Model Hybridisation and Visualisation Techniques for the Investigation of Complex Disease Processes

James Andrew Butler

PhD

University of York

Biology

September 2016

Abstract

Tertiary lymphoid tissues (TLT) develop ectopically in most autoimmune disorders, their presence is strongly correlated with disease prognosis. The autoantibody response driven by germinal centres within TLT is an important driver of autoimmunity in Sjörgen's syndrome, for which there currently lacks any adequate therapy beyond palliative care. The cellular and molecular processes driving lymphoid neogenesis have remained elusive despite intense scrutiny utilising gene knock-out mice, lineage specific reporter mice, gene expression analysis, immunohistochemistry and flow cytometry. These approaches permit a thorough understanding of the formation of secondary lymphoid tissues. However, the mechanisms driving the formation and function of tertiary lymphoid tissues have proven to be more controversial and enigmatic, principally due to differences between experimental models and human disease pathology. A set of hypotheses describing a potential theory of TLT formation is developed from a combination of *in vitro* and *in vivo* data. This is then described as a mathematical and computational model through which the veracity of the hypotheses may be determined.

A framework is developed for the description of hybridised models combined of many constituent sub-models utilising different mathematical approaches, and implemented this in the development of a hybrid agent-based model incorporating Markov models, differential equations, cellular automata and generative grammar into a useful results model for understanding the effects of biologics on TLT formation. The simulation also serves as case study for other disorders, utilising the frameworks developed herein. Our simulation, in combination with a 'Model-driven Experimentation' paradigm, has demonstrated the possibility of using such tools for predictive purposes within the emerging field of quantitative systems pharmacology.

Utilising a hybrid agent-based complex systems model, we demonstrate that TLT formation can be described in terms of lymphocyte-stromal crosstalk and the differential response of B and T cells to chemokines CXCL13 and CCL19 respectively. We evaluated the potential therapeutic effects of three biologics *in silico*: anti-TNF, anti-LTβ, and, anti-VLA4. *In vivo* validation is provided regarding the predicted efficacy of anti-VLA4.

Contents

List of Tables

List of Figures

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.* (2013), Greaves *et al.* (2013), Butler *et al.* (2014), 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

Complex systems modelling and simulation have proven to be extremely valuable techniques in biology (?), that can augment and extend experimental biology through hypothesis testing and generation (?), providing insight that would not be possible through in vivo *and* in vitro *models alone. However, as datasets grow increasingly large and greater computational resources become available, it is necessary to develop means of maintaining tractability in simulations that capture phenomena at an ever-finer granularity while ensuring that models can be effectively leveraged to achieve maximum impact; this requires complex models and their results to be communicable across disciplines and developed according to well-defined, principled frameworks. This ensures that a model's fitness-for-purpose can be demonstrated through exposition of the biological data used to construct the model, the assumptions and abstractions that have been made, and the mathematical and computational methods used in their implementation as simulations.*

This chapter presents the current state-of-the-art in complex systems modelling in biology and pathophysiology, particularly with respect to the provision of principled frameworks for defining models, simulations and their results. Particular emphasis is placed on modelling lymphoid tissues and their role in immune disorders. The modelling and simulation techniques developed and applied herein utilise a case study of tertiary lymphoid tissue development during Sjögren's syndrome. Novel simulation visualisation and quantification methodologies are developed using an existing

simulation of secondary lymphoid tissue formation in the mouse gut as an initial case study, and these are subsequently applied to the development of simulation of tertiary lymphoid tissue development. The literature is reviewed for hybrid models which integrate more than one modelling methodology into a simulation, of disease models that aim to develop and evaluate the efficacy of therapeutic interventions, and of model visualisation techniques and their use in simulation analysis. The motivation underlying this work is defined, and the thesis aims are discussed.

Prior to discussing the role of mathematical and computational modelling in understanding immune function and disease pathophysiology, it is necessary to introduce the reader to a few key concepts in immunology. This section provides a very brief overview of the mammalian immune system followed by a more detailed review of the development and role of lymphoid tissues in providing a specialised tissue environment for effective immune responses. This then permits a discussion of the state-of-the-art in immune system modelling and the identification of open questions and means of providing improved methodologies for describing complex immune processes that occur over many time and length scales.

1.1   Introducing the Immune System

The mammalian immune system is comprised of two fundamental components: innate and adaptive immunity. The cells that collectively constitute the immune system are termed leukocytes, otherwise known as white blood cells. Innate immunity has no memory of previous infection but responds rapidly with a large and complex array of methods for destroying pathogenic microbes including bacteria and viruses; additionally, innate immune cells such as macrophages or dendritic cells present antigen (small fragments from foreign bodies) on their surface, thus permitting adaptive immunity to respond through direct cellular activity and via humoral responses, in which antibodies are produced specific to the invading pathogen. The adaptive immune system generates memory cells following antigen challenge, thus enabling a quicker and more effective response if the same pathogen is encountered

again in the future, this provides the basis for protective vaccines (**?**).

T and B lymphocytes are the principal constituents of the adaptive immune response. B cells differentiate from haematopoietic stem cells and mature initially in the bone marrow followed by further development in the spleen (**?**), and produce antibodies when antigen binds to their B-cell receptor (the humoral response), B cells also secrete cytokines (signalling molecules) and can act as 'professional' antigen-presenting cells. T cells also develop from haematopoietic stem cells in the bone marrow, but mature in the thymus where auto-reactive T cells are eliminated in a process known as 'thymus education' (**?**); there are many T cell subsets with different functions (**?**), including *i.* T helper cells that assist in the maturation of B cells into plasma (antibody secreting) cells and memory B cells, *ii.* cytotoxic T cells that target virus infected cells, and *iii.* regulatory T cells (T$_{regs}$) that maintain immunological tolerance, with the capacity to suppress T-cell mediated immune responses as a pathogenic threat has been cleared, or to suppress autoreactive T cells that escaped the thymus (**??**). Following their respective maturation processes, B and T lymphocytes enter circulation and migrate to secondary lymphoid tissues such as lymph nodes, that are strategically located to provide an efficient adaptive immune response to local infections.

1.1.1 Lymphoid Tissues: Form and Function

Lymphoid tissues are indispensable for the organisation and function of the adaptive immune system. The mechanisms that drive the development of these tissues have been under intense scrutiny for hundreds of years. This section provides an account of the historical literature and the subsequent modern understanding the development and function of secondary lymphoid tissues.

The first account of lymphoid tissue pathophysiology belongs to the Hippocratic Corpus (**?**), entitled *Peri adenôn* (*About Glands*). The year of composition is disputed, but it is likely to have been penned circa the fourth century BCE. The insight into lymph node anatomy and function held within this ancient treatise is

astonishing, especially considering that the first microscope was not to be invented for another 1900 years. In the hippocratic corpus, lymph nodes are described as 'adenes, or glands, which become swollen and inflamed in the presence of disease, resulting in illness and fever. In addition to prefiguring modern immunology, the text described lymphatic vessels and the role of LNs in filtering lymph back into the blood via the lymphatic vasculature (**?**). Little changed in the understanding of the lymphatic system until 1622, when Pavian surgeon, Gasparo Aselli, discovered the mesenteric lymphatic vasculature and began to document the lymphatic system in significant detail (**?**). There was little progress in understanding of lymphoid tissues following the work of Aselli until the late eighteenth century, with experiments by William Hunter into the role of lymphatics in draining lymph from tissues (the 'absorbent theory') (**?**), and the development of histology in 1836 lead by Camillo Golgi (**??**), for which he shared the 1906 Nobel Prize in Physiology or Medicine. The earliest attempt at describing the formation of lymph nodes was in **?**, which linked the origin of lymph nodes with lymph sacs, and by the early 20th century, the stages of lymph node anlagen development were characterised in humans by **?**, who acknowledged their 'protective role' in adults, subsequent work in swine (**?**) and rabbit (**?**) bolstered this understanding of the development of lymphoid tissues and their importance in providing immunity.

A modern understanding of the molecular and cellular mechanisms driving lymphoid tissue formation and maintenance has developed from the application of state of the art imaging technologies, genetic tools and cell culture techniques – and increasingly, the application of mathematical and computational models (**????**). Lymphoid tissues are divided into primary, secondary and tertiary lymphoid organs. The bone marrow and thymus constitute the primary lymphoid tissues; these are responsible for the selection and differentiation of immature T progenitor cells into naive CD8+ and CD4+ T cells. CD4+ T cells bind antigen presented in MHC-II by Antigen-Presenting Cell (APC), and CD8+ T cells bind antigen presented in MHC-I. The secondary lymphoid organs (SLO) provide a highly specialised micro-
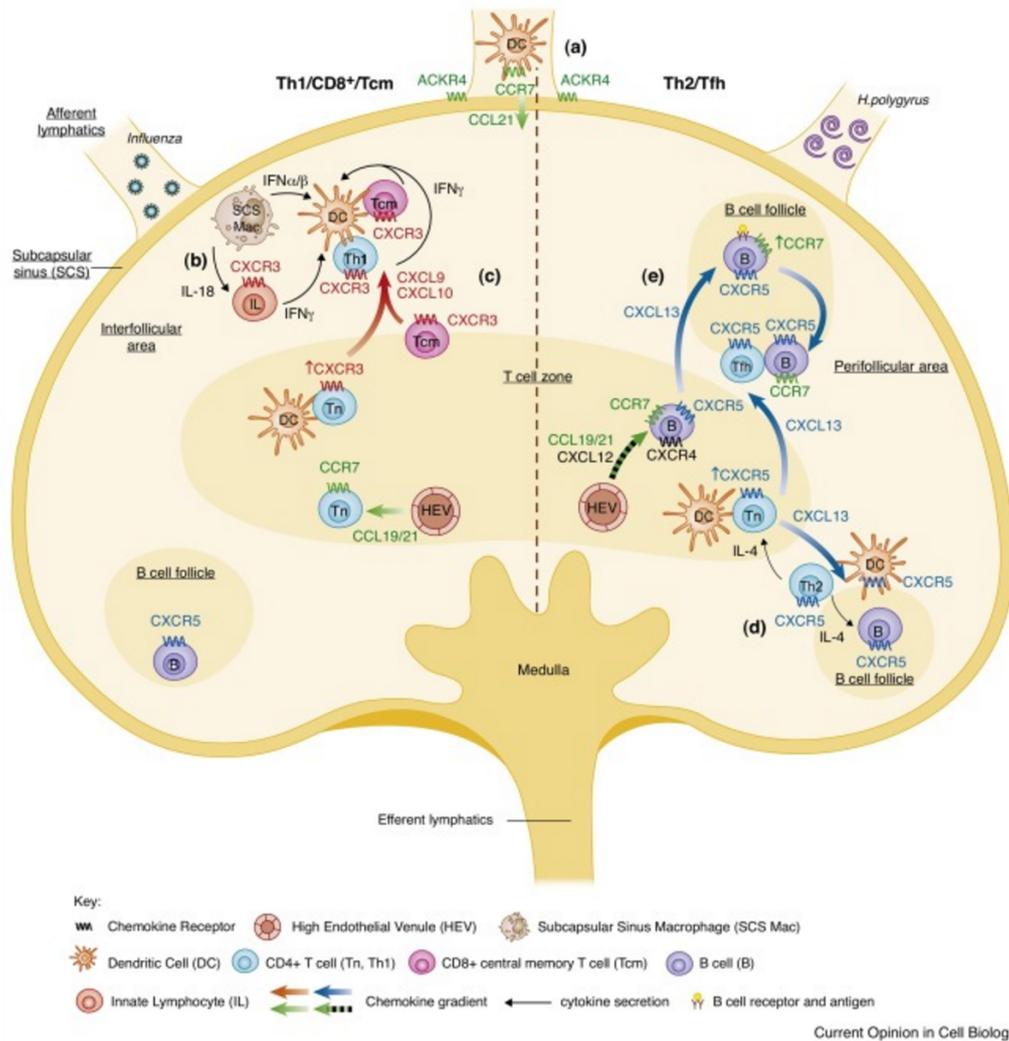
Figure 1.1: Specialised niches develop in the lymph node through lymphoid chemokine-mediated orchestration. (a). Antigen arrives via the lymphatics carried by DCs or in soluble form. (b) Subcapsular sinus macrophages take up antigen from the lymphatics, attracting and priming CD4+ T helper cells. (c) Naive T cells enter the LN through high endothelial venules within the T cell zone and are primed by antigen-presenting DCs prior to migration toward the periphery. (d) Activated naive T cells upregulate CXCR5 and migrate along a CXCL13 gradient into the perifollicular regions adjacent to B cell follicles. (e) Naive B cells are recruited into the follicle via CXCL13 through CXCR5 expression, and upon activation upregulate CCR7, chemotactically migrating to the edge of the follicle permitting interaction with T follicular helper cells to undergo class switching. Germinal centres develop within the B cell follicle resulting in the humoral immune response. Taken from **?**.

environment with subdivisions ('niches') specific for different lymphocyte types (**?**). Niches provide environments for the differentiation and activation of immune effector cells; SLOs capture pathogens from peripheral tissues through a process of dendritic cell (DC) recruitment. The SLO microanatomy then facilitates contact with their cognate APC, resulting in activation of the adaptive immune response. The process of lymphocyte migration into and within SLOs is orchestrated by lymphoid chemokines, principally CCL19/21 and CXCL13 expressed in the T cell zone and B cell follicles respectively (**?**), chemokine gradients are sculpted by atypical 'scavenging' chemokine receptors such as CCRL1 (**?**) to provide optimal gradients for migration along the sub-capsular sinus. This process of chemokine-mediated self-organisation within the lymph node, and the overall resulting structure is shown in Figure 1.1.

SLOs are mainly comprised of lymph nodes, Peyers patches (PP), the white pulp of the spleen and the mucosal-associated lymphoid tissue (MALT). MALT has important roles in the innate immune response through the secretion of type I interferon and other innate immune mediators (**?**), which when combined with the strategic placing of macrophages within tissues, results in a reduction of pathogen load in the host. This is crucial in allowing time for the adaptive immune system to successfully mount a response (**?**). In healthy specimens, evidence suggests that peripheral tissues are patrolled by relatively low numbers of mature lymphocytes (**?**), indicating that SLOs are necessary for an effective adaptive immune response by facilitating the detection of microorganisms, acting both as a garrison for immune cells and a sink for invading pathogens (**?**). Leukocytes including lymphocytes and phagocytes migrate between tissues by recirculating through the lymphatic circulatory system (Figure 1.2).
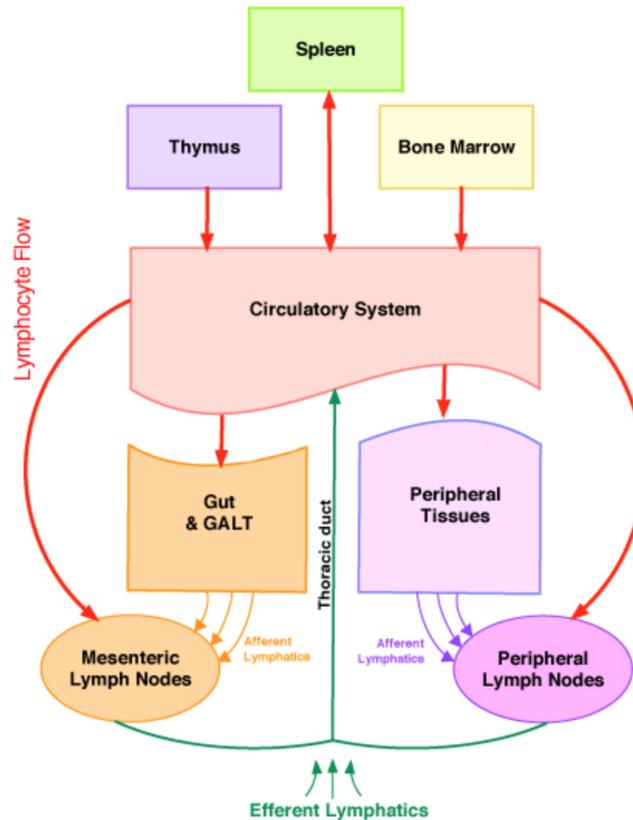
Figure 1.2: Certain lymphocytes and phagocytes are capable of recirculation, moving between lymphoid and non-lymphoid tissue. This is crucial in allowing lymphocytes to perceive their cognate antigen, and allows effector cell populations to access the tissues in which they are needed. Arrows in the figure indicate direction of blood and lymph.

1.1.2 Tertiary Lymphoid Tissues: Lymphoid Neogenesis

Tertiary Lymphoid Tissues (TLT) are similar in organisation to lymph nodes, but form during chronic inflammation, due to infection, autoimmune disease and some cancers, and their presence is generally associated with a poorer prognosis (**?**). The principal components that drive the formation of TLT a process termed lymphoid neogenesis, are poorly understood – what the minimum requirements are for TLT induction is an open question in immunology. There are a vast number of signalling molecules and cell types implicated, and these are subtly different when tissues develop in different pathologies (**?**). An introduction to TLT form and function is provided here, as TLT formation is used as the basis for a case study in the development of a novel hybrid multiscale model and simulation in Chapters 3 and 4.

While lymph nodes develop during embryogenesis in pre-determined anatomical positions, lymphoid aggregates with very similar morphology to lymph nodes, containing organised lymphocytes niches have been observed to develop ectopically during periods of chronic inflammation, due to persistent infection, neoplasticity in cancer, or in autoimmune disease. Chronic inflammation involves complex pathophysiology comprising of a large accumulation of peripheral blood mononuclear cells (PBMCs) including macrophages, T cells, B cells and dendritic cells, which establish a self-perpetuating process of stimulation, recruitment and colonisation termed *lymphoid neogenesis.* This phenomenon was first established by **?**, in which it was proposed that lymphoid organogenesis and chronic inflammation share a common physiological basis, and that lymphotoxin induced chronic inflammation has the characteristics of organised secondary lymphoid tissue and can respond to antigen. The earliest reference in the literature to lymphoid neogenesis can be found in **?**, regarding the observation that bronchus-associated lymphoid tissue is not present at birth, unlike lymph nodes and Peyer's patches, nor is it present in every specimen. The role of stroma in the orchestration of lymphoid neogenesis was initially suggested in **?** with respect to non-Hodgkin's lymphoma, wherein Follicular Dendritic Cells (FDC) were found in mucosa-associated lymphoid tissue (MALT) and extranodal non-MALT.

When TLT forms during chronic infection, it permits local antigen presentation at the site of infection and enables priming of naïve lymphocytes (**?**), which may assist in clearing the infection (**?**). However, it has also been noted that the presence of TLT can be detrimental in effect, leading to chronic hepatic inflammation and ultimately malignant neoplasia. In cancers, the presence of infiltrating lymphoid tissue has been found in some cases to both promote tumour growth (**??**), and also to contribute towards anti-tumour immunity (**?**), for example, in ductal breast carcinomas (**??**).

During autoimmune disease, TLT formation often occurs within disease-associated tissues, and its presence is generally associated with a poorer prognosis (**?**), although
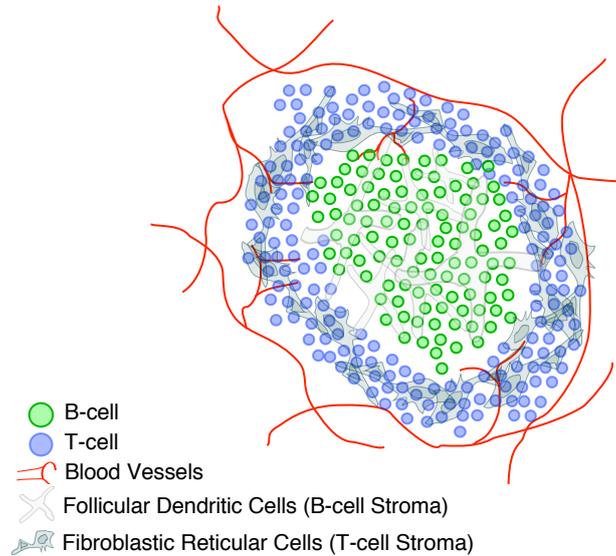
Figure 1.3: Schematic diagram of tertiary lymphoid tissue, illustrating separate B and T cell zones and distinct stromal network phenotypes. B cells form follicular niches supported by stroma of an FDC-like phenotype, surrounded by T cells supported by Fibroblastic Reticular Cells (FRC)-like phenotype stroma. Unlike in LNs, where vasculature penetrates into the cortical interface, within TLT, blood vessels surround the tissue externally. Another notable difference to LN organisation is that T cells are present prior to tissue formation with B cells recruited later during lymphoid neogenesis, however in LN organogenesis, T cells must undergo education in the thymus and migrate to the LN anlage following B cell colonisation.

their precise involvement in disease pathology is not well understood. There is evidence to suggest that in autoimmune disease, the presence of TLT correlates with local levels of antibody production (**?**), and the promotion of cytotoxic T-cell responses (**?**). In rheumatoid arthritis, it has been noted that individuals with the most highly developed TLT, with well-developed FDCs, germinal centres (GCs) and clearly defined B/T-cell zones, are associated with greater disease activity and increased severity of symptoms (**?**). Local antibody production correlation with TLT presence has also been identified in several other autoimmune diseases, including Sjörgen's syndrome (SS) (**??**). A simplified schematic of TLT structure can be seen in Figure 1.3, illustrating B and T cell segregation, supported by differential stromal phenotypes that are analogous to those found in the lymph node (FDCs for B cells, and FRCs for T cells).

1.1.3   Sjörgen's Syndrome

Sjörgen's Syndrome (SS) is an autoimmune disorder with a prevalence of about 1% of the population, characterised principally by inflammation and subsequent tissue damage in the exocrine glands, particularly the salivary gland and tear ducts. However, as a systemic condition, it can affect the function of exocrine glands in the vaginal tract, skin, trachea, nose and gastrointestinal tract (**?**). Due to the systemic inflammatory nature of SS, clinical presentation can extend beyond the exocrine glands resulting in fatigue and cachexia (weakness and wasting) (**?**).

Treatment of SS is limited to palliative measures, in particular the use of drops (pilocarpine and cevimeline) to treat dryness in the eyes, mouth, and other glands with "sicca" features (loss of fluid secretion) (**?**), and also surgical excision of tissue (**?**). More recently, large molecule biologic therapies have been employed in a research setting, however clinical efficacy has yet to be clearly demonstrated with any biologic therapy (**?**). Anti-TNFα, a biologic intervention that binds TNFα (Tumour Necrosis Factor alpha) in order to block its activity, was hypothesised to demonstrate efficacy due to its known, central role in SS pathogenesis. Unfortunately, there was no significant improvement in saliva production. There is early evidence that rituximab (anti-CD20) is an effective treatment for SS including improvement in both saliva production and overall measures of fatigue (**??**). Widespread adoption of rituximab is however restricted due to potentially dangerous side-effects including immunosuppression through loss of B cells, and the high economic cost of the therapy. This thesis identifies the likely reasons for this and evaluates other potential biologic therapies that may be more tolerable, using a computational model describing the inflammatory processes driving SS induction during pathogenesis, described in chapter 6.

It is well-established that TLT has an important role in SS pathology. The formation of TLT structures within exocrine glands through lymphocyte recruitment via chemokines results in established lymphoid follicular structures with active GC reactions, resulting in lymphocyte and antibody-driven damage to the exocrine glands

in which they are present (**??**).

1.2 Motivation for Mathematics, Modelling and Simulation in Immunology

There has traditionally been a great divide between the biological and physical sciences – it was argued by **?** that many immunologists regard mathematical science and biology to be somehow fundamentally different, a position possibly engendered by their artificial segregation in academia. The proceeding decade has lead to the emergence of a vast expansion in the use of both 'top-down' (bioinformatics and statistical modelling of large datasets) and 'bottom-up' (mechanistic models and complex systems analysis) approaches to the mathematics and modelling of biological systems (**?**), with systems biology emerging as an integral aspect of modern biology. It should be noted that mathematics has long since had a significant role in biology, in the application of statistical techniques to experimental results, in the development of experimental methodologies, and more recently in the high-throughput analysis of 'omics' data including gene expression analysis (**?**), metabolomics (**?**) and proteomics (**?**): this thesis is concerned with the application of mathematics in describing immunological processes, particularly for the purpose of understanding complex pathophysiology, and it is mathematical modelling toward describing biological systems mechanistically with which this section is concerned.

