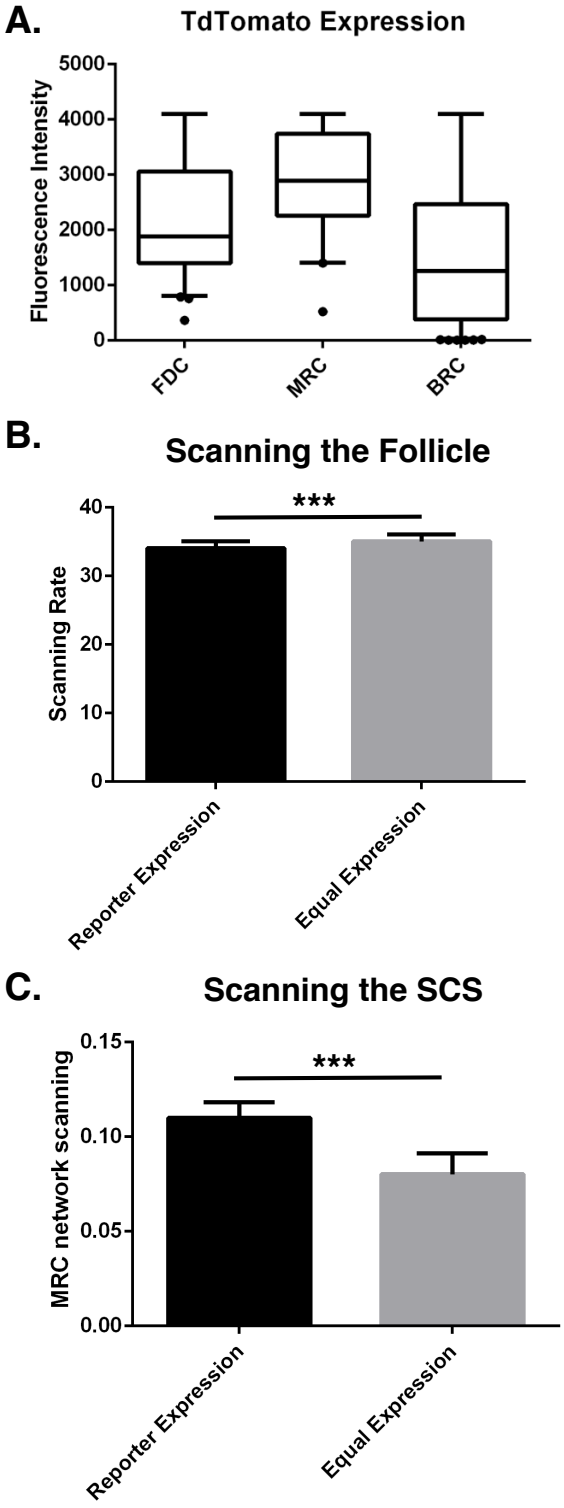

Figure 6.3. Mapping fluorescent reporter expression *in silico.* (A) TdTomato intensity of the different stromal subsets based on histological analysis of TdTomato expression. Horizontal lines represent median values with boxplot height representing the interval between the 5th and 95th percentile. Figure A provided by N.P. and M.N. (B) scanning rates for the entire follicle where relative fluorescent intensity values from (A) were used as inputs into the simulator (C) scanning rates for the proportion of the SCS network scanned where relative fluorescent intensity values from (A) were used as inputs into the simulator. For (B) and (C) median values are shown with error bars representing the I.Q.R. Significance assessed using the Mann-Whitney test.

6.3. Assessing the Solubility of CXCL13 Within the *In Silico* Follicle

A question regarding CXCL13 bioavailability is whether it acts in principally an immobilized or a soluble form. To assess this we investigate two models: (i) a scenario where CXCL13 binds to extracellular matrix components creating short sharp gradients proximal to the cells that secrete it; and (ii) a model where CXCL13 is largely soluble and diffuses more freely throughout the tissue (**Figure 6.4**).

Firstly, we assessed the sensitivity of the model to these parameters using an OAT analysis where either the decay constant or diffusion coefficient are changed while holding all other parameter values held at their baseline calibrated value. This analysis suggests that perturbing each parameter in isolation leads to a modest change in scanning rates with high values for the decay constant promoting the highest scanning rates (**Figure 6.5a-b**).

Subsequently, we assessed the combinatorial effects of these parameters. Specifically, we generate short sharp gradients proximal to the stromal network (**model 1**) through having a low diffusion coefficient and high decay rate (1.0 $\mu m^2\ s^{-1}$ and 0.05 s^{-1}) and diffuse gradients (**model 2**) with a high diffusion coefficient and low decay rate (15 $\mu m^2 s^{-1}$ and 0.0035s^{-1}). Both parameter sets give a total concentration of ~18nM CXCL13 within the follicle. Consistent with results from the OAT analysis **model 1** yields higher scanning rates than **model 2** (**Figure 6.5c**). The scanning rates obtained using model 1 were greater than those achieved through OAT perturbation suggesting a synergistic effect between diffusion and decay rates. This synergy between parameters is consistent with results from the eFAST global sensitivity analysis (**Chapter 5.6**).

To map the chemotactic landscape generated by each model we measured the spatial autocorrelation of chemokine concentrations *in silico* (**Figure 6.6**). This analysis shows that with short sharp gradients chemokine concentration is correlated over short distances (~60 μm) and then becomes significantly uncorrelated, while for more diffuse distributions chemokine concentrations were correlated over distances of 100 μm.

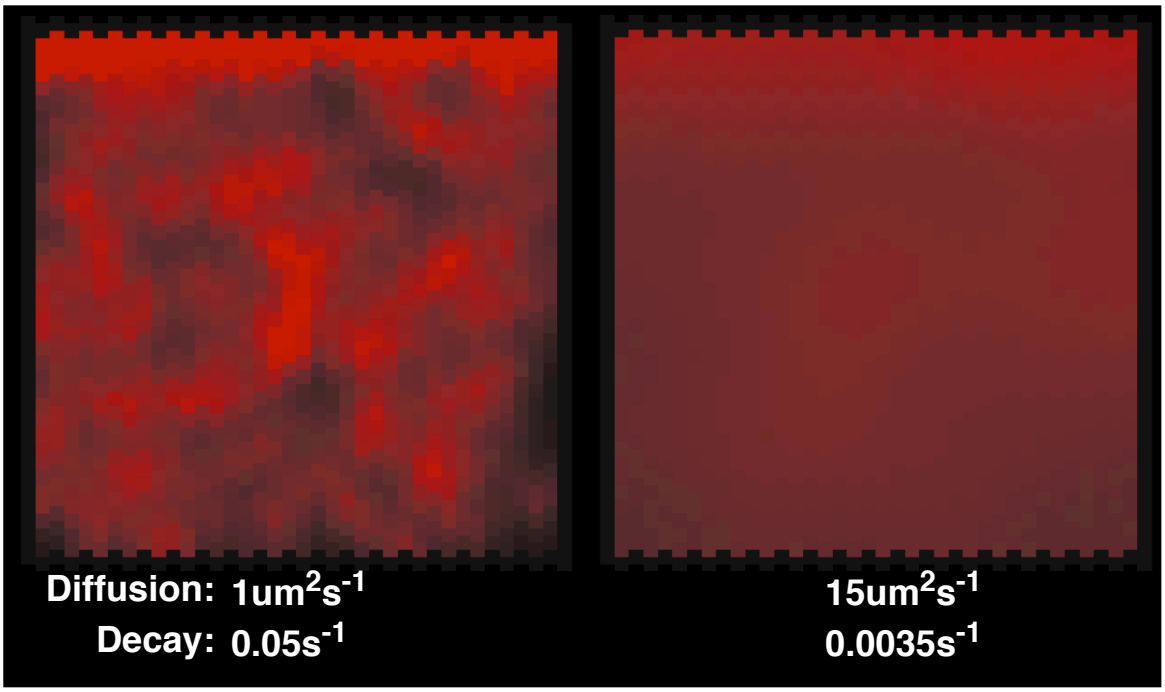

Figure 6.4 Comparison of two competing models for CXCL13 bioavailability: Model 1 assumes a scenario where CXCL13 binds strongly to extracellular matrix components creating short sharp gradients proximal to the cells that secrete it and (ii) a model where CXCL13 is largely soluble and diffuses more freely throughout the tissue

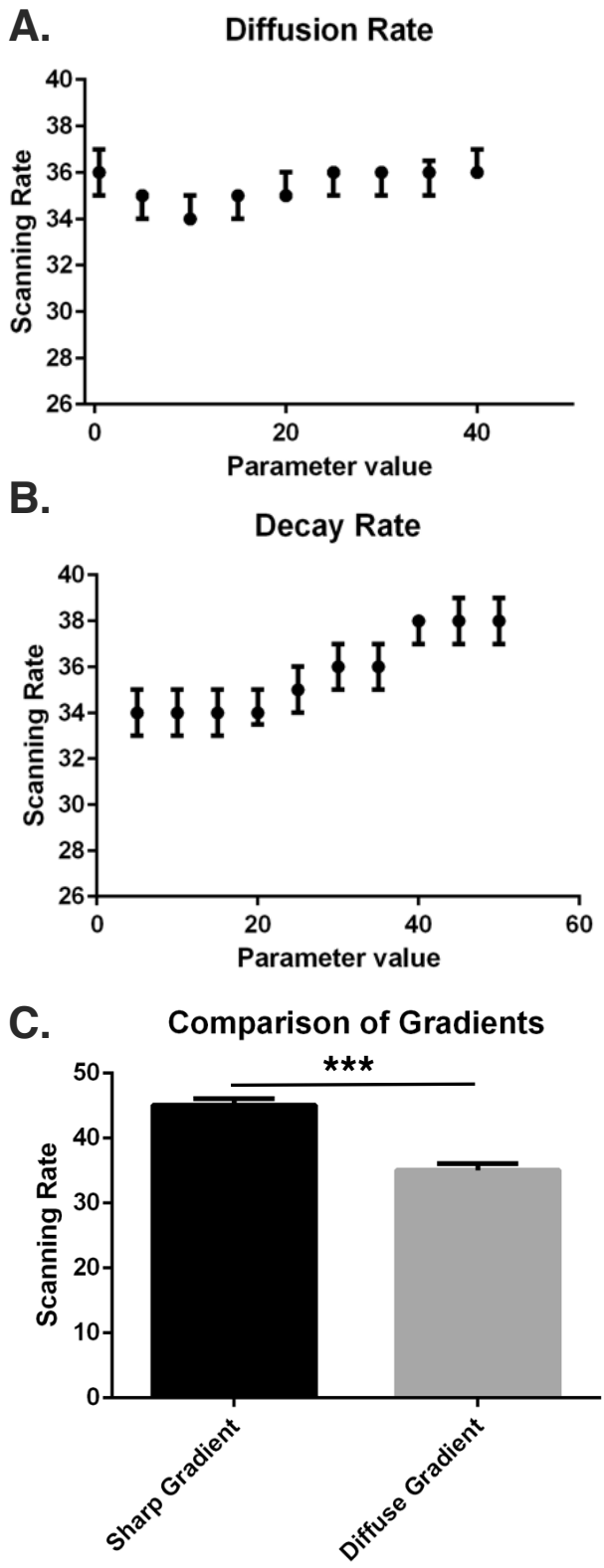

Figure 6.5 Robustness of scanning rates to different spatial distributions of CXCL13. (A) Perturbation of CXCL13 diffusion rates holding all other parameters fixed at baseline calibrated values. (B) Perturbation of CXCL13 decay rates holding all other parameters fixed at baseline calibrated values. (C) Comparison of scanning rates obtained using model 1 and model 2. In each plot median values are shown with error bars representing the I.Q.R.

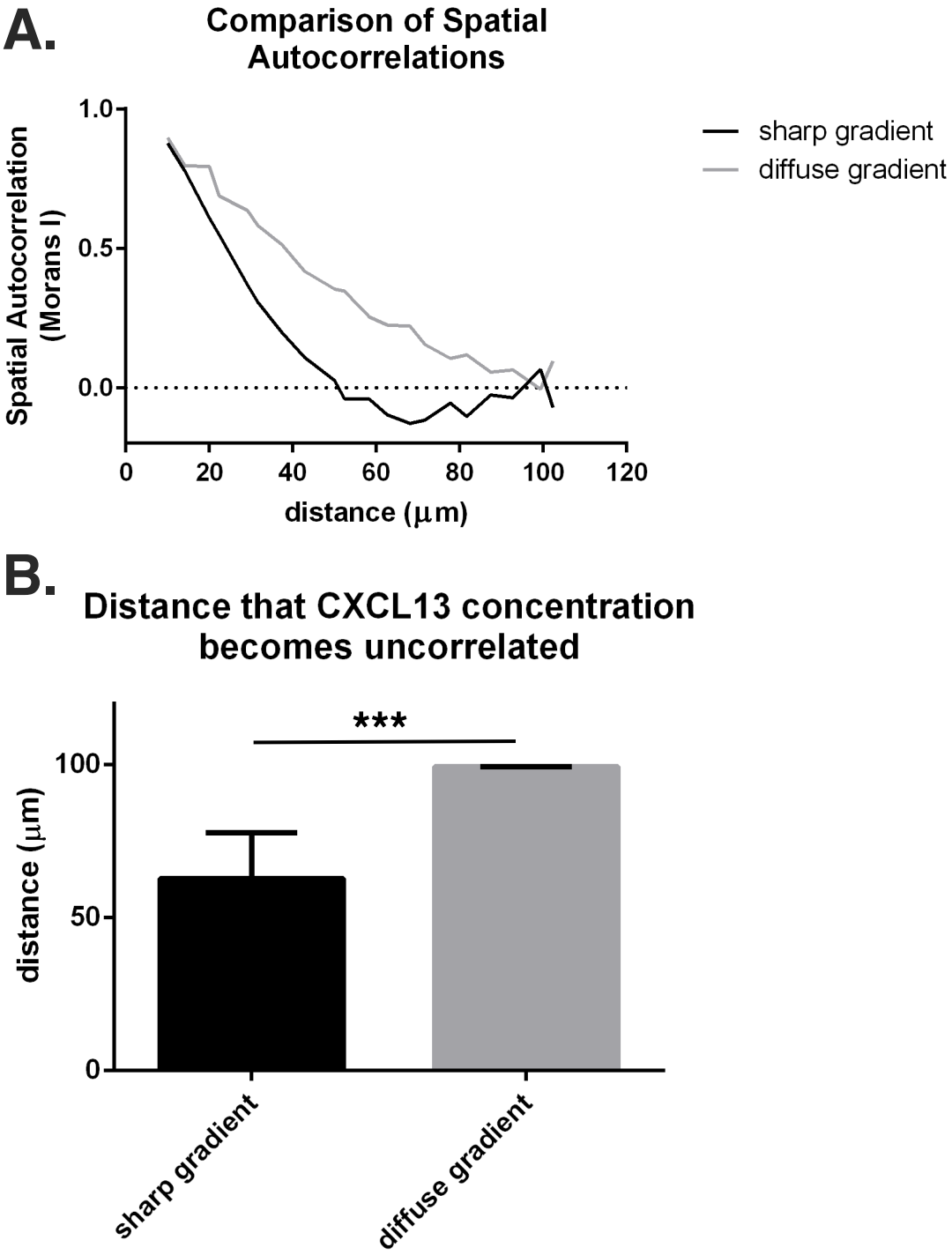

Figure 6.6 Comparison of the spatial autocorrelation of *in silico* gradients. (A) an example correlogram showing how chemokine concentrations are spatially correlated in silico. (B) The distances at which there is no significant spatial autocorrelation in chemokine concentrations. Significance was determined through permutation testing using the ncf package in R. Using model (i) we generate spatial distributions where chemokine concentrations are correlated over distances of 60 μm, while for model (ii) a more diffuse situation we see that chemokine concentrations are correlated over longer distances (~ 100 μm). In figure (B) median values are shown with error bars representing the I.Q.R. Significance assessed using a Mann-Whitney test.

6.4. Determining the Optimal Configuration of CXCL13 Within the Primary Follicle

Simulation analyses performed in the previous sections favour a model whereby CXCL13 is less diffuse. However, this data cannot be used to determine what spatial configuration of CXCL13 promotes the highest rates of antigen scanning. To address this we perform multiobjective optimisation of the CXCL13emulator using the non-dominated sorting genetic algorithm II (NSGA-II) ref Deb using the mco package in R. The four objectives to be assessed by the algorithm were to: minimize the root mean squared error between emulator and simulator responses for cell speed, meandering index and motility coefficient; and maximize scanning rates (**Figure 6.7**).

This analysis shows that our objectives are conflicting, with increased scanning rates leading to poorer agreement between emergent cell behaviours *in silico* and laboratory measures (**Figure 6.7**). Analysis of the parameter distributions corresponding to the population of Pareto optimal solutions shows that diffusion constants are highly constrained to low values, with FDC secretion rates and decay rates constrained to high values (**Figure 6.7**). This is consistent with previous analyses (**section 6.3**) that favour a model of CXCL13 where diffusion is constrained by the local tissue anatomy.

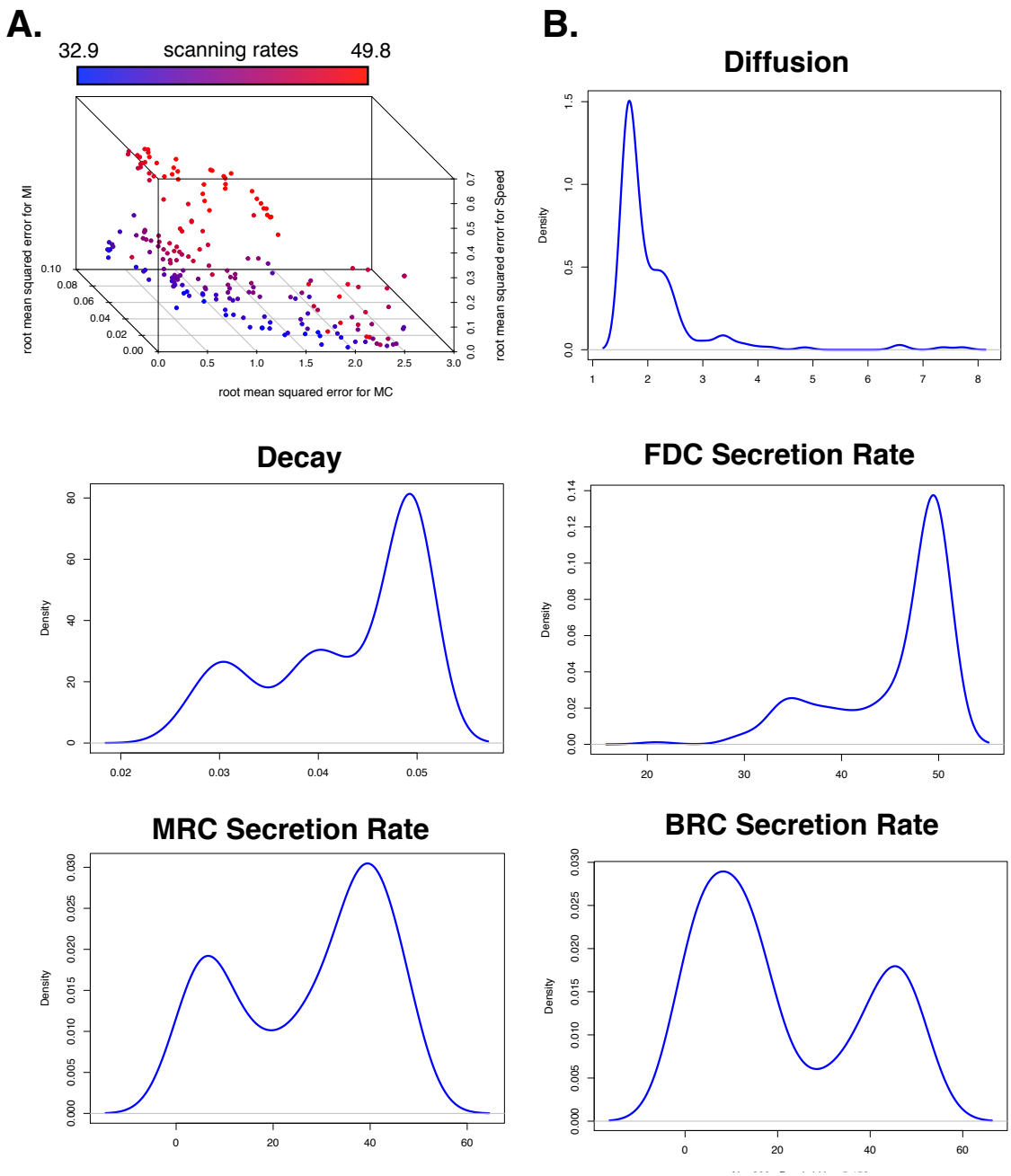

Figure 6.7 Pareto optimal solutions obtained by NSGA-II on CXCL13emulator. (A) Pareto front of solutions representing the optimal trade off in performance between cell behaviours (r.m.s.e of motility coefficient not shown) and scanning rates, using NSGA-II. (B) Parameter distributions corresponding to the Pareto optimal solutions shown in (A).

6.5. Experimental Assessment of CXCL13 Solubility

Simulation analysis predicts a model where CXCL13 diffusion is constrained, leading to short sharp gradients proximal to the stromal cells that secrete CXCL13. To assess the veracity of this prediction we performed an immunohistochemistry analysis of CXCL13 expression in FDC networks within human tonsils and LNs.

In this experiment we stained frozen tissue sections with antibodies against the FDC-specific marker CD35 and for CXCL13. This analysis shows that CXCL13 is tightly associated with stromal cells (**Figure 6.8-6.9**) with foci of both high and low fluorescence, consistent with **model 1**. To quantify this observation we perform a spatial autocorrelation analysis whereby each image is discretised into 14.4μm^2 bins and the fluorescent intensity in each is compared to all other bins. Lymph node samples had a large proportion of immature FDC networks with little CXCL13 signal detected (**Figure 6.8**). Consequently, it was not possible to perform this analysis on lymph nodes, and so this analysis was performed on human tonsil sections **Figure 6.9.** Correlograms show that the correlation in fluorescence intensity decays as a function of distance (**6.11a**), with intensities becoming significantly uncorrelated (detailed in **Chapter 2.1.8.4**) over short distances (order of 40 μm). Comparing this result to those obtained for **model 1** and **model 2** shows that model 1 produced a chemokine field with no statistically significant difference to human tonsil sections, while model 2 did display a statistically significant difference (**Figure 6.11b**).