This biology–mathematics divide continues to exist in spite of significant, early contributions from mathematics to biology including the seminal work by **?** on the 'Chemical Basis of Morphogenesis' which continues to have relevance today, despite being overlooked or even dismissed previously by experimentalists (**?**). Turing demonstrated, critically, that instability in a biochemical system could arise from the complex interaction of many stabilising components, and therefore, that complex spatially inhomogeneous 'profile' formation has to be considered within the context of interacting processes; Turing noted that gastrulation – the process during

early embryogenesis wherein a homogeneous blastula is driven by complex cellular dynamics to begin to develop into an organism of differentiated, specialised cell types, could be described in terms of a uniform spatial steady-state that becomes unstable upon the introduction of a diffusive process, which, arising from symmetry breaking, permits complex spatial profiles developing from initially homogeneous, steady-state conditions; this occurring despite diffusion being understood to act as a stabilising process in itself (leading to spatial homogeneity at its limit). This leads to fundamental insights and principles that drive modern mathematical and systems biology today: the notion that chaotic, non-linear behaviour of individual biological processes can result in *emergent properties* that cannot be understood from consideration of each individual component in isolation. Subsequent work failed to acknowledge the fundamental insights afforded by such work; for instance, **?** wrote that Turing's model 'is quite unable to generate regulative systems', while failing to acknowledge the fundamental insights that Turing's notion of *diffusion-driven instability* and supporting toy models afforded to modern understanding complexity in biology, with Turing's morphogens essentially predicting the existence of cytokines and growth hormones (which he termed 'morphogens') that drive cellular differentiation and tissue growth, many examples of systems that behave in a manner similar to his morpogens are now known (**???**).

? noted that, despite the fascinating discoveries and insights provided by experimental immunology, mere exposition of experimental results cannot quantitatively answer crucial questions such as what determines the viral load or cellular turnover rate in HIV infection, and shrewdly acknowledged that the common refrain of biological processes being too complicated to be suited for mathematical investigation, becomes a more logical argument when reversed: that it is precisely their complexity that invites the incorporation mathematical analysis. Indeed, the complexity of the immune system necessitates a mathematics suited to investigating chaotic, non-linear systems composed of many individual components; that is, immune function must be considered within the context of complex systems analysis. It is this

non-linearity in immune function that requires mathematical analysis for true, quantitative understanding.

To address the application of complex systems analysis, mathematical modelling and simulation to immunology, we will require precise definitions of a 'complex system', 'complex systems analysis' and the notion of chaotic, non-linear processes within biology, discussed in Section 1.2.1. The state of the art in the application of mathematical modelling in addressing important immunological questions may then be explored (Section 1.2.3).

1.2.1   Complex Systems Analysis and Biological Processes

A 'complex system' can be defined as one driven principally by the behaviour of many *animate* constituent components that individually may be simple but whose collective interactions are too non-linear, chaotic, and possibly ill-defined, to admit precise mathematical analysis (**?**). Complex systems are characterised by emergent properties, that is, the interaction of many individual components results in the collective emergence of a particular phenomenon, pattern or system behaviour that cannot be intuitively understood from complete knowledge of each constituent component nor derived from an analytical mathematical analysis. Complex systems analysis became a practical reality with the advent of two key advances in the latter half of the nineteenth century: improved experimental methods providing accurate, quantitative data on the constituent components of complex systems through reductionism, and the development of digital computing subsequently permitting abstracted descriptions of these constituent components *in silico* enabling computational simulations through which complex systems can be systematically explored. The ever-increasing availability of computational resources has allowed systems of greater complexity to be analysed and understood.

Biological systems are inherently complex; millions of years of evolution has resulted in highly complex, robust chemical systems involving phenomena spanning nanometres to metres, and nanoseconds to years. It is not surprising that the ma-

jority of complex systems that have been studied are grounded in biology or are otherwise anthropogenic – the collective behaviour of many millions of atoms or molecules yield to mathematical analysis through application of statistical mechanics due to their adherence to well-defined, inalienable 'laws' of physics. However, the collective behaviour of nucleic acids and the large number of different proteins that give rise to biological cells, and the many heterogeneous cell populations that give rise to complex organisms including mammals, rely on ever increasing levels of abstraction such that it is not practicable nor desirable to attempt to apply reductionist principles *ad infinitum* to derive analytic solutions of their behaviour at the most fundamental levels, an idea most eloquently described by **?**. Such analyses would be absurdly complicated and not amenable to developing a true understanding of a complex system and the factors that affect its behaviour.

1.2.2 Non-linearity in immune function

It is often said that immune function, and many other biological processes beside, are 'highly non-linear' (**????**), but what is precisely meant by non-linearity in immunology, and why is it important? The mathematical definition of a non-linear process in one in which the output is not directly proportional to the input. The example, well-known to immunologists, of T-cell responses to different concentrations of interleukin-2 was given by **?** (Figure 1.4). At low IL2 concentrations, there is little or no proliferative response, however as IL-2 levels increase, T-cell proliferation increases exponentially before plateauing and then decreasing. Such a dose-response curve does not easily admit to mathematical definition, and the response by different cell types to the myriad molecules that drive behaviour is different in each case.

What may not be immediately obvious is the complex, chaotic behaviour that results from several interacting non-linear processes, including counter-intuitive effects. An example of this would be the role of IL10 in rheumatoid arthritis, it is known to both promote accumulation of IgM-B cells within the synovium and promotes rheumatoid factor (autoantibody) presence (**?**), but also acts in a regulatory
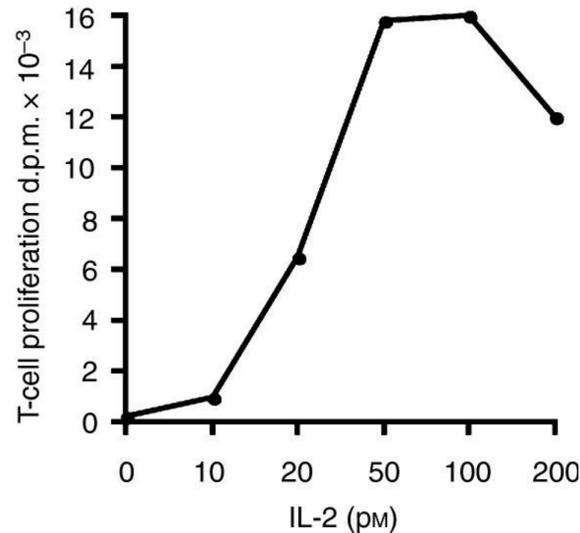
Figure 1.4: An example of non-linearity in immunology: the response of IL-2-dependent T-cells to increasing doses of IL-2. There is little or no proliferative response at low doses, but as the IL-2 concentration is increased, T-cell proliferation grows exponentially before plateauing (and in some cases, then decreasing). Taken from **?**

capacity as an inhibitor of synovial inflammation (**?**) – one soluble molecule responsible for apparently contradictory roles. A corollary, but more generic example, is the unpredictability of the effects of biologic blockade (e.g. fusion proteins and monoclonal antibody therapies) on disease pathology. For instance, multiple sclerosis onset is associated with anti-TNFα therapy (**?**) despite the pivotal role of TNFα in driving neuronal death within multiple sclerosis (**?**). These counter-intuitive, often contradictory, roles of biomolecules within immunology and biology are the result of interacting non-linear systems driving chaotic behaviour that cannot be understood through reductionist experimentation alone.

Through the rapid development of systems immunology, it has been established that many of the 'high-level' properties of the immune system are a result of positive/negative feedback controls, amplification methods, cellular heterogeneity and a reliance on stochastic events (**?**). These properties rely on many individual, interacting entities and are therefore amenable to computational and mathematical analyses, permitting a greater understanding of immune function, dysregulation and identification of potential approaches for modulating the immune system as a ther-

apeutic intervention strategy.

1.2.3   Modelling Immunology as a Complex Dynamical System

The development of models that capture the essential, emergent behaviour of specific biological processes, with extraneous components excluded, enables understanding of how the fundamental non-linear systems govern the process(es) of interest, and can therefore lead to new insight or quantitative prediction (**?**). Often, such models will exclude components known to be involved in the system under study in some capacity – the inclusion of as much biological detail as possible may obscure the key entities involved. 'Top-down' models that seek to include all available data are essentially phenomenological descriptions of the biological system derived from bioinformatics analysis of high-dimensional datasets. While such models may more accurately fit experimental data and are useful for identifying the importance of particular components within a system, such as genes (**?**) or enzymatic processes (**?**), greater insights into minimum requirements, prerequisites and governing dynamics arise from models constructed from the 'bottom-up' designed to address well-defined, pre-specified research questions (**?**).

A rich body of literature has developed applying mathematical and computational modelling to address immunological questions, using both analytical and numerical approaches encompassing techniques covering the gamut of mathematical techniques. This section introduces generally a selection of analytical mathematical models and computer simulations which have been used to address various immunological questions, prior to Section 1.2.4 reviewing the literature surrounding lymphoid tissue development.

One aspect of immunology that has been the subject of significant mathematical modelling is the quantification of thymic output (**?**), which is critical for the understanding of healthy homeostasis and disease dysregulation in the peripheral T-cell compartment, such as is the case in HIV infection. The thymus clearly has

a key role in immunity as it provides functional T cells to the periphery in both
children and adults, and is necessary for immune function reconstitution during im-
munodeficiency conditions such those caused by HIV infection. This is an excellent
example of a crucial immunological problem for which experimentation alone was
insufficient, eventually yielding to a mathematical analysis. Quantifying mammalian
thymic output is difficult to determine experimentally as there is no known surface
protein of recent thymic emigrant T cells with sufficiently short half-life that would
permit calculation of emigration rate. **?** proposed that T-cell receptor excision cir-
cles (TRECs) could be used to quantify thymic output. TRECs are a by-product of
TCR rearrangement, made up of episomal DNA circles. During T-cell thymic de-
velopment, the TCR is rearranged through recombination of its variable and joining
regions on chromosome 14, and involves excision of the δ locus that resides between
the V and J segment genes on the chromosome α chain. As this excised DNA con-
tains specific sequences common in 70% of $\alpha\beta$ T cells, are exclusively in those T cells
of thymic origin (**?**), and do not divide (**?**), these TRECs are suitable candidates for
use as markers of recent thymic emigrant T cells.

Several studies attempted to quantify thymic output, particularly in the con-
text of HIV infection and the role of a reduced thymic output on disease outcomes,
and arrived at substantially different conclusions regarding the suitability of using
TREC as a phenotypic marker for measuring CD4+ lymphocyte depletion. TREC
frequency was identified as decreasing with age, lower in HIV-infected individuals,
but that it was significantly higher in those individuals treated with antiretroviral
drugs (**??**). It was further determined in one study that TREC frequency could
be used as a predictor for HIV-1 disease progression, beyond viral load and CD4+
T-cell count (**?**). Following this early research, many more studies were conducted
attempting to quantify thymic output using TRECs (**???**), including some early
work exploring mathematical models as a tool for the characterisation of factors
leading to TREC concentration changes (**?**). Such factors influencing interpretation
of TREC frequency include its variation in response to changes in cells both with

and without TRECs, therefore biodynamical processes beyond thymic output into the periphery (such as T-cell proliferation or death) could influence observed TREC frequency changes. The reader is referred to the literature for each respective TREC model for further details with regard to each individual model's parameterisation and validation. Given that it is well-known that T-cell proliferation increases in HIV infection, this presents significant difficulty in the interpretation of how TREC frequency relates to thymic output in these individuals (**?**). Mathematical modelling is well-suited to potentially providing insights regarding such complexities. In seminal work by **?**, it was argued that mathematical modelling must be utilised to properly analyse these data and determine what biological parameters needed measuring. This study (**?**) utilised a very simple approach using linear ordinary differential equations to quantify TREC-containing cells (C) and total T cell population size (T), defining the following equations:

$$\frac{dC}{dt} = \alpha\sigma - dC \tag{1.1}$$

and

$$\frac{dT}{dt} = \sigma + (p - d)T \tag{1.2}$$

let σ be thumic output, α be the fraction of thymic emigrants containing a TREC, finally let p and d be the T cell proliferation and death rate. These linear ODEs are sufficient to describe the experimental data obtained in previous studies, and can obtain an equation describing TREC frequency, $F(t)$), utilising Equations 1.1 and 1.2, applying the quotient rule to express $F(t)$ as a function of $\frac{C(t)}{T(t)}$:

$$\frac{dF}{dt} = \alpha\frac{\sigma}{T} + (p - \frac{\sigma}{T})F. \tag{1.3}$$

We see in Equation 1.3 that TREC frequency dynamics are dependent upon not only thymic output, but also on the proliferation rate. Therefore, the assumption that TREC frequency could be used as a surrogate measure for thymic output

was indeed incorrect (**?**). This insight led **?** to analyse the dynamics of TREC in HIV-infected individuals in the light of this discovery, and determined that the reduction in TREC frequency in such individuals reported in previous studies (**???**) is potentially caused by immune 'hyper-activation' reflected by an increase in T-cell proliferation and further concluded that measuring changes TREC frequency was not sufficient to provide evidence of thymic impairment in an individual.

This simple mathematical model provided an enormous contribution by determining that TREC frequency could not be a direct measurement of thymic output, and any analyses that assumed it was would lead to contradictory or incorrect results. Furthermore, Equation 1.3 establishes that TREC concentration, rather than frequency, could indeed permit direct measurement of thymic output. These insights may appear obvious in the light of the mathematical model, however intuition alone was insufficient to arrive at these conclusions. This provides an excellent example of using mathematical modelling to evaluate the veracity of a biological hypothesis, even though the model described in Equations 1.1—1.3 is a highly simplified description of a complex, non-linear emergent biological phenomena. This model led to various *in vivo* experiments that permitted direct quantification of thymic output, for instance in **?** the daily number of CD8+ and CD4+ T cells exported from the thymus in rhesus macaques was determined. The simple model described above was later developed upon in **?**, wherein a non-linear model of thymopoesis was derived (Figure 1.5). Implementation of the thymic involution model described for humans in Figure 1.5 permitted simulations to determine and quantify age-dependent changes of recent thymic emigrants produced per day (Figure 1.6), which could not be otherwise measured experimentally. Furthermore, the model was able to explore whether thymic concentration is an accurate predictor of thymic function, as suggested by the simpler model described in **?**.

The model described in Figure 1.5 predicts that both thymic output and peripheral T cell division could equally affect TREC concentration for individuals at any age, with a lesser contribution from T cell death – broadly in agreement with the
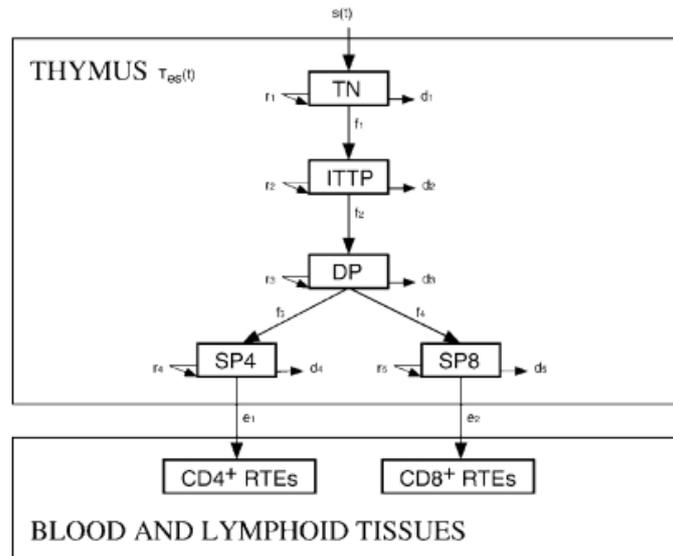
Figure 1.5: A multi-compartment model describing thymopoiesis (**?**) including five thymocyte subsets (TN, ITTP, DP, SP4 and SP8). Model includes thymocyte growth (γ_i) limited by maximum thymocyte population size within the thymic epithelial space ($T_{es}(t)$), thymocyte differentiation (f_i), death (d_i) and emigration from thymus as recent thymic emigrants (e_1, e_2). Taken from **?**.

earlier model by **?**, however noting that this model assumed that division of naive T cells does not occur and neglected the possibility of intracellular TREC degradation. Therefore, **?** proposed that T cell division and death should be examined in an experimental system before relying upon TREC concentration as a direct correlate of thymic output, and that their model can be used as an integrated system paired with experimentation to determine thymic output through integration of additionally measured T cell dynamics. The notion of combining a measure of both TREC concentration and a methodology for estimating intrathymic precursor T-cell proliferation rates through quantification of distinct TREC molecules in peripheral blood cells was later developed in **?**, providing a reliable measure of thymic function in HIV-infected patients undergoing anti-retroviral therapy.

The experimental (**?????**) and modelling (**?????**) work that developed our understanding of measuring thymic output, with important consequences for monitoring HIV-infected individuals, spans over a decade – from early experiments predicting that TREC frequency was a direct measure of not only thymic output, but
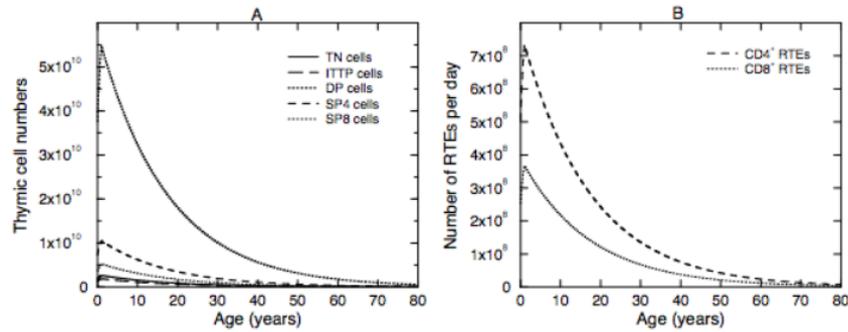
Figure 1.6: Recent thymic emigrant dynamics for five thymocyte subsets over an 80-year lifespan. Panel A shows simulations of five thymocyte subsets, and Panel B shows the number of emigrants output from the thymus per day. Results from the model described in **?** as reproduced in **?**.

HIV-infection disease progression, to simple mathematical modelling implicating the role of T-cell proliferation (**?**) and later, complex non-linear modelling (**?**) further refining this understanding to ultimately enable a reliable experimental framework for measuring thymic output as a function of TRECs (**?**). This development of the literature in this area provides a clear demonstration of the value of mathematical modelling in the interpretation of *in vivo* datasets that cannot be intuitively understood due to complex non-linearities and unappreciated subtleties within the data.

1.2.4   Modelling Lymphoid Tissue Development *In Silico*

The molecular and cellular mechanisms driving the formation of secondary lymphoid tissues have been extensively studied using a combination of mouse knockout studies, lineage specific reporter mice, microarray gene expression analysis, immuno-histochemistry, imaging and flow cytometry. However, the molecular and cellular processes which drive the formation and maintain the function of tertiary lymphoid tissue (TLT) have proven to be more controversial and enigmatic, principally due to differences between experimental animal models and human disease pathology.

Murine models of immune-mediated inflammatory diseases are acute and fail to replicate the chronic human disease generally characterised by cycles of "flare"

in disease activity and subsequent remission, limiting their translational capacity to human disease. Infection and tumour models in mice either resolve too quickly for any chronic pathology to be established, or necessitate euthanasia for welfare purposes prior to the establishment of TLT pathology. This contrasts sharply with ostensibly similar human pathology: humans may live the rest of their life with chronic disease pathology, particularly in the context of treatment with biologics (e.g. monoclonal antibody therapies) and small molecule drugs, and this permits pathology the opportunity to evolve from localised to systemic inflammation, including fibrotic tissue failure, and autoimmunity working in synchrony to prevent disease resolution. Human three-dimensional tissue culture models (**???**) containing both stroma and lymphocytes have become increasingly common and useful in understanding molecular mechanisms driving TLT formation. However, it is not currently possible to represent the full complexity of chronic human pathology *in vitro*, and additional means of understanding these complex disease processes are required.

In humans, tertiary lymphoid tissues (TLT) are found in inflammatory immune responses associated with chronic pathologies related to hip joint replacements, keloids, tissues in autoimmune disease (e.g. the salivary gland in Sjögren's syndrome or articular joint synovial lining in rheumatoid arthritis) to solid tumours and follicular lymphomas in the bone marrow (**????**). Although the role of specific cell types has been controversial, a paradigm has begun to emerge centred on a multi-step process in which localised inflammation induces stromal cell activation in a lymphocyte-independent process, thus leading to localised micro-environments permissive for T and B cells entry.

Computational and mathematical modelling of lymphoid tissues may be broadly broken down into a several key areas, each of which has the ability to address crucial questions required to develop a comprehensive mechanistic understanding of tissue form and function - these areas include models concerned with: cell motility and the reticular network, B and T lymphocyte dynamics (proliferation, activation, de-

velopment of effector functions), tissue organisation into anatomical niches, and the role of lymphoid tissues within the context of infection or other disease processes. This section aims to briefly consider significant and recent work addressing each of these areas using a variety of mathematical and computational modelling techniques paired with experimentation, model hybridisation is then considered within the context of lymphoid tissue modelling and QSP (quantitative systems pharmacology): the incorporation of different modelling methodologies to capture, with varying degrees of abstraction, multiple aspects of a complex system to address difficult questions and drive understanding.

Two-photon imaging permitted a new approach to generate data on *in vivo* cell motility, as it allows tracking of cell motion within living organs with minimum disturbance from external factors, and as such, observed behaviour can be considered to be representative of genuine *in vivo* behaviour. Two-photon tracking of B and T lymphocytes in mouse lymph nodes (**?**) appears to favour a lymphocyte motility model in which a random walk dominates cell motion (**?**), despite the lymphocytes' expression of chemokine receptors and the abundance of chemokine within the lymph node micro-environment (**??**). To address these conflicting data, **?** developed a theoretical model of lymphocyte motility using the extended Potts model in 2D, a technique well-suited to describing cells at the sub-cellular level including cell morphology, surface molecule expression and additional internal structure utilising a lattice of volumetric elements with a defined spin state – this may generally be described as a potential-based thermodynamic model in which cell motion (or indeed any lattice changes) are determined by a Boltzmann law in which contributions to the energy term of the Boltzmann exponential define cellular dynamics and interactions. A broader description of simulating cellular dynamics using extended Potts models may be found in **?**. Upon development of an extended Potts model describing lymphocyte motility that broadly considered orientation persistence in combination with a random walk description, and assumed motility induced cell elongation, **?** interpreted the two-photon data from **?** within the framework of this

theoretical model, and found that lymphocytes exist in a single velocity state with a highly variable, stochastic, cellular diameter, and that their assumptions were consistent with the experimental data.

More recently, this view of lymphocyte motility within lymphoid tissues was extended (**?**), particularly with respect to B and T lymphocyte migration within GCs, in light of additional two-photon imaging data (**?**). The model presented in **?** still quantitatively explains imaging data in terms of random walks with persistence of motion, as in the earlier model described in **?**, however it further predicts that chemotaxis has an active role in maintaining GC dark and light zones and predicts that chemokine sensitivity is quickly down-regulated through receptor internalisation. The resulting model combines a general random walk with persistence with an important role for chemotaxis in maintaining zoning within the GC – complicating the more simplistic understanding developed originally (**?**). The pairing of mathematical modelling and two-photon imaging in **?** identified some statistical limitations for the interpretation of two photon cell motility measurements generally, thus indicating that the reality is likely more complex than first predicted. This migration model integrates multiple mathematical techniques that have been applied to modelling GC reactions, including temporal ODE systems that describe lymphocyte recycling rates between GC zones (compartments), antigen uptake by centrocytes, and T—B lymphocyte interaction kinetics (**???**) and computational models addressing the spatiotemporal aspects of T—B lymphocyte kinetics, based primarily on the cellular automaton model described by **?** hybridised with the cell motility extended Potts model described by **?**.