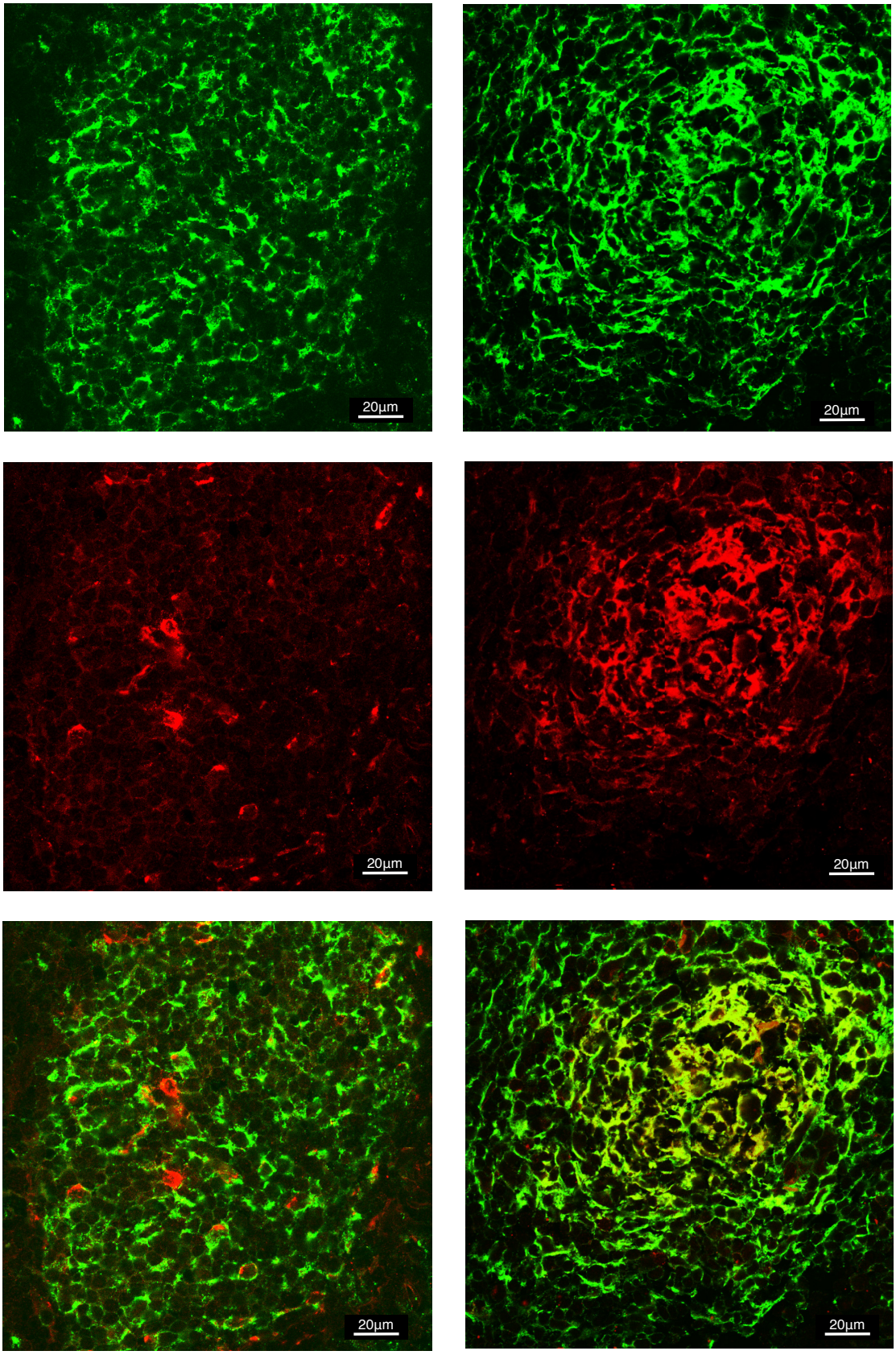

Figure 6.8 CXCL13 expression in human lymph nodes. The left hand side shows an immature FDC network with relatively low expression of CD35 (green) and CXCL13 (red). The right hand side shows a mature FDC network with upregulated CD35 and CXCL13. Scale bar = 20 µm.

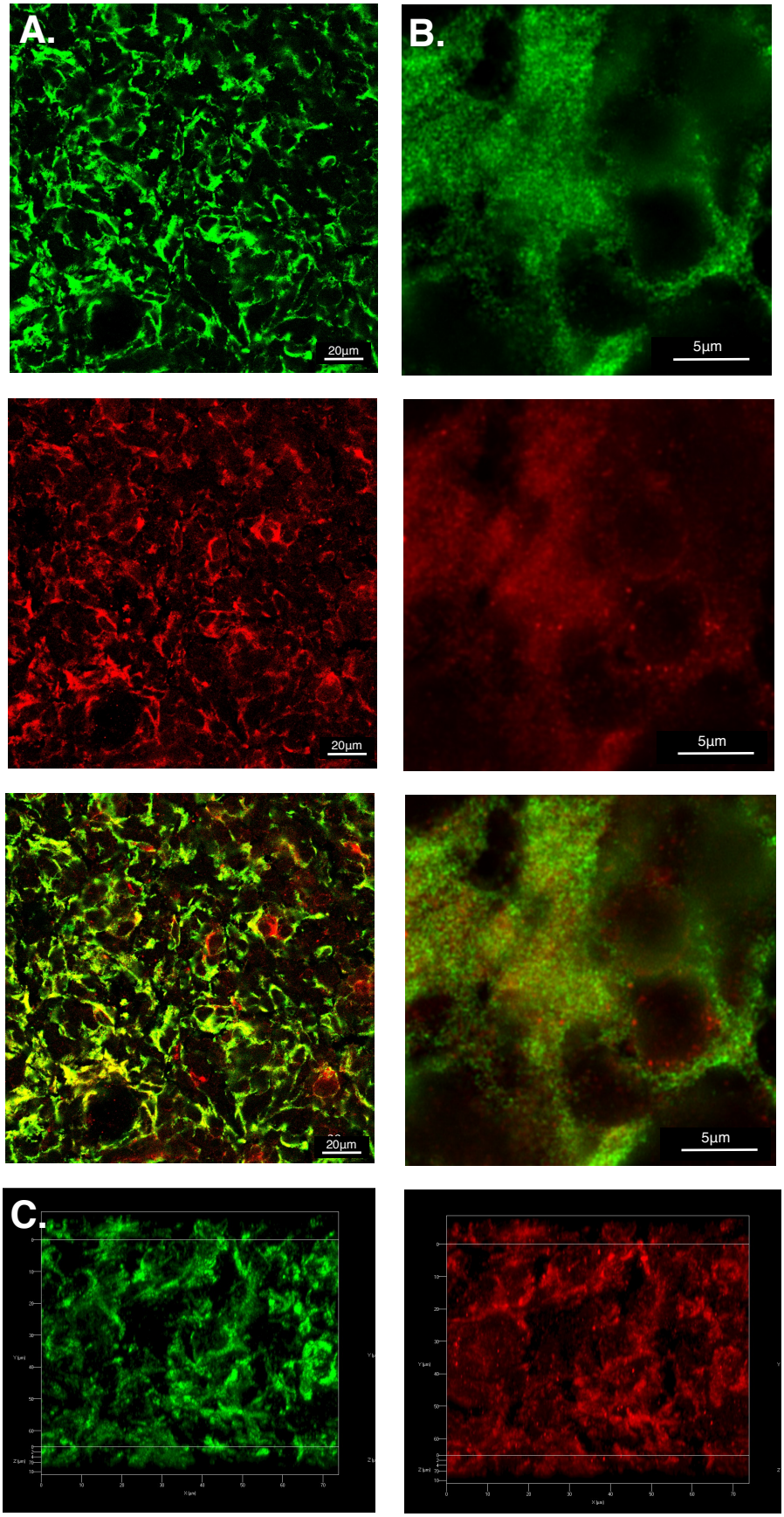

Figure 6.9 3D image of CXCL13 expression in human tonsils. (A) CD35 (green) and CXCL13 (red) expression in human tonsils, scale bar = 20 µm. (B) Close up of CXCL13 and CD35 co-expression, scale bar = 5 µm (C) Z-stack of CD35 and CXCL13 in human tonsils.

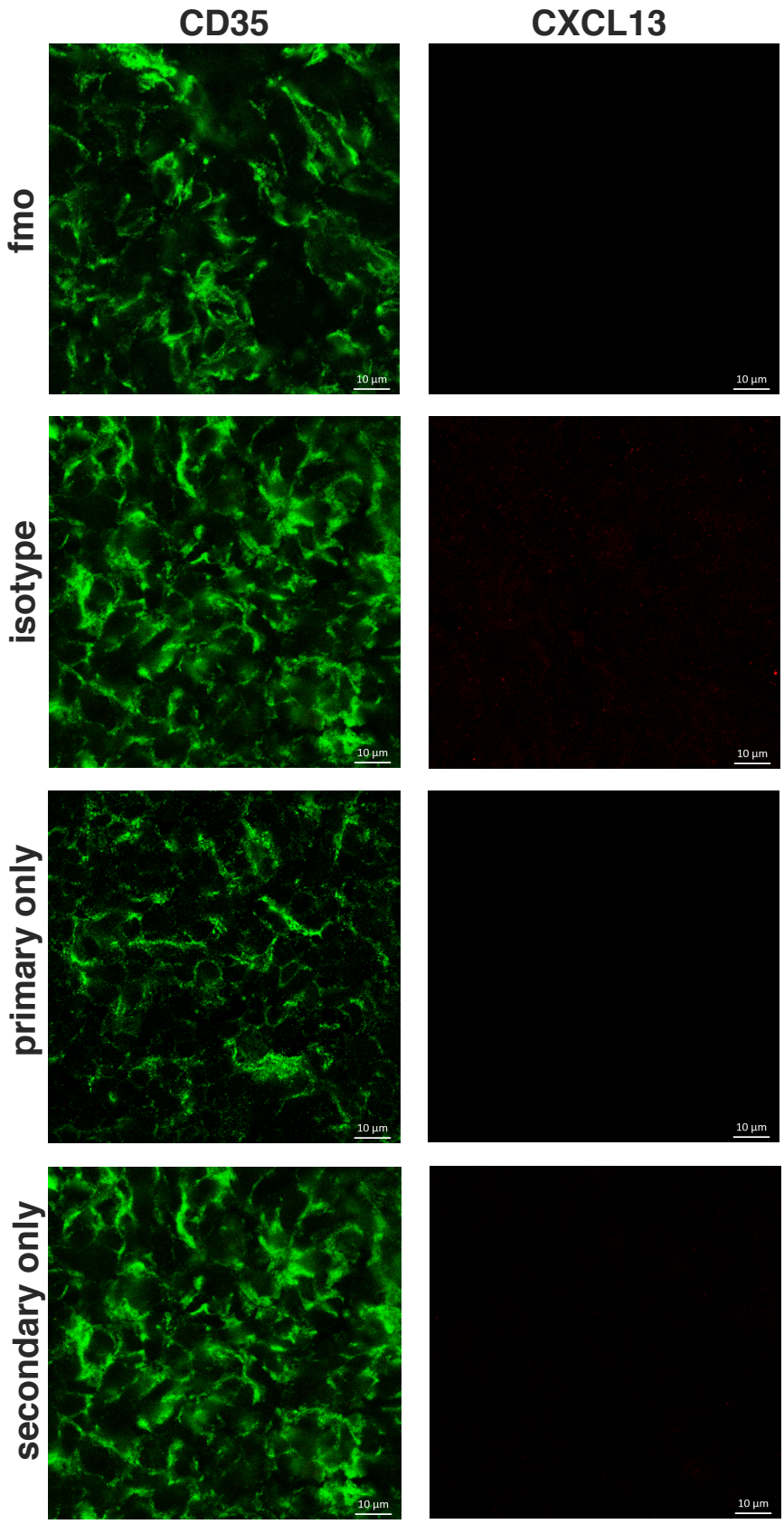

Figure 6.10 IHC controls for CXCL13 expression in tonsils. FMO refers to the fluorescence minus one control. Scale bar = 10µm

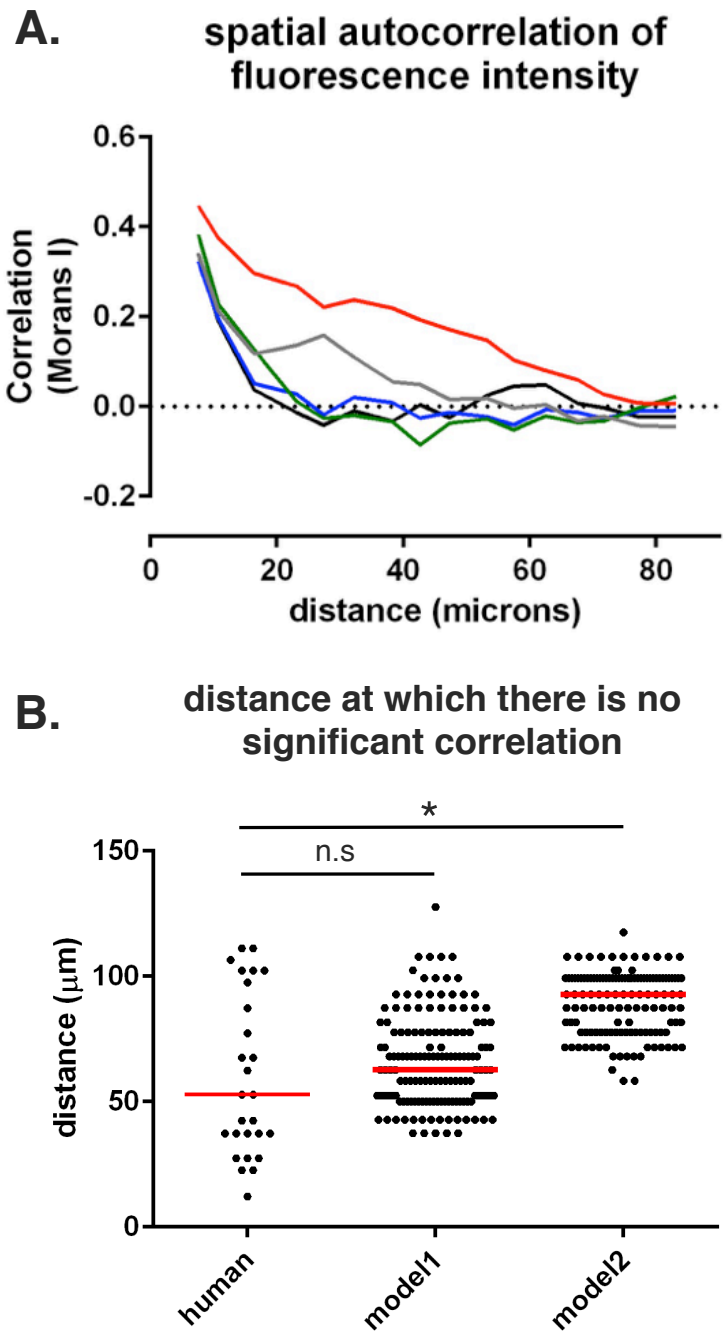

Figure 6.11 Calculating the spatial autocorrelation of fluorescent intensities in the CXCL13 channel from confocal images of frozen tonsil sections. (A) Shows the spatial autocorrelations obtained for 5 FDC networks taken from the same patient. (B) Indicates the distances at which the spatial autocorrelation is no longer statistically significant spatial correlation. No statistically significant difference was found in the distances between model 1 and the human tonsil data but was found for model 2.

6.6 Perturbing CXCR5 Expression *In Silico*

In vivo analyses of B-cell migration using multiphoton imaging showed that $CXCR5^{-/-}$ B cells had altered migration characteristics compared to wild-type B cells. Analysis of turning angle distributions and meandering indices show that CXCR5 deficient B cells migrate in a less directed manner while analysis of speed and motility coefficients show that cells move more slowly (Coelho et al., 2013). To assess the consequences of CXCR5 deficiency on the induction of humoral immune responses we quantified scanning rates in $CXCR5^{-/-}$ B cells *in silico*. This analysis predicts that CXCR5 deficiency leads to a significant decrease in the ability of cognate B-cells to scan the follicle for antigen (**figure 6.12a**). While complete loss of the receptor yielded a strong phenotype, perturbations to total numbers of CXCR5 (between 10-100,000 receptors) led to modest changes in the baseline rate of network scanning (**Figure 6.12b**).

Overall scanning rates were robust to one-at-a-time perturbation to both K_{on} and K_{off} rates suggesting that scanning is robust to the affinity of binding (defined as $1/K_d$ or $K_{on}/K_{off.}$) over the ranges examined, suggesting the scanning can occur over a broad range of concentrations (**Figure 6.13b & 6.13e**). This is consistent with multiphoton measurements following viral infection, where despite a 20-fold reduction in chemokine expression CXCR5 remained essential for B cells scanning of antigen-presenting cells (Coelho et al., 2013) . However, OAT perturbations to the rate of desensitisation did alter scanning rates as did perturbation of K_i and K_r rates (**Figure 6.13a,c and d**). This result, and results from an eFAST global sensitivity analyses suggesting parameter synergy (**Chapter 5.6**), led to the hypothesis that the dynamic modulation of CXCR5 signalling is a key regulatory mechanism of antigen scanning.

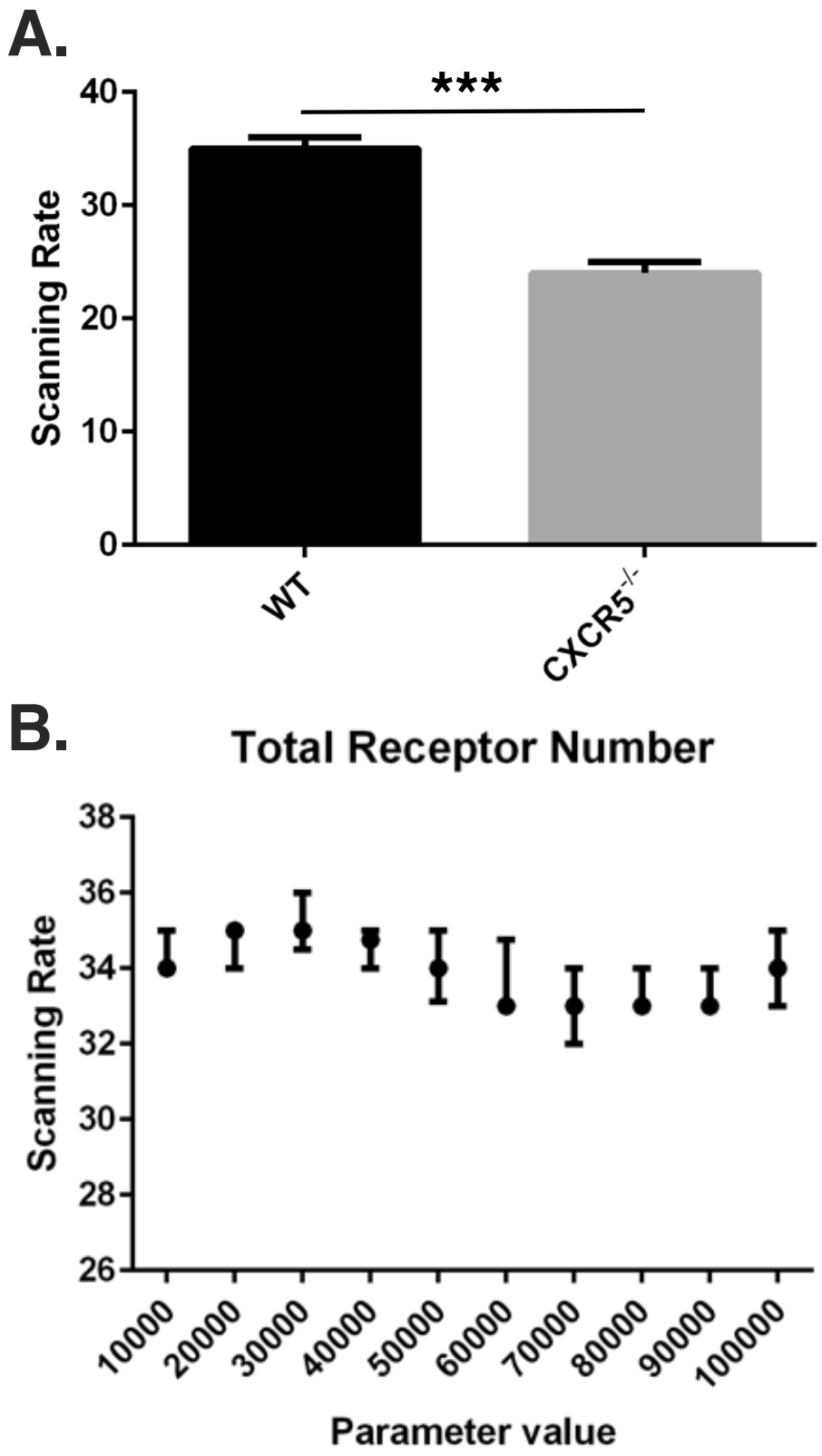

Figure 6.12 Robustness of scanning rates to CXCR5 Expression *in silico*. (A) quantification of scanning rates in wild-type and CXCR5 deficient B cells. (B) Quantifying scanning rates when perturbing total receptor numbers holding all other parameters fixed at baseline values. Median values are shown with error bars representing the I.Q.R. Significance assessed using a Mann Whitney test.

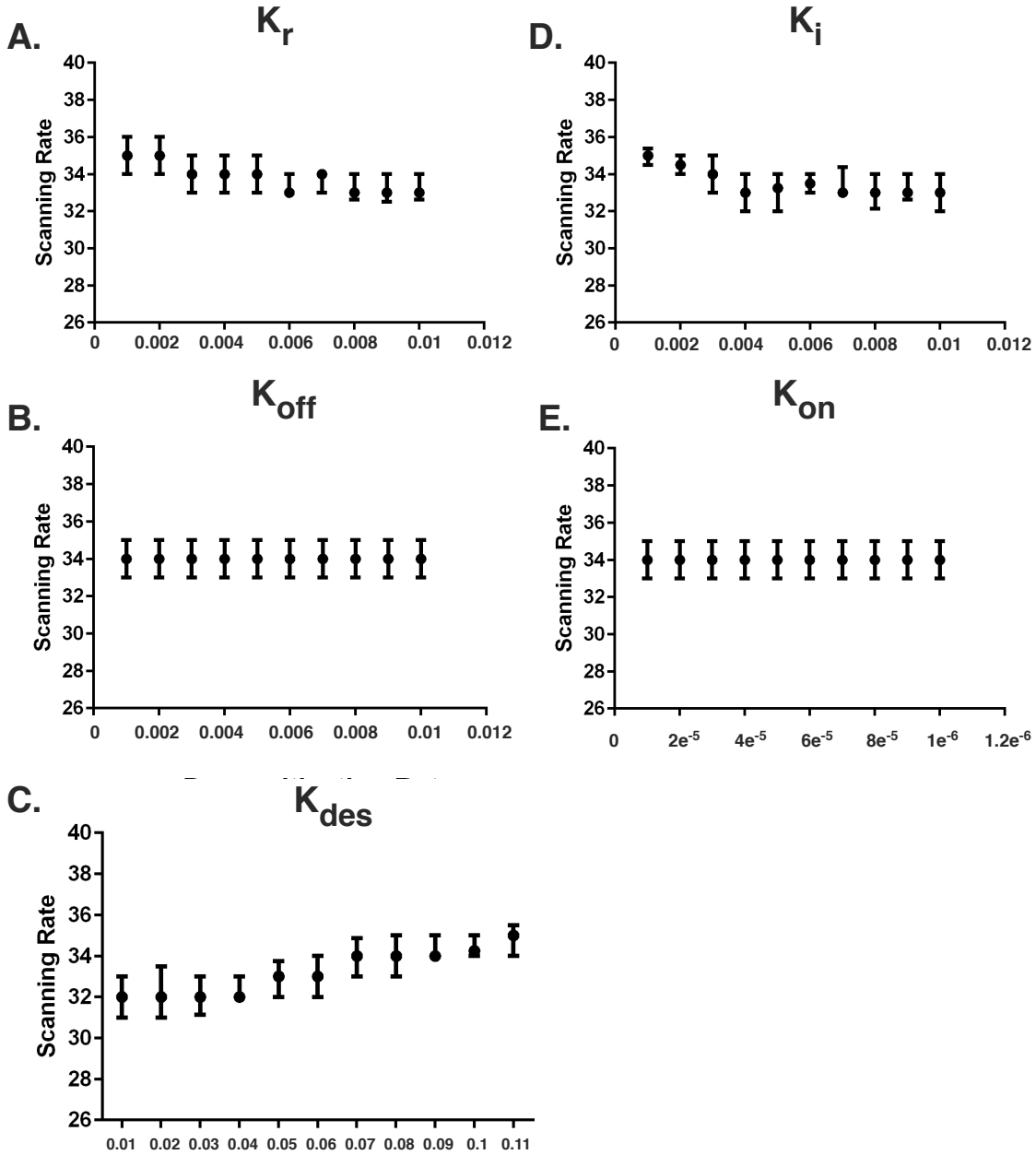

Figure 6.13. **Perturbing parameters that regulate CXCR5 signalling.** (a) perturbing the recycling rate of CXCR5 and assessing the effect on *in silico* scanning rates. (b) perturbing the off rate of CXCR5 and assessing the effect on *in silico* scanning rates. (c) perturbing the desensitisation rate of CXCR5 and assessing the effect on *in silico* scanning rates. (d) perturbing the internalisation rate of CXCR5 and assessing the effect on *in silico* scanning rates. (e) perturbing the on rate of CXCR5 and assessing the effect on *in silico* scanning rates. Median values are shown with error bars representing the I.Q.R.