This earlier work modelling GC reactions in lymphoid tissues (**????**) eventually led to an integrated, functional theory and model of B-cell selection, division, and exit within GCs, supported by T-B lymphocyte interactions (**?**) incorporating models describing two-photon motility data including both 'random walk with persistance' models, chemotactic responses to chemokines, and chemokine desensitisation through receptor internalisation and many other phenomena (**?**), with the
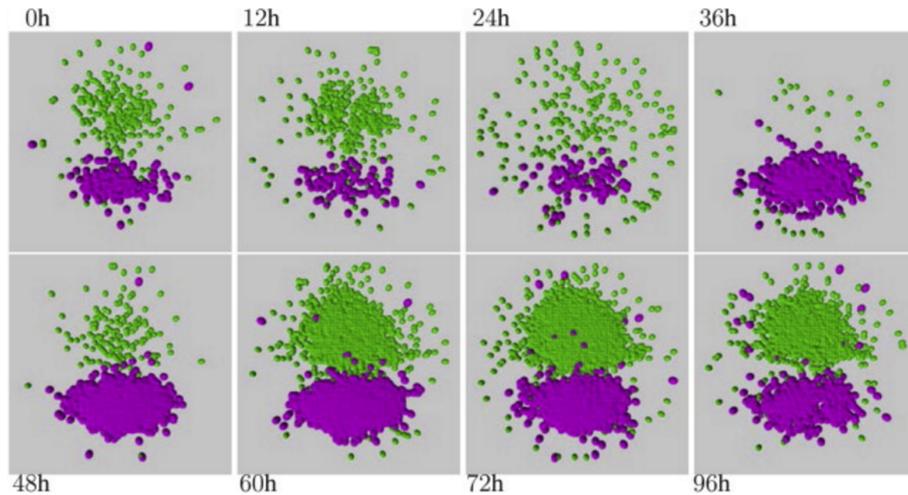
Figure 1.7: Visualisation over 4 days of an instantiated GC in the model described by **?** following adjuvant challenge, wherein the model assumes cyclic re-entry is the dominant pathway of positively selected B cells supported by affinity-dependent activation by T follicular helper cells . Dividing B cells are coloured magenta and non-dividing B cells are depicted in green. Adapted from **?**

resulting B cell projection visualisation over 4 days demonstrating emergent organisation properties arising from this model (Figure 1.7). This model incorporated a novel 'stochastic event generator' (**?**) which represents each cell within the model explicitly in both space and time, with discrete implementations of the previously developed differential equations permitting heterogeneous responses within the model. The fundamental cell states permitted in this model were differentiation, mutation, division, and spatial position, with lymphocyte shuttling and maintenance of GC zoning emerging from interactions of many non-linear interactions. The 'stochastic event generator' in combination with the mathematical descriptions derived from the extended cellular Potts model essentially results in a highly implementation-specific instantiation of what may be considered broadly equivalent to an 'agent-based' model (although it is not described as such), relying on mathematical descriptors of cell behaviour rather than state machine descriptions, yet representing heterogeneous cellular dynamics explicitly in time and space. It is important to note that this model of GC maintenance may be considered an early example of 'large-scale' model hybridisation, although in this case much of the integration between model types is implicit as there lacks a clear segregation of each 'sub-model' and the

means by which they were integrated into a single coherent simulation. The notion of 'hybridised' models composed of discrete, integrated 'sub-models' is introduced in Section 1.3.2. It was noted by **?** that the 'basic philosophy of the model is to implement known mechanisms and to use the mathematical model for the analysis of the whole interacting system. Thus, the read-outs of the system are not set by hand but instead emerge': a philosophy at the core of using agent-based simulation to understand complex dynamical biosystems (**?**).

Secondary and tertiary lymphoid tissues are supported by a 'scaffold' of mesenchymal stromal cells which form a collagen-based reticular network that provides for antigen-transport, an environment conducive for leukocyte adhesion enabling lymphoid-stromal crosstalk, and a support structure to assist in cell migration within the tissue. The stromal network is dynamic and undergoes remodelling in response to infection, allowing lymph nodes to adaptively grow and shrink in size and number of follicles, and is therefore crucial in the maintenance of immune homoeostasis **?**.

Given the crucial importance of the stromal network in lymphoid tissue function, it is not surprising that a number of mathematical and computational models have been developed to address various open-questions regarding its precise role. A 3D dual cellular automaton model was developed by **?** which describes cell movements and interactions along the FRC network ('T-cell stroma') within lymphatic tissues, the authors define one automaton which creates and stores the location of FRCs and their connecting edges – defining the network, and another that stores the location of DCs and naive T cells, and allows them to move in 3D such that T—DC interactions can be measured (Figure 1.8). Stromal network density changes significantly during LN development and remodelling, and the FDC network supporting B-cell follicles is much denser than the surrounding FRC network (**?**). It was unknown to what extent the change in network density modified the dynamics of cell—cell interactions within the tissue, as naive T cells scan for antigen within a secondary lymphoid tissue it is important that they can efficiently cover as much of the APC population as possible, particularly as the cognate TCR clonotype for a specific antigen challenge
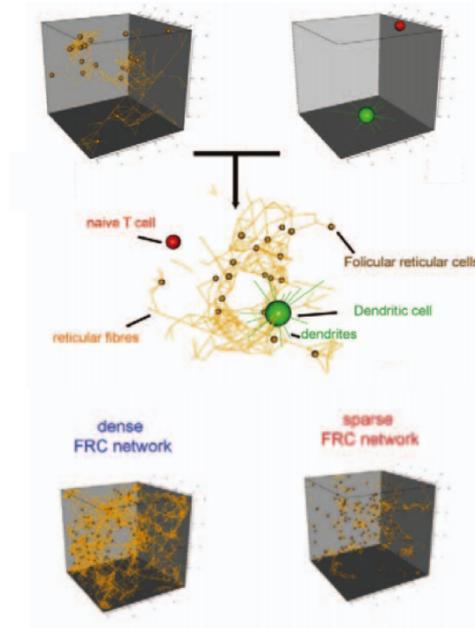
Figure 1.8: Model and simulation of cell—cell interactions on the FRC network (**?**). The model consists of two cellular automata, one defining FRCs and the connecting edges that create the network structure, and one defining the location of DCs and antigen-scanning naive T cells. These automata interact as the naive T cell moves within the 3D network space. The bottom 3D plots show FRC networks generated with different densities, enabling the rate of cell—cell interactions as a function of network density to be quantified mathematically. Adapted from **?** (Figure 1 A and B).

may be extremely rare. Such questions are well-suited to mathematical analysis, as network structures with varying densities may be constructed using a broad array of algorithms; in addition, the inherent heterogeneity of stromal network densities *in vivo* and their dynamic nature makes this a difficult question to address experimentally (*praeter silico* - with techniques besides computational modelling). There has however been some success in constructing FRC networks *in vitro* by seeding onto a polyurethane macroporous scaffold (**?**), which could potentially offer a means for independent validation of theoretical FRC network models by modifying the FRC seed dosage to generate tissue-engineered scaffolds of different FRC densities.

The 3D cellular automaton model (Figure 1.8) was able to demonstrate that the presence of an FRC network increased the probability of two cells coming into contact by 25%, or as much as 40% under the assumption that the FRC network

contributes to the motile cell velocities. Surprisingly however, FRC network density was found to have only a small effect on the probability that a motile cell (e.g. a naive T cell) will come into contact with either a static or motile 'target' cell (e.g. an APC). This probability is reflective of scanning efficiency, higher contact probabilities would suggest an antigen-scanning T cell is more likely to successfully activate and promote an effective immune response. However, it was found that disruption to the FRC network (a loss of edges between nodes resulting in reduced network connectivity) may have a detrimental effect at physiologically relevant FRC network densities, specifically, those found in a naive murine spleen, but still found this to be relatively minor. Since **?** determined network density is unlikely to have a biologically significant effect on cell—cell interactions, the authors observed that the mean diameter between two filaments connecting FRCs is approximately equal to the mean T cell diameter, and therefore hypothesised that the presence of the network encourages efficient T-cell flow through the spatial environment by essentially creating channels within the LN paracortex. This notion is consistent with the earlier observation by **?** that T cells tended to move in parallel planes during multi-photon imaging of cell motility, used as justification for describing the dynamics in a 2D plane.

1.2.5 Quantitative Systems Pharmacology in Simulation Modelling

Quantitative Systems Pharmacology has emerged in recent years as an approach that aims to inform 'bottom-up' modelling and simulation with "Big Data", producing a synergistic relationship between traditional pharmacology and computational biology to investigate complex pathological processes and develop novel therapeutic interventions. It was argued by **?** that this synergy is one of compromise, that models should seek to be as least complex as possible, with fewest parameters, with respect to the data available. That is, models should be precisely as complex as is necessary and not more so. It is this concept that drives the development of the TLT

formation model outlined in Chapters 3 and 4 - the aim is to identify the *minimum requirements* capable of describing the process, rather than attempt to include all known molecular pathways, factors, and cell phenotypes involved.

The model-based drug discovery and development paradigm is gaining traction in the pharmaceutical industry. There has been a near constant flow of new terms introduced into the literature (**?**) in an attempt to capture this phenomenon: 'MBDD' (**?**), 'model-facilitated/informed drug development' (**?**), 'Quantitative and Systems Pharmacology' (QSP) (**?**), and 'pharmacometrics'; furthermore, large pharmaceutical companies have begun to review, quantify, and report the successes derived from the adoption of a model-based strategy, providing a thorough description of its implementation and impact (**???**). The US Food and Drug Administration (FDA) recently utilised mechanistic model-based methodologies to design a post-marketing clinical trial (**?**), providing a clear indicator to practitioners within computational biology and QSP of the increasing confidence in, and prevalence, of model-based techniques in pharmacology, drug-development, and biomedical research generally. Although improvements in our ability to demonstrate evidence of fitness-for-purpose within computational models taking cues from safety-critical systems (**?**), there is space for vast improvement in our ability to describe, implement and experiment with complex models of even more complex pathological processes. The increasing availability of large human data sources permit a lesser reliance on translation from *in vivo* to human models of disease, as the many issues associated with this have been well-addressed in the literature, for detail review on this topic the reader is referred to **???**.

Forming a key component of the 'learn and confirm' drug discovery and development paradigm, many of the methodologies utilised to support the deployment of modelling and simulation-based strategies are recurrent across the pharmaceutical industry (**?**). A non-exhaustive list of such methodologies include pharmacokinetics and pharmacodynamics modelling (PK/PD), statistical design methods, and signalling network reconstruction methods, and increasingly - complex systems and

agent-based approaches (**?**). Such techniques can be applied across many stages of the drug development process, with the capacity to inform experimental design, 'go or no-go' decisions, preclinical development, and optimal portfolio prioritization. Wider adoption of *in silico* modelling for novel therapeutic design requires the development of techniques capable of assessing whether a putative target will yield a desired disease outcome (**?**). Describing heterogeneous biological systems with genomic, transcriptomic, metabolomic-scale events occurring over length and time scales spanning orders of magnitude, within an individual model is an important challenge in model-driven target evaluation and selection, and must ultimately remain a key end-goal within QSP as an emerging field. It is argued in **?** that QSP should become a the 'central pillar of translational medicine', incorporating multi-scale spatiotemporal models with genomic medicine, developing new approaches to understand drug mechanisms of action spanning different levels of detail, complexity, and scale: becoming the key driver in advancement of drug discovery and development through an integrative multidisciplinary approach. The data in Figure 1.9, taken from **?** show over a ten year period common causes of attrition in the drug discovery and development process. These data demonstrate a decline in the effectiveness of PK in addressing toxicity and efficacy, and the authors argue overall success of the "traditional" PK approach had decreased from 1991 to 2000. Given the ever increasing complexity that modern QSP and multi-scale modelling approaches are revealing, and the increasingly difficult task of identifying therapeutics that are safe and efficacious, it is clear that hybrid complex systems modelling approaches will be crucial in maintaining the historical success of medicine and pharmacology in developing therapeutic interventions.

One of the key advantages of applying multi-scale, hybridised modelling approaches permits capture of a wide range of phenomena that occur at differing time and length scales, including different cellular phenotypes and interactions, inflammatory molecules, chemotactic signals in the context of complex tissue micro-environments in a format that permits *in silico* experimentation through the tem-
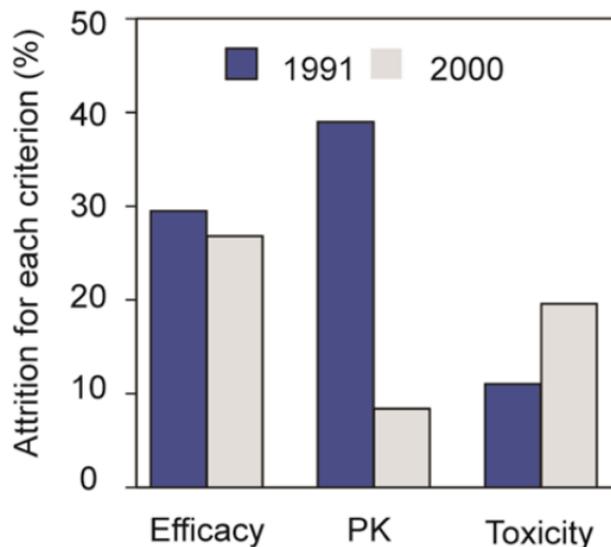
Figure 1.9: Attrition causes in drug discovery are illustrated over a ten year period, 1991-2000. While great success has been found with PK approaches incorporating modelling to address issues of drug bioavailability, attrition in efficacy and toxicity are not being adequately addressed by traditional PK modelling approaches. This figure was taken from **?** whom adapted it from **?**; the reader is referred to the latter reference for a complete account of the data from which this figure was compiled.

poral inhibition of different signalling processes during key stages in TLT pathology using various statistical and data driven approaches for the determination of effective means of modulating TLT pathology within immune-mediated inflammatory disease. This permits identification of pathways that could be targeted to induce resolution rather than formation. We have previously addressed methods of determining the roles and modulating the presence of signalling pathways of therapeutic interest within the context of established, calibrated models of pathological processes (**?**). Subsequently, this led to **?** presenting a methodology grounded in agent-based computational modelling using models that capture the heterogeneous cellular processes which drive disease pathology and resolution, specifically in relation to simulations of inflammatory disease manifestation in order to determine the likely, dose-dependent efficacy of potential intervention strategies *in silico*.

A large number of novel antibody therapies, biologics and small molecule inhibitors have been developed to target immune function for the treatment of immune-mediated inflammatory diseases (**???**). These therapies are unlikely to show maxi-

mal efficacy against existing tissue pathology when used as mono-therapies, use of selective therapeutic combinations derived from a detailed mechanistic understanding of the disease pathology are rather more likely to show clinical efficacy. The principal challenge is determination of the most effective and safe therapeutic combinations, performing a clinical trial for the thousands of possible combinations is intractable, however integrated computational modelling and experimentation based approaches provide a rational approach for providing the identification of candidate therapeutic regimes with a body of evidence to support their use, allowing the most effective candidates to be translated into clinical use.

1.2.5.1 Evaluating Efficacy of Therapeutic Interventions Incorporated into *in silico* Models

Utilising agent-based and other complex systems modelling approaches, there has been some success in the development of mechanisms to implement and evaluate putative therapeutics *in silico*. In one example, utilising ARTIMMUS (**??**) , an agent-based simulation of murine experimental autoimmune encephalomyelitis (EAE), a murine disease model used for the study of multiple sclerosis in humans . We explored the potential cell-level consequences of CD200 regulation of dendritic cells (DC) (**?**)[1]. Recent evidence suggests a potentially significant role for the membrane glycoprotein CD200 in regulating T-cell priming outside the central nervous system (CNS) (**?**). However, the downstream effects of CD200 down-regulation on DC behaviour remain unclear. Two mechanisms were investigated in response to CD200 signalling: a reduction in DC priming capacity of T cells, and, the promotion of DC type II cytokine secretion. It was determined that CD200 down-regulation of DC priming capacity promotes autoimmune-mediated disease processes. This hypothetical mechanistic consequence of CD200 down-regulation does not support recovery from autoimmunity individually, however, it was identified that resulting DC cytokine profile switching substantially promotes type II deviation of the au-

---

[1]My contribution to the work described in **?** was the development of a novel model of two modes of action CD200 and its principled incorporation into the extant ARTIMMUS simulation

toimmune response, wherein CD8 Tregs are unable to directly apopotse Th1 cells. Thus, CD200-supported a type II deviation and, in the context of the ARTIMMUS model of T cell signalling in EAE, strongly supports recovery from disease pathology. These two potential mechanisms, while in isolation drive the system toward increased pathological activity or disease resolution respectively, in combination they support the view that CD200 regulation can promote recovery from autoimmunity – a position supported by additional literature (**???**).

This approach to identifying and exploring drug targets was extended when the role of anti-CD3 biologic therapy was assessed in EAE, described in **?**. Additionally, knock-out experiments can be replicated *in silico* to further investigate the role of key signalling pathways and disease mechanisms, for example key knock-out experiments were reproduced using PPSim, an agent-based model describing murine Peyer's patch formation (**?**) and using ARTIMMUS (**????**). Such methods of investigation has become a mainstay for the exploration of disease pathology within the context of agent-based models, however there lacks an integrated approach to identifying the relative efficacy of multiple candidate therapies, although there have been important developments in the capacity to assess temporal effects within ABMs (**?**).

1.3 Determining Appropriate Modelling Methodologies

The capacity of various modelling techniques to capture explicit notions of space and cellular heterogeneity is illustrated in Figure 1.10. Ordinary Differential Equations (ODEs) and Physiologically-Based Pharmacokinetic (PBPK) models cannot spatially resolve systems, although multi-compartment models exist in which several spatially connected regions are connected, they rely on the abstract notion of well-mixed space within each compartment. Partial Differential Equations (PDEs), and thus also systems of coupled ODEs, are capable of spatial resolution. To capture heterogeneous cellular phenotypes, however, 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. Heterogeneity within mathematical and computational models can be created through the introduction of stochasticity, or "randomness", for example by defining events probabilistically rather than through purely deterministic models.

| | | SPATIALLY RESOLVED | |
		NO	YES
CELLULAR HETEREOGENEITY	**NO**	**PBPK ODE**	**PDE**
	YES	**State-Based Model**	**ABM HYBRID-ABM**

Figure 1.10: Illustrative table describing the capacity for various model forms to describe spatial resolution and cellular heterogeneity. In determining appropriate modelling methodology, it is important to consider spatiotemporal scales relevant to the system and establish any requirement for heterogeneity across model entities. Figure previously published in (**?**).

1.3.1 Describing Spatiotemporal Systems using Agent-Based Models

An agent uses a pre-defined rule-set to assess its internal state in response to factors in the agents local environment or neighbourhood. Should an agent be in a

situation where the requirement of a rule is met, whether due to a change in the agents attributes or within a set location in the environment, the state of that agent is changed. An agents rule set can range from simple Boolean statements operating over the agents attributes, to more sophisticated mechanisms that relate agent inputs and outputs through differential equations (**?**) and metabolic models (**?**). Agent rules also offer a means of introducing stochasticity through probabilistic events utilising Monte Carlo methods, allowing for an approximation of behaviours in systems whose complexity precludes deterministic modelling. The rules governing an agent's behaviour may be described using finite state machines (FSM) expressed using the Unified Modelling Language (UML).

1.3.2 Multiscale Modelling and Model Hybridisation

Multiscale modelling permits one to describe different aspects of the domain on time and length scales of different orders of magnitude. In 2013, Martin Karplus was awarded the Nobel Prize in Chemistry for 'the development of multiscale models for complex chemical systems' (**?**). This work in multiscale mathematical modelling formed the basis for important theoretical developments in molecular physics (**??**), chemistry (**??**) and biochemistry (**?**), leading to the development of increasingly accurate models – capable of describing phenomena from both the quantum regime and classical mechanics. This work led to the genesis of a large-scale collaboration in interdisciplinary science applying the methods and concepts behind the work of Karplus to fields as diverse as economics (**?**), sociology (**?**), and as computational resources became more abundant, sophisticated complex systems models of biological phenomena (**??**).

A major contribution in multiscale modelling came from **?** in the development of a lysozyme, an important catalytic enzyme for the destruction of bacteria cellular membranes, in which it cleaves a glycoside chain. A purely classical molecular mechanics model of this system was insufficient to describe the function of the lysozyme enzyme, however a full description in quantum mechanical terms was com-

putationally intractable in the 70's. Therefore, Warshel *et al.* developed a modelling methodology that permitted description of most of the molecule using classical mechanics, while critical parts of the system were treated quantum mechanically, this 'QM/MM' (Quantum Mechanics)/Molecular Mechanics) molecular dynamics simulation constitutes the first form of hybridised modelling. A cartoon of this can be seen in Figure 1.11.
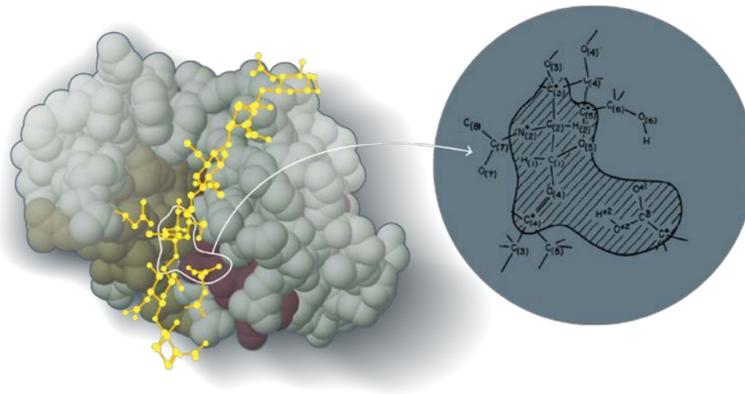
Figure 1.11: A cartoon of lysozyme cleaving a glycoside chain. The region shown in the circle on the right is a magnified part of the lysozyme; the shaded region is treated quantum mechanically, whilst the remainder of the system is treated using classical mechanics. These two theoretical systems describe phenomena on different scales, and therefore constitute an early multiscale model. The classical regime is computationally more tractable, and is essentially the limit of quantum theory as objects become sufficiently macroscopic. Figure taken from (**?**)

Model hybridisation has superficial similarities to multiscale modelling, insomuch as developing multiscale models may require different techniques to capture systems on different spatiotemporal scales **??**. However, this is not a prerequisite in the development of a multiscale model. For example, the QM/MM lysozyme model described by **?** combined molecular mechanics and quantum mechanics, but both of these theoretical frameworks are expressed using partial differential equations. Conversely, hybrid models combine different modelling techniques, or modalities, into one overarching model.

There has been important work as early as (**?**) ('Mycobacterium tuberculosis as viewed through a computer'), and additionally in (**?**), (**?**), (**?**), (**?**) and (**?**); hybrid
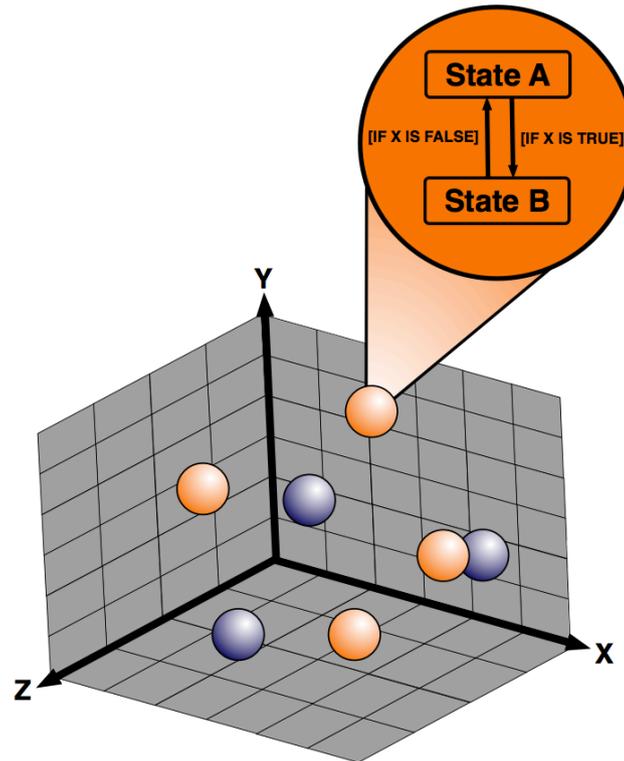
Figure 1.12: Structure of an Agent Based Model: Agents (shown as blue and orange spheres) are individual entities capable of modulating their associated states and attributes with respect to their local environment and pre-defined rules governing agent behaviour. Agents are expressed explicitly in space (grey grid lines), and captured at a specific time, which provides a context for their interactions. The aggregate behaviours of the agents can then lead to the emergence of complex patterns and behaviours. Taken from **?**

modelling methodologies are principally concerned with describing differing levels of abstraction at multiple time and length scales simultaneously.