6.7 The Spatiotemporal Dynamics of CXCR5 Expression *In Silico*

To further explore the hypothesis that the dynamic expression of CXCR5 at the cell surface and the dynamic modulation of signal transduction may impact scanning within the follicle we follow the temporal (**Figure 6.14**) and spatial (**Figure 6.15**) dynamics of CXCR5 expression. Each cell was unique in its temporal dynamics (**Figure 6.14a**). Analysis of signalling and free receptors on the cell surface shows an oscillatory pattern of expression suggesting that signalling within complex tissues is dynamic (**Figure 6.14b**). Analysis of each receptor subset shows that desensitised receptors were the highest CXCR5 subset followed by internalised receptors and relatively few free and signalling receptors on the cell surface (**Figure 6.14c**). Interestingly, CXCR5 expression was spatially regulated with highest levels of signalling occurring at the subcapsular sinus (**Figure 6.15**). This was associated with low numbers of free receptors and high numbers of internalised receptors (**Figure 6.15**).

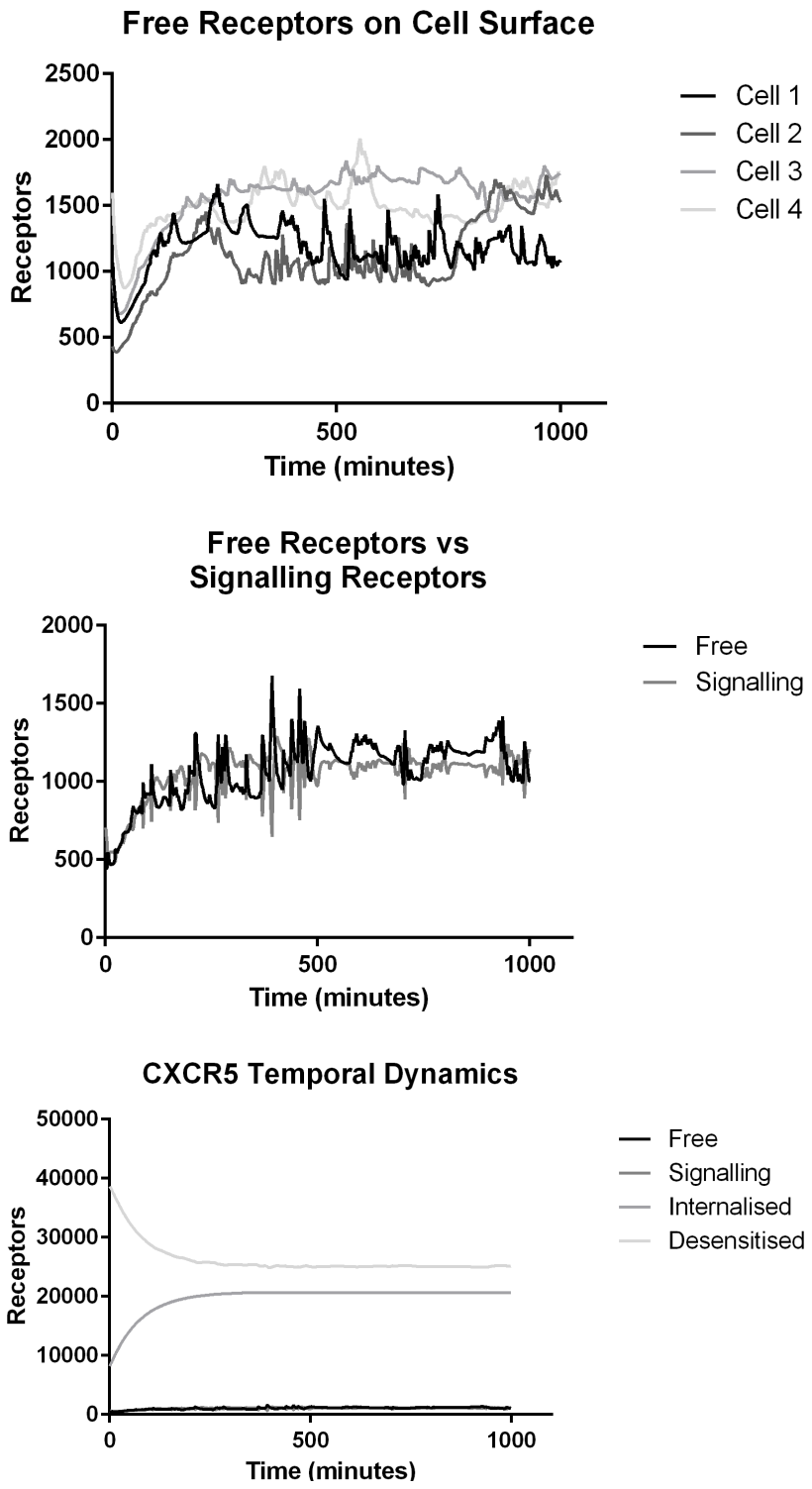

Figure 6.14 Single cell tracking *in silico* to assess the temporal dynamics of CXCR5 expression and signalling. (A) single-cell tracking of CXCR5 expression on the cell surface for 4 different B cells. Each line represents a distinct B cell within the same simulation run. (B) Comparison of free and receptor signalling dynamics within a single cell. (C) Comparison of free, signalling, internalised and desensitised receptor dynamics within a single cell.

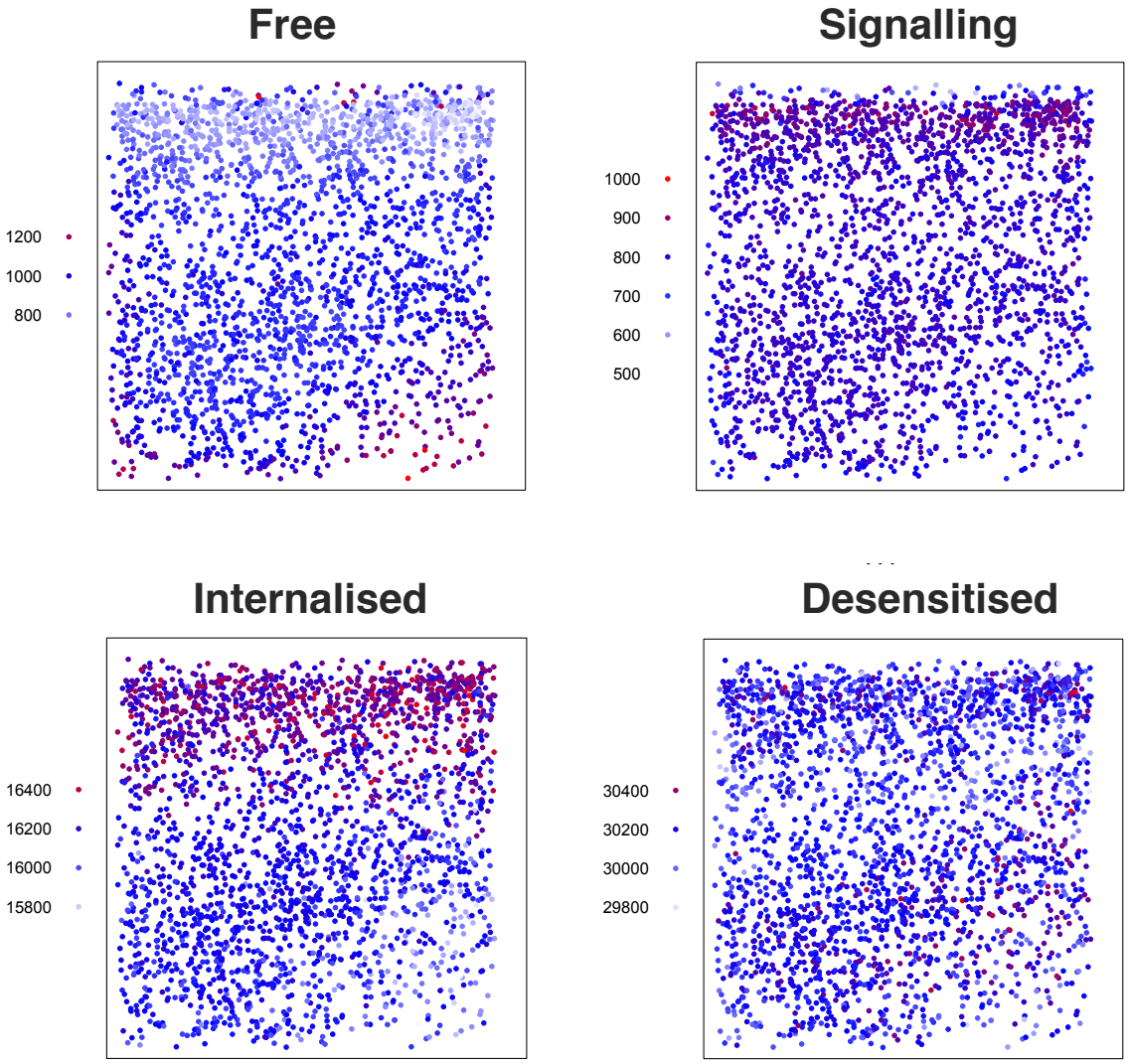

Figure 6.15. **Spatial dependence of CXCR5 signalling within the follicle.** Each dot in the diagram represents the X and Y coordinates of a B cell agent in the simulator. The top of each square diagram is the subcapsular sinus. Each agent is coloured by the amount of receptors (as indicated by the title of each plot) with red representing high values and blue representing low values. (top left) the spatial distribution of free receptors within the follicle. (top right) the spatial distribution of signalling receptors within the follicle. (bottom left) the spatial distribution of internalised receptors within the follicle (bottom right) the spatial distribution of internalised receptors within the follicle. Under baseline parameter values receptor signalling is highest at the SCS (the top of each square diagram). Free receptors on the cell surface are low, signalling and internalised receptors are high at the SCS.

6.8 Determining the Optimal Configuration of CXCR5 Signalling

Simulation analyses performed in the previous sections suggest that dynamic modulation of signalling rather than overall receptor numbers are a key determinant of CXCR5 efficacy. However, this data cannot be used to determine what configuration of CXCR5 promotes the highest rates of antigen scanning. To address this we perform multiobjective optimisation of the CXCL13emulator as per **section 6.4**.

This analysis shows that our objectives are conflicting, with increased scanning rates leading to poorer agreement between emergent cell behaviours *in silico* and laboratory measures (**Figure 6.16**). Analysis of the parameter distributions corresponding to the population of Pareto optimal solutions shows that K_{on}, K_{des}, K_i and K_r are highly constrained at high values, consistent with the hypothesis that dynamic modulation of signalling through rapid receptor turnover and desensitisation promotes effective migration.

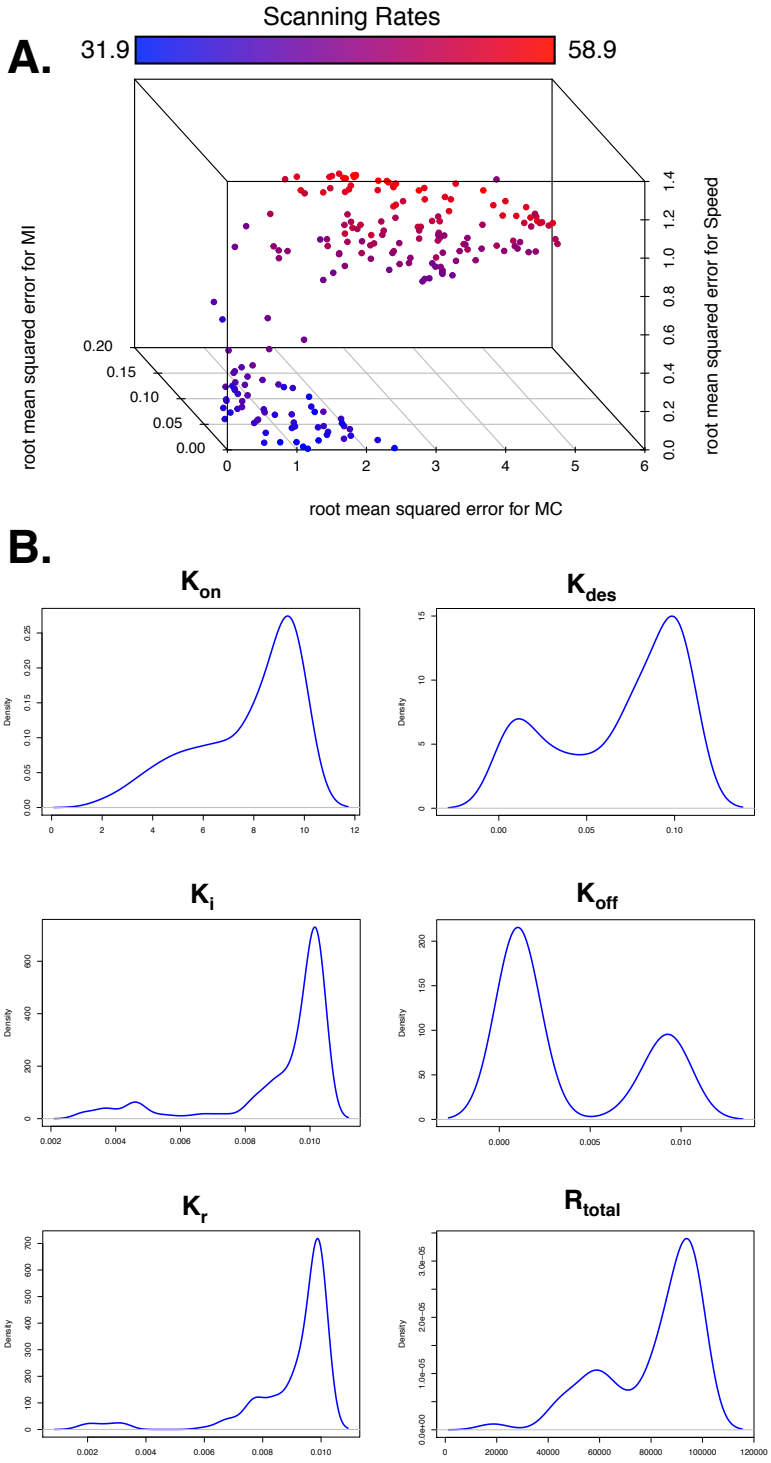

Figure 6.16 Optimising CXCR5 signalling *in silico*. (A) Pareto front of solutions representing the optimal trade off in performance between cell behaviours with scanning rates colour coded, using NSGA-II. (B) Parameter distributions corresponding to the Pareto optimal solutions shown in A.

6.9 Discussion

In this chapter we used the *in silico* tools (CXCL13Sim and CXCL13emulator) developed in **Chapter 5** in conjunction with experimental approaches to study CXCL13-mediated cross talk between B cells and stromal cells within lymphoid tissues. In **section 6.2** we assess cell specific contributions to follicular scanning rates and show that FDCs are a key determinant of scanning within the primary lymph node follicle. Loss of CXCL13 from BRCs and MRCs did not significantly alter scanning rates within the entire follicle. This result was consistent with an OAT analysis in which secretion rates were perturbed for each cell type. Interestingly, using the fluorescent intensity values for each cell types to inform relative secretion rates yielding a CXCL13 landscape which promoted scanning at the subcapsular sinus, a site where large antigen enters the LN. This configuration may also promote shuttling of antigen by naïve B cells to the FDC network for long-term storage.

Following analysis of cell-specific contributions to CXCL13 secretion we assessed the sensitivity of scanning rates to perturbations in either the decay constant or the diffusion constant. Scanning rates were more sensitive to changes in the decay constant than the diffusion constant but this analysis did not look at non-linear effects between these parameters. To address this we assessed 2 putative models for CXCL13 bioavailability within the follicle (**section 6.3**). Model 1 describes a situation whereby CXCL13 interacts strongly with the ECM yielding a complex patterning at the molecular scale with the emergence of non-uniform gradients. Model 2 assumes that CXCL13 is soluble and does not interact as strongly with the ECM. Consistent with results from the OAT analysis model 1 yields higher scanning rates than model 2 (**Figure 6.5c**), a prediction that was supported by IHC analysis and measures of the spatial autocorrelation of CXCL13 expression. Our data suggests that chemokine concentrations exist in short sharp gradients and are correlated over short distances (on the order of ~60 μm) and then become significantly uncorrelated. Using an MOEA approach to determine the spatial configuration of CXCL13 that promotes the best scanning rates further supports this model, with short sharp distributions making up a large proportion of the Pareto optimal set of solutions. This model is supported by additional mathematical analyses suggesting that gradient formation is increased when chemokines are secreted in matrix-binding form as compared to a non matrix-interacting form. In addition, this spatial profile may explain

the sharp bordering between B cell and T cell zones in lymphoid tissues (**Figure 6.17**) and between dark and light zones in GCs.

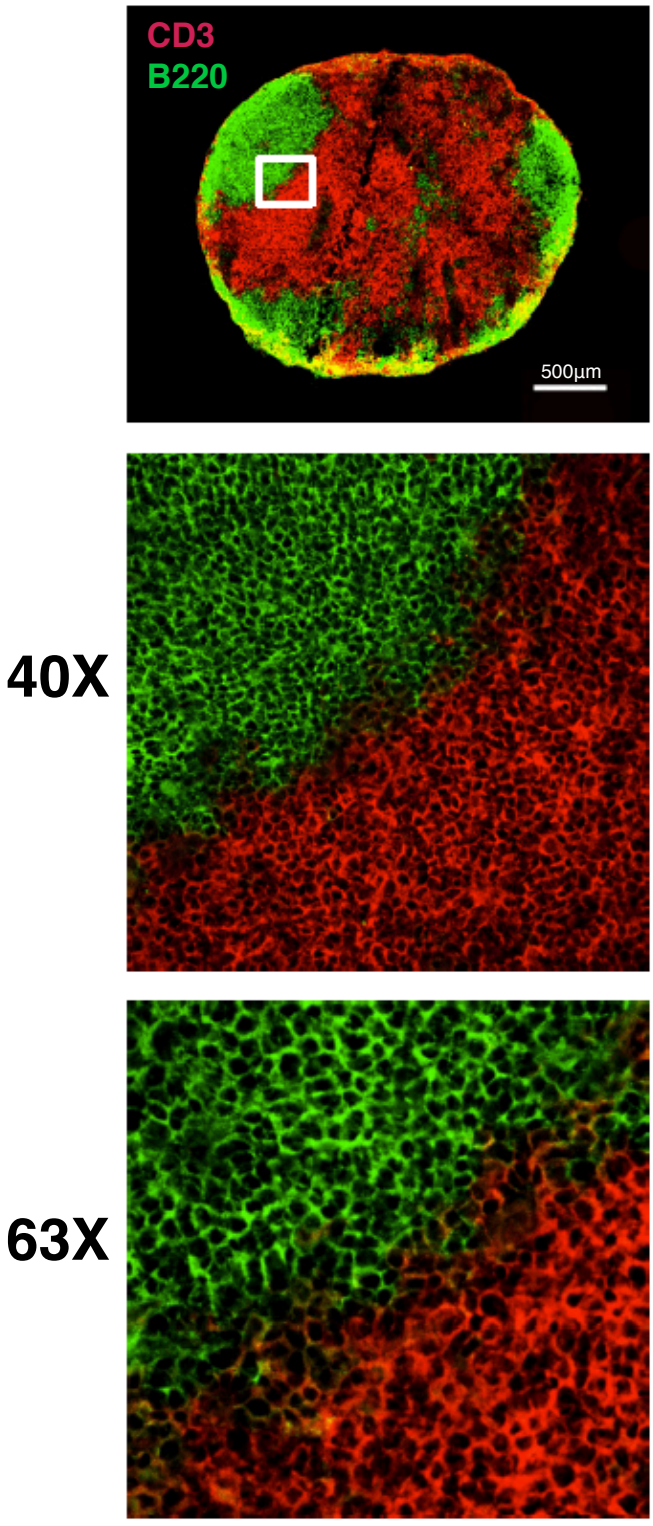

Figure 6.17.**Bordering between B and T cell zones of murine lymph nodes.** Data provided by Dr. Anne Thuery

Other factors that may also alter gradient formation include the propensity of chemokines to form oligomers, and the flow of lymph through the lymph node. The chemokine interactome has recently been mapped showing that chemokines have the capacity to form homo and heterodimers, which may alter their diffusivity and propensity to bind components of the ECM (Bennett et al., 2011; Hundelshausen et al., 2017). The impact of lymph flow on chemokine availability has been assessed in tissue engineered lymph node models where increased flow leads to increases in fluid pressure and shear stress reported to affect the maturation of stromal cells and the secretion rates of CCL21 (He et al., 2002; Matsumoto et al., 1990; Swartz et al., 2008; Tomei et al., 2009). Theoretical studies suggest that the flow of lymph could affect the spatial distribution of chemokines, with subtle flows in lymph driving significant asymmetry in protein concentrations (Fleury et al., 2006)

In addition to CXCL13 gradient formation we use CXCL13Sim and CXCL13emulator to assess CXCR5 signalling. Our data supports a model where CXCR5 mediated signals are dynamically regulated at the cell surface, with rapid desensitisation and turnover of receptors affording a high degree of temporal sensitivity. In this MOEA approach our objectives are highly conflicting, with improvements in scanning rates gained at the expense of fitting experimentally determined migration metrics. In this system good consistency with empirical measures yields scanning rates of 31-40 unique gridspaces reached within a simulation run representing 24 hours of real time.

Strikingly, as little as 0.1nM difference in concentration across the length of the cell is sufficient to induce chemotaxis (**Appendix 2**); a feature that allows B cells to detect even subtle asymmetries in the localized chemotactic landscape. This is consistent with studies that suggest a difference of 10 signalling receptors across a cell is sufficient to induce chemotaxis along a gradient (Herzmark et al., 2007; Zigmond, 1981), theoretical calculations (**Appendix 2**) and *in vitro* migration assays (Gunn et al., 1998a).

Chapter 7 Discussion

In this thesis we aimed to address the following hypothesis: CXCL13 gradients are a key determinant of the onset of humoral immune responses and are determined by the localised tissue microenvironment. This hypothesis was tested through the following objectives:

1. Use visual notations to scope key components of the biological system and quantify these components using an ensemble of imaging and cytometry approaches.
2. Consolidate this data through development of a 3D hybrid multiscale simulator, CXCL13Sim.
3. Use CXCL13Sim to simulate CXCL13 gradient formation and associated B-cell responses.

This chapter provides a reflection of how the data provided in this thesis addresses these objectives in light of our research question. Key contributions to the field of systems biology are highlighted, with particular focus on the development of multiscale models. In addition, we provide a perspective on future directions for systems-based approaches in biomedical research. Finally the results presented in this thesis are placed in the context of the wider literature, with an updated schematic for CXCL13-mediated regulation of B-cell migration and how this model applies more broadly to morphogens.

7.1 Summary of Key Biological Findings and their Relevance

7.1.1 Addressing Objective 1: Quantitative Characterisation of the Homeostatic Follicle

To address objective 1 we have developed a conceptual model of CXCL13-mediated regulation of B-cell migration in the primary LN follicle as a means of consolidating understanding in the existing literature (**Chapter 3**). This process was used to scope the research context, identifying key entities and time points of interest. In addition, the process identified a number of knowledge gaps in the literature, namely a lack of quantitative analyses of CXCL13-mediated cross talk between B cells and stromal cells.

A key challenge was documenting this specification in a way that it is clear to an interdisciplinary team, yet concise enough to be practical. Formal mathematical notation is a powerful approach to summarise current understanding but can alienate team-members who lack formal training in mathematics and can make it difficult to separate conceptual models from implementation models. Visual notations are a useful intermediary that afford the following advantages: (i) can be interpreted by non-specialists (provided they are familiar with the syntax); (ii) explicit (can be interpreted objectively, *not* subjectively); and (iii) can concisely describe large amounts of information (Cosgrove et al., 2015; Read, 2011). In light of this we describe both Domain and Platform models using an adaptation of the UML described by Read *et al (2011)*.

This process of developing a Domain Model and performing a functional requirements analysis identified the need to obtain key quantitative datasets to inform and validate our theoretical model. These datasets can be technically challenging to acquire, requiring the combined expertise of multiple research laboratories, spanning several scientific disciplines (**Chapter 4**). To study the structural organisation of the follicle we characterised the key topological properties of $CXCL13^{+}$ stromal cell networks in collaboration with the Ludewig group. Data from the Ludewig group is consistent with previous studies showing that follicular stroma, although interconnected, are heterogeneous. Using the novel CXCL13-EYFP reporter mouse we find at least 3 distinct $CXCL13^{+}$ subsets: FDCs, BRCs and MRCs. Notably, FDCs have a very unique

morphology – maximising their surface area to create single cell niches within which B-cells floss the network for immune complexes.

Topological analysis of the CXCL13^{+} follicular stromal cells (**Chapter 4**) identified some of the key structural properties that underpin the B-cell niche. The complete network displays a high clustering coefficient and low shortest path length, conferring small-world properties that may promote information transfer throughout the follicle (Watts and Strogatz, 1998). Within the follicle, information transfer can refer to either antigen, which must shuttle from the SCS to the maturing FDC network (Carrasco and Batista, 2007; Phan et al., 2007) or B cells which must migrate to different niches within the B-cell microenvironment (Pereira et al., 2010). In future work it may be possible to test this hypothesis experimentally by comparing the dynamics of ovalbumin acquisition by a B-cell receptor transgenic strain of mice (Carrasco and Batista, 2007) in homeostatic stromal networks or those remodelled through TLR-4 adjuvants using multiphoton imaging.

As the FDC network is much denser with many finer processes it was not possible to perform a morphological reconstruction using the IMARIS software package. Additionally, we only sample a section of tissue from the centre of the follicle and so our samples may not be representative of the entire follicle. To improve the resolution of confocal microscope images it may be possible to use an Airyscan detector while to avoid sampling biases light sheet microscopy could be used to image an entire follicle. Mapping an entire follicle at high resolution could potentially generate datasets that can be directly read into a simulator without the need to generate *in silico* networks algorithmically. Future work entails looking at how these topological networks change following immunisation and infection, such as in a secondary follicle containing GCs. It is known that the primary follicle reorganises during this process to generate the dark zone where GC B-cells undergo proliferation and mutation and the light zone where GC B-cells undergo selection (Allen et al., 2004, 2007). It is however unclear how the topology of stromal cells is implicated in this process. In the context of autoimmune disease it would be interesting to assess whether stromal network topology is reconfigured to facilitate the sustained presentation of self-antigen (Salomonsson et al., 2003).

While topological mapping provided an insight into the tissue level processes that govern CXCL13 bioavailability, a quantitative understanding of chemokine mobility was lacking.

To address this we measured the dynamics of CXCL13 diffusion in collagen matrix and lymphoid tissues using single molecule imaging (**Chapter 4**). This was achieved through collaboration with the Leake lab, with expertise in combining high-speed narrowfield microscopy with bespoke analysis techniques, to develop an assay capable of probing the heterogeneity of molecular diffusion in complex tissues. This approach is capable of single-molecule sensitivity and tracking across multiple length scales, enabling ~40nm spatial precision, quantitative stoichiometry and mobility determination with sub-millisecond sampling. We show using BSA as an exemplar case study that it outperforms competing molecular mobility methods (FCS and FRAP). Strikingly, empirical values were 2 orders of magnitude lower than those derived using the Einstein stokes scheme in which we modelled the stokes radius of the molecule and assumed the viscosity equalled that of homogeneous intracellular media. This large difference between free diffusion and diffusion in tissue can be attributed to biochemical interactions within the localised microenvironment that give rise to multimodal diffusion patterns.

Our data demonstrates that single molecule approaches, which afford high spatiotemporal resolution, are particularly suited to looking at molecular mobility in complex tissues. While we demonstrate the efficacy of the approach on chemokines, this is a proof-of-concept for a more general scheme that could be applied to lipids and cytokines. In future studies it would be interesting to compare the diffusivity of proteins to signalling lipids, both of which regulate B-cell migration. Single molecule approaches have already yielded significant insights into diffusion of proteins within plasma membranes, identifying 3-5 distinct modes of diffusion. These observations led to the development of updated theoretical tools with older models viewing the membrane as roughly homogenous replaced by a fence and picket model where the actin-based membrane skeleton partitions the membrane into smaller sub compartments (Ritchie et al., 2003). Increased research focus on single molecule measurements of chemokines may also power updated theoretical models to understand and better predict chemokine bioavailability.

In addition to quantifying follicular stromal cell network topology and CXCL13 diffusivity we used flow cytometry to quantify the cellular density of the follicle while our collaborators provided multiphoton imaging data to quantify the dynamics of B-cell migration. To place these measurements in the context of the wider literature we derive a quantitative description of the canonical lymph node follicle to increase understanding of

CXCL13 bioavailability at tissue, cellular and molecular levels of organisation. However bridging understanding between these diverse scales was challenging, limiting our ability to precisely map CXCL13 *in situ*. To address this, the *in silico* tool developed in **Chapter 5** was used to investigate the dynamics of CXCL13 gradient and associated B-cell responses (**Chapter 6**).

7.1.2 Addressing Objective 2: Development of an *In Silico* Follicle

The work undertaken in chapters 3,4 and 5 culminated in the development of an *in silico* platform to simulate chemokine gradient formation and associated cell migration patterns, a process requiring extensive experimental input and the consolidation of a number of existing models within the literature. To promote model parsimony and reduce execution time, we hybridised a number of different techniques within an agent-based framework. Specifically, we extend upon previous models of chemotaxis *in vitro*, and *in vivo* models of lymph node stromal cells to develop a tool capable of examining the migration of heterogeneous cell types within a complex tissue.

The emergent cell behaviours of the simulator are consistent with our *in vivo* datasets with predictive capabilities confirmed through analysis of CXCR5 deficient migration. The outputs of the simulator are non-deterministic, producing a diverse set of outputs for a fixed parameter input. To mitigate this aleatory uncertainty we empirically determined the number of replicate runs required to evaluate a parameter set (**Chapter 5**). In addition, we quantify parametric uncertainty using local and global analyses to facilitate translation of simulator-derived outputs to the biological system (**Chapter 5**). Key design decisions, data used to inform model development and test data is provided as arguments over evidence to facilitate model evaluation and repurposing.

The system complements experimental work, given the relative speed at which data is acquired, the ability to perform single cell analyses, and the ability to assess how perturbations at one level of organisation manifest at different scales. The use of ordinary differential equations within each agent permits single-cell analyses of CXCR5 expression with high spatiotemporal resolution. This key advantage allows the tool to serve as an adjunct to experimental approaches that are not capable of performing single cell analyses *in situ.* Flow cytometry approaches require cells to be isolated from a tissue,

while confocal microscopy takes a static snapshot of a dynamic process. While single-cell sequencing is becoming more prevalent, linking this information back to spatial positioning within a tissue is challenging, with need for a robust panel of markers that differentiate cell subsets on the basis of spatial positioning or the use of micro-dissection. Similar issues arise when experimentally measuring CXCL13 within a tissue: fluorescent reporter systems cannot be used to tag the molecule itself, as the fluorescent label would drastically outweigh the molecule and alter its diffusive and binding characteristics.

This model accounts for one chemokine, a key limitation given CCL19/21, CXCL12 and 7,25α hydroxycholesterol also regulate the migration of naïve B cells (Pereira et al., 2010). The decision to omit these factors from the model was influenced by the lack of quantitative data for these molecules. Including these factors with little data to constrain parameter values and calibrate model behaviours would introduce too much uncertainty to create a robust model. Consequently, a design decision was made to focus on one chemokine and to spend the additional time making key quantitative measures for CXCL13. Future work to incorporate these additional factors would require additional data on the stromal cells that secrete these molecules, their concentrations *in vivo*, and their mobility characteristics.

A key unknown in this system is whether a cell's exposure to simultaneous conflicting gradients leads to a balanced amalgamation of signals where each is equally potent, or a hierarchical system where dominant signals actively downmodulate others, through reduced affinity for ligand or through receptor desensitization and downregulation. This is particularly interesting for B cells that respond to both protein and lipid based morphogens. We hypothesise that lipids are much more diffuse and thus prone to form long shallow gradients in contrast with chemokines. This may alter gradient steepness and potentially bias signalling. To address this question it would be necessary to quantify the migration profiles of WT, CXCR5$^{-/-}$ and EBI2$^{-/-}$ mice using multiphoton microscopy. If sufficient information were available to constrain input parameters then it would be possible to develop 3 competing models to explain this data: (i) CXCR5 and EBI2 signals are equally potent (ii) CXCR5 signals dominate, and (iii) EBI2 signals dominate. Use of a multiobjective evolutionary algorithm could then be used to determine the optimal trade-off between different migration metrics allowing for assessment of which model can best reproduce experimentally measured kinetics, an approach illustrated by (Read et al.,

2016). Once a model has been selected, it would be necessary to generate predictions that can be used to experimentally rule out other competing theories.

7.1.3 Addressing Objective 3: Analysis of Chemokine Gradient Formation

Using CXCL13Sim, we first assessed the cell-specific contributions to gradient formation. To address this we systematically perturbed CXCL13 secretion rates using a one-at-a-time approach. This analysis suggests that FDCs are a key contributor to CXCL13 scanning rates, even within the primary follicle. This result is supported by analyses where CXCL13 secretion was blocked in each specific cell type, with FDC-specific blockade yielding significantly different scanning rates.

Despite this progress, the relative secretion rate of each cell type was poorly understood. Interestingly, we noted that CXCL13 reporter expression was not homogeneous between the different stromal cell subsets with MRCs showing highest expression, followed by FDCs and MRCs. To assess the functional consequences of this spatial configuration of chemokine we used fluorescent intensities to inform relative secretion rates. While this did not increase overall scanning rates, scanning at the SCS was significantly increased. This may be important functionally as the SCS is a major site of antigen entry into the follicle.

A key knowledge gap we wished to assess was whether CXCL13 acts in principally a soluble or immobilised form. To address this we exploited the emergent properties of agent-based models, generating different spatial configurations and seeing which configuration yielded the highest emergent scanning rates, assuming that CXCL13 is spatially configured to promote scanning *in vivo.* Specifically, we assessed the veracity of two different models: in model 1 we assume that CXCL13 interacts strongly with ECM components and is thus limited in its diffusivity. In model 2 we assume that CXCL13 is more soluble and diffuses more readily through the tissue. Comparison of these two models suggests that model 1 leads to higher scanning rates, in agreement with IHC and associated spatial autocorrelation analysis. Interestingly, OAT perturbation analysis showed only a modest difference in scanning rates compared to combinatorial

perturbations, a result suggesting that there is a synergistic effect between these two parameters.

While this analysis shows that short sharp gradients are more efficient in regulating scanning than long diffuse patterns it still does not determine the optimal configuration of CXCL13 within the primary follicle. To address this we employed a MOEA to determine which parameter configurations yielded the highest scanning rates. This analysis was consistent with previous findings with parameter sets in the Pareto front heavily skewed towards low diffusion constants and high decay rates. Taken in concert our data suggests that the mobility of CXCL13 *in vivo* is constrained and leads to an irregular patterning. This unique configuration can be attributed to the topology of the stromal cell networks and interactions with the ECM. In addition, the spatial distribution and not just total concentration of CXCL13 regulates the induction of humoral immune responses.

In future work we would like to assess the effect that the ECM has in regulating gradient formation experimentally. To assess if the ECM alters CXCL13 diffusivity we propose treating lymphoid tissue sections with enzymes that can disrupt the extracellular matrix and quantifying the mobility of fluorescently labelled chemokine molecules. It may also be of interest to assess proteases including matrix metalloproteinases, aminopeptidases and members of the cathepsin family which have been reported to cleave chemokines and alter their signalling potency (Mortier et al., 2008, 2012; Wolf et al., 2008).

This study also illustrates the complexity of the chemotactic landscapes encountered *in vivo* and the need to better model these landscapes *in vitro* to better study migration in the context of lymphoid tissues. Simulation analyses could be used to calculate the key properties of *in vivo* landscapes to inform *in vitro* assays.

7.1.4 Addressing Objective 3: Analysis of CXCR5 Dynamics

CXCR5 is a key molecular player in CXCL13-mediated cross talk between B cells and stromal cells. *In vitro* studies of GPCR-mediated migration have highlighted a highly dynamic and intricate regulation network at the cell surface that allows immune cells to dynamically perceive the localised environment. Translating these findings *in vivo* is

challenging due to the limitations of currently available experimental approaches. To address this we explored the robustness of B cell scanning to perturbations in CXCR5 signalling and expression *in silico*.

In silico scanning of the LN follicle was robust to changes in total receptor numbers while complete loss of the receptor yielded a significant phenotype. This led to the hypothesis that the dynamic modulation of CXCR5 signalling through rapid turnover and desensitisation of receptors is a key determinant of antigen scanning. To explore our prediction we used an MOEA approach to determine what combination of CXCR5 associated parameters gave rise to the highest scanning rates. Consistent with our hypothesis, this analysis supports a model whereby dynamic turnover of receptors and modulation of signalling are critical. Parameter sets that gave rise to Pareto optimal solutions were characterised by highly constrained values of internalisation, recycling, on, and desensitisation rates. In addition, we performed single cell tracking experiments that indeed identify surface as oscillatory, an emergent property that has been reported in other theoretical studies but has not been verified experimentally (Chan et al., 2013). Interestingly, CXCR5-mediated signalling was also spatially regulated – an emergent property where highest rates of signalling were observed at the SCS, close to the site of antigen entry into the lymph node parenchyma. Given our data, we hypothesise that dynamic modulation of receptors at the cell surface leads to fine-tuning of migratory responses within the B-cell niche. However, further analyses are required to determine the contribution of other chemotactic molecules and adhesion molecules to this phenomenon.

7.2 Summary of Contributions to the Field of Systems Biology

7.2.1 Agile Development of Multiscale Models

For the work conducted in this thesis, we tightly integrated theoretical and experimental approaches. Consequently, the simulator was constantly evolving over the course of the project; as new data became available from the laboratory or from simulation analysis our conceptual understanding of the system was refined. To address these issues, we have updated the current instantiation of the CoSMoS process, drawing inspiration from the Chaste project to incorporate agile techniques into the development workflow (**Chapter**

3) (Pitt-Francis et al., 2009). Novel features include a modular design and a hierarchical testing strategy that facilitates software testing at both the validation and verification level. The separation and encapsulation of functionality through modular design allowed us to refactor software subcomponents while mitigating disruption to other components. This approach was complemented by the use of regression testing, an automated approach to assess whether new functionality disrupted the existing codebase.

Through incorporation of agile development approaches, based on the observation that project requirements can change over time, we were able to adapt to changing circumstances rather than assume that model development is a predictable activity. With further case studies and refinement, agile approaches may become industry standard best practises for model qualification in the pharmaceutical industry.

7.2.2 Emulation and Optimisation of Multiscale Simulators.

Despite the demonstrated success of our agile CoSMoS instantiation, we encountered significant bottlenecks at the Results Model phase of development. Multiscale simulators often require repeated execution and extensive sampling of high-dimensional parameter spaces to fully analyse and evaluate. While a range of statistical analysis techniques can be applied to understand and mitigate parametric and aleatory uncertainty, their application becomes limited by time and resource constraints. As the execution time for a simulation increases, it becomes less tractable to perform these statistical analyses within a time frame that can run parallel to laboratory or clinical studies. In chapters 5 and 6 we employ emulation and heuristic algorithms to address this issue.

Using machine-learning algorithms we were able to emulate the behaviours of CXCL13Sim. This significantly reduced the time taken to evaluate the simulator, permitting rapid prototyping and identification of errors in model design, parameterization, and software infrastructure. Due to the complexity of scientific software it can be challenging to locate errors before running time intensive statistical analyses. Emulation allowed for early detection of errors reducing the time penalties incurred if found at later stages of development. In addition, it was possible to perform an eFAST analysis that was previously intractable due to time constraints. In future work it would be

of interest to assess different machine learning algorithms and to combine them using an ensemble approach. Aside from emulation we note 2 additional technologies which could be used to reduce the time taken to evaluate model inputs: (i) tuneable resolution, where the granularity of a simulator is dynamically adapted within a single execution run; (ii) parallelisation, where distributed computing systems can reduce execution time.

Using CXCL13Sim we wished to determine the optimal configuration of CXCR5 and CXCL13 within the primary follicle with respect to scanning. However, two key technical challenges had to be addressed: (i) multiple conflicting objectives (ii) a large complex search space. To address this we combined our emulator with the MOEA NSGAII. This rapid evaluation of parameter sets in conjunction with an efficient search strategy, making the analysis feasible. MOEAs use a similar principle to the germinal center, using mutation and selection to determine a set of solutions (a Pareto front) where improvement in one objective cannot be obtained without compromising performance for another objective. In the GC this trade-off occurs because a mutation that improves binding for one epitope on a complex microbial antigen may reduce binding to other epitopes, reducing overall binding avidity. In **sections 6.4** and **6.8** we see that highest scanning rates were obtained at the expense of matching with experimentally measured cell kinetics highlighting how our objectives are conflicting.While NSGAII has been applied to calibrate model behaviours to experimental data this may be one of the first instances of using a MOEA to optimise an immune system mechanism. Thus the model acts as a non-linear set of constraints within which we seek to obtain a desired response. It would be interesting to explore this approach in a therapeutic setting in future work.

Given the success of these approaches we propose that future updates to the agile CoSMoS instantiation incorporate techniques that facilitate rapid prototyping and identification of errors in model design, parameterization, and software infrastructure. In reducing the time taken to evaluate parameter inputs, these techniques permit the application of a wider range of analysis techniques to enrich understanding of the simulator and to facilitate translation of simulation outputs back to the biological system. This ultimately could have a significant impact on the outcomes of a model-informed biological research project.

7.3 Looking ahead: Emerging Systems Biology Technologies and Applications

In this thesis we showcase the use of systems approaches to consolidate understanding across broad spatiotemporal scales in the context of the immune system. However, systems biology is a rapidly evolving field, with demonstrated value across a range of biological sub-disciplines, such as synthetic biology (Purnick and Weiss, 2009), microbiology (Vieites et al., 2009), integrative biology (Karr et al., 2012b), and biomedicine (Winter et al., 2012). In the following section we provide a perspective on emerging technologies and applications of systems-based approaches in the biomedical sciences.

7.3.1 Model-Driven Drug Discovery

The model-driven drug discovery paradigm provides a "quantitative framework centred on knowledge and inference generated from integrated models of compound, mechanism and disease level data to improve the quality, efficiency and cost effectiveness of decision making" *(EFPIA MID3 Workgroup et al., 2016)*. The pharmaceutical industry has been a relatively late adopter of model-driven approaches despite the demonstrated value of modelling and simulation across a range of industries and disciplines, such as finance, aerospace, and weather forecasting. However, the approach is now gaining significant traction in the pharmaceutical industry with key milestones including the use of mechanistic modelling approaches to design a post marketing clinical trial by the US Food and Drug Administration (FDA) (Peterson and Riggs, 2015), a study showing that human *in silico* modelling outcompetes animal models in predicting drug-induced cardiotoxicity (Passini et al., 2017), and the use of an integrative systems approach to quantify anticancer drug synergy in imatinib-resistant chronic myeloid leukaemia (Winter et al., 2012).

Increasingly, models are being used in all stages of the drug development process, with the capacity to inform experimental design, "go/no-go" decisions, portfolio prioritization, and to bridge understanding between experimental animal models and human disease (Allerheiligen, 2010; Lalonde et al., 2007). In the context of drug discovery, modelling approaches can be broadly classified as follows: (1) mechanistic "bottom" up approaches that integrate existing knowledge to inform mechanistic target evaluation; (2) Data-driven

pharmacokinetics and pharmacodynamics modelling (PK/PD) that link drug concentration to a therapeutic response, such as the change in biomarker expression and (3) quantitative systems pharmacology approaches that combine aspects of the previous 2 categories to assess the mechanism of action at a target incorporating physiological processes and different dosing strategies to focus further experimental studies (**Figure 7.1**). Currently, much focus is placed on defining a well-defined set of industry standards for model qualification such that regulatory bodies can better assess and evaluate model and simulation derived results. With the introduction of standardised practises, pharmaceutical companies are now beginning to review, quantify, and report the efficacy of model-based strategies (Milligan et al., 2013; Visser et al., 2014).