1.4 Principled Approaches to Modelling and Simulation

A model is only useful if one has confidence in its results, and if the evidence-based for confidence in the model and its results can be demonstrated and communicated to others. Therefore, in using a simulation as a tool for understanding biological systems, it is important that the relationship between the simulation and the sys-

tem it is designed to describe is appreciated. Due to their nature, ABMs or highly hybridised models cannot be expressed as succinctly as other modelling approaches such as sets of differential equations, this leads to the need for appropriate documentation describing the model from the biological system it encapsulates, the implementation and parameterisation decisions taken, and the supporting evidence-base. Producing documentation that is both clear to an interdisciplinary team, yet concise enough to be practicable is non-trivial. The challenge of communicating agent-based models to non-specialists has been highlighted in the literature (**??**). To ameliorate this, visual notations can ensure that the design process is...

1. ...simple for both domain experts and model developers to interpret, with a short learning curve.

2. ...explicit and formal such that it may be interpreted objectively, not subjectively. Model parameterisation should be clearly justified.

3. ...accessible: model documentation should be made available, with supporting documentation.

Utilising a principled model design and development framework, with commonly agreed means of specifying models throughout the development process, ensures that evidence for confidence can be provided. Unfortunately, there is not yet a single agreed upon modelling framework, however several have been proposed in recent years, some of which are discussed in Section 1.4.1. When employing a principled framework, the data used to inform the model must be appropriately justified for the research context. Different transgenic models of disease can shed light on key processes and interactions but may have altered dynamics which are not a true representation of the underlying disease. A rigorous justification process can be used to examine the suitability of data used to inform a model (**?**), this is particularly pertinent in cases where the modeller does not have an in-depth understanding of the nuances associated with different disease models. Such a modelling development

process also exposes the decisions made in its design and implementation to the community for scientific scrutiny.

1.4.1 Model and Simulation Development Frameworks

When developing a model of a complex biological system, the relationship between the simulation implementation and the biological model needs to be rigorously defined, described and clearly communicated. This is necessary to enable robust justification of the means of implementing specific aspects of the biology and abstractions made during model development. Crucially, a *domain* model should be defined solely in the terms of biology, be it embedded within diagrams or as prose, before a specification for implementation as an executable model is developed, and should be developed to answer a specific *a priori* defined research question or questions.

Various frameworks have been established that aim to enforce a robust model development methodology and emphasise the need to identify a specific research question for which a model is designed to answer. ODD (Overview, Design concepts, and Details) has been proposed as a three-block standardised protocol for describing mathematical and computational models (**?**). The three blocks are sub-divided into seven stages: Purpose, State variables and scales (including parameterisation), Process overview and scheduling, Design concepts, Initialization, Input, and Submodels. These stages are well-defined in terms of when and how each aspect of the model should be defined, and their development provided an important first step towards establishing a common framework for ABM development. However, although ODD focuses strongly on rigour in describing the model, it does not address the need for similar rigour in model implementation, verification, parameterisation and parameter calibration, validation, evidencing fitness-for-purpose and *in silico* experimentation. Parameter calibration in the context of complex-systems agent-based models is the process by which parameter values are found that permit emergence of expected behaviours of the system at the macro level. Recent developments are exploring means of 'automated calibration' (**?**), applying multi-objective optimisa-

tion specifying experimental data as objectives against which parameter values are optimised for using linear programming or genetic algorithm approaches. This is however a recent and experimental development, with the mainstay of agent-based model calibration to date being dependent on 'by hand' perturbation of parameters to calibrate the model to expected behaviours (**??**).

A framework for the development of complex systems models, independent of both domain or modelling techniques, termed the CoSMoS (Complex System Modelling and Simulation) process (**?**) has since evolved. The CoSMoS process comprises an iterative process of model refinement and implementation, in collaboration between domain expert and modeller. The process broadly decomposes an entire model into four separate entities, termed 'domain model, 'platform model, 'simulation platform, and 'results model. The domain model describes the aspects of the biological system required to produce observed phenomena, explicitly in only biological terms and without implementation considerations such as code or mathematical equations. The domain model will contain some degree of abstraction from the biological system under study, but these should be to permit a self-contained description of the system and not to permit tractability *in silico*. The platform model describes how the domain model is to be implemented as an executable simulation, with many similarities to a software specification. The simulation platform itself represents the executable simulation - a specific instantiation of the platform model. The understanding derived as a result of the interpretation of the collective results from *in silico* experimentation using the simulation tool comprises the results model. Figure 1.13 shows a cartoon summary of the CoSMoS framework (rectangles) and tools that support its use (ovals). Parameterisation within the CoSMoS process occurs predominantly at the 'domain' and 'platform' stages of modelling. As an iterative process, the way in which these models are parameterised may change as necessary to ensure the developed simulation platform is fit-for-purpose in addressing the *a priori* identified research questions.

The CoSMoS framework provides a rigorous approach to developing a model
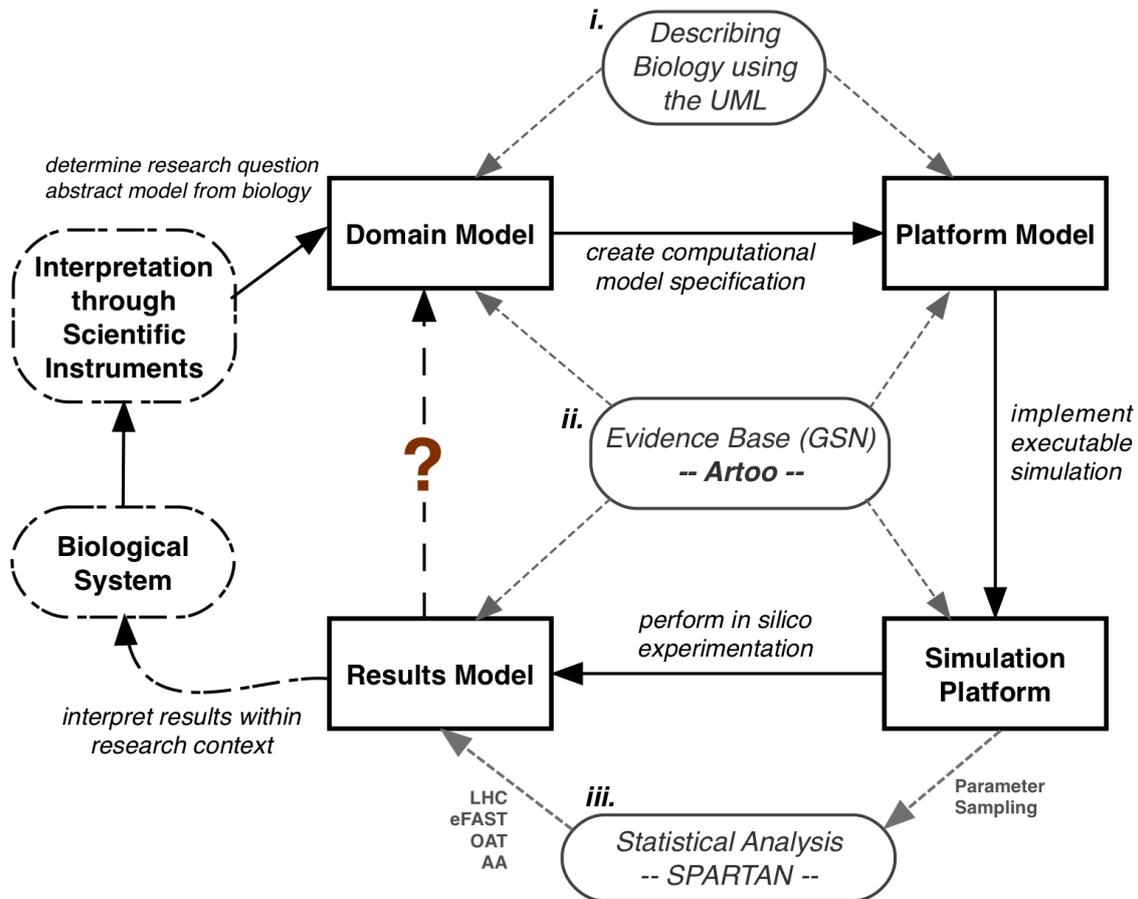
Figure 1.13: The CoSMoS framework is depicted with the tools currently used to support its adoption within a model. The framework consists of the iterative process of Domain-Platform-Simulation-Results modelling. The ovals show tools and techniques used to support the utilisation of this framework during biological model development: i) describing biological systems using the unified modelling language (**?**), ii) describing the evidence base and providing an argumentation structure that the model is fit for purpose using GSN (**?**) with *Artoo* (**?**), iii) statistical analysis of simulation results using the SPARTAN package (**?**) in R.

with confidence in its results, and places strong emphasis on designing models to address specific aims or research questions defined *a priori*. Thus, a model should not attempt to be a general purpose description of a system, as research context is crucial in determining what aspects of the system to abstract and by what means model entities should be implemented. A model ought never be re-purposed beyond its original scope or within a different research context, unless a principled methodology for model extension is followed (**?**). For a comprehensive description of the CoSMoS process, the reader is directed to **?**. The model used as a case study in Chapter 2

developed by **?**, and the model developed in Chapters 3 and 4, both utilise the CoSMoS process in their development and analysis.

1.4.1.1   Model Identifiability and Parameter Estimation

Model identifiability is concerned with parameter estimation, and therefore model calibration. A model is deemed as identifiable if it is possible to precisely infer the true value of the model parameters, and for this to be the case there must be a unique solution, and changes in parameter values must result in changes in observed probability distributions. There is a wealth of literature concerning model identification in dynamical systems across many domains. **?** describe model identifiability in the context of non-linear ODE models of viral dynamics, in which identifiability analysis is discussed as an important and necessary process to determine unknown parameters in ODE models based on experimental data, however the authors note that such analysis techniques for non-linear ODE models are still under development.

In the context of systems biology, there is some literature concerning model identifiability, for example, in gene regulatory networks (**?**), biochemical networks (**?**) and signal transduction networks (**?**). In **?**, the authors argued that parameter estimation in complex systems biology models is a "major obstacle" to model development, largely due to difficulties in measuring *in vivo* parameters and the large uncertainties yielded by fitting to experimental data. **?** assessed 17 systems biology models, and found that the obtained parameters were 'sloppy', or not well-defined, and therefore the models were universally unidentifiable. However, an argument was also made that obtaining precise parameter values is not required to draw meaningful biological conclusions.

There is a dearth of literature on model identifiability in agent-based models, but important early work has been undertaken to improve the process of parameter estimation, for instance through the application of multi-objective optimisation (**?**), and using genetic algorithm approaches (**?**). Such approaches require that *in vivo* data are available for which precisely matched model outputs can be produced. This

is especially challenging in the case of TLT formation, as enormous heterogeneity is observed in the size, frequency, organisation of these tissues, and number and density of lymphoid follicles that form within them. Such data are not available, and the requirements for the formation of TLT are unknown. Therefore, for the model of TLT formation developed in this thesis (Chapters 3 and 4), it can be stated *a priori* that the model will be unidentifiable. The model must therefore be treated as a theoretical construct aimed at exploring whether the hypotheses developed in Chapter 3 are generally consistent with the observable *in vivo* organisation that takes place over a 15-day period, in a generally qualitative sense. Thus, the model outlines will be semi-quantitative, in that outputs produced are numerical and lead to testable predictions *in vivo*, but that model identification and precise parameterisation will not be possible.

1.4.1.2   Validating Theoretical Models

A complex biological model may be entirely theoretical in nature, that is, driven by a novel theoretical framework that is capable of describing the emergence of a complex biological phenomena in a manner somewhat analogous to theoretical physics. Validating such models, in addition to the use of argument-driven validation and principled development frameworks, is also a case of identifying whether the theory, model and resulting simulation produces *a priori* defined expected behaviours. For the TLT model developed in Chapters 3 and 4, much validation comes from domain expert observation of simulation results with respect to available experimental evidence (as outlined above in Section 1.4.1.1). This is particularly the case when it is not possible to define any metric for model "precision" due to lack of knowledge of variability of the domain-under-study. For example, the variation in number, size and distribution of B-cell follicles in TLT is currently unknown, as are cell population sizes, and so on, and thus validation relies somewhat on observation that the theoretical system produces behaviour similar to that which is observed *in vivo* from the perspective of "opinion leaders" and domain experts. Such validation can

be enhanced by making novel predictions *in silico* about the domain-under-study using the theoretical model and then performing *in vivo* experimentation to evaluate if system responses *in silico* and *in vivo* are qualitatively or semi-quantitatively similar.

1.4.1.3 Demonstrating Model Fitness-For-Purpose Through Argument-Driven Validation

A computational model of biological phenomena may be considered as analogous to a safety-critical system, given that at some point decisions regarding interventions, clinical trials or future research direction could be based on results from the model. It is important therefore to develop an argumentation structure that can present a case that the simulation is valid and appropriate given the experimental scope. Goal Structuring Notation (GSN) is a technique developed originally in the context of arguing safety of aircraft by **?**, and was later introduced to biological simulations by **?**. Figure 1.14 illustrates a sample GSN argument, with each element describing its purpose in the context of arguing fitness.

1.5   Thesis Aims

This thesis is concerned with the development of complex systems models of immune pathophysiology; in particular, the development of techniques to improve model granularity, visualisation and quantification methods to permit intuitive understanding of many highly complex multidimensional datasets. The development of visualisation techniques consists of two strategies. Firstly, we look toward experimental biology to develop emulations of *praeter silico* (non-computational) experimental techniques and data interpretation; this is explored in Chapter 2. Secondly, unsupervised machine learning analysis techniques are used to explore model response to therapeutic intervention through manifold learning with Kohonen networks, these are introduced in Chapter 5 and applied to therapeutic efficacy in
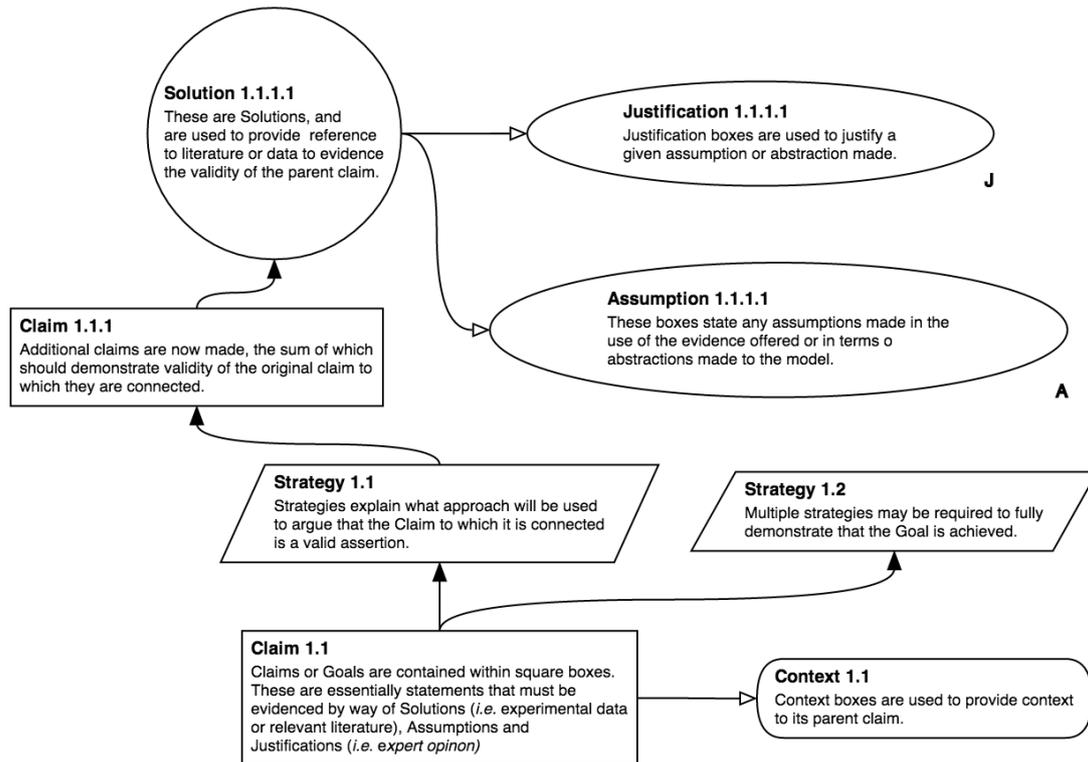
Figure 1.14: An example GSN argumentation structure. All GSN drawing elements have been included, with the contents of each explaining its purpose within the argument-driven validation framework.

Chapter 6, permitting visualisation of the heterogeneity of simulation outputs with different therapeutic intervention strategies. We aim to present a flexible methodology and scheme for integrating multiple modelling techniques into an overarching hybrid multiscale model and executable simulation that reproduces the domain system at the molecular, cellular and tissue level across spatiotemporal scales spanning orders of magnitude. Furthermore, novel approaches to producing simulation outputs that are analogous to those used in experimental biology are described, helping to create a well-defined link between the *results model* and the *domain model*, improving validation and assisting in effective communication of model results.

The following list enumerates the principal aims of this thesis, and is succeeded by further exposition regarding each item including where within this thesis evidence and discussion may be found fulfilling the aim.

1. Derive a hypothesis-driven model of tertiary lymphoid tissue (TLT) formation

during autoimmune disease for use as a case study in applying the result of aims 2 and 3 below.

2. Development of a schema and novel paradigm for developing a highly integrated hybrid multiscale model.

3. Development of improved methods for visualising simulation outputs that are analogous to those utilised in experimental biology.

4. Determine the veracity of the model hypothesis describing the minimum requirements for TLT formation.

5. Utilize the TLT formation model to evaluate the efficacy of therapeutic interventions for Sjögren's syndrome *in silico*.

1.5.1   Thesis Structure

The aims described in Section 1.5 are addressed over seven chapters, as set out below:

Aim 1: Derive a hypothesis-driven model of tertiary lymphoid tissue (TLT) formation during autoimmune disease for use as a case study in applying the result of aims 2 and 3

Chapter 3 presents a TLT formation domain model, and Chapter 4 describes the Platform Model/simulation implementation. The model is analysed and predictions derived thereof in Chapters 5 and 6.

Aim 2: Development of a schema and novel paradigm for developing highly integrated hybrid multiscale models

Chapter 4 describes the development of an approach to succinctly yet briefly summarise complex hybrid models sharing inputs and ouputs. When attempting to describe a mathematical model constructed from 'sub-models', with each representing a specific biological entity or process, it quickly becomes difficult to understand

and mentally maintain the model network topology and we lack notation to succinctly describe the model structure and organisation. Therefore it is contended that the most informative representation should be centred on information flow between sub-models. Each input/output across the various sub-models can be defined using arrows between sub-models defined within boxes. An example of such a diagram is presented in Figure 4.2.

Aim 3: Development of improved methods for visualising simulation outputs that are analogous to those utilised in experimental biology

Chapter 2 describes the development of a set of methods for emulating immunohistochemistry, flow cytometry, and gene and protein expression analysis. These are applied in a case study using a simulation of murine Peyer's patch development during organogenesis. A demonstration is provided in Section 2.3.1 that model predictions requiring complex statistical analyses during 'exploratory' *in silico* experimentation can be quickly identified using emulations of protein/gene expression analysis. It is argued that the techniques described are an important additional approach in the development of simulations with results models that are effectively communicable across interdisciplinary teams, and can quickly identify aspects of the model worthy of further interrogation using established statistical analysis techniques that require significantly greater computational expense.

Aim 4: Determine the veracity of the model hypothesis describing the minimum requirements for TLT formation

Chapter 5 deals with analysis of the simulation and hypothesis veracity verification through comparison with *in vivo* descriptions of the TLT induction process.

Aim 5: Utilize the TLT formation model to evaluate the efficacy of therapeutic interventions for Sjögren's syndrome *in silico*

Chapter 6 deals with the evaluation of intervention strategies, and the use of fea-

ture mapping in Kohonen networks to provide surrogate measures of clinical efficacy.

1.6 Summary of Novel Contributions

This thesis presents new approaches to visualising simulation outputs and quantifying model responses in biological sciences, based on the concept that emulating experimental techniques will provide additional insight and improve interdisciplinary collaboration by increasing the communicative capacity of models to those not from a modelling background. These developments are described in Chapter 2, applied using a case study extending a pre-existing agent-based simulation of Peyer's patch formation. In Chapter 3 a new theoretical model of TLT formation derived from human *in vitro* and murine *in vivo* experimental data, and frames this within the context of Sjörgen's syndrome pathophysiology. A semi-formal 'Domain Model' is described and parameterised following the CoSMoS process outlined in Section 1.4.1. A formal mathematical and computational model is developed in Chapter 4 from the 'Domain Model'. This 'Platform Model' is formulated in a platform-agnostic manner prior to implementation using Java as a software simulation tool that has been termed 'NeoSim' (Lymphoid *Neo*genesis *Sim*ulator). The platform model is an abstract description of my hypothesised minimum requirements in TLT formation, intended to determine whether the dynamics of lymphocyte–stroma crosstalk and chemokine induction as described in the model are sufficient to produce structures that are qualitatively similar to those found in Sjörgen's syndrome in humans and murine models.

This thesis fundamentally demonstrates that relatively low-dimensional data combined with hypotheses of how biological systems function, when implemented as a model describing the fundamental processes in an abstract, theoretical manner through hybridisation of multiple mathematical techniques, is capable of demonstrating hypothesis veracity or consistency. Furthermore, through the generation of high-dimensional data from simulation, such models lend themselves well to high-

throughput analysis and have significant predictive capacity.

NeoSim is designed as an abstract theoretical model intended to evaluate whether TLT formation is consistent with the hypothesis formulated in Chapter 3.NeoSim is found to recapitulate the 15-day TLT formation process, with randomly entering lymphocytes stimulating stromal cells to differentially express chemokines resulting in a self-organisation process leading to B-cell follicles supported by FDC-like ('B-Cell') stroma surrounded by T cells and FRC-like ('T-cell') stroma. Analysis of the model and simulation in Chapters 5 and 5 investigated whether NeoSim responded as clinical trial data suggest when biologic therapies are introduced as various time-points and dosages. This investigation ultimately led to the discovery that anti-VLA4 (e.g. natalizumab), an adhesion molecule blockade, can prevent the formation of B-cell follicles responsible for driving Sjörgen's syndrome pathology. This was then tested *in vivo* using an acute murine model, and near complete ablation of B-cell follicular structures was found, suggesting that natalizumab, currently used in the treatment of irritable bowel disease, may also be a viable treatment for Sjörgen's syndrome.

A more detailed summary of original contributions is presented in Chapter 7 (Section 7.1), following presentation of the work itself in the proceeding chapters.

Chapter 2

Simulation Analysis utilising Visualisation and Emulation of Experimental Techniques

The use of modelling and simulation as a predictive tool for research in biology is becoming increasingly popular. However, outputs from such simulations are often abstract and presented in a very different manner to equivalent data from the biological domain. The development of a flexible tool-chain for emulating various biological laboratory techniques to produce biologically homomorphic outputs in computer simulations is presented and applied to simulations in case studies. These emulations include immunohistochemistry, microscopy, flow cytometry, and quantifying gene and protein expression levels. Pre-natal lymphoid organ development is utilised as a case study of the application of this tool chain to simulations without affecting their extant behaviour. A pre-existing simulation of Peyer's patch (PP) formation in the mouse gut is extended to produce various new outputs aimed at improving insight into simulation behaviour, and at accelerating the rate of discovery.

This chapter demonstrates that application of the tool-chain can provide additional, biologically relevant data, that are inaccessible with pre-existing methodologies for analysis of simulation results. It is argued that experimental techniques

borrowed from in vivo *animal models and* in vitro *cell culture systems are an important additional approach to the analysis of simulations in computational biology, and might furthermore inspire confidence in simulation results from the perspective of experimental biologists, and improve cross-disciplinary model communication.*

2.1   Introduction

A typical approach to experimentation with an existing computational biology simulation is to perform various statistical analyses on abstract simulation outputs while varying one or more parameter values. These multivariate analyses are then used in an attempt to make determinations regarding the fitness-for-purpose of the model and to make predictions about the biological system which it represents. These predictions may lead to further *in silico* experimentation within the simulation, and also predictions that could be tested in the wet laboratory. However, there does not exist a principled approach to linking the results model of a simulation back to the original biological domain model. Model predictions often arise from observations of how the system responds to parameter perturbation in terms of resulting changes to abstract data structures representative of biological entities; *in vivo* confirmation of these predictions therefore require an *ad hoc* transformation, firstly, of what the observed change in model output permits one to infer about the system under study, and secondly, a prediction of what one would expect to be observable in either *in vitro* or *in vivo* models of the same system presuming the original prediction is indeed correct. The differing abstractions and limitations that affect these extremely different model systems present significant confounding factors with respect to what may be construed as sufficient evidence that a prediction is generally accepted to be true. *In silico* models which frame outputs within the context of generally accepted constructs within the field permit *directly* testable predictions without pre-requisite inference of what may be considered to construe a suitable test.

The rapidly increasing popularity of mechanistic *in silico* models of highly complex biological phenomena, including tissue organogenesis and immune processes,

is resulting in an ever-growing population of researchers with diverse backgrounds collaborating on multidisciplinary projects toward common goals. Just as, now long established, it is known that a model must be entirely transparent and demonstrated to be fit-for-purpose using sufficiently rigorous argumentation tools and model development frameworks (such as CoSmoS), we must also acknowledge that models need to be *sufficiently communicable*; a model whose results cannot be readily understood and assimilated by the scientific community at large (that is, beyond mathematicians and computer scientists) is unlikely to have maximum impact, regardless of its fitness-for-purpose or predictive power.

A variety of theoretical advancements and practical tools have been developed in recent years that improve understanding of interdisciplinary communication aimed at enabling domain experts to adequetely identify flaws in model implementation such as unsuitable abstractions or, domain misunderstanding by non-expert model developers. Key examples of such advancements being the development of the use of 'expected behaviours' diagrams, first utilised in **?**, as a means to broadly communicate a model overview, including associated aims and hypotheses, and the adaptation of the UML to permit the description of massively parallel biological systems such as cellular signalling networks, through a technique that biologists and other domain experts can quickly learn to understand (**?**). However, there has been little to advance interdisciplinary communication and discussion of the *results* model, in which inferences about the domain are often made using techniques that are not easily communicable to domain experts and interested third parties. This may lead to *in silico* models being overlooked in favour of other modelling approaches that are accepted within the dogma of a particular scientific field.

We propose the creation of a stronger link between biological models and executable simulations in general by developing simulations that produce outputs that can map to the types of data produced and used by experimental biologists. Model developers should aim to design outputs that are amenable to analysis methodologies that are commonly applied to data derived from *praeter silico* (besides *in*
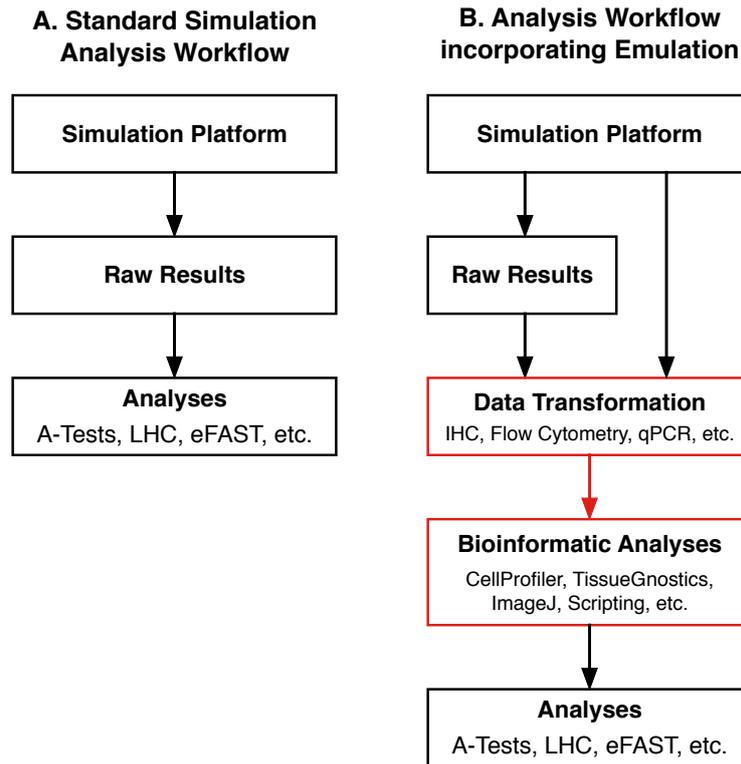
Figure 2.1: A typical approach to performing and analysing simulations within computational biology. The red boxes represent an extended work-flow designed to provide outputs that directly map to experimental data constructs, which enables additional analysis of model dynamics and model predictions to be framed within a *praeter silico* experimental context.

silico) models by experimentalists. It is to this end that software and protocols for the production of such data structures that can be applied to pre-existing and new simulations have been developed, and enable computational models to be better interpreted within the context of the biology that they represent. Figure 2.1 presents a typical simulation analysis work-flow (black rectangles) and incorporates the proposed additional steps (red rectangles) for creating models and simulations that better integrate with experimental biology. The net result of these additional processes is a wider range of model outputs that can be utilised during simulation calibration and validation, to improve model communication, and to aid development of directly testable predictions in terms of *in vivo* and *in vitro* model-based experimentation.

2.1.1 Taking Cues from Experimental Biology During *In Silico* Model Development

To develop a quantitative understanding of enormously complex biological processes, such as those involved in immunity, immunologists and other life-scientists have developed and utilised a combination of technologies to quantify molecular and cellular mechanisms during immune responses. These include flow cytometry, immunohistochemistry, western blotting, qPCR (quantitative polymerase chain reactions), ELISA (enzyme-linked immunosorbent assay) and *in vitro* functional assays; however these technologies are, in isolation, inherently limited in spatial, temporal or cellular resolution and insufficient alone to understand the underlying physics and biology of immune responses. The development of myriads of biologically-compatible chemical dyes and fluorescent proteins, in combination with advances in imaging technologies, have provided the potential to image immune function at the single-cell level in three-dimensions over time.

Multi-photon confocal imaging has provided a unique tool to quantify immune responses providing accurate quantitative data on cell migration and interactions that are key to the parameterisation of mathematical and computational models of immune responses (**?**). Such models have been used to understand the biophysical mechanisms of cell migration and interactions through analysis of data sets and resolved how three dimensional cellular topologies dictate immune cell signalling (**??**). Bottom-up *in silico* models are ultimately dependent on the rich array of techniques at the disposal of experimental biology, whether the model is data-driven, utilising these data directly, or fundamentally theoretical and mechanistic, making indirect use of insights afforded by experimental biology. While many of these techniques are designed to quantify specific entities and from a computational modelling perspective are principally useful in model design and parameterisation, others offer experimentalists a highly intuitive means of exploring highly multi-dimensional datasets and visualising spatiotemporal organisation of complex tissue micro-environments, and highly specialised software exists to aid interpretation of these data. By emulating

experimental techniques, beyond the other benefits described herein, a modeller may leverage these tools (such as flow cytometry and microscopy image analysis software tools) to gain additional insights from their computational models.

For a given technique to be useful within the context of emulation, it must present data in a manner that intuitively allows inference of how an aspect of the biological system behaves or provide a high-dimensional dataset generally understood within the field. For instance, it would make no sense to discuss western blot emulation – the presence of proteins within a simulation result can be simply established through direct interrogation of the internal data structure. However, emulating means of quantifying gene or cell surface protein expression such that the relative expression of multiple elements over time and space can be succinctly visualised could add significant value to a simulation. Table 2.2 describes three experimental approaches that are potentially useful within simulations and typically produce data that are replicable within many agent-based or hybrid *in silico* models, either through the existence of direct correlates within model data structures, through the definition of abstract quantities that permit equivalent visualisations to be developed or through data transformation achieved by combining entities from multiple sub-models within an overarching hybridised model (discussed in Section 4.8 following development of a hybridisation framework in Chapter 4). In unidentifiable models, in which multiple candidate models exist, these techniques provide an additional platform for comparing model emergent phenomena across candidate models to data from experimental biology.

The development of multi-colour flow cytometry and higher resolution thin sectioning of immune tissues combined with an ever-expanding number of antibodies and fluorochromes drove rapid progress in immunology, due to the capacity of these technologies to quantify gene expression and classify live immune populations and their effector function at the single cell level. Such technologies have been fundamental to the modern understanding of immune responses, and have led to new models of the key events in antigen initiated immune responses (**?**). The ubiquitous

Flow Cytometry

Cellular Analysis	Description	Cells are input via fluidic channels and individually struck by coherent light, measured through photodetectors to determine the intensity of their fluorescent antibody stained surface.
	Output	FCS Formatted Files / Multi- dimensional expression data for dot-plot production.
	Illustrates	Identification of cell populations, correlations in protein expression.

Immunohistochemistry & Microscopy

Tissue Analysis	Description	Sections of tissue are fluorescent-antibody stained, then imaged with confocal microscopy.
	Output	Simulation visualisations that are comparable to tissue IHC micrographs, particularly with image post-processing.
	Illustrates	Cell/Protein Co-localisation and tissue structure; amenable to high-throughput image analysis.

Gene/Protein Expression Analysis

Molecular Analysis	Description	Deep-sequencing or qPCR type analyses of simulation state.
	Output	Spatial, temporal & combined spatiotemporal heat-maps
	Illustrates	Dominance of specific genes in controlling particular processes, changes in expression over time.

Figure 2.2: Table describing the experimental techniques to be emulated in this case study, the outputs they produce and their illustrative and communicative potential.

presence of flow cytometry, immunohistochemistry and confocal microscopy in modern immunology makes them well-suited for use as a basis for designing simulation data structures that can be readily interpreted by domain experts.

2.2 Case Study: Simulating Peyer's Patch Formation

A pre-existing simulation, designed as a predictive tool for exploring mechanisms that drive pre-natal lymphoid organ development within the mouse gut, was selected as a case study in applying the techniques described herein to extend the model's outputs and predictive ability without modifying simulation behaviour. This model and simulation has aided the generation of testable biological hypotheses concerning the complex cellular interactions leading to the generation of organs that trigger adaptive immune responses: interactions which cannot currently be fully explored using laboratory techniques. A brief overview of the model is provided here, however the reader is directed to previously published work detailing the simulation design, implementation and analysis (**???**). The tool captures the 72 hour period of tissue development in pre-natal mice. Populations of haematopoietic cells, known as Lymphoid Tissue Initiator (LTin) and Lymphoid Tissue Inducer (LTi) cells, migrate into the developing gut, with data from laboratory observations suggesting these cells follow a random motion. Both cell populations express receptors for the adhesion molecule VCAM-1, expressed by stromal Lymphoid Tissue Organizer (LTo) cells residing in the gut wall. VCAM-1 causes cells expressing the cognate receptor to adhere to the VCAM-1 expressing cell, thereby restricting its movement. Contact between a haematopoietic cell and LTo cell triggers the LTo cell to differentiate (become more specialised), leading to increased adhesion molecule expression. In addition, LTo cell differentiation increases chemokine secretion, creating a chemokine gradient that promotes migration of the LTi cell population towards the differentiated LTo cell. In the vicinity of LTo cells, movement of LTin and LTi cells will
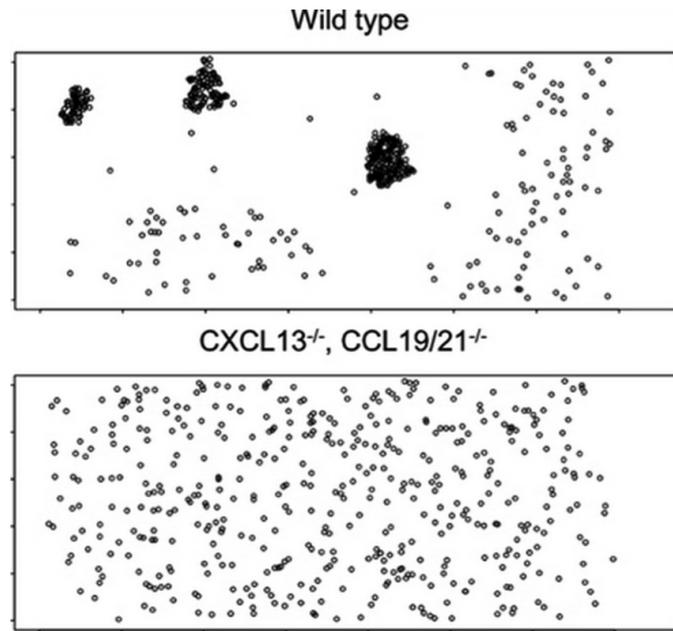
Figure 2.3: Shows output of one simulation run of PPSim illustrating a 10% section of mouse gut showing LTi aggregations leading to PP formation in wild type mice after 72 hours (top), and a lack of PP formation in CXCL13$^{-/-}$ and CCL19$^{-/-}$/21$^{-/-}$ mice (bottom).

be restricted by adhesion factors (VCAM-1 and others), forming aggregations of haematopoietic cells around LTo cells at the end of the 72 hour period. These aggregations later mature into lymphoid organs called Peyers patches (PPs), which are capable of initiating immune responses against pathogenic bacteria encountered in the gut. The visual output of the simulation at the end of the 72 hour period from the extant simulation as described **?** is shown in Figure 2.3.

It has been previously shown that the emergent cell behaviour observed in laboratory experimentation is statistically similar to that observed in the simulation, and sensitivity analysis techniques were utilised to explore the simulated biological pathways to reveal those which have a significant impact on simulation response (**??**). The output from these statistical techniques provided evidence that our simulation is fit for the purpose of aiding biologists in their exploration of the system. However, it is proposed that confidence in the simulation would be further increased by providing experimental biologists with simulator outputs that are comparable to primary laboratory data, which can be more intuitively interpreted. Furthermore,

such outputs can provide additional insight into the simulation dynamics, enable additional exploratory experimentation *in silico* and furnish mechanistic detail not readily accessible with descriptive nor inferential statistical analyses.

2.2.1   Emulating Flow Cytometry

Flow cytometry is a technology with broad application in biology and medicine, particularly used as a means of cell sorting, cell counting, and the detection of biomarkers through fluorescent labelling. With a large number of measurable variables and production of highly multi-parametric datasets, flow cytometric analysis is performed using specialised software such as FlowJo (TreeStar) and WEASEL (Walter and Eliza Hall Institute of Medical Research).

The data generated by flow cytometers may be plotted as a histogram, in one dimension, or as dot plots in two or three dimensions. A sequence of subset extractions (termed 'gates') may be applied to sequentially separate regions of dot plots based on MFI (Mean Fluorescent Intensity), FSC (Forward Scatter – proportional to cell volume), SSC (Side Scatter – indicative of morphological complexity), and a range of other parameters. Such gated plots are useful for identifying cell populations and sub-populations based on biomarker expression, permitting quantification of the phenotypic make-up of tissues. Datasets produced by modern flow cytometers are written in the ISO Flow Cytometry Standard (**?**) file format using the .fcs file extension.

2.2.1.1   Flow Cytometry Emulation Methodology

For each time-point a flow cytometry analysis is to be performed, expression levels of each protein of interest are placed into a CSV column and multiplied by a scaling factor to bring the values into an appropriate order of magnitude for Flow Cytometry data ($s = 105$ in this case) prior to bring cast as integers. A value of '105' was selected for this dataset as it was the minimum number that re-scaled outputs to the same order of magnitude observed in data-files taken from actual flow cytometry

experiments, the value should be selected by interrogating typical count numbers from similar 'real' flow cytometry experiments. This ultimately enables flow cytometry software to interpret the values as fluorescent intensities, which are proportional to expression level. Following this, the CSV data are transformed into fcs format compliant files according to the ISO standard described in ?, which is universally used by modern flow cytometers to store acquisition data. This is an important step as it allows the data to be interrogated using flow cytometry software, specifically designed for use with biological data and is familiar to immunologists.

In order to emulate flow cytometry, a simulation requires explicit values for cell surface expression of proteins. Absolute values are not important, rather, it is the relative differences in expression levels that enables the insight afforded by this technique. In the Peyers patch simulation, expression is strictly binary, such that each cell in the simulation is either expressing a protein or not. An increase in protein expression is achieved by changing the parameters of abstract mathematical functions that determine behaviour. For example, LTo chemokine expression levels are abstracted as a sigmoidal cumulative probability density function, sampled at each time step by every LTi cell responsive to chemokine. An increase in chemokine expression level is represented as a reduction in the standard deviation (or tightness) of a sigmoid curve, as detailed in ?. Expression of VCAM-1 by LTo cells is handled in a similar manner, according to a truncated linear function, this is illustrated in Figure 2.4. Therefore, it is necessary to modify the simulation to provide relative quantitative expression levels without compromising extant simulation dynamics. In these cases, a new agent property needs to be created that represents relative expression of the factor concerned, but is not used by the simulation for decision making purposes (i.e. state changes). This value is obtained by incrementing (or decreasing) the property each time step in direct proportion to the change in the mathematical function responsible for the protein. For instance, each time the standard deviation of the function responsible for describing chemotactic response is reduced, an integer is incremented by one. This results in an integer value associated
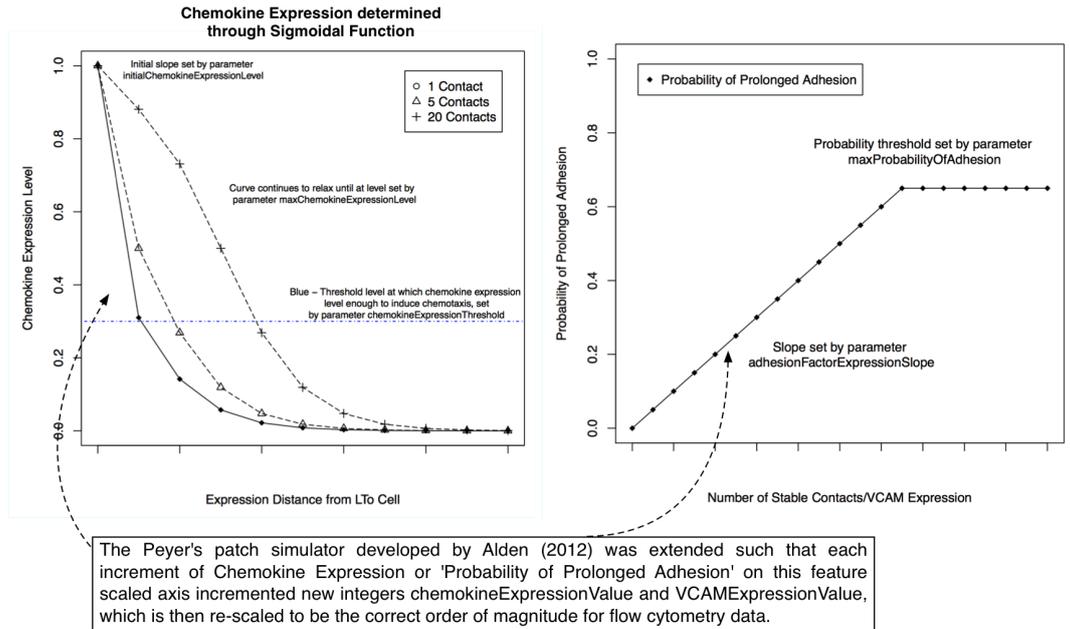
Figure 2.4: Graphs show the sigmoidal and linear equation that represent chemokine expression, and adhesion probability (via VCAM-1), respectively in the Peyer's patch simulator developed by **?**. To capture a numerical value of adhesion molecule expression, an integer is incremented for each stromal cell, every time this discrete linear equation increments. The two charts in this figure were taken from **?**.

with each LTi cell within the simulation that may be considered directly proportional to CCL19 expression, allowing analyses that require relative cell surface or gene expression values to be utilised.

The forward scatter (FSC) of light that occurs when lasers strike the cell in the flow cytometer, which is proportional to the volume of the cell, was also emulated. In the simulation, cells of a given phenotype have the average diameter of those cells as measured experimentally. This was adapted in the simulation such that the radius is sampled from a truncated Gaussian (normal) distribution about the mean experimentally measured radius, with a standard deviation of 1m. This enables identification of cell populations based on both their size and their expression of VCAM-1. In order to avoid modification of the validated extant simulation behaviour, the original average diameter is still used for model decision making processes, for example, cell-cell contact detection, and the randomly sampled cell radius was recorded for transformation into the FCS file for flow cytometry-based analyses.
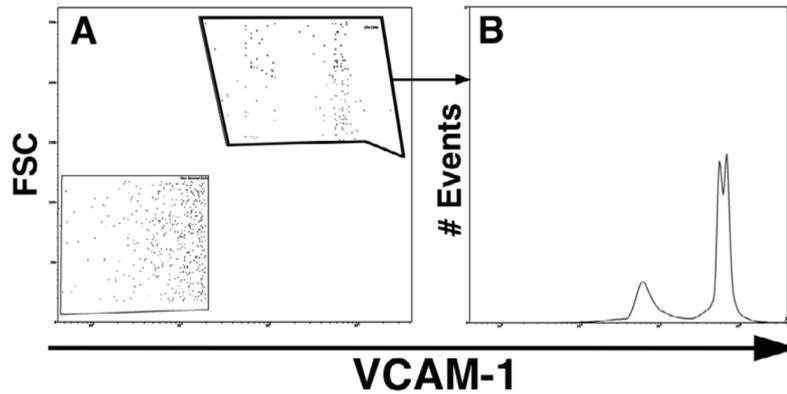
Figure 2.5: **(A)** Emulated flow cytometry dot-plot of VCAM-1 Mean Florescent Intensity (MFI, arbitrary units) vs Forward Scatter (FSC). This is then gated on the VCAM-1 positive cells with high forward scatter, which are the large stromal cells. From this subset, **(B)** shows a histogram of VCAM-1 MFI (log scale, arbitrary units) vs. the number of recorded events (cells). These plots were generated using WEASEL flow cytometry data analysis and display software (Walter and Eliza Hall Institute of Medical Research), intended for use with biological data from flow cytometers.

2.2.1.2 Flow Cytometry Emulation Results

Figure 2.5A shows a dot-plot of adhesion molecule VCAM-1 and forward scatter (FSC) for all cells in the simulation at the end time-point of 72 hours. This is gated on VCAM^{hi} cells to produce a histogram that illustrates the distribution of VCAM-1 expression fluorescent intensities.