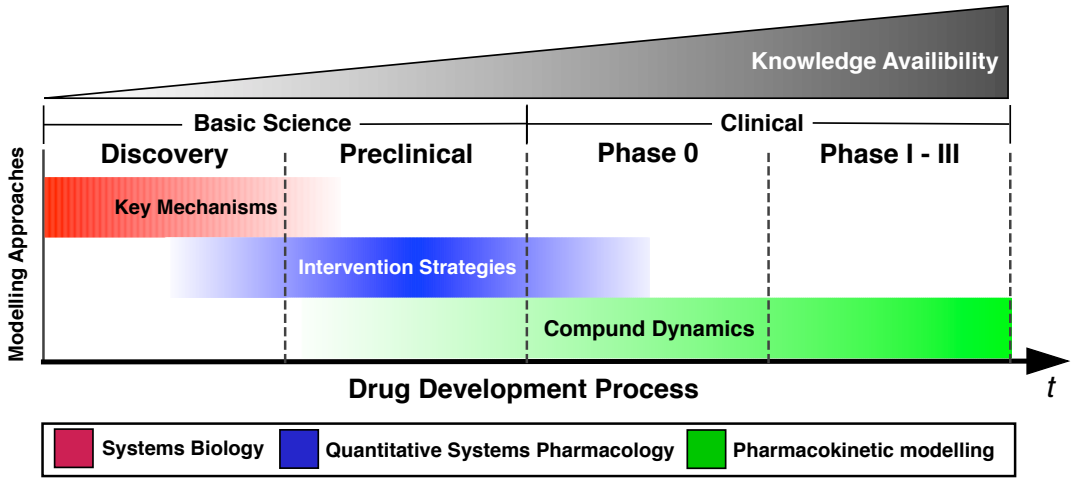

Figure 7.1 **Integration of systems-based approaches into the drug development workflow**

A key area of drug development where modelling approaches may have significant impact is person-centred healthcare. Personalised healthcare approaches use patient specific information to (i) accurately diagnose disease and to (ii) tailor treatment plans to maximise efficacy (Chan and Ginsburg, 2011). Advances in personalised medicine approaches to date have largely been driven by technological advances in –omics technologies (Fernald et al., 2011). However, making the transition from the laboratory to clinical practise is challenging; relative to experimental animal models, such as age and sex matched mice from the same genetic background, patients vary widely. Thus modelling approaches are required to analyse noisy multivariate datasets to determine what key parameters can aid with diagnosis and tailoring treatment plans. In this context, bottom up approaches have been used to stratify patients on the basis of mechanistic understanding of spinal cord (SCI) and vocal cord pathologies. Both diseases display a

high degree of patient variability can make it difficult to predict patient-specific disease progression and treatment responses. In the context of SCI, simulation analysis predicted a higher propensity for ulcer formation in SCI patients, thus identifying a high-risk patient subset (Solovyev et al., 2013), while for vocal cord pathology simulation analysis identified optimised treatment strategies (Li et al., 2008). Aside from academic studies, there are already many examples of systems-based approaches used in clinical practise such as use of a pharmacogenetic dosing algorithm for warfarin (Sagrieya et al., 2010) and identification of susceptible genotypes to reduce the incidence of hypersensitivity reactions to the HIV-1 medication abacavir (Hetherington et al., 2002).

Machine learning is a highly active research area that also has the capacity to significantly impact the drug development process. Over the past decades computers have automated tasks that programmers could describe with clear rules and algorithms; machine-learning techniques allow us to do the same for tasks where such rules are difficult to define. This attribute makes machine-learning approaches well placed to impact personalized medicine with demonstrated efficacy in the diagnosis of complex diseases. Recently, a deep neural network to diagnose skin cancers from clinical images, with performance comparable to that of dermatologists (Esteva et al., 2017). However, a limitation of machine learning approaches is that they are considered "black-box" and may fail to provide mechanistic explanations for the disease-associated patterns that are discovered. To address this, closer integration of bottom-up mechanistic based with top-down data-driven approaches may be required to develop personalized medicine strategies. For example, molecular biomarkers could be determined obtained using machine learning approaches and subsequently used as inputs to simulate patient-specific disease trajectories and/or intervention strategies.

7.3.2 An Executable Lymph Node Atlas

Lymph nodes are highly complex 3D structures supported through networks of mesenchymal stroma, vasculature and lymphatics (Junt et al., 2008). As such it is necessary to develop an accurate 3D atlas of lymph node microanatomy that incorporates current understanding. We argue that this should be an *in silico* platform acting as an executable knowledge repository. This approach allows researchers from different

disciplines and geographical locations to have the most up to date understanding of the physical system available as an executable software package. The software would serve as a useful adjunct to laboratory studies, as resource and ethical constraints ultimately limit the amount of experiments that can be performed. Such an approach affords the following additional advantages**:** (i) Quantitative modelling offers a principled framework for combining disparate sources of information; (ii) the identification of knowledge gaps; (iii) the discovery of new organizing principles of biological systems; and (iv) the development of sufficiently accurate computational models to facilitate decision making during various phases of drug discovery and bioengineering applications.

However, to realize this ambition there a number of technical considerations to be addressed: (i) what data is required to inform and validate such a model; (ii) how to describe the underlying conceptual model and implementation blueprint in a way that is accessible to a large interdisciplinary consortium; (iii) how should the resultant simulation platform be implemented and tested; (iv) what granularity is required for each system component; (v) how to deal with processes occurring on different timescales (vi) how to reduce computational expense; (vii) how to deal with phenomena that are poorly characterized and have no quantitative data available.

Given currently available technologies we propose that the following datasets could be incorporated in the first instance: (i) lymph node size, volume and cell density; (ii) 3D organisation of stromal, vascular and lymphatic networks; (iii) concentrations of key soluble factors; (iv) cellular mass, volume, shape and migration characteristics; (v) transcript data from key cell types; (vi) metabolic fluxes for each key cell type. Key considerations in terms of data acquisition include the need for rigorous protocols and standards, a non-trivial exercise given that different cell types may not be amenable to certain types of analysis: lymphocytes are small and thus are more difficult to analyse using single-cell RNA-sequencing; FDCs are a rare cell type and are thus difficult to isolate without enzymatic digestion of the tissue.

Following the initial data acquisition phase modelling efforts could then be used to determine additional key measures to obtain in subsequent iterations, in line with the model-driven approach to experimentation (Butler et al., 2016b). Following characterisation of the homeostatic lymph node it may be of interest to perform similar

analyses following immunisation, bacterial infection and viral infection. The value of a phased approach was illustrated by the Human Genome Project, which defined key milestones along the way (genetic maps, physical maps, rough-draft sequence, finished sequence). Another key consideration is providing metadata in an accessible format that is clear and unambiguous but also accessible to an interdisciplinary team; the use of visual notations may assist with this. Focus on improving visualisation of simulation outputs through emulation of experimental approaches or GUIs may expedite the process and facilitate communication of model-derived results (Butler et al., 2014).

As the complexity of a simulator is dependent on the research context, a lightweight modular framework that would allow for researchers to reconfigure the platform to suit a particular application is warranted. This is an important point as to be useful; an atlas must also be an abstraction—comprehensively representing certain features, while ignoring others. Such an approach would allow for models of differing granularities to be used for each module with a bespoke architecture created for each research application. Such an approach would allow for features to be assessed at different levels of magnification, with high-dimensional information collapsed into simpler views. Such an approach would require well-defined interfaces and design patterns to link modules effectively. This may also facilitate interfacing with existing tools such as SBML based models to enrich cell behaviours and computational fluid dynamics software to assess how lymphatic flows may alter the distribution of molecules throughout the tissue and how this may influence emergent cellular behaviours. Another interesting extension would be to add intelligent agents into the *in silico* lymph node, where through reinforcement learning, pathogen agents and immune cell agents would co-evolve strategies to outcompete one another.

Ultimately, such large-scale models must offer fundamentally new insights and experimentally testable predictions to justify the resources required to build them. Despite the challenges listed above, this approach has already demonstrated value through a whole cell computational model of the human pathogen *Mycoplasma genitalium (Karr et al., 2012b).* Despite being a nascent technology, experimental analysis directed by model predictions identified previously undetected kinetic parameters and biological functions while simulation analyses predicted an inverse relationship between the durations of DNA replication initiation and cell replication rates (Karr et al., 2012b).

7.4. Closing Remarks and an Updated Model of CXCL13 Gradient Formation

The data presented in this thesis shows that $CXCL13^+$ follicular stromal cells are not randomly orientated but organise into small-world networks. At the molecular level we show that CXCL13 diffusion is highly constrained with simulation analyses suggesting that this creates a complex patterning at the molecular scale. Perturbation of CXCL13 bioavailability was shown to affect scanning rates *in silico* with optimisation analysis suggesting that this configuration is designed to promote the onset of humoral immune responses. We conclude that this data is consistent with our hypothesis that lymphoid tissues are essential for the formation of CXCL13 gradients, and perturbation of the spatial distribution of CXCL13 can affect the onset of humoral immune responses. In light of our results we propose the following model of CXCL13 bioavailability and associated B cell responses in lymphoid tissues (**Figure 7.2**):

The spatial distribution of CXCL13 is regulated at tissue, cellular and molecular levels of organization. At the tissue and cellular level, the topological organization of $CXCL13^+$ stromal cells is configured to promote trafficking of B cells and antigen where stromal cell processes act as guidance structures for adhesion factor expressing immune cells. At the molecular level, this architecture, in conjunction with the localized tissue microenvironment, creates a complex patterning. CXCL13 diffusion is highly constrained by localized tissue microanatomy, undergoing transient interactions with other chemokines, GPCRs and the ECM. This leads to the formation of complex CXCL13 fields with non-linear gradients regulating the precise localization of CXCR5 positive immune cells.

Surface immobilized CXCL13 may require proteolytic processing into a functional form. Once liberated, CXCL13 mediated signalling through CXCR5 polarises the actin cytoskeleton orientating the cell along a gradient. At the cell surface, CXCR5 mediated signals are transient with rapid desensitisation and turnover of receptors affording a high degree of temporal sensitivity. Strikingly, as little as 0.1nM difference in concentration across the length of the cell is sufficient to induce chemotaxis; a feature that allows B cells to detect even subtle asymmetries in the localized chemotactic landscape.

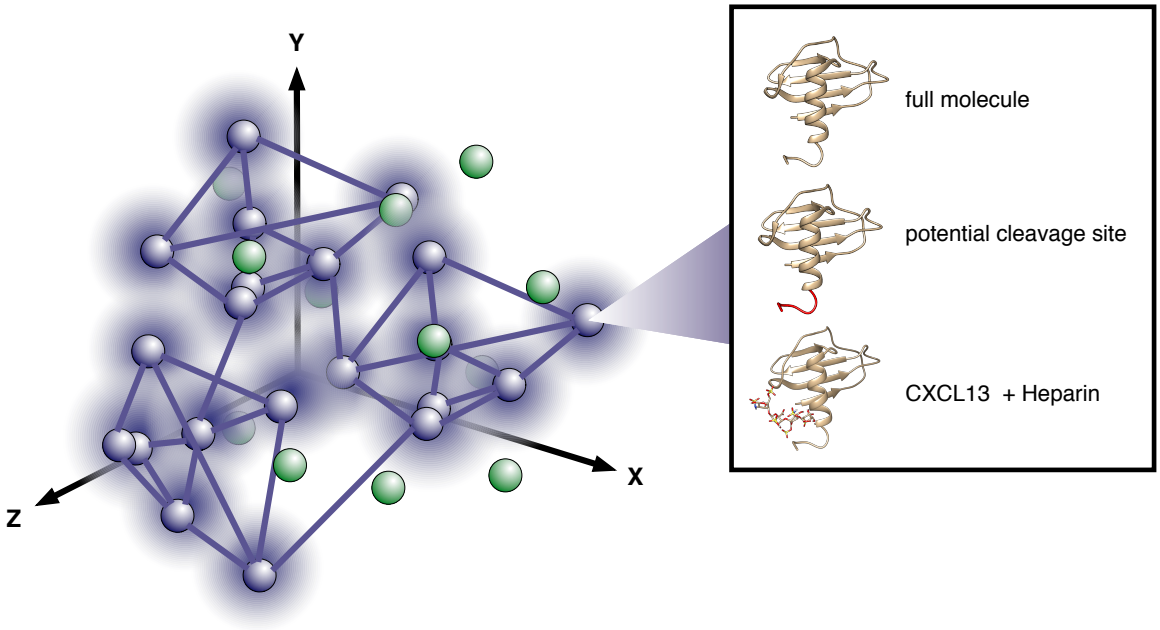

Figure 7.2. Updated scheme for CXCL13 bioavailability within lymphoid tissues as described in section 7.4. CXCL13 is secreted by a dynamic stromal network that displays a small world-network topology. At a molecular level, mobility is constrained by transient molecular interactions giving rise to complex non-linear gradients that are largely immobilized. Immune cells rapidly modulate cell surface and intracellular signalling components to interpret these signals with fine spatiotemporal sensitivity. In addition to signalling based mechanisms, proteolytic processing may be required to liberate chemokine into a functional form. On the figure we highlight a potential cleavage site that would remove the carbohydrate binding domain at the heparin-binding C-terminus of the molecule (red) (Monneau et al., 2017).

Appendix

Appendix 1. Derivation of the Diffusion Constant for Chemokines in Water.

The diffusion constant of a molecule can be estimated using the Einstein Stokes relation.

$$D = kT/\ 6\pi\eta r$$

Where k is the Boltzmann constant, T is the temperature, η is the viscosity of the medium in which the molecule is travelling, and r is the Stokes radius of the molecule. Once this constant is determined we can then predict the mean squared displacement of the molecule over time using the following expression:

$$< x^2 > = 2dDt$$

Where $<x^2>$ is the mean distance squared, d is the number of dimensions. To estimate the radius of the molecule we model it as a sphere (as it is globular) with a MW of 10kDa (1.66053×10^{-20} grams).

$$\text{Volume} = \text{mass} / \text{density}$$

The density of a protein can be assumed to be constant at 1.37:

$$\text{volume} = \frac{1.66053 \times 10^{-20}}{1.37} = 1.212065 \times 10^{-20}\ \text{g cm}^{-3}$$

$$\frac{4}{3}\pi r^3 = 1.212065 \times 10^{-20}\ \text{g cm}^{-3}$$

$$r = 1.42 \times 10^{-7}\ \text{cm} = 1.42\text{nm} = 1.42 \times 10^{-9}\ \text{m}$$

$$D = \frac{kT}{6\pi\eta r} = \frac{(1.381 \times 10^{-23})(310)}{6\pi(0.00101)(1.42 \times 10^{-9})} = 1.583596 \times 10^{-10}\ \text{m}^2\ \text{s}^{-1}$$

$$D = 158.35\ \mu\text{m}^2\ \text{s}^{-1}$$

Now that we have an estimate of the diffusion coefficient we can estimate how long on average it would take to diffuse a distance of 200μm:

$$t = \frac{\langle x^2 \rangle}{6D} = \frac{200^2}{(6)(158.35)} = 47\ \text{seconds}$$

Appendix 2. Deriving a Spatial Limit of Gradient Detection.

Studies have suggested that as little as 10 signalling receptors are sufficient to induce chemotaxis. We use this value to derive a detection threshold for immune cells using the following expression:

$$\frac{dC}{dL} = \frac{\text{number of molecules per cell volume}}{\text{length of the cell}}$$

Where C is the concentration of chemokines and L is the length of the cell. The volume of an immune cell is 206μm^3 (or 2.06e^7 μl). The mass of a chemokine is 10kDa; the mass of 10 molecules is 0.00017fg.

$$\text{moles of chemokine} = \frac{\text{mass (grams)}}{\text{molecular weight (kDa)x } 10^3}$$

$$\frac{1.7 \text{ x } 10^{19} \text{ grams}}{10 \text{ x } 10^3 \text{ kDa}} = 1.7\text{x}10^{-23} \text{ moles}$$

Thus the molarity of 10 molecules per cell volume is 825pM. The Detection threshold of an immune cell can then be calculated as such:

$$\frac{dC}{dL} = \frac{825\text{pM}}{7\mu\text{m}} = 117\text{pM} \approx 0.1\text{nM } \mu\text{m}^{-1}$$

Appendix 3. Example Software Tests

The automated regression testing framework used to develop and validate CXCL13Sim employs a number of Java methods that perform tests at the unit, module and system level (**Chapter 3**). Each test involves generating some test inputs, running a method or sequence of methods and evaluating the outputs. This is performed internally within the testing suite with a test report generated in Java to indicate which tests pass and the extent of code coverage. To illustrate how this pipeline works we provide an exemplar test at each level of testing.

Unit test: This unit test asserts that the method getCollisionClass returns the correct enum value for a B cell object.

```
public void testGetCollisionClass() {
	assertEquals(bc.getCollisionClass(), CLASS.LYMPHOCYTE);
}
```

Module test: This module checks that the correct boolean is returned for B-cell interactions with stromal cells. This is achieved by placing a stroma edge at a fixed location in space (defined by the variables loc1 and loc2). We then set the location of BC at loc1 and expect a true to be returned and then set at a location not on the stromal cell and expect a false to be returned.

```
public void testCollideStromaEdge() {

	BC bc = new BC();

	Double3D loc1 = new Double3D(0, 0, 0);
	Double3D loc2 = new Double3D(1, 1, 1);
	bc.setObjectLocation(loc1);
	StromaEdge se = new StromaEdge(loc1,
	loc2,StromaEdge.TYPE.FDC_edge);
	bc.getM_d3aMovements().add(new Double3D(loc2));

	// assert that the stroma and BC collide
	boolean test = bc.collideStromaEdge(se, 1);
	assertEquals(true, test);

	// assert that the stroma and BC don't collide
	Double3D loc3 = new Double3D(5, 5, 5);
	bc.setObjectLocation(loc3);
	boolean test2 = bc.collideStromaEdge(se, 1);
```

```
        assertEquals(false, test2);
    }
```

Integration test: In this test we assess whether BCs acquire antigen over the course of a simulation.

```
    public void testShouldAcquireAntigen() {
        // set up the simulation
        long steps = 0;
        long seed = System.currentTimeMillis();
        SimulationEnvironment.simulation = null;
        SimulationEnvironment.simulation= new SimulationEnvironment(
                    seed,

        IO.openXMLFile("LymphSimParameters.xml"));
        SimulationEnvironment.simulation.start();

        // run the simulation for 400 steps
        do {
              steps =
              `SimulationEnvironment.simulation.schedule.getSteps();
              if
              (!SimulationEnvironment.simulation.schedule.step(SimulationE
              nvironment.simulation))
                    break;
              } while (steps < 400);

        // get all cognate B-cells
        Bag cells = BC.bcEnvironment.allObjects;

        // counter for the number of primed cells
        int primedCount = 0;

        // count the number of primed b cells
        for (int i = 0; i < cells.size(); i++) {
              if(cells.get(i) instanceof cognateBC){
                    cognateBC cBC = (cognateBC) cells.get(i);

                    if (cBC.type == TYPE.PRIMED) {
                          primedCount += 1;
                    }
              }
        }
        // assert that at least 20 of the cells have been primed
        assertThat(primedCount, greaterThan(20));
    }
```

Appendix 4. Lymphocyte Interactions With Stroma

To account for dense lymph node environment lymphocyte migration must take into account interactions with other cell types. To model these interactions we treat both the movement vector of lymphocytes and the edge of the stromal cell as lines (**Figure X**). To determine if the two agents are interacting we calculate whether the closest point between the two lines is less than the sum of their diameters. To determine the closest point, we define the lines L1 and L2 as follows:

$$\begin{aligned} \mathrm{L1(s)} &= \mathrm{P1} + \mathrm{sd1}, \text{where } \mathrm{d1} = \mathrm{Q1} - \mathrm{P1} \\ \mathrm{L2(t)} &= \mathrm{P2} + \mathrm{td2}, \text{where } \mathrm{d2} = \mathrm{Q2} - \mathrm{P2} \end{aligned}$$

Then for some pair of values for s and t, L1 (s) and L2 (t) correspond to the closest points on the lines, and v(s,t) describes a vector between them. The closest point between the two lines is obtained when the vector is perpendicular to both lines (Figure X) i.e. when the dot product of the two vectors is equal to zero or

$$\begin{aligned} \mathrm{d1} * \mathrm{v(s,t)} &= 0 \\ \mathrm{d2} * \mathrm{v(s,t)} &= 0 \end{aligned}$$

Using Cramer's rule we can then solve this system of equations to determine which values s and t where L1(s) and L2(t) correspond to the closest points on the lines.

Substituting the parametric equation for v(s, t) gives:

$$\begin{aligned} \mathrm{d1} * (\mathrm{L1(s)} - \mathrm{L2(t)}) &= \mathrm{d1} * ((\mathrm{P1} - \mathrm{P2}) + \mathrm{sd1} - \mathrm{td2}) = 0 \\ \mathrm{d2} * (\mathrm{L1(s)} - \mathrm{L2(t)}) &= \mathrm{d2} * ((\mathrm{P1} - \mathrm{P2}) + \mathrm{sd1} - \mathrm{td2}) = 0. \end{aligned}$$

This can be expressed as a system of linear equations:

$$\begin{aligned} (\mathrm{d1} * \mathrm{d1})\mathrm{s} - (\mathrm{d1} * \mathrm{d2})\mathrm{t} &= -(\mathrm{d1} * (\mathrm{P1\text{-}P2})) \\ (\mathrm{d2} * \mathrm{d1})\mathrm{s} - (\mathrm{d2} * \mathrm{d2})\mathrm{t} &= -(\mathrm{d2} * (\mathrm{P1\text{-}P2})) \end{aligned}$$

which we can write in matrix form as:

$$\begin{bmatrix} a & -b \\ b & -e \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} -c \\ -f \end{bmatrix}$$

We can solve this using Cramer's rule. Cramer's rule is an explicit formula for the solution of a system of linear equations with as many equations as unknowns, valid whenever the system has a unique solution

$$s = (bf - ce)/d$$
$$t = (af - bc)/d$$

Appendix 5. Absolute numbers of CXCR5 and CXCL13 molecules in the lymph node follicle

To approximate the number of CXCR5 receptors in the follicle we assume that there are 30,000 [10000-50000] receptors per B cell and 48,000 B cells per follicle then:

$$\text{Number of CXCR5 molecules per follicle} = 1.4 \times 10^9 \ [4.8 \times 10^8 - 2.4 \times 10^9]$$

To approximate the amount of CXCL13 we use measures of CXCL13 obtained by western blotting suggesting that there is 0.065ng cxcl13 per mg of tissue (Luther et al., 2002). The mass of a pLN is 1.0124mg ~ 1mg (mean of 5 datapoints) with all CXCL13 located in the B-zones comprised of 15 follicles then:

$$0.065\text{ng} / 15 = 0.01666 \text{ ng per follicle}$$

$$\text{Molar concentration} = \text{moles} / \text{Litre} = 34.64\text{nM}$$

$$34.64\text{nM [10-50nM] in a } 1.25 \times 10^8 \text{ L follicle} = 2.6 \times 10^8 \ [7.5 \times 10^7 - 3.7 \times 10^8] \text{ molecules per follicle}$$

Glossary

APC	Antigen Presenting Cell
APRIL	A Proliferation-Inducing Ligand
BAFF	B-cell Activating Factor
Bcl6	B-cell lymphoma 6
BCR	B-cell Receptor
BEC	Blood Endothelial Cell
BSA	Bovine Serum Albumin
BSF	Biological Services Facility
CR1/CR2	Complement Receptor 1 and 2
DAPI	4',6'-diamidino-2-phenylindole
DC	Dendritic Cell
EDTA	Ethylenediaminetetraacetic Acid
ER-TR7	Reticular Fibroblasts and Reticular Fibres
FCS	Foetal Calf Serum
FDC	Follicular Dendritic Cell
FRC	Fibroblastic Reticular Cell
FSC	Forward Scatter
GC	Germinal Centre
Gp38	Podoplanin
HEV	High Endothelial Venule
ICAM	Intercellular Adhesion Molecule
ICOS	Co-stimulatory molecule inducible co-stimulator
IFN	Interferon
IL-7	Interleukin-7
LEC	Lymphatic Endothelial Cell
LN	Lymph Node
LPS	Lipopolysaccharide
LTa1b2	Lymphotoxina1b2
LTB2	Lymphotoxin B receptor
MAdCAM-1	Mucosal Vascular Addressin Cell Adhesion Molecule-1
MHC-II	Major Histocompatibility Complex II

MI	Meandering Index
MC	Motility Coefficient
MRC	Marginal Reticular Cell
OVA	Ovalbumin
PBS	Phosphate Buffered Saline
pLN	Popliteal LN
RANK-L	Receptor Activator of Nuclear Factor kappa-B ligand
S1P	Sphingosine-1-Phosphate
S1PR1	Sphingosine-1-Phosphate Receptor 1
SBML	Systems Biology Markup Language
SCS	Subcapsular Sinus
SEM	Standard Error of the Mean
SLO	Secondary lymphoid Organ
SSC	Side scatter
SSM	Subcapsular Sinus Macrophage
TLR	Toll-Like Receptor
TNFa	Tumour Necrosis Factor a
VCAM	Vascular Cell Adhesion Molecule
WT	Wild-Type
DE	an equation describing how the rate of change of a variable is related to other variables and parameters.
ODE	a model that predicts changes to components over one independent variable (e.g. time) on a continuous scale.
ABM	Agent based model. A modelling approach in which individual entities are explicitly represented and maintain a set of attributes
ATDD	Acceptance Driven Development
Simulation	An executable implementation of a model
UML	A visual notation used in software engineering ,can be adapted to describe biological systems
A-Test	An effect magnitude test to examine and quantify the difference between two distributions
PRCC	Partial Rank Correlation Coefficient. Assesses the correlation between a parameter input and a model output with respect to the

	other parameters
Platform model	An intermediate model which describes how a conceptual model is to be implemented
Latin-hypercube	A parameter sampling approach to efficiently sample large parameter spaces
In vitro	An experimental procedure performed outside of an organism such as a test tube
In vivo	An experiment performed on a living organism
In silico	An experiment performed using a simulator
Domain	The biological system under study
Domain model	A description of current understanding of the domain.
GSN	A visual notation designed to present an argument over evidence.

Bibliography

Acton, S.E., Astarita, J.L., Malhotra, D., Lukacs-Kornek, V., Franz, B., Hess, P.R., Jakus, Z., Kuligowski, M., Fletcher, A.L., Elpek, K.G., et al. (2012). Podoplanin-Rich Stromal Networks Induce Dendritic Cell Motility via Activation of the C-type Lectin Receptor CLEC-2. Immunity *37*, 276–289.

Aguzzi, A., Kranich, J., and Krautler, N.J. (2014). Follicular dendritic cells: origin, phenotype, and function in health and disease. Trends Immunol. *35*, 105–113.

Albert, R. (2005). Scale-free networks in cell biology. J. Cell Sci. *118*, 4947–4957.

Alden, K., Timmis, J., Andrews, P.S., Veiga-Fernandes, H., and Coles, M.C. (2012a). Pairing experimentation and computational modeling to understand the role of tissue inducer cells in the development of lymphoid organs. Front. Immunol. *3*, 172.

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

Alden, K., Read, M., Timmis, J., Andrews, P.S., Veiga-Fernandes, H., and Coles, M. (2013). Spartan: a comprehensive tool for understanding uncertainty in simulations of biological systems. PLoS Comput. Biol. *9*, e1002916.

Alden, K., Andrews, P.S., Polack, F.A.C., Veiga-Fernandes, H., Coles, M.C., and Timmis, J. (2015). Using argument notation to engineer biological simulations with increased confidence. J. R. Soc. Interface *12*, 20141059.

Allen, C.D.C., Ansel, K.M., Low, C., Lesley, R., Tamamura, H., Fujii, N., and Cyster, J.G. (2004). Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. Nat. Immunol. *5*, 943–952.

Allen, C.D.C., Okada, T., and Cyster, J.G. (2007). Germinal-Center Organization and Cellular Dynamics. Immunity *27*, 190–202.

Allerheiligen, S.R.B. (2010). Next-Generation Model-Based Drug Discovery and Development: Quantitative and Systems Pharmacology. Clin. Pharmacol. Ther. *88*, 135–137.

Althaus, C.L., and Boer, R.J.D. (2008). Dynamics of Immune Escape during HIV/SIV Infection. PLOS Comput. Biol. *4*, e1000103.

Ansel, K.M., Ngo, V.N., Hyman, P.L., Luther, S.A., Förster, R., Sedgwick, J.D., Browning, J.L., Lipp, M., and Cyster, J.G. (2000). A chemokine-driven positive feedback loop organizes lymphoid follicles. Nature *406*, 309–314.

Antia, R., Bergstrom, C.T., Pilyugin, S.S., Kaech, S.M., and Ahmed, R. (2003). Models of CD8+ responses: 1. What is the antigen-independent proliferation program. J. Theor. Biol. *221*, 585–598.

Astarita, J.L., Cremasco, V., Fu, J., Darnell, M.C., Peck, J.R., Nieves-Bonilla, J.M., Song, K., Kondo, Y., Woodruff, M.C., Gogineni, A., et al. (2015). The CLEC-2-podoplanin axis controls the contractility of fibroblastic reticular cells and lymph node microarchitecture. Nat. Immunol. *16*, 75–84.

Athale, C.A., and Deisboeck, T.S. (2006). The effects of EGF-receptor density on multiscale tumor growth patterns. J. Theor. Biol. *238*, 771–779.

Axelsson, I. (1978). Characterization of proteins and other macromolecules by agarose gel chromatography. J. Chromatogr. A *152*, 21–32.

Baldwin, C.Y., and Clark, K.B. (2000). Design Rules, Vol. 1: The Power of Modularity (Cambridge, Mass: The MIT Press).

Barabási, A.-L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101–113.

Barinov, A., Luo, L., Gasse, P., Meas-Yedid, V., Donnadieu, E., Arenzana-Seisdedos, F., and Vieira, P. (2017). Essential role of immobilized chemokine CXCL12 in the regulation of the humoral immune response. Proc. Natl. Acad. Sci. *114*, 2319–2324.

Barmore, A.J., Castex, S.M., Gouletas, B.A., Griffith, A.J., Metz, S.W., Muelder, N.G., Populin, M.J., Sackett, D.M., Schuster, A.M., and Veldkamp, C.T. (2016). Transferring the C-terminus of the chemokine CCL21 to CCL19 confers enhanced heparin binding. Biochem. Biophys. Res. Commun. *477*, 602–606.

Barnes, A.L., Genever, P.G., Rimmer, S., and Coles, M.C. (2016). Collagen-Poly(N-isopropylacrylamide) Hydrogels with Tunable Properties. Biomacromolecules *17*, 723–734.

Barroso, R., Martínez Muñoz, L., Barrondo, S., Vega, B., Holgado, B.L., Lucas, P., Baíllo, A., Sallés, J., Rodríguez-Frade, J.M., and Mellado, M. (2012). EBI2 regulates CXCL13-mediated responses by heterodimerization with CXCR5. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. *26*, 4841–4854.

Basso, K., and Dalla-Favera, R. (2015). Germinal centres and B cell lymphomagenesis. Nat. Rev. Immunol. *15*, 172–184.

Batista, F.D., and Harwood, N.E. (2009). The who, how and where of antigen presentation to B cells. Nat. Rev. Immunol. *9*, 15–27.

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

Begley, C.G., and Ellis, L.M. (2012). Drug development: Raise standards for preclinical cancer research. Nature *483*, 531–533.

Bennett, L.D., Fox, J.M., and Signoret, N. (2011). Mechanisms regulating chemokine receptor activity. Immunology *134*, 246–256.

Bishop, C.M. (1996). Neural Networks for Pattern Recognition (Oxford : New York: Clarendon Press).

Bjørnstad, null, Ims, null, and Lambin, null (1999). Spatial population dynamics: analyzing patterns and processes of population synchrony. Trends Ecol. Evol. *14*, 427–432.

Bocharov, G., Danilov, A., Vassilevski, Y., Marchuk, G.I., Chereshnev, V.A., and Ludewig, B. (2011). Reaction-Diffusion Modelling of Interferon Distribution in Secondary Lymphoid Organs. Math. Model. Nat. Phenom. *6*, 13–26.

Bown, J., Andrews, P.S., Deeni, Y., Goltsov, A., Idowu, M., Polack, F.A.C., Sampson, A.T., Shovman, M., and Stepney, S. (2012). Engineering simulations for cancer systems biology. Curr. Drug Targets *13*, 1560–1574.

Breiman, L. (2001). Random Forests. Mach. Learn. *45*, 5–32.

Brosch, R., Pym, A.S., Gordon, S.V., and Cole, S.T. (2001). The evolution of mycobacterial pathogenicity: clues from comparative genomics. Trends Microbiol. *9*, 452–458.

Brown, K.S., Hill, C.C., Calero, G.A., Myers, C.R., Lee, K.H., Sethna, J.P., and Cerione, R.A. (2004). The statistical mechanics of complex signaling networks: nerve growth factor signaling. Phys. Biol. *1*, 184.

Butler, J., Sawtell, A., Jarrett, S., Cosgrove, J., Leigh, R., Timmis, J., and Coles, M. (2016a). Imaging Immunity in Lymph Nodes: Past, Present and Future. In Biophysics of Infection, M.C. Leake, ed. (Cham: Springer International Publishing), pp. 329–346.

Butler, J.A., Cosgrove, J., Alden, K., Timmis, J., and Coles, M.C. (2016b). Model-driven Experimentation: A new approach to understand mechanisms of tertiary lymphoid tissue formation, function and therapeutic resolution. Front. Immunol. *7*.

Calizo, R.C., and Scarlata, S. (2013). Discrepancy between fluorescence correlation spectroscopy and fluorescence recovery after photobleaching diffusion measurements of G-protein-coupled receptors. Anal. Biochem. *440*, 40–48.

Carrasco, Y.R., and Batista, F.D. (2007). B Cells Acquire Particulate Antigen in a Macrophage-Rich Area at the Boundary between the Follicle and the Subcapsular Sinus of the Lymph Node. Immunity *27*, 160–171.

Carrasco, Y.R., Fleire, S.J., Cameron, T., Dustin, M.L., and Batista, F.D. (2004). LFA-1/ICAM-1 interaction lowers the threshold of B cell activation by facilitating B cell adhesion and synapse formation. Immunity *20*, 589–599.

Chan, I.S., and Ginsburg, G.S. (2011). Personalized medicine: progress and promise. Annu. Rev. Genomics Hum. Genet. *12*, 217–244.

Chan, T.D., and Brink, R. (2012). Affinity-based selection and the germinal center response. Immunol. Rev. *247*, 11–23.

Chan, C., Billard, M., Ramirez, S.A., Schmidl, H., Monson, E., and Kepler, T.B. (2013). A model for migratory B cell oscillations from receptor down-regulation induced by external chemokine fields. Bull. Math. Biol. *75*, 185–205.

Chang, J.E., and Turley, S.J. (2015). Stromal infrastructure of the lymph node and coordination of immunity. Trends Immunol. *36*, 30–39.

Chen, N., Glazier, J.A., Izaguirre, J.A., and Alber, M.S. (2007). A parallel implementation of the Cellular Potts Model for simulation of cell-based morphogenesis. Comput. Phys. Commun. *176*, 670–681.

Coelho, F.C., Codeço, C.T., and Gomes, M.G.M. (2011). A Bayesian Framework for Parameter Estimation in Dynamical Models. PLoS ONE *6*, e19616.

Coelho, F.M., Natale, D., Soriano, S.F., Hons, M., Swoger, J., Mayer, J., Danuser, R., Scandella, E., Pieczyk, M., Zerwes, H.-G., et al. (2013). Naive B-cell trafficking is shaped by local chemokine availability and LFA-1–independent stromal interactions. Blood *121*, 4101–4109.

Cohen, J.N., Guidi, C.J., Tewalt, E.F., Qiao, H., Rouhani, S.J., Ruddell, A., Farr, A.G., Tung, K.S., and Engelhard, V.H. (2010). Lymph node–resident lymphatic endothelial cells mediate peripheral tolerance via Aire-independent direct antigen presentation. J. Exp. Med. *207*, 681–688.

Coles, M.C., Veiga-Fernandes, H., Foster, K.E., Norton, T., Pagakis, S.N., Seddon, B., and Kioussis, D. (2006). Role of T and NK cells and IL7/IL7r interactions during neonatal maturation of lymph nodes. Proc. Natl. Acad. Sci. U. S. A. *103*, 13457–13462.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. Mach. Learn. *20*, 273–297.

Cosgrove, J., Butler, J., Alden, K., Read, M., Kumar, V., Cucurull-Sanchez, L., Timmis, J., and Coles, M. (2015). Agent-Based Modeling in Systems Pharmacology. CPT Pharmacomet. Syst. Pharmacol. n/a-n/a.

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. Science *353*, aaf1420.

Cremasco, V., Woodruff, M.C., Onder, L., Cupovic, J., Nieves-Bonilla, J.M., Schildberg, F.A., Chang, J., Cremasco, F., Harvey, C.J., Wucherpfennig, K., et al. (2014). B cell homeostasis and follicle confines are governed by fibroblastic reticular cells. Nat. Immunol. *15*, 973–981.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal *Complex Systems*, 1695.

Cyster, J.G. (2005). Chemokines, Sphingosine-1-Phosphate, and Cell Migration in Secondary Lymphoid Organs. Annu. Rev. Immunol. *23*, 127–159.

Cyster, J.G., Dang, E.V., Reboldi, A., and Yi, T. (2014). 25-Hydroxycholesterols in innate and adaptive immunity. Nat. Rev. Immunol. *14*, 731–743.

Deb, K., and Kalyanmoy, D. (2001). Multi-Objective Optimization Using Evolutionary Algorithms (New York, NY, USA: John Wiley & Sons, Inc.).

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. *6*, 182–197.

Di Niro, R., Lee, S.-J., Vander Heiden, J.A., Elsner, R.A., Trivedi, N., Bannock, J.M., Gupta, N.T., Kleinstein, S.H., Vigneault, F., Gilbert, T.J., et al. (2015). Salmonella Infection Drives Promiscuous B Cell Activation Followed by Extrafollicular Affinity Maturation. Immunity *43*, 120–131.

Dong, M., Martinez, M.M., Mayer, M.F., and Pappas, D. (2012). Single molecule fluorescence correlation spectroscopy of single apoptotic cells using a red-fluorescent caspase probe. The Analyst *137*, 2997–3003.

Donovan, G.M., and Lythe, G. (2012). T-cell movement on the reticular network. J. Theor. Biol. *295*, 59–67.

Druzd, D., Matveeva, O., Ince, L., Harrison, U., He, W., Schmal, C., Herzel, H., Tsang, A.H., Kawakami, N., Leliavski, A., et al. (2017). Lymphocyte Circadian Clocks Control Lymph Node Trafficking and Adaptive Immune Responses. Immunity *0*.

Dyer, D.P., Salanga, C.L., Volkman, B.F., Kawamura, T., and Handel, T.M. (2016). The dependence of chemokine–glycosaminoglycan interactions on chemokine oligomerization. Glycobiology *26*, 312–326.

Eden, E., Geva-Zatorsky, N., Issaeva, I., Cohen, A., Dekel, E., Danon, T., Cohen, L., Mayo, A., and Alon, U. (2011). Proteome half-life dynamics in living human cells. Science *331*, 764–768.

EFPIA MID3 Workgroup, Marshall, S., Burghaus, R., Cosson, V., Cheung, S., Chenel, M., DellaPasqua, O., Frey, N., Hamrén, B., Harnisch, L., et al. (2016). Good Practices in Model-Informed Drug Discovery and Development: Practice, Application, and Documentation. CPT Pharmacomet. Syst. Pharmacol. *5*, 93–122.

El Shikh, M.E.M., El Sayed, R.M., Sukumar, S., Szakal, A.K., and Tew, J.G. (2010). Activation of B cells by antigens on follicular dendritic cells. Trends Immunol. *31*, 205–211.

Elemans, M., Basatena, N.-K.S. al, Klatt, N.R., Gkekas, C., Silvestri, G., and Asquith, B. (2011). Why Don't CD8+ T Cells Reduce the Lifespan of SIV-Infected Cells In Vivo? PLOS Comput. Biol. *7*, e1002200.

Erdős, P., and Rényi, A. (1960). On the Evolution of Random Graphs. In Publication of the Mathematical Institute of the Hungarian Academy of Sciences, pp. 17–61.

Erika M. Adkins, ‡, Devadoss J. Samuvel, §, Jacob U. Fog, ‡, Jacob Eriksen, ‡, Lankupalle D. Jayanthi, §, Christian Bjerggaard Vaegter, ‡, Sammanda Ramamoorthy,

,§ and, and Ulrik Gether, ‡ (2007). Membrane Mobility and Microdomain Association of the Dopamine Transporter Studied with Fluorescence Correlation Spectroscopy and Fluorescence Recovery after Photobleaching†.

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature *542*, 115–118.

Evans, S., Alden, K., Cucurull-Sanchez, L., Larminie, C., Coles, M.C., Kullberg, M.C., and Timmis, J. (2017). ASPASIA: A toolkit for evaluating the effects of biological interventions on SBML model behaviour. PLOS Comput. Biol. *13*, e1005351.

Farah, M., Birrell, P., Conti, S., and Angelis, D.D. (2014). Bayesian Emulation and Calibration of a Dynamic Epidemic Model for A/H1N1 Influenza. J. Am. Stat. Assoc. *109*, 1398–1411.

Fernald, G.H., Capriotti, E., Daneshjou, R., Karczewski, K.J., and Altman, R.B. (2011). Bioinformatics challenges for personalized medicine. Bioinformatics *27*, 1741–1748.

Figge, M.T., Garin, A., Gunzer, M., Kosco-Vilbois, M., Toellner, K.-M., and MEYER-HERMANN, M. (2008). Deriving a germinal center lymphocyte migration model from two-photon data. J. Exp. Med. *205*, 3019–3029.

Finke, D., Acha-Orbea, H., Mattis, A., Lipp, M., and Kraehenbuhl, J. (2002). CD4+CD3- cells induce Peyer's patch development: role of alpha4beta1 integrin activation by CXCR5. Immunity *17*, 363–373.

Fletcher, A.L., Malhotra, D., and Turley, S.J. (2011a). Lymph node stroma broaden the peripheral tolerance paradigm. Trends Immunol. *32*, 12–18.

Fletcher, A.L., Malhotra, D., Acton, S.E., Lukacs-Kornek, V., Bellemare-Pelletier, A., Curry, M., Armant, M., and Turley, S.J. (2011b). Reproducible Isolation of Lymph Node Stromal Cells Reveals Site-Dependent Differences in Fibroblastic Reticular Cells. Front. Immunol. *2*.

Fletcher, A.L., Acton, S.E., and Knoblich, K. (2015). Lymph node fibroblastic reticular cells in health and disease. Nat. Rev. Immunol. *15*, 350–361.

Fleury, M.E., Boardman, K.C., and Swartz, M.A. (2006). Autologous Morphogen Gradients by Subtle Interstitial Flow and Matrix Interactions. Biophys. J. *91*, 113–121.

Förster, R., Mattis, A.E., Kremmer, E., Wolf, E., Brem, G., and Lipp, M. (1996a). A putative chemokine receptor, BLR1, directs B cell migration to defined lymphoid organs and specific anatomic compartments of the spleen. Cell *87*, 1037–1047.

Förster, R., Mattis, A.E., Kremmer, E., Wolf, E., Brem, G., and Lipp, M. (1996b). A Putative Chemokine Receptor, BLR1, Directs B Cell Migration to Defined Lymphoid Organs and Specific Anatomic Compartments of the Spleen. Cell *87*, 1037–1047.

Förster, R., Schubel, A., Breitfeld, D., Kremmer, E., Renner-Müller, I., Wolf, E., and Lipp, M. (1999). CCR7 coordinates the primary immune response by establishing functional microenvironments in secondary lymphoid organs. Cell *99*, 23–33.

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

Garin, A., Meyer-Hermann, M., Contie, M., Figge, M.T., Buatois, V., Gunzer, M., Toellner, K.-M., Elson, G., and Kosco-Vilbois, M.H. (2010). Toll-like Receptor 4 Signaling by Follicular Dendritic Cells Is Pivotal for Germinal Center Onset and Affinity Maturation. Immunity *33*, 84–95.

Gatto, D., and Brink, R. (2013). B cell localization: regulation by EBI2 and its oxysterol ligand. Trends Immunol. *34*, 336–341.

Gatto, D., Paus, D., Basten, A., Mackay, C.R., and Brink, R. (2009). Guidance of B cells by the orphan G protein-coupled receptor EBI2 shapes humoral immune responses. Immunity *31*, 259–269.

Gilks, W.R., Richardson, S., and Spiegelhalter, D. (1995). Markov Chain Monte Carlo in Practice (CRC Press).

Gong, C., Mattila, J.T., Miller, M., Flynn, J.L., Linderman, J.J., and Kirschner, D. (2013). Predicting lymph node output efficiency using systems biology. J. Theor. Biol. *335*, 169–184.

Grajdeanu, A., and Grajdeanu, A. MODELING DIFFUSION IN A DISCRETE ENVIRONMENT.

Green, J.A., Suzuki, K., Cho, B., Willison, L.D., Palmer, D., Allen, C.D.C., Schmidt, T.H., Xu, Y., Proia, R.L., Coughlin, S.R., et al. (2011). The sphingosine 1-phosphate receptor $S1P_2$ maintains the homeostasis of germinal center B cells and promotes niche confinement. Nat. Immunol. *12*, 672.