Unlike flow cytometry performed experimentally, in which the sample is destroyed during measurement, emulated flow cytometry data can be produced at any desired frequency during a simulation run. Crucially, this permits time-series flow cytometric analysis in which the same 'sample' can be observed over time. To illustrate this, the VCAM-1 histogram from Figure 2.5B is shown at 4-hour intervals from the 12th to 72nd hour of the formation process in Figure 2.6. The time-dependent development of the VCAM-1+ LTo population is clearly visible.
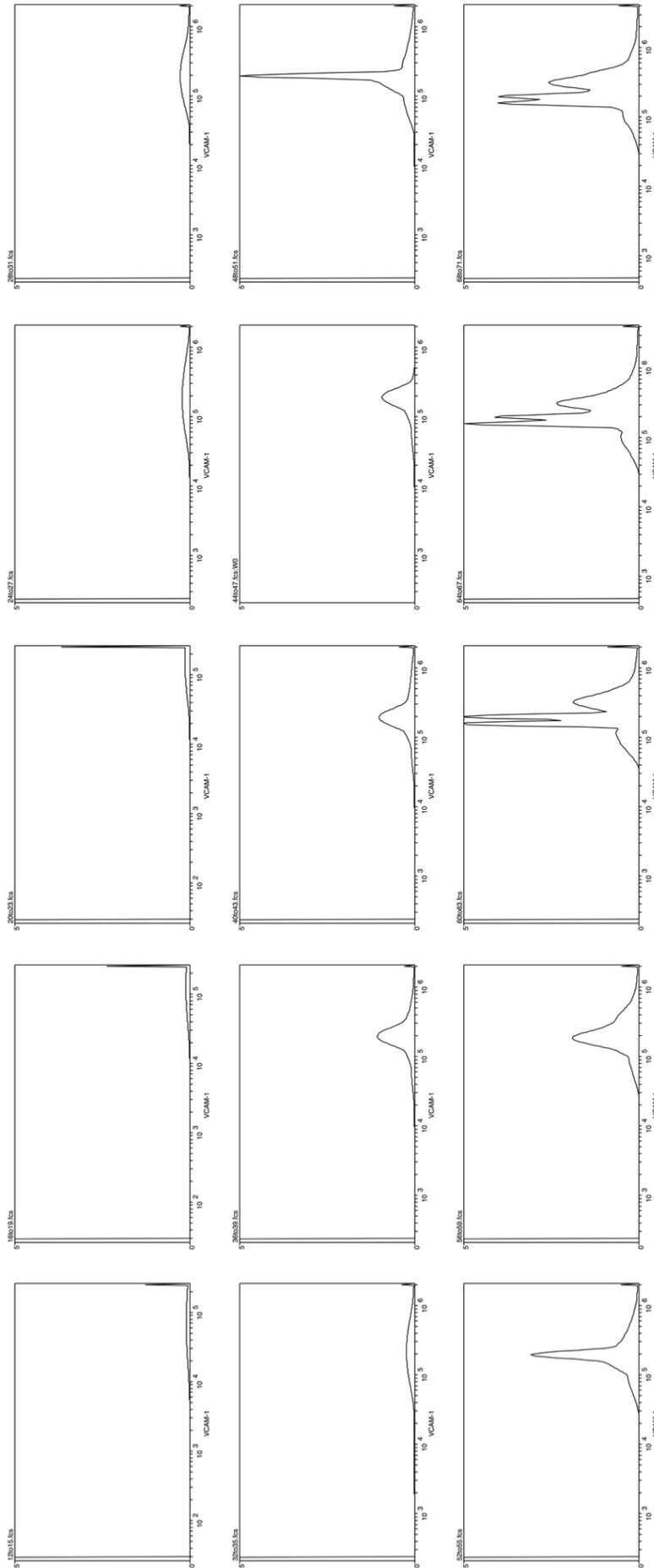
Figure 2.6: Time-series flow cytometry data from hour 12 to 72 in 3 hour increments showing the development of VCAM$^+$ stromal populations. These are flow cytometry histograms showing MFI (log scale) vs frequency (number of recorded, emulated, detection events). While grounded in techniques of experimental biology, time-series flow cytometry is impossible in the domain as samples are necessarily destroyed during measurement. Simulated time-series flow permits observation of how cell populations emerge over time.

2.2.2 Emulating Immunohistochemistry and Imaging Techniques

Immunohistochemistry is the process of antigen detection within tissue sections through the use of cognate antibodies. Immunoflourescence is often used, in which an antibody is conjugated to a fluorophore that emits photons of a specific wavelength when excited with an appropriate laser. Sections of antibody-tagged tissues are placed in a microscope in order to visualise the spatial location of tagged proteins and therefore the tissue organisation. Confocal microscopy is widely used, in which the image is constructed only from light caused by fluorescence very close to the focal plane.

2.2.2.1 IHC Emulation Methodology

As seen in Figure 2.3, the spatial environment captured by PPSim is a 2-dimensional (2D) rectangular grid that represents a 10% length of the mouse gut. This is toroidal about the Y-axis, and is thus an abstract representation of an 'un-folded' section of intestinal tissue. LTo cells in the simulation are stored in a 2D discrete space grid. All haematopoetic cells in the simulation (LTi and LTin cells) are stored in a 2D continuous grid data structure provided by the MASON agent-based simulation toolkit (Luke et al., 2005), a library for use with the Java programming language. Continuous space is represented by storing each cell in a discretised grid and associating them with two double precision floating point numbers corresponding to their 'true' position within the grid. This is illustrated in Figure 2.7. To produce images analogous to those produced by IHC and confocal microscopy, each cell object is extracted from this grid and then drawn to on a canvas object according to the properties of that cell, as seen in typical immunohistological imaging.

Discrete Space - LTo Cells
Cells stored as objects in a 2D array, cell location stored as integer tuplet.

Continuous Space - LTi & LTin Cells
Cells stored as objects in a 2D array. These are associated with a tuplet of double precision floating point values: X position & Y position.
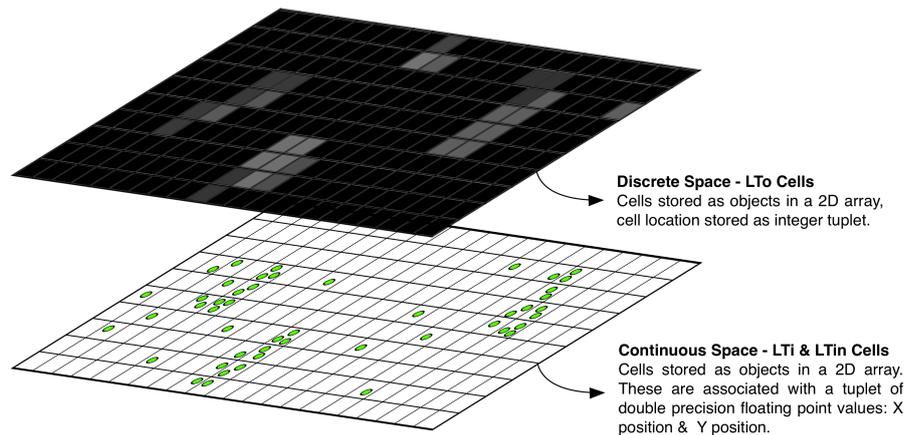
Figure 2.7: Discrete and continuous data structures used to store cell locations in PPSim. Continuous space is represented through the coupled association of a cell with a discrete grid element and a set of Cartesian co-ordinates.

2.2.2.2   IHC Emulation Results

Figure 2.8 presents new visualisation approaches derived from the existing simulation of Peyers patch formation. In Figure 2.8A, LTi cells are drawn as green circles ((0,255,0) in the RGB colour space) on a black canvas, in a manner similar to the appearance of GFP (green fluorescent protein)-stained cells when imaged with confocal microscopy. Overlapping cells produce a region of green with a higher alpha (transparency) value, to enable determination of the density of a region by measuring the level of alpha in that region as compared with the base level assigned to individual cells. LTo cells are illustrated in Figure 2.8B, drawn red (255,0,0), with an alpha value corresponding to VCAM-1 expression level. LTo cells without sufficient VCAM-1 expression (beyond the threshold required for adhesion to be possible within the model specification) have been coloured grey (128,128,128) for the purposes of the figure so as to be rendered visible to the reader.

The two canvases undergo several stages of post-processing before combination into Figure 2.8C as an emulation of IHC and microscopy that illustrates the co-localisation of LTi cells and VCAM-1 expression on LTo cells. The VCAM-1 canvas undergoes Gaussian blur and posterisation, to provide an interpolated, continuous approximation of VCAM-1 expression within Peyers patches. Furthermore, a canvas
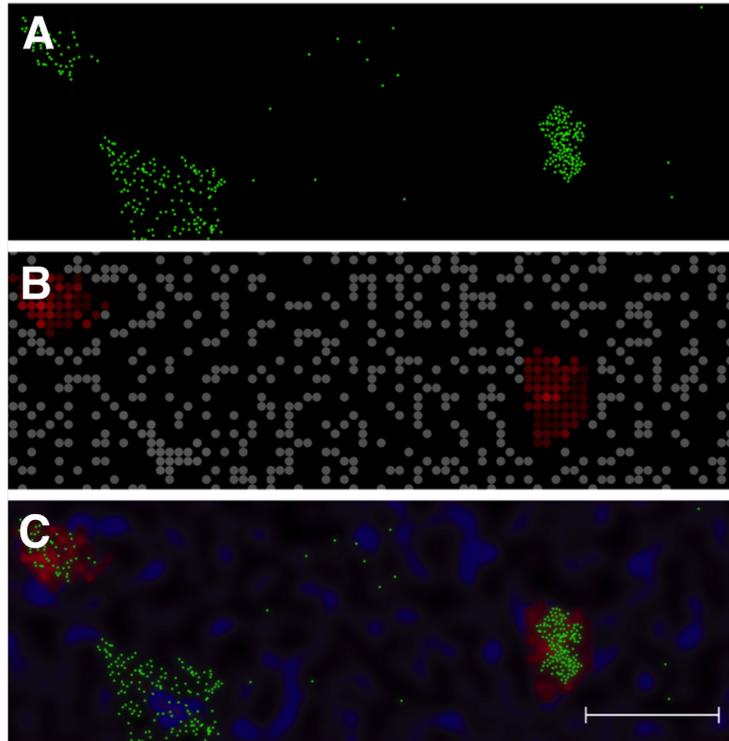
Figure 2.8: Simulation visualisation after 72 hours. **(A)** LTi cells coloured green to simulate GFP in microscopy. **(B)** LTo cells in grey, and level of red proportional to VCAM-1 expression level. **(C)** Complete emulated histology and microscopy image showing two Peyers patches. LTi cells are stained green, LTo cells stained blue, with additional VCAM-1 staining in red. Scale bar 175 microns.

with dark blue circles drawn at 25% opacity representing all LTo cells undergoes a Gaussian blur and is placed beneath the red VCAM-1+ LTo cell canvas. The LTi cell layer is duplicated, with the lower layer undergoing a Gaussian blur. Posterisation is a process in which continuous tonal gradations are reduced to a smaller number of tones. A Gaussian blur is a common image filtering technique that blurs an image through application of a Gaussian function, for a full description of the method the reader is referred to **?**. This was found to produce a VCAM-1 LTo canvas that while more diffuse, reduced extension of LTo cells at the periphery and allowed each distinct cell to be clearly identifiable despite the application of blurring.

The *in silico* images can be written from the simulation at any pre-specified resolution, with the minimum usable value linearly-dependent on the diameter of the smallest visualised object. The images can be automatically analysed en masse
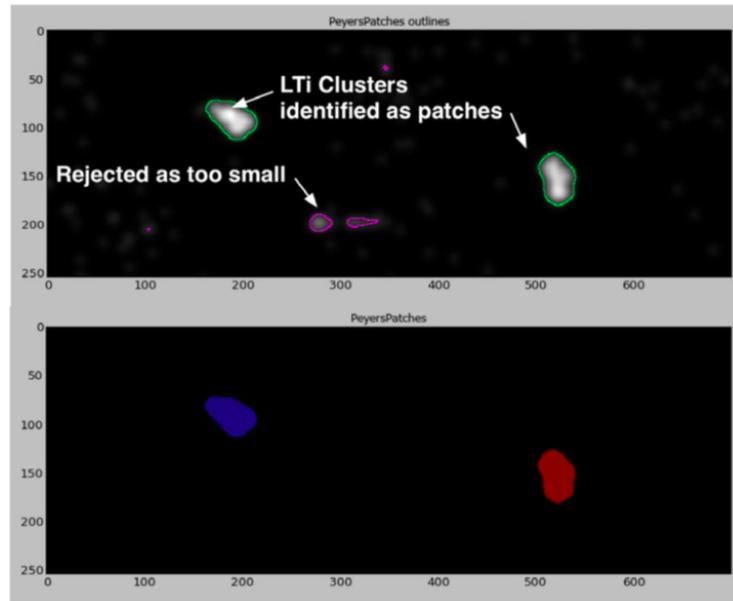
Figure 2.9: CellProfiler output for one simulation run in which two Peyers Patches are identified, and one aggregate of LTi cells is discarded due to occupying an insufficient area. This analysis can be automated over many thousands of runs to identify trends in patch size in terms of area, density and compactness as model parameters are perturbed (however this figure illustrates only patch detection and evaluation with a baseline sample run of PPSim).

to detect the presence of Peyers patches using CellProfiler (**?**), pipeline-based image analysis software for automated quantification of cell phenotypes from imaging data. This is illustrated in Figure 2.9 for a single simulation run, but may be applied to an arbitrarily large dataset. Prior to the development of this CellProfiler pipeline, identification and quantification of patch formation has proven difficult within the simulation, as it is achieved in the domain through manual analysis of histology (that is, 'by eye') (**?**).

2.2.3 Producing Heat-maps Illustrating Gene or Protein Expression

Methods of presenting and analysing genomic or transcriptomic data acquired from DNA microarrays, deep sequencing (e.g. RNA-seq) and related technologies are useful for understanding spatiotemporal cellular dynamics in terms of gene expression

and protein synthesis. Heat-maps that illustrate differential expression profiles over time can provide important clues as to the order in which events take place during biological processes (**?**).

2.2.3.1 Gene and Protein Emulation Methodology

Mean protein expression levels can be output from the simulation both over time and across space. This permits the creation of heat maps that capture cell phenotypes spatiotemporally, and can show the progression of protein expression over time and space. Figure 2.10A shows a sample of 9 cells and their expression of VCAM-1 and chemokine CCL19 at three different time-points. The expression levels used to generate these heat-maps were generated in an identical manner to the method used for the flow cytometry emulation.

Depending on the level of abstraction within a simulation, expression levels may represent cell-surface expression level and gene expression level specifically and independently, or abstract these into a single quantitative value. Within PPSim, the values generated may be thought of as representing the relative cell-surface expression level of VCAM-1 and the relative gene expression or secretion rate of CCL19. The heat-maps themselves were drawn manually using EazyDraw (Dekorra Optics LLC), with the exception of the spatial heat-map which was generated in Java and extracted prior to post-processing to produce the emulated micrograph images as can was shown above in Figure 2.8 (B).

2.2.3.2 Gene and Protein Emulation Results

Figure 2.10A shows a sample of 9 cells and their expression of VCAM-1 and chemokine CCL19 at three different time-points. In Figure 2.10B, the spatially-distributed expression of VCAM-1 is shown at 72 hours for three isolated Peyers patches that formed during one simulation execution. The change over time of the spatial distribution of VCAM-1 within the simulation can be visualised using a 3D spatial heat-map in which each 'slice' of the z-axis illustrates the distribution of VCAM-1,
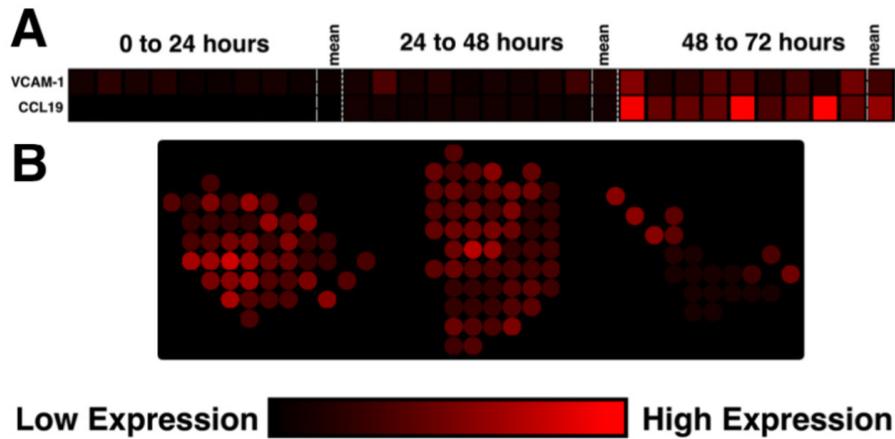
Figure 2.10: Heat-maps generated from simulation data over both time and space, to illustrate change in protein expression levels. **(A)** VCAM-1 and CCL19 expression over time for a subset of individual LTo Cells and then the mean level over that subset. **(B)** Spatially-resolved heat-map showing VCAM-1 distribution in three Peyers Patches that formed during one simulation run.
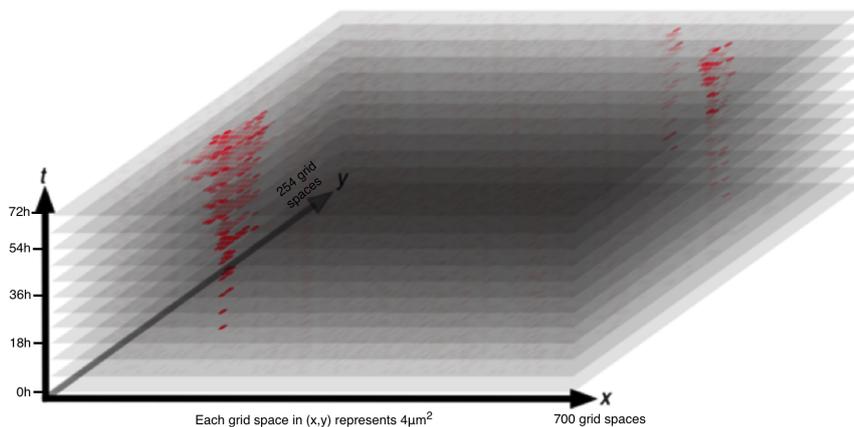
Figure 2.11: 3D heatmap illustrating change of VCAM-1 concentration over time (vertical axis), demonstrating the formation of regions of high VCAM-1 levels over 72h. Each of the 12 slices in t represent a 6 hour time increment. This visualisation clearly shows the non-linear increase in area of VCAM-1^{hi} regions over time, providing a unique means of visualising the spatiotemporal patch formation process in one plot.

as shown in Figure 2.11. Each slice in the z-axis represents a 7.2 hour increment over the 72 hour PP formation process. The heat-maps in Figures 2.10 and 2.11 were generated by exporting the contents of the discrete stromal cell grid from the Peyer's patch simulator at desired time-points during a simulation, and coloured

according to VCAM-1 expression calculated as described in section 2.2.1.1.

2.3   Discussion

It is proposed herein that a simulation may be considered *effectively communicable* if the predictions derived from it can be understood not just by the modeller, but by domain experts generally who may lack significant experience in computer science or *in silico* modelling approaches, without requiring significant exposition of the analysis methodology utilised. This widens possible participation by enabling models to be easily understood across disciplines. For example, the prediction regarding the time-dependent roles of VCAM-1 and CCL19 can be illustrated using either A-test plots or heat-maps of expression levels over time. Whilst the former may be necessary to have confidence in the statistical significance of the effect, the *biological significance* is readily demonstrated using the heat-map of Figure 2.10A. The techniques described in this chapter for grounding computational models within experimental biology, while not intended to replace existing analysis methodology, are proposed to be a valuable tool for communicating model outcomes to the wider scientific community and as a guide for early prediction identification prior to exploratory statistical analyses.

It has been suggested variously throughout development of the field of computational biology, that a 'reverse Turing test' could present an ultimate validation test for computer simulations of biological phenomena (**??**), in which both biological data and simulated data are presented to a biologist in order to observe whether the simulated dataset can be readily identified, and how. This has largely not been feasible to date because of the very different manner in which simulation results are presented, and the difficulty presented in buffering these differences from biological domain experts. Now, with the availability of simulated histology and flow cytometry data, computational biology and QSP is approaching the point at which this is becoming feasible. For instance, prior to the development of the CellProfiler pipeline demonstrated in Figure 2.9, there was no reliable, automated means

by which to identify, count and quantify Peyers patches over many simulation run results, and simple dots were used for each cell in the simulation visualisation layer. Furthermore, there are no other means except emulated flow cytometry to perform an in-depth analysis of cell populations and their properties within a simulation in a manner indistinguishable from *in vivo* animal models and *in vitro* cell culture system analyses. The possibility of simulation and experimental model results becoming indistinguishable is an exciting prospect in terms of accelerating scientific progress in biology and medicine.

2.3.1 Gene and Protein Expression Analysis Emulation

Figure 2.10A demonstrates the potential for expression heat-maps to illustrate the differential temporal roles of proteins or genes of interest; in this example, notable VCAM-1 expression is seen much earlier than chemokine CCL19 expression, however, by day 3 expression of CCL19 is clearly dominant. The dominance of adhesion molecules early in the formation process and chemotaxis later on was a prediction generated through statistical analysis from the original simulation in **?**, as shown in Figure 2.12. The original determination of this prediction required extensive exploratory analysis of sensitivity to model parameters and significant computational expense. Individual simulation runs consistently show such phenomena in a visually and immediately-apparent manner. This suggests there is significant value in generating heat-maps of time-series expression data to identify potential predictions and guide the statistical analyses undertaken.

The temporally and spatially-resolved heat-maps presented in Figures 2.10 and 2.11 have the potential to be combined in various novel ways to demonstrate the development of biological structures through changing gene expression across time and space. The 2D spatiotemporal heat-map in Figure 2.11 presents an intuitive means of visualising how factor expression within a tissue changes over time. This predicts that VCAM-1 distribution in patches begins at a central point and spreads radially outwards from that point over time.
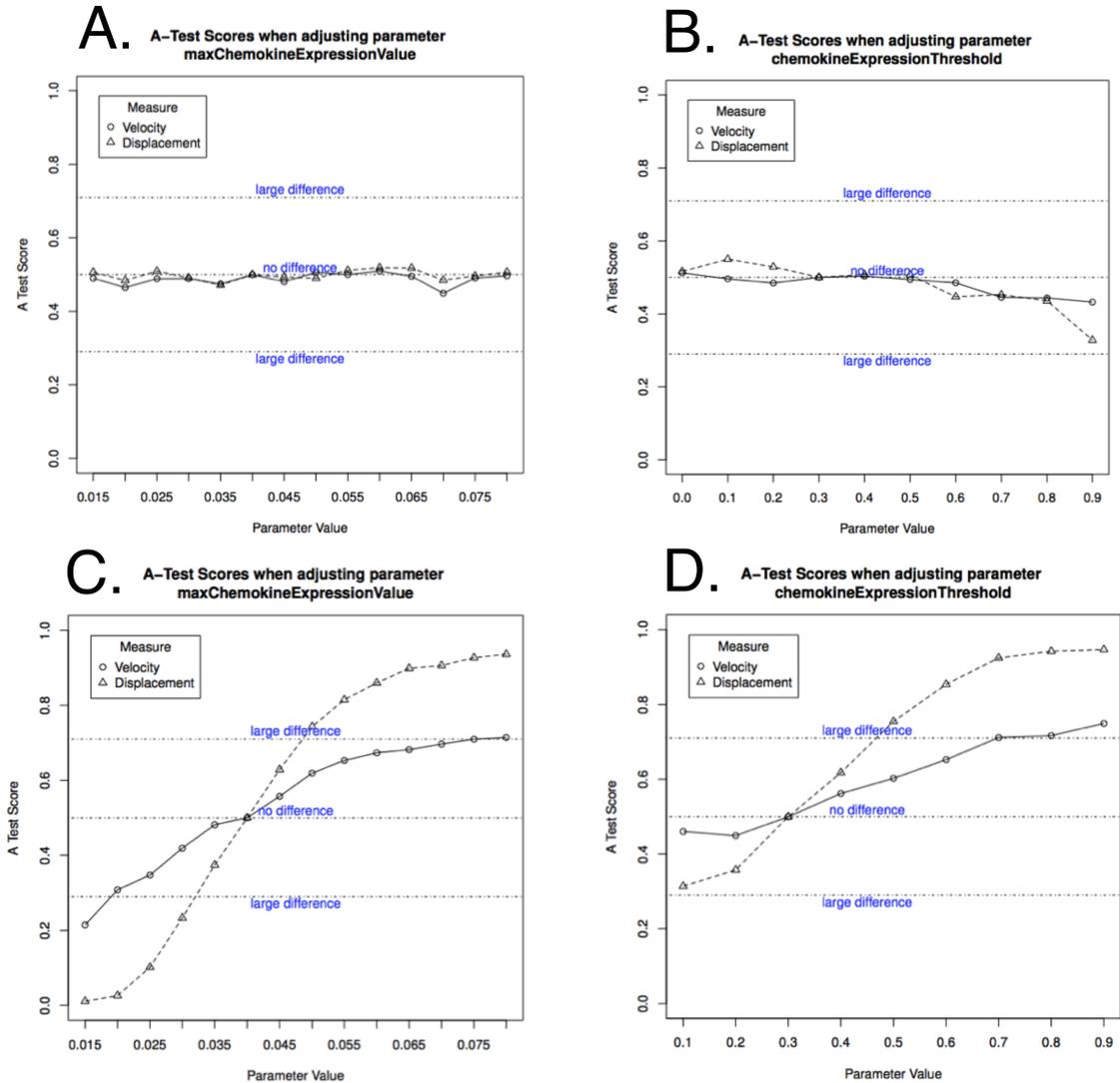
Figure 2.12: The observation that LTi cell velocity and displacement were not sensitive to key parameters controlling chemotaxis during the first 24 hours of patch development (A and B), but highly sensitive during the final 24 hours (C and D), with the opposite effect found for adhesion molecules, led to the prediction that adhesion molecules are responsible for driving early patch development, while chemokines are dominant later on in the process. (A and C. The maximum chemokine expression level per LTo; B and D. The minimum chemokine expression level required for LTi chemotaxis)

2.3.2 IHC Emulation

Emulating IHC and microscopy permits automated analysis with software such as CellProfiler, and also provides visual insight into emergent structures. Although the rudimentary image transformations applied to the spatial information contained within the simulation are presently easily distinguishable from genuine micrographs, it is conceivable that further development would diminish observed differences. Despite this, the realism afforded by the present emulation methodology is sufficient to apply image analysis techniques to simulation results. For comparison purposes, Figure 2.13 shows the emulated micrograph alongside a similar micrograph on the same scale obtained from a developing murine lymph node.