Grigorova, I.L., Panteleev, M., and Cyster, J.G. (2010). Lymph node cortical sinus organization and relationship to lymphocyte egress dynamics and antigen exposure. Proc. Natl. Acad. Sci. *107*, 20447–20452.

Gross, J.A., Johnston, J., Mudri, S., Enselman, R., Dillon, S.R., Madden, K., Xu, W., Parrish-Novak, J., Foster, D., Lofton-Day, C., et al. (2000). TACI and BCMA are receptors for a TNF homologue implicated in B-cell autoimmune disease. Nature *404*, 995–999.

Gryczynski, Z. (Karol) (2008). FCS Imaging—A Way to Look at Cellular Processes. Biophys. J. *94*, 1943–1944.

Gunn, M.D., Ngo, V.N., Ansel, K.M., Ekland, E.H., Cyster, J.G., and Williams, L.T. (1998a). A B-cell-homing chemokine made in lymphoid follicles activates Burkitt's lymphoma receptor-1. Nature *391*, 799–803.

Gunn, M.D., Tangemann, K., Tam, C., Cyster, J.G., Rosen, S.D., and Williams, L.T. (1998b). A chemokine expressed in lymphoid high endothelial venules promotes the adhesion and chemotaxis of naive T lymphocytes. Proc. Natl. Acad. Sci. *95*, 258–263.

Guo, Z., and Tay, J.C. (2008). Granularity and the Validation of Agent-based Models. In Proceedings of the 2008 Spring Simulation Multiconference, (San Diego, CA, USA: Society for Computer Simulation International), pp. 153–161.

Guo, L., Har, J.Y., Sankaran, J., Hong, Y., Kannan, B., and Wohland, T. (2008a). Molecular Diffusion Measurement in Lipid Bilayers over Wide Concentration Ranges: A Comparative Study. ChemPhysChem *9*, 721–728.

Guo, Z., Sloot, P.M.A., and Tay, J.C. (2008b). A hybrid agent-based approach for modeling microbiological systems. J. Theor. Biol. *255*, 163–175.

Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., and Sethna, J.P. (2007). Universally Sloppy Parameter Sensitivities in Systems Biology Models. PLoS Comput. Biol. *3*.

Handel, T.M., Johnson, Z., Crown, S.E., Lau, E.K., Sweeney, M., and Proudfoot, A.E. (2005). Regulation of Protein Function by Glycosaminoglycans—as Exemplified by Chemokines. Annu. Rev. Biochem. *74*, 385–410.

Hasan, M., Najjam, S., Gordon, M.Y., Gibbs, R.V., and Rider, C.C. (1999). IL-12 is a heparin-binding cytokine. J. Immunol. Baltim. Md 1950 *162*, 1064–1070.

He, C., Young, A.J., West, C.A., Su, M., Konerding, M.A., and Mentzer, S.J. (2002). Stimulation of regional lymphatic and blood flow by epicutaneous oxazolone. J. Appl. Physiol. Bethesda Md 1985 *93*, 966–973.

Heesters, B.A., Chatterjee, P., Kim, Y.-A., Gonzalez, S.F., Kuligowski, M.P., Kirchhausen, T., and Carroll, M.C. (2013). Endocytosis and Recycling of Immune Complexes by Follicular Dendritic Cells Enhances B Cell Antigen Binding and Activation. Immunity *38*, 1164–1175.

Helton, J.C. (2008). Uncertainty and Sensitivity Analysis for Models of Complex Systems. In Computational Methods in Transport: Verification and Validation, F. Graziani, ed. (Springer Berlin Heidelberg), pp. 207–228.

Henderson, D.A., Boys, R.J., Krishnan, K.J., Lawless, C., and Wilkinson, D.J. (2009). Bayesian Emulation and Calibration of a Stochastic Computer Model of Mitochondrial DNA Deletions in Substantia Nigra Neurons. J. Am. Stat. Assoc. *104*, 76–87.

Herzmark, P., Campbell, K., Wang, F., Wong, K., El-Samad, H., Groisman, A., and Bourne, H.R. (2007). Bound attractant at the leading vs. the trailing edge determines chemotactic prowess. Proc. Natl. Acad. Sci. U. S. A. *104*, 13349–13354.

Hess, E., Duheron, V., Decossas, M., Lézot, F., Berdal, A., Chea, S., Golub, R., Bosisio, M.R., Bridal, S.L., Choi, Y., et al. (2012). RANKL induces organized lymph node growth by stromal cell proliferation. J. Immunol. Baltim. Md 1950 *188*, 1245–1254.

Hetherington, S., Hughes, A.R., Mosteller, M., Shortino, D., Baker, K.L., Spreen, W., Lai, E., Davies, K., Handley, A., Dow, D.J., et al. (2002). Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. Lancet Lond. Engl. *359*, 1121–1122.

Hirose, J., Kawashima, H., Swope Willis, M., Springer, T.A., Hasegawa, H., Yoshie, O., and Miyasaka, M. (2002). Chondroitin sulfate B exerts its inhibitory effect on secondary lymphoid tissue chemokine (SLC) by binding to the C-terminus of SLC. Biochim. Biophys. Acta *1571*, 219–224.

Huff, J. (2015). The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution. Nat. Methods *12*.

Humphries, M.D., and Gurney, K. (2008). Network "small-world-ness": a quantitative method for determining canonical network equivalence. PloS One *3*, e0002051.

Hundelshausen, P. von, Agten, S.M., Eckardt, V., Blanchet, X., Schmitt, M.M., Ippel, H., Neideck, C., Bidzhekov, K., Leberzammer, J., Wichapong, K., et al. (2017). Chemokine interactome mapping enables tailored intervention in acute and chronic inflammation. Sci. Transl. Med. *9*, eaah6650.

Inza, I., Calvo, B., Armañanzas, R., Bengoetxea, E., Larrañaga, P., and Lozano, J.A. (2010). Machine learning: an indispensable tool in bioinformatics. Methods Mol. Biol. Clifton NJ *593*, 25–48.

Irla, M., Guenot, J., Sealy, G., Reith, W., Imhof, B.A., and Sergé, A. (2013). Three-dimensional visualization of the mouse thymus organization in health and immunodeficiency. J. Immunol. Baltim. Md 1950 *190*, 586–596.

Jafarnejad, M., Woodruff, M.C., Zawieja, D.C., Carroll, M.C., and Moore, J.E. (2015). Modeling Lymph Flow and Fluid Exchange with Blood Vessels in Lymph Nodes. Lymphat. Res. Biol. *13*, 234–247.

James Butler, Jon Timmis, and Mark Coles (2014). Novel Approaches to the Visualization and Quantification of Biological Simulations by Emulating Experimental Techniques. ALIFE 14 Fourteenth Conf. Synth. Simul. Living Syst. *14*.

Jarjour, M., Jorquera, A., Mondor, I., Wienert, S., Narang, P., Coles, M.C., Klauschen, F., and Bajénoff, M. (2014). Fate mapping reveals origin and dynamics of lymph node follicular dendritic cells. J. Exp. Med. *211*, 1109–1122.

Johnson, C.G., Goldman, J.P., and Gullick, W.J. (2004). Simulating complex intracellular processes using object-oriented computational modelling. Prog. Biophys. Mol. Biol. *86*, 379–406.

Jr, C.A.J., Travers, P., Walport, M., Shlomchik, M.J., Jr, C.A.J., Travers, P., Walport, M., and Shlomchik, M.J. (2001). Immunobiology (Garland Science).

Junt, T., Scandella, E., and Ludewig, B. (2008). Form follows function: lymphoid tissue microarchitecture in antimicrobial immune defence. Nat. Rev. Immunol. *8*, 764–775.

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

Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Assad-Garcia, N., Glass, J.I., and Covert, M.W. (2012b). A whole-cell computational model predicts phenotype from genotype. Cell *150*, 389–401.

Kasturi, S.P., Skountzou, I., Albrecht, R.A., Koutsonanos, D., Hua, T., Nakaya, H.I., Ravindran, R., Stewart, S., Alam, M., Kwissa, M., et al. (2011). Programming the magnitude and persistence of antibody responses with innate immunity. Nature *470*, 543–547.

Katakai, T. (2012). Marginal reticular cells: a stromal subset directly descended from the lymphoid tissue organizer. Front. Immunol. *3*, 200.

Kazmi, N., Hossain, M.A., and Phillips, R.M. (2012). A hybrid cellular automaton model of solid tumor growth and bioreductive drug transport. IEEEACM Trans. Comput. Biol. Bioinforma. IEEE ACM *9*, 1595–1606.

Kelly, L.M., Pereira, J.P., Yi, T., Xu, Y., and Cyster, J.G. (2011). EBI2 guides serial movements of activated B cells and ligand activity is detectable in lymphoid and nonlymphoid tissues. J. Immunol. Baltim. Md 1950 *187*, 3026–3032.

Kennedy, M.C., and O'Hagan, A. (2001). Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B Stat. Methodol. *63*, 425–464.

Kirschner, D.E., Hunt, C.A., Marino, S., Fallahi-Sichani, M., and Linderman, J.J. (2014). Tuneable resolution as a systems biology approach for multi-scale, multi-compartment computational models. Wiley Interdiscip. Rev. Syst. Biol. Med. *6*, 289–309.

Kislitsyn, A., Savinkov, R., Novkovic, M., Onder, L., and Bocharov, G. (2015). Computational Approach to 3D Modeling of the Lymph Node Geometry. Computation *3*, 222–234.

Kitano, H. (2002). Systems Biology: A Brief Overview. Science *295*, 1662–1664.

Kitano, M., Moriyama, S., Ando, Y., Hikida, M., Mori, Y., Kurosaki, T., and Okada, T. (2011). Bcl6 protein expression shapes pre-germinal center B cell dynamics and follicular helper T cell heterogeneity. Immunity *34*, 961–972.

Kohout, T.A., Nicholas, S.L., Perry, S.J., Reinhart, G., Junger, S., and Struthers, R.S. (2004). Differential desensitization, receptor phosphorylation, beta-arrestin recruitment, and ERK1/2 activation by the two endogenous ligands for the CC chemokine receptor 7. J. Biol. Chem. *279*, 23214–23222.

Kondo, S., and Miura, T. (2010). Reaction-Diffusion Model as a Framework for Understanding Biological Pattern Formation. Science *329*, 1616–1620.

Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D., and Vajda, S. (2017). The ClusPro web server for protein-protein docking. Nat. Protoc. *12*, 255–278.

Kumar, V., Scandella, E., Danuser, R., Onder, L., Nitschké, M., Fukui, Y., Halin, C., Ludewig, B., and Stein, J.V. (2010). Global lymphoid tissue remodeling during a viral infection is orchestrated by a B cell–lymphotoxin-dependent pathway. Blood *115*, 4725–4733.

Lalonde, R.L., Kowalski, K.G., Hutmacher, M.M., Ewy, W., Nichols, D.J., Milligan, P.A., Corrigan, B.W., Lockwood, P.A., Marshall, S.A., Benincosa, L.J., et al. (2007). Model-based Drug Development. Clin. Pharmacol. Ther. *82*, 21–32.

Legler, D.F., Loetscher, M., Roos, R.S., Clark-Lewis, I., Baggiolini, M., and Moser, B. (1998). B cell-attracting chemokine 1, a human CXC chemokine expressed in lymphoid tissues, selectively attracts B lymphocytes via BLR1/CXCR5. J. Exp. Med. *187*, 655–660.

Li, N.Y.K., Verdolini, K., Clermont, G., Mi, Q., Rubinstein, E.N., Hebda, P.A., and Vodovotz, Y. (2008). A Patient-Specific in silico Model of Inflammation and Healing Tested in Acute Vocal Fold Injury. PLoS ONE *3*.

Lillemeier, B.F., Mörtelmaier, M.A., Forstner, M.B., Huppa, J.B., Groves, J.T., and Davis, M.M. (2010). TCR and Lat are expressed on separate protein islands on T cell membranes and concatenate during activation. Nat. Immunol. *11*, 90–96.

Lin, F., and Butcher, E.C. (2008). Modeling the Role of Homologous Receptor Desensitization in Cell Gradient Sensing. J. Immunol. Baltim. Md 1950 *181*, 8335–8343.

Link, A., Vogt, T.K., Favre, S., Britschgi, M.R., Acha-Orbea, H., Hinz, B., Cyster, J.G., and Luther, S.A. (2007). Fibroblastic reticular cells in lymph nodes regulate the homeostasis of naive T cells. Nat. Immunol. *8*, 1255–1265.

Liu, C., Yang, X.V., Wu, J., Kuei, C., Mani, N.S., Zhang, L., Yu, J., Sutton, S.W., Qin, N., Banie, H., et al. (2011). Oxysterols direct B-cell migration through EBI2. Nature *475*, 519–523.

Luke, S., Cioffi-Revilla, C., Panait, L., Sullivan, K., and Balan, G. (2005). MASON: A Multiagent Simulation Environment. SIMULATION *81*, 517–527.

Luther, S.A., Bidgol, A., Hargreaves, D.C., Schmidt, A., Xu, Y., Paniyadi, J., Matloubian, M., and Cyster, J.G. (2002). Differing Activities of Homeostatic Chemokines CCL19, CCL21, and CXCL12 in Lymphocyte and Dendritic Cell Recruitment and Lymphoid Neogenesis. J. Immunol. *169*, 424–433.

Ma, B., Jablonska, J., Lindenmaier, W., and Dittmar, K.E.J. (2007). Immunohistochemical study of the reticular and vascular network of mouse lymph node using vibratome sections. Acta Histochem. *109*, 15–28.

Macal, C.M., and North, M.J. (2010). Tutorial on agent-based modelling and simulation. J. Simul. *4*, 151–162.

Macháň, R., Foo, Y.H., and Wohland, T. (2016). On the Equivalence of FCS and FRAP: Simultaneous Lipid Membrane Measurements. Biophys. J. *111*, 152–161.

Maiuri, P., Rupprecht, J.-F., Wieser, S., Ruprecht, V., Bénichou, O., Carpi, N., Coppey, M., De Beco, S., Gov, N., Heisenberg, C.-P., et al. (2015). Actin Flows Mediate a Universal Coupling between Cell Speed and Cell Persistence. Cell *161*, 374–386.

Malhotra, D., Fletcher, A.L., Astarita, J., Lukacs-Kornek, V., Tayalia, P., Gonzalez, S.F., Elpek, K.G., Chang, S.K., Knoblich, K., Hemler, M.E., et al. (2012). Transcriptional profiling of stroma from inflamed and resting lymph nodes defines immunological hallmarks. Nat. Immunol. *13*, 499–510.

Mandel, T.E., Phipps, R.P., Abbot, A.P., and Tew, J.G. (1981). Long-term antigen retention by dendritic cells in the popliteal lymph node of immunized mice. Immunology *43*, 353–362.

Marino, S., Hogue, I.B., Ray, C.J., and Kirschner, D.E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. J. Theor. Biol. *254*, 178–196.

Matloubian, M., Lo, C.G., Cinamon, G., Lesneski, M.J., Xu, Y., Brinkmann, V., Allende, M.L., Proia, R.L., and Cyster, J.G. (2004). Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. Nature *427*, 355–360.

Matsumoto, N., Koike, K., Yamada, S., and Staub, N.C. (1990). Caudal mediastinal node lymph flow in sheep after histamine or endotoxin infusions. Am. J. Physiol. *258*, H24-28.

McKay, M.D., Beckman, R.J., and Conover, W.J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics *21*, 239–245.

Meinhardt, H., and Gierer, A. (2000). Pattern formation by local self-activation and lateral inhibition. BioEssays News Rev. Mol. Cell. Dev. Biol. *22*, 753–760.

Meyer-Hermann, M.E. (2006). An analysis of B cell selection mechanisms in germinal centers. Math. Med. Biol. *23*, 255–277.

Meyer-Hermann, M.E., and Maini, P.K. (2005). Interpreting two-photon imaging data of lymphocyte motility. Phys. Rev. E *71*, 61912.

Meyer-Hermann, M., Mohr, E., Pelletier, N., Zhang, Y., Victora, G.D., and Toellner, K.-M. (2012). A Theory of Germinal Center B Cell Selection, Division, and Exit. Cell Rep. *2*, 162–174.

Miller, H., Zhou, Z., Wollman, A.J.M., and Leake, M.C. (2015a). Superresolution imaging of single DNA molecules using stochastic photoblinking of minor groove and intercalating dyes. Methods San Diego Calif *88*, 81–88.

Miller, H., Zhou, Z., Wollman, A.J.M., and Leake, M.C. (2015b). Superresolution imaging of single DNA molecules using stochastic photoblinking of minor groove and intercalating dyes. Methods *88*, 81–88.

Milligan, P.A., Brown, M.J., Marchant, B., Martin, S.W., van der Graaf, P.H., Benson, N., Nucci, G., Nichols, D.J., Boyd, R.A., Mandema, J.W., et al. (2013). Model-based drug development: a rational approach to efficiently accelerate drug development. Clin. Pharmacol. Ther. *93*, 502–514.

Milo, R., and Phillips, R. (2015). Cell Biology by the Numbers (New York, NY: Garland Science).

Mirsky, H.P., Miller, M.J., Linderman, J.J., and Kirschner, D.E. (2011). Systems biology approaches for understanding cellular mechanisms of immunity in lymph nodes during infection. J. Theor. Biol. *287*, 160–170.

Monneau, Y., Arenzana-Seisdedos, F., and Lortat-Jacob, H. (2016). The sweet spot: how GAGs help chemokines guide migrating cells. J. Leukoc. Biol. *99*, 935–953.

Monneau, Y.R., Luo, L., Sankaranarayanan, N.V., Nagarajan, B., Vivès, R.R., Baleux, F., Desai, U.R., Arenzana-Seidedos, F., and Lortat-Jacob, H. (2017). Solution structure of CXCL13 and heparan sulfate binding show that GAG binding site and cellular signalling rely on distinct domains. Open Biol. *7*.

Monroe, J.G., and Cambier, J.C. (1983). Sorting of B lymphoblasts based upon cell diameter provides cell populations enriched in different stages of cell cycle. J. Immunol. Methods *63*, 45–56.

Moore, P.A., Belvedere, O., Orr, A., Pieri, K., LaFleur, D.W., Feng, P., Soppet, D., Charters, M., Gentz, R., Parmelee, D., et al. (1999). BLyS: Member of the Tumor Necrosis Factor Family and B Lymphocyte Stimulator. Science *285*, 260–263.

Moran, P.A.P. (1950). Notes on Continuous Stochastic Phenomena. Biometrika *37*, 17–23.

Mori, S., Nakano, H., Aritomi, K., Wang, C.-R., Gunn, M.D., and Kakiuchi, T. (2001). Mice Lacking Expression of the Chemokines Ccl21-Ser and Ccl19 (plt Mice) Demonstrate Delayed but Enhanced T Cell Immune Responses. J. Exp. Med. *193*, 207–218.

Mortier, A., Van Damme, J., and Proost, P. (2008). Regulation of chemokine activity by posttranslational modification. Pharmacol. Ther. *120*, 197–217.

Mortier, A., Van Damme, J., and Proost, P. (2012). Overview of the mechanisms regulating chemokine activity and availability. Immunol. Lett. *145*, 2–9.

Muppidi, J.R., Schmitz, R., Green, J.A., Xiao, W., Larsen, A.B., Braun, S.E., An, J., Xu, Y., Rosenwald, A., Ott, G., et al. (2014). Loss of signaling via Gα13 in germinal center B cell-derived lymphoma. Nature *516*, 254–258.

Muppidi, J.R., Lu, E., and Cyster, J.G. (2015). The G protein–coupled receptor P2RY8 and follicular dendritic cells promote germinal center confinement of B cells, whereas S1PR3 can contribute to their dissemination. J. Exp. Med. *212*, 2213–2222.

Murphy, K.M. (2011). Janeway's Immunobiology (Garland Science).

Natkanski, E., Lee, W.-Y., Mistry, B., Casal, A., Molloy, J.E., and Tolar, P. (2013). B Cells Use Mechanical Energy to Discriminate Antigen Affinities. Science *340*, 1587–1590.

Nibbs, R.J.B., and Graham, G.J. (2013). Immune regulation by atypical chemokine receptors. Nat. Rev. Immunol. *13*, 815–829.

Novkovic, M., Onder, L., Cupovic, J., Abe, J., Bomze, D., Cremasco, V., Scandella, E., Stein, J.V., Bocharov, G., Turley, S.J., et al. (2016). Topological Small-World Organization of the Fibroblastic Reticular Cell Network Determines Lymph Node Functionality. PLOS Biol *14*, e1002515.

Nowosad, C.R., Spillane, K.M., and Tolar, P. (2016). Germinal center B cells recognize antigen through a specialized immune synapse architecture. Nat. Immunol. *17*, 870–877.

Onder, L., Mörbe, U., Pikor, N., Novkovic, M., Cheng, H.-W., Hehlgans, T., Pfeffer, K., Becher, B., Waisman, A., Rülicke, T., et al. (2017). Lymphatic Endothelial Cells Control Initiation of Lymph Node Organogenesis. Immunity *47*, 80–92.e4.

Otero, C., Groettrup, M., and Legler, D.F. (2006). Opposite fate of endocytosed CCR7 and its ligands: recycling versus degradation. J. Immunol. Baltim. Md 1950 *177*, 2314–2323.

Oyler-Yaniv, A., Oyler-Yaniv, J., Whitlock, B.M., Liu, Z., Germain, R.N., Huse, M., Altan-Bonnet, G., and Krichevsky, O. (2017). A Tunable Diffusion-Consumption Mechanism of Cytokine Propagation Enables Plasticity in Cell-to-Cell Communication in the Immune System. Immunity *46*, 609–620.

P.Andrews J. Timmis, F. Polack, S. Stepney, and A. Sampson (2010). The CoSMoS Process, Version 0.1: A Process for the Modelling and Simulation of Complex Systems.

Park, C., Hwang, I.-Y., Sinha, R.K., Kamenyeva, O., Davis, M.D., and Kehrl, J.H. (2012). Lymph node B lymphocyte trafficking is constrained by anatomy and highly dependent upon chemoattractant desensitization. Blood *119*, 978–989.

Passini, E., Britton, O.J., Lu, H.R., Rohrbacher, J., Hermans, A.N., Gallacher, D.J., Greig, R.J.H., Bueno-Orovio, A., and Rodriguez, B. (2017). Human In Silico Drug Trials Demonstrate Higher Accuracy than Animal Models in Predicting Clinical Pro-Arrhythmic Cardiotoxicity. Front. Physiol. *8*.

Patel, A., Harker, N., Moreira-Santos, L., Ferreira, M., Alden, K., Timmis, J., Foster, K., Garefalaki, A., Pachnis, P., Andrews, P., et al. (2012). Differential RET signaling pathways drive development of the enteric lymphoid and nervous systems. Sci. Signal. *5*, ra55.

Patel, D.D., Koopmann, W., Imai, T., Whichard, L.P., Yoshie, O., and Krangel, M.S. (2001). Chemokines have diverse abilities to form solid phase gradients. Clin. Immunol. Orlando Fla *99*, 43–52.

van de Pavert, S.A., and Mebius, R.E. (2010). New insights into the development of lymphoid tissues. Nat. Rev. Immunol. *10*, 664–674.

de Paz, J.L., Moseman, E.A., Noti, C., Polito, L., von Andrian, U.H., and Seeberger, P.H. (2007). Profiling heparin-chemokine interactions using synthetic tools. ACS Chem. Biol. *2*, 735–744.

Pellegrini, L., Burke, D.F., Von, D., Mulloy, B., and Blundell, T.L. (2000). Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. Nature *407*, 1029–1034.

Pereira, J.P., Kelly, L.M., Xu, Y., and Cyster, J.G. (2009). EBI2 mediates B cell segregation between the outer and centre follicle. Nature *460*, 1122–1126.

Pereira, J.P., Kelly, L.M., and Cyster, J.G. (2010). Finding the right niche: B-cell migration in the early phases of T-dependent antibody responses. Int. Immunol. *22*, 413–419.

Perelson, A.S., and Ribeiro, R.M. (2013). Modeling the within-host dynamics of HIV infection. BMC Biol. *11*, 96.

Peterson, M., and Riggs, M. (2015). FDA Advisory Meeting Clinical Pharmacology Review Utilizes a Quantitative Systems Pharmacology (QSP) Model: A Watershed Moment? CPT Pharmacomet. Syst. Pharmacol. *4*, 189–192.

Phair, R.D., and Misteli, T. (2001a). Kinetic modelling approaches to in vivo imaging. Nat. Rev. Mol. Cell Biol. *2*, 898–907.

Phair, R.D., and Misteli, T. (2001b). Kinetic modelling approaches to in vivo imaging. Nat. Rev. Mol. Cell Biol. *2*, 898–907.

Phan, T.G., Grigorova, I., Okada, T., and Cyster, J.G. (2007). Subcapsular encounter and complement-dependent transport of immune complexes by lymph node B cells. Nat. Immunol. *8*, 992–1000.

Phan, T.G., Green, J.A., Gray, E.E., Xu, Y., and Cyster, J.G. (2009). Immune complex relay by subcapsular sinus macrophages and non-cognate B cells drives antibody affinity maturation. Nat. Immunol. *10*, 786–793.

Pitt-Francis, J., Bernabeu, M.O., Cooper, J., Garny, A., Momtahan, L., Osborne, J., Pathmanathan, P., Rodriguez, B., Whiteley, J.P., and Gavaghan, D.J. (2008). Chaste: using agile programming techniques to develop computational biology software. Philos. Trans. R. Soc. Lond. Math. Phys. Eng. Sci. *366*, 3111–3136.

Pitt-Francis, J., Pathmanathan, P., Bernabeu, M.O., Bordas, R., Cooper, J., Fletcher, A.G., Mirams, G.R., Murray, P., Osborne, J.M., Walter, A., et al. (2009). Chaste: A test-driven

approach to software development for biological modelling. Comput. Phys. Commun. *180*, 2452–2471.

Plank, M., Wadhams, G.H., and Leake, M.C. (2009). Millisecond timescale slimfield imaging and automated quantification of single fluorescent protein molecules for use in probing complex biological processes. Integr. Biol. *1*, 602–612.

Poole, D., and Raftery, A.E. (2000). Inference for Deterministic Simulation Models: The Bayesian Melding Approach. J. Am. Stat. Assoc. *95*, 1244–1255.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (2007). Numerical Recipes 3rd Edition: The Art of Scientific Computing (Cambridge, UK ; New York: Cambridge University Press).

Pruett, W.A., and Hester, R.L. (2016). The Creation of Surrogate Models for Fast Estimation of Complex Model Outcomes. PLOS ONE *11*, e0156574.

Purnick, P.E.M., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. Nat. Rev. Mol. Cell Biol. *10*, 410–422.

Putnam, F.W. (Frank W. (1975a). The Plasma proteins : structure, function, and genetic control / edited by Frank W. Putnam (New York: Academic Press).

Putnam, F.W. (Frank W. (1975b). The Plasma proteins : structure, function, and genetic control / edited by Frank W. Putnam (New York: Academic Press).

Qi, H., Kastenmüller, W., and Germain, R.N. (2014). Spatiotemporal Basis of Innate and Adaptive Immunity in Secondary Lymphoid Tissue. Annu. Rev. Cell Dev. Biol. *30*, 141–167.

Qian, H., and Elson, E.L. (1990). Distribution of molecular aggregation by analysis of fluctuation moments. Proc. Natl. Acad. Sci. U. S. A. *87*, 5479–5483.

Rahman, Z.S.M., Rao, S.P., Kalled, S.L., and Manser, T. (2003). Normal induction but attenuated progression of germinal center responses in BAFF and BAFF-R signaling-deficient mice. J. Exp. Med. *198*, 1157–1169.

Randall, T.D., Carragher, D.M., and Rangel-Moreno, J. (2008). Development of secondary lymphoid organs. Annu. Rev. Immunol. *26*, 627–650.

Rantakari, P., Auvinen, K., Jäppinen, N., Kapraali, M., Valtonen, J., Karikoski, M., Gerke, H., Iftakhar-E-Khuda, I., Keuschnigg, J., Umemoto, E., et al. (2015). The endothelial protein PLVAP in lymphatics controls the entry of lymphocytes and antigens into lymph nodes. Nat. Immunol. *16*, 386–396.

Rashidian, M., Keliher, E.J., Bilate, A.M., Duarte, J.N., Wojtkiewicz, G.R., Jacobsen, J.T., Cragnolini, J., Swee, L.K., Victora, G.D., Weissleder, R., et al. (2015). Noninvasive imaging of immune responses. Proc. Natl. Acad. Sci. *112*, 6146–6151.

Ray, J.C.J., Flynn, J.L., and Kirschner, D.E. (2009). Synergy between individual TNF-dependent functions determines granuloma performance for controlling Mycobacterium tuberculosis infection. J. Immunol. Baltim. Md 1950 *182*, 3706–3717.

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

Read, M., Andrews, P.S., Timmis, J., and Kumar, V. (2012). Techniques for grounding agent-based simulations in the real domain: a case study in experimental autoimmune encephalomyelitis. Math. Comput. Model. Dyn. Syst. *18*, 67–86.

Read, M., Andrews, P.S., Timmis, J., Williams, R.A., Greaves, R.B., Sheng, H., Coles, M., and Kumar, V. (2013a). Determining Disease Intervention Strategies Using Spatially Resolved Simulations. PLoS ONE *8*, e80506.

Read, M., Tripp, M., Leonova, H., Rose, L., and Timmis, J. (2013b). Automated calibration of agent-based immunological simulations. (MIT Press), pp. 874–875.

Read, M., Andrews, P.S., Timmis, J., and Kumar, V. (2014). Modelling biological behaviours with the unified modelling language: an immunological case study and critique. J. R. Soc. Interface *11*, 20140704.

Read, M.N., Bailey, J., Timmis, J., and Chtanova, T. (2016). Leukocyte Motility Models Assessed through Simulation and Multi-objective Optimization-Based Model Selection. PLOS Comput Biol *12*, e1005082.

Reyes-Lamothe, R., Sherratt, D.J.D.J., and Leake, M.C.M.C. (2010). Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. Science *328*, 498–501.

Ritchie, K., Iino, R., Fujiwara, T., Murase, K., and Kusumi, A. (2003). The fence and picket structure of the plasma membrane of live cells as revealed by single molecule techniques (Review). Mol. Membr. Biol. *20*, 13–18.

Robson, A., Burrage, K., and Leake, M.C. (2013). Inferring diffusion in single live cells at the single-molecule level. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *368*, 20120029.

Rodda, L.B., Bannard, O., Ludewig, B., Nagasawa, T., and Cyster, J.G. (2015). Phenotypic and Morphological Properties of Germinal Center Dark Zone Cxcl12-Expressing Reticular Cells. J. Immunol. Baltim. Md 1950 *195*, 4781–4791.

Roozendaal, R., and Mebius, R.E. (2011). Stromal cell-immune cell interactions. Annu. Rev. Immunol. *29*, 23–43.

Roozendaal, R., Mebius, R.E., and Kraal, G. (2008). The conduit system of the lymph node. Int. Immunol. *20*, 1483–1487.

Roozendaal, R., Mempel, T.R., Pitcher, L.A., Gonzalez, S.F., Verschoor, A., Mebius, R.E., von Andrian, U.H., and Carroll, M.C. (2009). Conduits mediate transport of low-molecular-weight antigen to lymph node follicles. Immunity *30*, 264–276.

Rot, A., and von Andrian, U.H. (2004). Chemokines in innate and adaptive host defense: basic chemokinese grammar for immune cells. Annu. Rev. Immunol. *22*, 891–928.

Royston, P. (1992). Approximating the Shapiro-Wilk W-test for non-normality. Stat. Comput. *2*, 117–119.

Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat. Methods *3*, 793–795.

Sáez de Guinoa, J., Barrio, L., Mellado, M., and Carrasco, Y.R. (2011). CXCL13/CXCR5 signaling enhances BCR-triggered B-cell activation by shaping cell dynamics. Blood *118*, 1560–1569.

Sagrieya, H., Berube, C., Wen, A., Ramakrishnan, R., Mir, A., Hamilton, A., and Altman, R.B. (2010). Extending and evaluating a warfarin dosing algorithm that includes CYP4F2 and pooled rare variants of CYP2C9. Pharmacogenet. Genomics *20*, 407–413.

Salomonsson, S., Jonsson, M.V., Skarstein, K., Brokstad, K.A., Hjelmström, P., Wahren-Herlenius, M., and Jonsson, R. (2003). Cellular basis of ectopic germinal center formation and autoantibody production in the target organ of patients with Sjögren's syndrome. Arthritis Rheum. *48*, 3187–3201.

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

Saltelli, A., Chan, K., and Scott, E.M. (2008). Sensitivity Analysis (New York; Chichester; Weinheim etc: Wiley-Blackwell).

Schmutz, S., Bosco, N., Chappaz, S., Boyman, O., Acha-Orbea, H., Ceredig, R., Rolink, A.G., and Finke, D. (2009). Cutting edge: IL-7 regulates the peripheral pool of adult ROR gamma+ lymphoid tissue inducer cells. J. Immunol. Baltim. Md 1950 *183*, 2217–2221.

Schneider, P., MacKay, F., Steiner, V., Hofmann, K., Bodmer, J.-L., Holler, N., Ambrose, C., Lawton, P., Bixler, S., Acha-Orbea, H., et al. (1999). BAFF, a Novel Ligand of the Tumor Necrosis Factor Family, Stimulates B Cell Growth. J. Exp. Med. *189*, 1747–1756.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Schwarz, J. (2016). A microfluidic device for measuring cell migration towards substrate-bound and soluble chemokine gradients. Sci. Rep. *6*, 36440.

Schwickert, T.A., Lindquist, R.L., Shakhar, G., Livshits, G., Skokos, D., Kosco-Vilbois, M.H., Dustin, M.L., and Nussenzweig, M.C. (2007). In vivo imaging of germinal centres reveals a dynamic open structure. Nature *446*, 83–87.

Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., and Friedman, N. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat. Genet. *34*, 166–176.

Sh, Z., Sj, S., and Da, L. (1982). Kinetic analysis of chemotactic peptide receptor modulation., Kinetic analysis of chemotactic peptide receptor modulation. J. Cell Biol. J. Cell Biol. *92*, *92*, 34, 34–43.

Shulman, Z., Gitlin, A.D., Targ, S., Jankovic, M., Pasqual, G., Nussenzweig, M.C., and Victora, G.D. (2013). T follicular helper cell dynamics in germinal centers. Science *341*, 673–677.

Shulman, Z., Gitlin, A.D., Weinstein, J.S., Lainez, B., Esplugues, E., Flavell, R.A., Craft, J.E., and Nussenzweig, M.C. (2014). Dynamic signaling by T follicular helper cells during germinal center B cell selection. Science *345*, 1058–1062.

Siegert, S., Huang, H.-Y., Yang, C.-Y., Scarpellino, L., Carrie, L., Essex, S., Nelson, P.J., Heikenwalder, M., Acha-Orbea, H., Buckley, C.D., et al. (2011). Fibroblastic Reticular Cells From Lymph Nodes Attenuate T Cell Expansion by Producing Nitric Oxide. PLoS ONE *6*.

Sklar, L.A., Finney, D.A., Oades, Z.G., Jesaitis, A.J., Painter, R.G., and Cochrane, C.G. (1984). The dynamics of ligand-receptor interactions. Real-time analyses of association, dissociation, and internalization of an N-formyl peptide and its receptors on the human neutrophil. J. Biol. Chem. *259*, 5661–5669.

Solovyev, A., Mi, Q., Tzen, Y.-T., Brienza, D., and Vodovotz, Y. (2013). Hybrid Equation/Agent-Based Model of Ischemia-Induced Hyperemia and Pressure Ulcer Formation Predicts Greater Propensity to Ulcerate in Subjects with Spinal Cord Injury. PLoS Comput Biol *9*, e1003070.

Sommerville, I. (2010). Software Engineering (Boston: Pearson).

Stein, J.V., Rot, A., Luo, Y., Narasimhaswamy, M., Nakano, H., Gunn, M.D., Matsuzawa, A., Quackenbush, E.J., Dorf, M.E., and Andrian, U.H. von (2000). The Cc Chemokine Thymus-Derived Chemotactic Agent 4 (Tca-4, Secondary Lymphoid Tissue Chemokine, 6ckine, Exodus-2) Triggers Lymphocyte Function–Associated Antigen 1–Mediated Arrest of Rolling T Lymphocytes in Peripheral Lymph Node High Endothelial Venules. J. Exp. Med. *191*, 61–76.

Swartz, M.A., Hubbell, J.A., and Reddy, S.T. (2008). Lymphatic drainage function and its immunological implications: from dendritic cell homing to vaccine design. Semin. Immunol. *20*, 147–156.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. *39*, D561–D568.

Telesford, Q.K., Joyce, K.E., Hayasaka, S., Burdette, J.H., and Laurienti, P.J. (2011). The Ubiquity of Small-World Networks. Brain Connect. *1*, 367–375.

Tilo Beyer, M.M.-H. (2007). Modeling emergent tissue organization involving high-speed migrating cells in a flow equilibrium. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *76*, 21929.

Tolar, P., Sohn, H.W., and Pierce, S.K. (2005). The initiation of antigen-induced B cell antigen receptor signaling viewed in living cells by fluorescence resonance energy transfer. Nat. Immunol. *6*, 1168–1176.

Tomei, A.A., Siegert, S., Britschgi, M.R., Luther, S.A., and Swartz, M.A. (2009). Fluid Flow Regulates Stromal Cell Organization and CCL21 Expression in a Tissue-Engineered Lymph Node Microenvironment. J. Immunol. *183*, 4273–4283.

Tomura, M., Yoshida, N., Tanaka, J., Karasawa, S., Miwa, Y., Miyawaki, A., and Kanagawa, O. (2008). Monitoring cellular movement in vivo with photoconvertible fluorescence protein "Kaede" transgenic mice. Proc. Natl. Acad. Sci. *105*, 10871–10876.

Tschumperlin, D.J. (2013). Fibroblasts and the Ground They Walk On. Physiology *28*, 380–390.

Turing, A.M., and S, F.R. (1952). The chemical basis of morphogenesis. Phil Trans R Soc Lond B *237*, 37–72.

Turner, V.M., and Mabbott, N.A. (2017). Structural and Functional Changes to Lymph Nodes in Ageing Mice. Immunology n/a-n/a.

Turner, J.S., Marthi, M., Benet, Z.L., and Grigorova, I. (2017). Transiently antigen-primed B cells return to naive-like state in absence of T-cell help. Nat. Commun. *8*, 15072.

Ulvmar, M.H., Hub, E., and Rot, A. (2011). Atypical chemokine receptors. Exp. Cell Res. *317*, 556–568.

Ulvmar, M.H., Werth, K., Braun, A., Kelay, P., Hub, E., Eller, K., Chan, L., Lucas, B., Novitzky-Basso, I., Nakamura, K., et al. (2014). The atypical chemokine receptor CCRL1 shapes functional CCL21 gradients in lymph nodes. Nat. Immunol. *15*, 623–630.

Vargha, A., and Delaney, H.D. (2000). A Critique and Improvement of the "CL" Common Language Effect Size Statistics of McGraw and Wong. J. Educ. Behav. Stat. *25*, 101–132.

Vernon, I., Liu, J., Goldstein, M., Rowe, J., Topping, J., and Lindsey, K. (2016). Bayesian uncertainty analysis for complex systems biology models: emulation, global parameter searches and evaluation of gene functions. ArXiv160706358 Q-Bio Stat.

Vicini, P. (2010). Multiscale Modeling in Drug Discovery and Development: Future Opportunities and Present Challenges. Clin. Pharmacol. Ther. *88*, 126–129.

Victora, G.D. (2014). SnapShot: The Germinal Center Reaction. Cell *159*, 700–700.e1.

Victora, G.D., and Mesin, L. (2014). Clonal and cellular dynamics in germinal centers. Curr. Opin. Immunol. *28*, 90–96.

Victoratos, P., and Kollias, G. (2009). Induction of Autoantibody-Mediated Spontaneous Arthritis Critically Depends on Follicular Dendritic Cells. Immunity *30*, 130–142.

Vieites, J.M., Guazzaroni, M.-E., Beloqui, A., Golyshin, P.N., and Ferrer, M. (2009). Metagenomics approaches in systems microbiology. FEMS Microbiol. Rev. *33*, 236–255.

Visser, S. a G., de Alwis, D.P., Kerbusch, T., Stone, J.A., and Allerheiligen, S.R.B. (2014). Implementation of Quantitative and Systems Pharmacology in Large Pharma. CPT Pharmacomet. Syst. Pharmacol. *3*, 1–10.

de Vries, I.J.M., Lesterhuis, W.J., Barentsz, J.O., Verdijk, P., van Krieken, J.H., Boerman, O.C., Oyen, W.J.G., Bonenkamp, J.J., Boezeman, J.B., Adema, G.J., et al. (2005). Magnetic resonance tracking of dendritic cells in melanoma patients for monitoring of cellular therapy. Nat. Biotechnol. *23*, 1407–1413.

Walker, D.C., Hill, G., Wood, S.M., Smallwood, R.H., and Southgate, J. (2004). Agent-based computational modeling of wounded epithelial cell monolayers. IEEE Trans. NanoBioscience *3*, 153–163.

Wang, Y., and Irvine, D.J. (2013). Convolution of chemoattractant secretion rate, source density, and receptor desensitization direct diverse migration patterns in leukocytes. Integr. Biol. Quant. Biosci. Nano Macro *5*, 481–494.

Wang, X., Cho, B., Suzuki, K., Xu, Y., Green, J.A., An, J., and Cyster, J.G. (2011a). Follicular dendritic cells help establish follicle identity and promote B cell retention in germinal centers. J. Exp. Med. *208*, 2497–2510.

Wang, X., Cho, B., Suzuki, K., Xu, Y., Green, J.A., An, J., and Cyster, J.G. (2011b). Follicular dendritic cells help establish follicle identity and promote B cell retention in germinal centers. J. Exp. Med. *208*, 2497–2510.

Wang, Y., Wang, J., Sun, Y., Wu, Q., and Fu, Y.X. (2001). Complementary effects of TNF and lymphotoxin on the formation of germinal center and follicular dendritic cells. J. Immunol. Baltim. Md 1950 *166*, 330–337.

Watts, D.J., and Strogatz, S.H. (1998). Collective dynamics of “small-world” networks. Nature *393*, 440–442.

Williams, R.A., Timmis, J., and Qwarnstrom, E.E. (2016). Statistical Techniques Complement UML When Developing Domain Models of Complex Dynamical Biosystems. PLOS ONE *11*, e0160834.

Winter, G.E., Rix, U., Carlson, S.M., Gleixner, K.V., Grebien, F., Gridling, M., Müller, A.C., Breitwieser, F.P., Bilban, M., Colinge, J., et al. (2012). Systems-pharmacology dissection of a drug synergy in imatinib-resistant CML. Nat. Chem. Biol. *8*, 905–912.

Wolf, M., Albrecht, S., and Märki, C. (2008). Proteolytic processing of chemokines: implications in physiological and pathological conditions. Int. J. Biochem. Cell Biol. *40*, 1185–1198.

Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. J. Theor. Biol. *25*, 1–47.

Woodfin, A., Voisin, M.-B., and Nourshargh, S. (2007). PECAM-1: a multi-functional molecule in inflammation and vascular biology. Arterioscler. Thromb. Vasc. Biol. *27*, 2514–2523.

Wrenshall, L.E., Platt, J.L., Stevens, E.T., Wight, T.N., and Miller, J.D. (2003). Propagation and control of T cell responses by heparan sulfate-bound IL-2. J. Immunol. Baltim. Md 1950 *170*, 5470–5474.

Wu, D., and Lin, F. (2011). Modeling Cell Gradient Sensing and Migration in Competing Chemoattractant Fields. PLoS ONE *6*, e18805.

Yang, C.-Y., Vogt, T.K., Favre, S., Scarpellino, L., Huang, H.-Y., Tacchini-Cottier, F., and Luther, S.A. (2014). Trapping of naive lymphocytes triggers rapid growth and remodeling of the fibroblast network in reactive murine lymph nodes. Proc. Natl. Acad. Sci. U. S. A. *111*, E109-118.

Yi, T., Wang, X., Kelly, L.M., An, J., Xu, Y., Sailer, A.W., Gustafsson, J.-A., Russell, D.W., and Cyster, J.G. (2012). Oxysterol Gradient Generation by Lymphoid Stromal Cells Guides Activated B Cell Movement during Humoral Responses. Immunity *37*, 535–548.

Yoshida, R., Nagira, M., Kitaura, M., Imagawa, N., Imai, T., and Yoshie, O. (1998). Secondary Lymphoid-tissue Chemokine Is a Functional Ligand for the CC Chemokine Receptor CCR7. J. Biol. Chem. *273*, 7118–7122.

Yu, L., Cecil, J., Peng, S.-B., Schrementi, J., Kovacevic, S., Paul, D., Su, E.W., and Wang, J. (2006). Identification and expression of novel isoforms of human stromal cell-derived factor 1. Gene *374*, 174–179.

Zabel, B.A., Zuniga, L., Ohyama, T., Allen, S.J., Cichy, J., Handel, T.M., and Butcher, E.C. (2006). Chemoattractants, extracellular proteases, and the integrated host defense response. Exp. Hematol. *34*, 1021–1032.

Zigmond, S.H. (1981). Consequences of chemotactic peptide receptor modulation for leukocyte orientation. J. Cell Biol. *88*, 644–647.

Zitzler, E., and Thiele, L. (1998). Multiobjective optimization using evolutionary algorithms — A comparative case study. In Parallel Problem Solving from Nature — PPSN V, (Springer, Berlin, Heidelberg), pp. 292–301.