The application of the IHC/microscopy emulation in combination with the Cell-Profiler pipeline illustrated in Figure 2.9 to PPSim has created potential for a set of *in silico* experiments that are not presently possible in the wet laboratory, not were previously feasible with the original PPSim model. These are to explore how patches change in response to parameter perturbation and gene-knockout. Previously, *in silico* experimentation within PPSim focused almost exclusively on velocity and displacement of LTi cells as the principle outputs, however it is now possible to
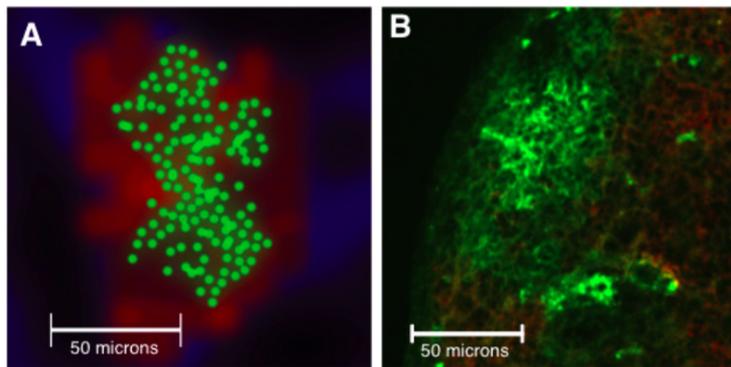
Figure 2.13: **A.** Emulated immunohistology and microscopy of a Peyers Patch at Embryonic Day 17.5, simulated *in silico*. LTi cells are green, the level of red is directly proportional to VCAM-1 expression on LTo cells, while blue indicates the presence of LTo cells not expressing significant levels of VCAM-1. **B.** Actual confocal microscopy image of an antibody stained B Cell follicle within a developing murine lymph node. The B Cells are fluorescing green (B220), the surrounding T cells are fluorescing red (CD3). Scale bars are 50 microns

measure a large range of parameters relating to both the structure of the patches such as morphological features, cell density and area, and also the spatial organisation of expressed proteins. Previously, experiments relating to patch number were intractable due to the large number of replicates per parameter sample required to ameliorate aleatory uncertainty and the lack of a suitable means of identifying and counting patches; clustering algorithms were used in an attempt to count patches based purely on the locations of LTi cells, however they were found to be unreliable with unacceptable detection errors that resulted in over-estimation of both PP area and number.

2.3.3 Flow Cytometry Emulation

Experimental biologists experienced with flow cytometry could identify important patterns, populations and other results that could go unnoticed in other approaches to simulation analysis, or the significance of which may not be noted in analysis performed by a model developer that is not a domain expert. Flow cytometry software such as FlowJo (TreeStar) and WEASEL (Walter and Eliza Hall Institute of Medical Research) enable gating of events based on the value of multiple parameters, permitting identification of different phenotypes, and sub-populations within those phenotypes, and an exploration of the properties of these populations. Although for the sake of simplicity and demonstration purposes flow cytometry emulation was performed utilising just two factors in this analysis, depending on the complexity of the simulation, this approach can be extended *in silico* to an arbitrarily large number of expressed proteins (as opposed to genuine flow cytometry, which is currently limited to approximately 10 lasers in commercial devices). The multi-dimensional nature of flow cytometry data coupled with software designed specifically for its analysis in a biological context is what makes this approach particularly appealing. Attempting to perform the same sort of analyses on cell populations and phenotypes using the simulation alone, without flow cytometry emulation, would be extremely non-trivial and time consuming.
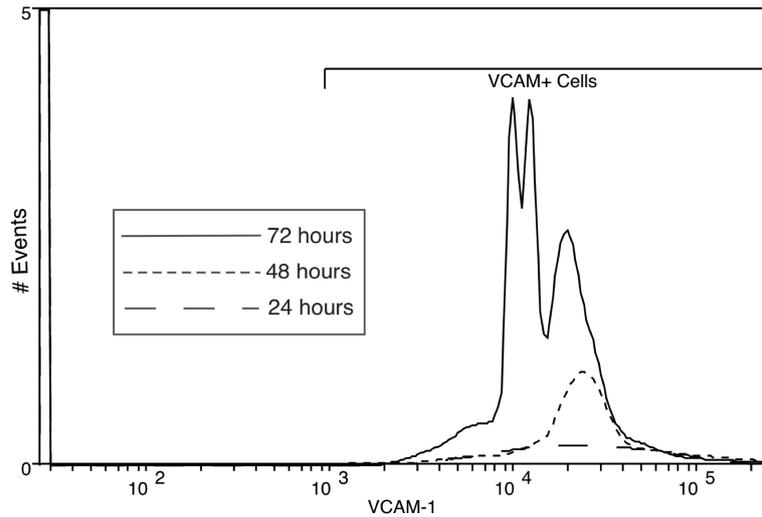
Figure 2.14: VCAM-1 mean fluorescent intensity histogram for all cells, taken at 24 (long dash), 48 (short dash), and 72 (solid line) hour time-points. The emergence of a prominent peak is visible after 48 hours, and by the end time-point this has diverged into several distinctly identifiable sub-populations, each corresponding to VCAM-1 expression levels within a specific Peyers patch.

New insight into the model dynamics arose from the simple emulated flow cytometry analysis shown in Figures 2.5 to 2.6. There is a clear emergence of two distinct populations of LTo cells in Figure 2.5A. Interestingly, it can be seen in Figure 2.14 that there is a sudden divergence after 48 hours that created these populations, before which expression levels conformed to a Gaussian distribution. The emergence of the Gaussian distribution over the VCAM-1+ population is encouraging, as this is indicative of adherence to the central limit theorem and would be expected in the domain. If the data were not normally distributed, it could be construed as evidence that the implementation of adhesion molecule expression within the model is an unsuitable abstraction. The divergence after 48 hours could either be a simulation artefact, or it could be that each patch has a relatively uniform distribution of VCAM-1 dependent on its unique properties, such as size and population size of co-localised LTi cells. Therefore this simple flow cytometry analysis affords a novel prediction regarding the dynamics of PP formation: multiple distinct populations of LTo cells emerge with large differences in VCAM-1 relative expression. This prediction is supported by the imaging emulation analysis in Section 2.2.2.2 and the

spatiotemporal heat-map shown in Figure 2.10B.

Any predictions arising from an emulated flow cytometry analysis is a prediction arising directly from the original simulation dynamics, insomuch that the techniques described herein do not effect extant simulation behaviour. Such predictions were merely unrecoverable with previously applied analytic techniques. Therefore, validation of the these techniques must be considered within the context of the simulation to which they are applied, as what is a suitable approach for one simulation may be wholly inappropriate when applied to another.

2.3.4 Simulation Visualisation and Emulation as Tools for Validation and Enhanced Prediction

Through the application of high performance, high-throughput computing, emulated experimental technique results from many simulations can be analysed and combined to provide a realistic quantitative analysis of biological simulations to provide data that can be more easily integrated into biological experiments, and enable direct comparison between computational and *in vivo* animal models and *in vitro* cell culture systems. This approach has the potential to somewhat simplify simulation calibration, given the data output from the simulation maps to the same data structures seen in biological data, the fitness function of the simulation effectively becomes the primary biological data, allowing systematic exploration of the parameter space to identify the points at which the simulation output is statistically no different from domain experimental data. Combined with genetic algorithms or other evolutionary computation approaches, the speed and accuracy by which simulations may be automatically calibrated by computer could be improved significantly.

The development of simulation outputs that directly reflect data obtained from the laboratory confers several advantages when used to augment current approaches to simulation analysis. Sensitivity analysis and other descriptive statistical methods are extremely important in evaluating simulations, and have proven useful in determining simulation robustness, aleatory uncertainty, and the roles of specific

parameters with respect to particular outputs (**???**). However, they do not offer the mechanistic insight into the spatial organisation of cells and the structures they form, or the changing cell phenotypes and emergent populations that may be observed using time-series emulated flow cytometry analysis, emulated histology and spatially resolved heat-maps. We therefore argue that current best-practice simulation analysis methodologies should be augmented with emulations of biological experimental techniques. Sensitivity analysis may then be performed that determines the effect magnitude of parameters on more biologically relevant outputs.

2.3.5 Feedback from the Life Sciences Community

Although a usability or human-computer interaction-type study has not been performed for use of these techniques and methods, the work described herein has been presented to those in the life-sciences at a number of international conferences, seminars and meetings. Feedback obtained from these has been overwhelmingly positive, particularly from those with no background in *in silico* modelling.

The author found that using these visualisation methods in the presentation of models significantly improved audience response and interaction at conferences with audiences consisting predominantly of experimental biologists. As such, the work described in this chapter has served as an incredibly useful tool for improving engagement with mathematical and computational models from the life sciences community. Additional value is gained in not having to spend presentation time providing exposition on model outputs that life scientists may be unfamiliar with, by focusing on the use of emulated flow cytometry, microscopy and gene expression images where possible. A final note in this area is that the caveat that *in silico* data are being presented is extremely important to avoid confusion with *in vitro* and *in vivo* models.

2.4 Future Developments

The principles and methodologies described herein, although broadly applicable to spatiotemporal *in silico* models, presently require unique manual implementation on a simulation-by-simulation basis. The development of a software library that permits automated generation of FCS files, emulated micrographs and spatial/temporal heat-maps of gene and protein expression, would permit quicker implementation of these techniques and therefore broader adoption of the approach. The method should easily generalise to other agent-based models through manual adoption of the techniques described herein, and to any system that encapsulates heterogeneity at cellular or molecular levels. However, one of the key challenges in developing an easily reusable software package would be not placing overly burdensome constraints on the data structures used to store location or expression data, important to ensure the tool-kit is not platform-dependent (i.e. does not depend on the modelling framework or programming language used) and to allow retroactive application to existing simulations. Mandating that cell locations and other properties be stored in highly specific data structures is likely to limit adoption, as such decisions often rely on structures provided by the modelling tool-kit being used and are heavily dependent on both the biological processes being modelled.

Extension of the DNA microarray/RNA-seq inspired heat-maps to include other aspects of bioinformatics analysis associated with these technologies would be a valuable addition to this work, and could be achieved through integration with existing software libraries such as the Biopython project (**?**) or PyCogent module (**?**) for the Python programming language.

Chapter 3

Developing a Domain Model of Tertiary Lymphoid Tissue Formation during Autoimmune Disease

This chapter describes the development of a hybrid, multi-scale model and computer simulation of tertiary lymphoid tissue (TLT) formation in autoimmune disease, using Sjögren's syndrome as an archetypical model within which to explore the formation process. Presently, there is no clear understanding of the fundamental, minimum requirements for the development of TLT. In vitro *models have provided important clues regarding the role of tissue-resident stromal progenitor cells but have proven inadequate for thoroughly evaluating hypotheses in TLT induction. Although there are a plethora of* in vivo *murine models of Sjögren's syndrome, many of which include TLT formation, these acute induction models do not closely reflect the process during chronic inflammation, and it is difficult to isolate the key cytokines, chemokines and cell types that fundamentally drive the process: while many signalling molecules may perturb the dynamics of formation, which differs across autoimmune diseases, cancers and chronic infections, identifying the princi-*

ple, non-redundant cellular and molecular mechanisms, and signalling feedback loops that drive lymphoid tissue neogenesis is a crucial step in understanding their larger role within autoimmunity. One of the largest challenges in understanding the formation and role of TLT in disease is that the mouse models used in experimentation are acute, and therefore not truly reflective of the chronic inflammatory environment within human disease. This chapter describes a hypothesis developed from human in vitro cell culture and in vivo mouse model data that aims to describe the minimum requirements for TLT formation to occur, and utilises state-of-the-art computational modelling methodology to confirm the veracity of this hypothesis. Mathematical and computational models are well-suited to investigating TLT formation due to their innate capacity for abstraction: to disregard mechanisms not thought to be pre-requisites for TLT formation, and to systematically explore the dynamics of a handful of signalling mechanisms and cell types hypothesised to be crucial to the development process.

Following a review of the relevant aspects of immunology and autoimmune pathology, a TLT formation domain model is presented: a formalised description of our hypothesis for TLT development, developed in collaboration with domain experts, experimental datasets, and further informed by existing literature. We then develop a platform-independent model; this describes how each aspect of the model may be implemented as a computer simulation. The TLT formation model integrates Markov chains, formal grammars, cellular automata, agent-based modelling, ordinary and partial differential equations into an executable simulation that captures molecular, cellular and tissue level phenomena at time-scales from one second to several minutes. Through identifying the most appropriate modelling technique for each individual model entity, it is ensured that the process is captured as accurately as possible, such that the model is fit-for-purpose and without necessitating the additional abstractions warranted if one was to attempt to capture all model components within the bounds of a single modelling technique. To achieve this, a novel framework and supporting schemata are developed and utilised for defining hybridised

models composed of many sub-models in terms of information flow between each of the sub-models. I further develop this framework to permit the production of biologically relevant outputs based on the methodologies described in Chapter 2 through combination and manipulation of data structures contained within each sub-model.

3.1   Tertiary Lymphoid Tissue: Form and Function

Lymphoid tissues are complex immunological structures, evolved to orchestrate key cell types and molecular interactions into functional immune responses. This is achieved through the development of a niche, which supports the retention, activation, and proliferation of immune cells. An individual lymphoid organ usually contains several microanatomical sub-compartments or niches - specialised areas for cellular entry and function. Lymphoid tissues contain distinct, segregated T and B cell zones with separate and supporting functions. The maintenance of a niche is managed by stromal cells, which provide survival, activation and migratory factors. Thus to fully understand lymphoid tissue function it is important to consolidate current understanding of how stroma regulates immune cells, either through direct interactions or via soluble signals. Experimental studies, typically in mice, have provided insights into the molecular and cellular mechanisms driving the development and maintenance of lymphoid organs (**???**).

3.2   TLT Domain Model Development

The dynamics of tertiary lymphoid tissue formation are highly multi-factorial, non-linear, inherently spatial, and involve phenomena occurring over time-scales from seconds to weeks; these features do not lend well to the utilisation of simple mathematical models, thus complex systems analysis must be applied, in which tissue formation is treated as an emergent phenomenon occurring due to the interaction of

many lower-level entities (i.e. molecular and cellular interactions). An introduction to the form and function of TLT was provided in Section 1.1.2; in this chapter, a combination of experimental data, existing literature and clinical studies are brought together to define a model of TLT formation that excludes many of the extraneous components of the system and strives to capture the essential signalling molecules and cell-types responsible for orchestrating the self-organisation process.

The hybrid-ABM paradigm, discussed in Section 4.2, lends itself well to knowledge integration due to its intrinsically modular organisation, capacity to describe phenomena occurring on distinct spatiotemporal scales simultaneously, and provides a highly visual output, particularly considering the visualisation and model communication methods discussed in Chapter 2 and means of defining them, to be discussed in Section 4.8. Such approaches afford a means of consolidating information at a systems level, supported by a strong evidence-base from biological experimentation demonstrated through the use of argument-driven validation (developed for the TLT model and simulation in Section 4.9) (**???**) derived from principles of safety-critical engineering (**?**). The resulting model and simulation provides an executable platform for hypothesis testing. The development of a hybrid model system of TLT formation can produce insights which otherwise may not have been reached *a priori*, and subsequently generate predictions which can be tested *in vivo* - leading to further model refinement; we term this paradigm *'model-driven experimentation'*, illustrated in Figure 3.1. There is in addition a clinical challenge in that there are already many thousands of possible different combinations using existing therapeutics (biologics and small molecule drugs) that would need to be trialled in order to find optimal targeting strategies to resolve TLT pathology. Thus MDE-based approaches as illustrated in Figure 3.1 provide a rational approach to identify novel combination therapeutic regimes that have a best potential in clinical trials.

Clinical data from a large cohort of Sjögren's syndrome patients have demonstrated that the presence of TLT in salivary glands is an effective predictor of more aggressive disease and development of B-cell lymphoma within the salivary gland
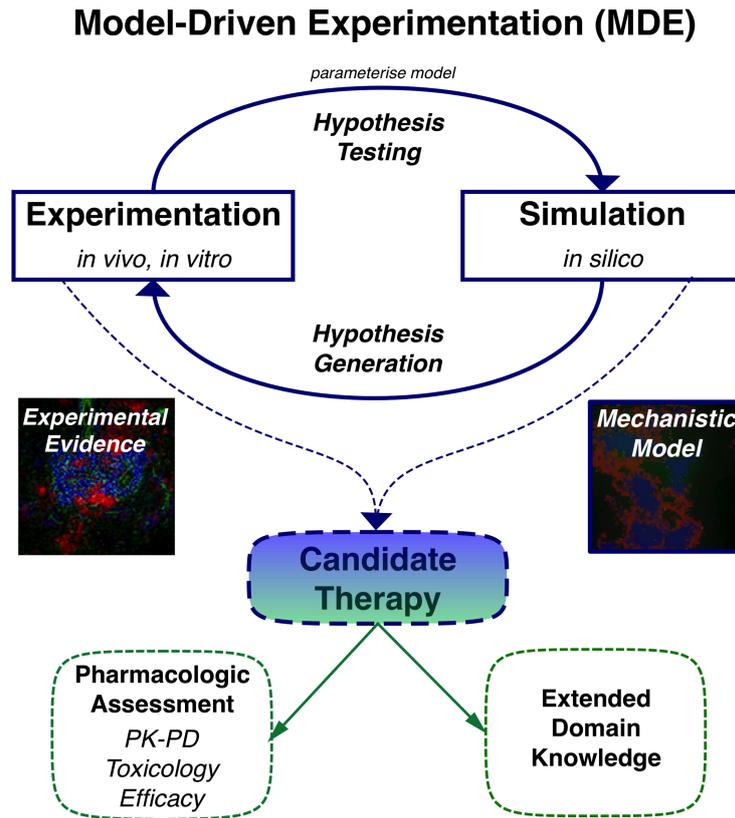
Figure 3.1: Model-driven Experimentation Paradigm - Figure adapted from **?**.

(**?**). The likely contributor to the worsened prognosis in the presence of TLT is the development of germinal centres (**?**) capable of selection and expansion of autoreactive B cells and plasma cell differentiation (**?**).

Taken together, these data suggest a potential means of therapeutic intervention in Sjögren's syndrome, by both preventing their formation and identifying means of dispersing established TLT structures. In order to achieve this, it is essential to understand what is necessary and sufficient for TLT formation and maintenance, that is, to identify the *minimum requirements*. This section presents a domain model of TLT formation in Sjögren's syndrome, integrating *in vitro* cell culture and *in vivo* mouse model datasets with observations from clinical studies.

To develop a useful domain model, at least one research question must be identified *a priori*, and a hypothesis derived that aims to answer the research questions, such that implementation of the model is capable of ultimately determining the

veracity or theoretical feasibility of the biological hypothesis. A 'bottom-up' simulation cannot definitively confirm that a hypothesis is true, but can demonstrate whether expected emergent phenomena emerge when tested *in silico*, and in combination with a strong argumentation case supported by clinical, *in vitro* and *in vivo* data, a strong argument can be made as to hypothesis veracity. The basic research questions underpinning this model may be defined as:

- What are the fundamental minimum requirements that enables TLT formation?

- What possible intervention strategies may prevent or reverse the process of TLT formation?

With our research questions in mind, the first step in developing a domain model is to identify key observable phenomena that occur during this process, and crucially, the time-points at which they occur. This allows the creation of an observable phenomena time-line that we must seek to encapsulate within our model, such that these phenomena emerge naturally from the underlying model entities (cells, molecules, etc.) when the system is ultimately implemented and simulated. Hypotheses may then be developed that explain each aspect of the formation process. The development time-line for our model of TLT formation is derived primarily from an adenovirus mouse model of Sjögren's syndrome, described in **?**, wherein TLT, with very similar characteristics and function, can be induced in murine salivary glands, developing over approximately a 15 day period before rapidly resolving. This is in contrast with the human disease in which there is no known role for infection as a trigger for TLT formation, and thus might develop due to the presence of chronic inflammation; furthermore, there is no evidence in humans of the resolution phase following development. The structure of submandibular gland TLT is shown in Figure 3.2, illustrating the co-localisation of B cells with their cognate chemokine CXCL13, and T cells with CCL21. This permits inference that a differential response to CXCL13 and CCL19/21 can be assumed for B and T cells.

By focusing on *minimum requirements*, avoiding incorporation of many molecular pathways known to be involved in TLT formation and Sjögren's syndrome, translational issues between murine and human models of disease can be minimised; this is particularly true due to the formation of the hypotheses as whether these system of differential chemokine feedback loops and stromal cross talk is *consistent* with the formation of self-organising TLT-like structures.

The formation of TLT pathology in autoimmune disease is associated with the formation of germinal centres, thought to be the primary source of auto-antibodies, (**????**), these structures are composed of B cells and require T-cell signalling for effective function. T and B cell organisation is dictated by lymphoid chemokines, thus any model capable of addressing key questions in TLT formation and function must simulate chemokine function. Mesenchymal stromal cells have an essential role in the development of lymphoid tissues, as discussed in Section 3.1, although the cellular origin and mechanisms driving their formation is less clear, recently it has been shown within the our laboratory (**?**) that mesenchymal progenitor cells can be induced to take on the different stromal cell fates *in vitro*, illustrated in Figure 3.3), events that have been verified using *in vivo* models.

The presence of the chemokines CCL19, CCL21 and CXCL13 has been observed in tertiary lymphoid tissues in many different animal models and in human pathology (**?**). CCL19 and CCL21 have potent chemotactic effects on T cells and activated dendritic cells, in contrast CXCL13 is a potent B-cell chemoattractant. It has previously been shown that localised inflammation is required to trigger tertiary lymphoid organs, this may result from infection, autoimmunity, neoplastic haematological cancers and solid tumours (**?**). The expression of hemostatic chemokines and their role in secondary lymphoid tissue formation suggests a significant role for chemokine production by ectopic lymphoid stroma in orchestrating spatiotemporal organisation mediating both B and T cell tissue colonisation and segregation. During secondary lymphoid tissue formation, local tissue-resident fibroblasts require interactions with specialised lymphocytes that regulate stromal cell differentiation
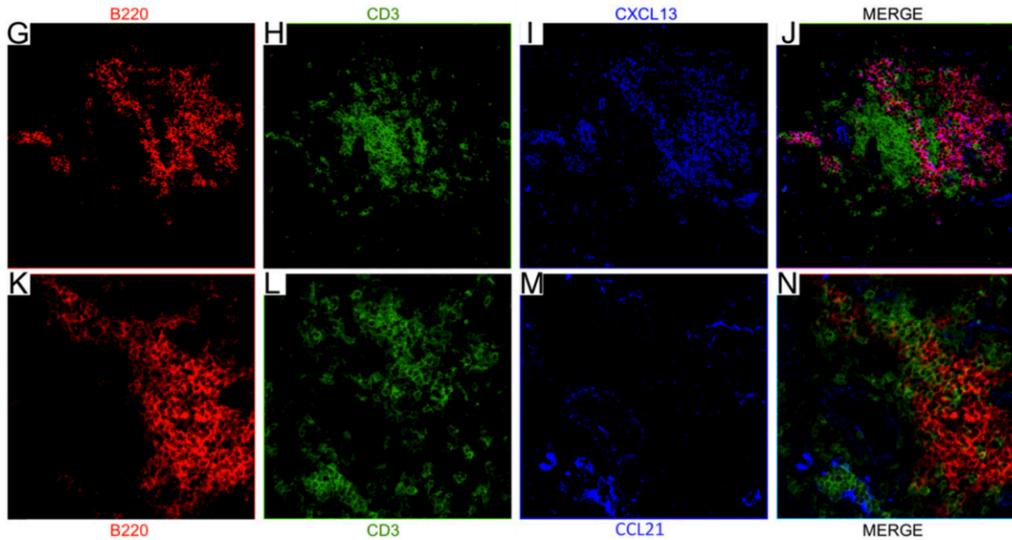
Figure 3.2: IHC confocal micrographs confirm expression of lymphoid chemokines CXCL13 (blue, I) and CCL21 (blue, M), also showing colocalisation of CXCL13 (B cell chemokine) within the B cell (red, CD19) populated region (GJ), and of CCL21 (T cell chemokine) within the T cell (green, CD3) populated area of the aggregate. Original magnification x200. Taken from **?**.

and acquisition of effector function stimulation by lymphocytes in order to adopt and maintain the lymphoid stromal phenotypes that produce the chemokines responsible for recruiting further lymphocytes and other mononuclear cells; this results in a positive feedback loop - the self-perpetuating process of lymphocyte recruitment and chemokine secretion by stroma. However, in contrast to secondary lymphoid tissue, the trigger for TLT development is inflammation, thus the initial recruitment of lymphocytes is likely triggered by the milieu of chemokines produced as part of the inflammatory process. This process has not been explicitly included in the model in an emergent sense, it is assumed lymphocytes enter the tissue compartment at the timepoints observed *in vivo*.

To identify the events that drive homeostatic chemokine production human adipose-derived stem cells (ADSC), adult pluripotent mesenchymal progenitor cells, were cultured in the presence inflammatory cytokines (replicating infection and inflammation in vitro) and the presence of lymphocytes. Using a combination of quantitative PCR and ELISA the expression of CCL19/21 and CXCL13, shown
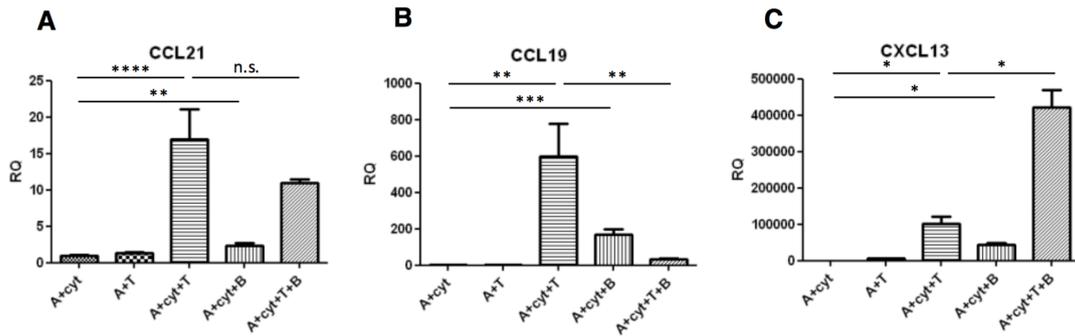
Figure 3.3: ADSCs co-cultured with inflammatory cytokines, cytokines and T cells, cytokines and B cells, and finally cytokines, T and B cells: qPCR relative expression of chemokines CCL19, CCL21 and CXCL13. Data published in **?**, experimentation performed by Bridget Glaysher.

in Figure 3.3. Expression of T cell lymphoid chemokines CCL19 and CCL21 was shown to be dependent on T cell co-culture with ADSCs that were pre-treated with cytokines (IL13/TNFα), in contrast the co-culture of B cells leading to the expression of CXCL13, consistent with their intrinsic expression of chemokine receptors for these chemokines. Interestingly, the presence of B cells suppressed *in vitro* the expression of CCL19/21. The mechanisms for this are unknown, however this might be related to the upregulation of CXCL13. In contrast to the T cell chemokines, CXCL13 expression is dependent on the presence of T cells. This implicates a role for T cell derived signalling in CXCL13 induction. To determine if these signals resulted from receptor-ligand interaction or indirect cytokine production by lymphocytes, ADSCs and lymphocytes were separated using a transwell that prevents physical interactions between the two cell types. No chemokines were induced in the absence of direct cell-cell contact. These results have been further validated using an *in vivo* model that demonstrates an absolute requirement for lymphocytes in chemokine induction, and identified LTβ (lymphotoxin beta) as having an essential role in chemokine induction.

In contrast to chemokines, upregulation of the key adhesion molecules VCAM and ICAM does not require the presence of lymphocytes, rather just stimulation by
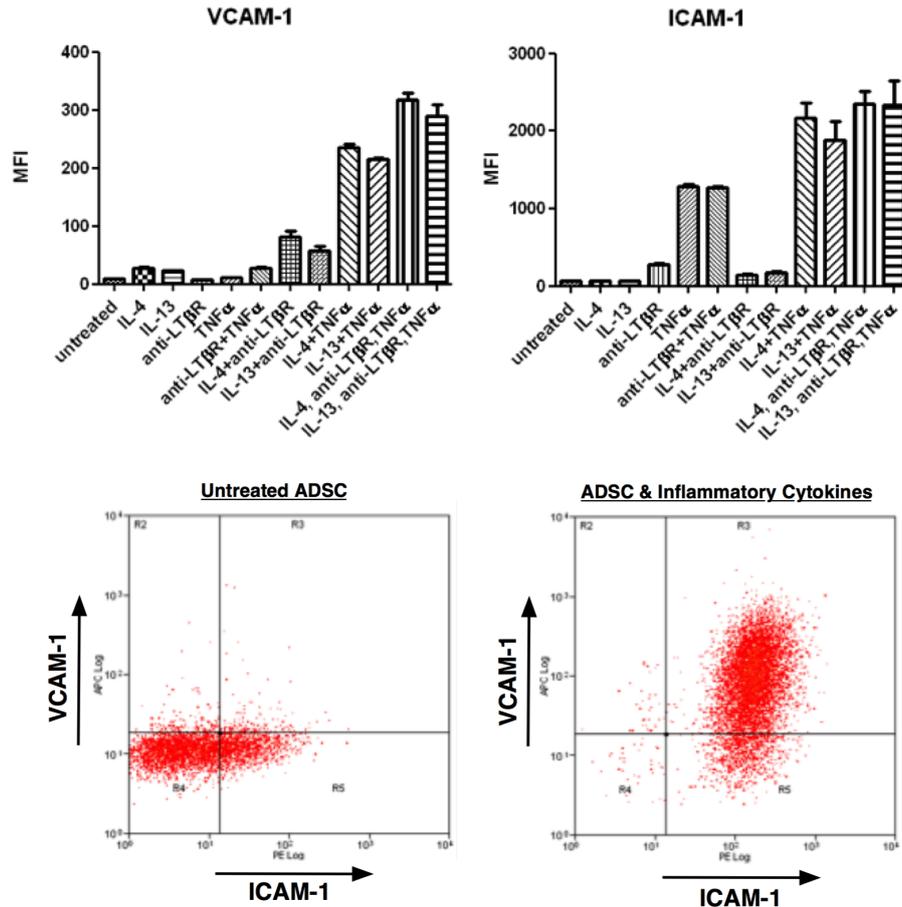
Figure 3.4: Inflammatory cytokines TNF-α and IL-13 induce adhesion molecule expression in ADSCs. (Previously Unpublished data from experiments by Bridget Glaysher.)

cytokines alone. Using the ADSC culture system, the addition of TNF and IL13 was sufficient to rapidly drive upregulation of ICAM and VCAM. Alone neither cytokine is sufficient to induce adhesion molecule upregulation, indicating a key role for cross-talk and synergy between the signalling pathways (Figure 3.4). This is further supported by *in vivo* data showing a key role for IL4 receptor signalling mediated by IL13 in the upregulation of adhesion molecules and expressing an activated stromal state including the upregulation of podoplanin. As observed *in vitro*, these events do not require the presence of lymphocytes, in Rag deficient mice, which lack both B and T cells, upregulation of these molecules occurs normally.

3.2.1 From Domain Data to developing a TLT Domain Model

Through considering a combination of the *in vitro* and *in vivo* datasets, and immunofluorescent staining of salivary gland biopsies from human samples, I have developed a basic model that describes the change in stromal cell 'state during the pathological disease process (Figures 3.3 and 3.4). Based on the data described in Section 3.2, I have developed a 'four state' model of stromal cell differentiation, in which a localised stromal precursor cell, following activation by inflammatory cytokines, is assumed to require a certain level of initially T cell, and later B cell contact, to stimulate pathological stromal development. Between each stage of development, further T and B cell stimulation that is not yet sufficient to cause a state change is assumed to accentuate the phenotype by further up/down-regulation of CCL19 and CXCL13 expression levels towards the limits defined by the data (Figure 3.3). The data from these co-culture experiments allow one to derive a simple four state model of stromal cell differentiation in response to three key 'triggers': the presence of soluble inflammatory cytokines causing resident stromal progenitor cells to differentiate into lymphoid stroma precursor cells, direct T cell contact resulting in the development of a fibroblastic reticular cell (FRC)-like phenotype, and direct B cell contact finally causing development of an FDC-like phenotype. This is encapsulated in Figure 3.5, and forms the basis of constructing a computational model that captures the dual chemokine feedback loop responsible for TLT formation.
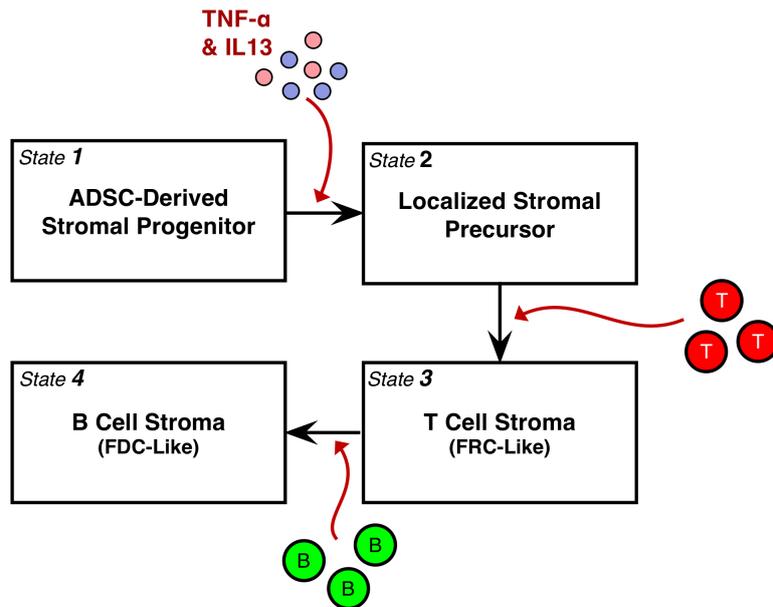
Figure 3.5: Simplified illustration of the 4-state stromal cell differentiation model. Stromal progenitors develop into lymphoid stroma in the presence of inflammatory cytokines IL-13 and TNF-α, adopt an FRC-like phenotype following T-cell stimulation and ultimately an FDC-like phenotype upon B-cell stimulation. The T and B cell factors responsible for causing the differentiation are unknown.

Taken together, the data in figures 3.3 and 3.4, along with the observations drawn from the *in vivo* model described by **?**, has allowed identification of low-level behaviours that drive emergent TLT formation, and to establish the time-scales over which the different process occurs. The relative expression levels of CCL19, CCL21 and CXCL13 allow thresholds for relative expression of these factors to be defined as biologically-derived parameter values. Furthermore, we can state a hypothesis that would be extremely difficult to test using *in vivo* animal models or *in vitro* cell culture systems, but may be tested *in silico* through the innate capacity of simulation to abstract factors not relevant to the research questions. Combining these data and observations, our principal hypothesis is stated in Figure 3.6.

3.2.1.1 Domain Model Limitations

Although this model provides capacity for analysis of the interactions between lymphocytes and localised mesenchymal cells to drive highly organised structures leading

---

Hypothesis

TLT formation occurs as a 'self-organising' emergent structure in Sjögren's syndrome following priming of mesenchymal stroma due to the presence inflammatory cytokines IL-13 and TNFα through two differential feedback loops of chemokine induction, lymphocyte migration and further stimulation by T and B cells, causing development of lymphoid stromal phenotypes from tissue-resident progenitors.

Figure 3.6: Central hypothesis that our TLT model aims to demonstrate: stromal cell chemokine induction due to chronic inflammation, resulting in lymphocyte recruitment and subsequent self-organisation. This is derived from a combination of *in vitro* cell culture data, *in vivo* mouse models and clinical datasets.

to the formation of germinal centres, it does not analyse the functional capacity of TLT to drive GC formation, the formation of autoantibodies, or the role of TLT itself in disease pathology as there is no formal evidence that it is required for disease formation and progression. However, developing a model of the induction and formation process will address key biological questions concerning mechanisms driving TLT organisation and formation, permitting analysis of therapeutic interventions that may modify these outcomes. As FDC development is critical in the formation of germinal centres and therefore autoantibody production (**?**), the change in ratios of stromal cell phenotypes as the model parameters are perturbed may be used as an appropriate surrogate for pathological outcomes, this is explored in Chapter 6.

3.2.2 Defining Expected Behaviours

Developing the domain model for TLT found in Sjögren's syndrome through a review of the relevant biological experiments and scientific literature, a cogent timeline was constructed that encapsulates the individual cellular and molecular model entities with the higher-level emergent phenomena of tissue development through self-organisation. The key observable phenomena resulting in the emergence of TLT formation are described in a cartoon based on multiple different experiments, shown in Figure 3.7 (**?**). This simplified multi-step process illustrates potential triggers including viral infection, although this is unlikely to be the only process involved in

TLT initiation. Localised inflammation drives the recruitment of leukocytes (Dendritic cells, T and B lymphocytes) to the inflammatory lesion, and stromal cells begin to adopt lymphoid phenotypes (principally an FRC-like phenotype). Finally, the structure forms, with clear separation within the niche for B and T cells, the B cell zone forming a follicular structure containing follicular dendritic cells that permit the generation of germinal centres capable of driving autoantibody responses. Recently, there has been growing clinical interest in TLT as a therapeutic target, with the development of biologic agents intended to both reduce or prevent their formation (e.g. lymphotoxin fusion protein (an antagonist for Lymphotoxin receptor β), Baminercept), and to induce TLT using lymphotoxin receptor agonists (i.e. LIGHT) in cancer (**?**).
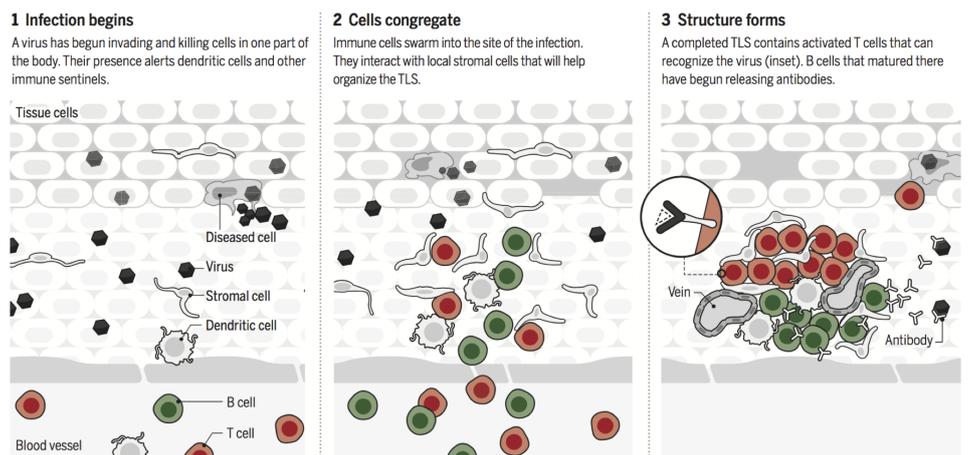
Figure 3.7: Figure describing the three stages of TLT formation: induction, colonisation, segregation. Taken from **?**.

The 'Expected Behaviours' diagram illustrated in Figure 3.8 develops upon the simple schema presented in Figure 3.7. This diagram aims to segregate but include logical links between what cellular and molecular processes take place, what may be directly observed through experimentation, and our hypotheses on how these processes result in what is observed. The 'expected Behaviours' diagram of a biological domain model is a useful tool for summarising the relevant model entities and the behaviour that is expected to manifest. The aggregate effects of these many, massively parallel, individual interactions, lead to the emergence of system wide

patterns and behaviours that are not explicitly encoded or intuitively understood from the defined interactions alone.

The stroma has been hypothesised to have an essential role in TLT formation, however how the interactions between lymphocytes and stroma lead to a highly organised and function tissue are unclear, thus I have proposed that two feedback loops, one via T cell signalling and chemokine-mediated recruitment, and the other via B cell signalling, and chemokine-mediated recruitment, outlined in the expected behaviours diagram (Figure 3.8). In this diagram, each coloured solid arrow indicates that it is responsible for the stromal differentiation shown by the dashed arrow of the same colour. Segregation occurs as B cells cause stromal progenitors to adopt an FDC phenotype, secreting predominantly CXCL13, in an area surrounded by FRC-like cells that predominantly secrete CCL19. Thus, the T cells migrate to the neighbouring FRC network as the B cells cluster tightly in a follicular structure supported by FDC-like cells. While this is a simple and intuitive idea, testing it experimentally has proven difficult thus far, and therefore a computational modelling approach has been employed herein to attempt to demonstrate the veracity of this hypothesis.

The simple model presented in this section and in Figures 3.8 and 3.7 may be somewhat confounded through the inclusion of receptor internalisation, which modulates both lymphocytes' ability to detect chemokine gradients, and the local levels of chemokine surrounding each cell. It was found in **?** that tissue pattern stability in secondary lymphoid tissues is highly dependent upon transmembrane chemokine receptors, following reports in **?** that receptor recycling has a significant effect on lymphocyte trafficking. It was however determined *in silico* (**?**) that receptor internalisation cannot account for regulation of B cell lymphoid follicle size despite experimental evidence of its involvement in secondary lymphoid tissue organogenesis (**???**). There is little experimental data in the literature regarding the role of chemokine receptor internalisation in lymphoid neogenesis, however there is evidence that it has important roles in the formation of *de novo* structures such as

tertiary lymphoid organs (**?**), and has possible roles in 'tuning' immune responses through chemokine gradient sculpting (**??**).

The actual significance of receptor internalisation and chemokine sequestration by lymphocytes in TLT formation and function is therefore largely unknown and likely to be highly dependent upon properties difficult to quantify experimentally including the relative difference between local chemokine concentrations, lymphocyte population size and the rate at which they internalise chemokine-ligand complexes. However the overall effect on development dynamics is likely to be generally similar in nature to those observed in secondary lymphoid tissue organogenesis, particularly in light of observations in **?**. These effects may be explored with an *in silico* model of TLT formation, therefore an ODE model describing the internalisation, recycling and decay of receptor-ligand complexes on an individual lymphocyte level has been developed and will be expanded upon in Chapter 4 (Section 4.5).
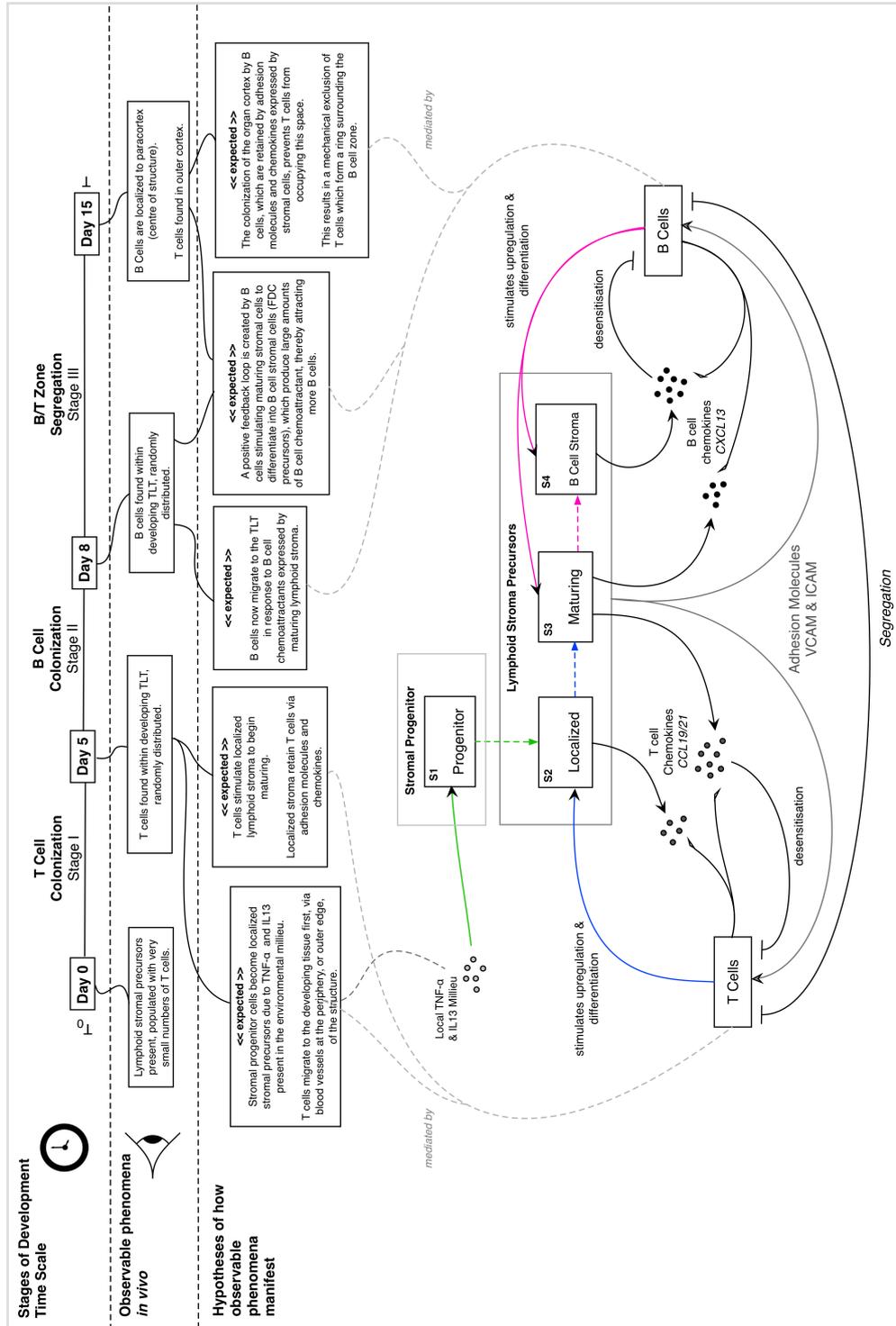
Figure 3.8: Expected Behavioursdiagram for TLT formation: this describes the observable phenomena during TLT formation, the entities that mediate the emergence of these phenomena and how micro–scale or 'lower–level' entities interact to produce the higher level behaviour of tissue formation.

3.2.3   Describing Cellular Behaviour

This section outlines a semi-formal description of lymphocyte and stromal cell behaviour within the domain, such that in Chapter 4, a computational and mathematical model of their dynamics can be appropriately derived.

3.2.3.1   Capturing Cellular Dynamics

The key cell types identified in the domain model, illustrated in the expected behaviours diagram (Figure 3.8), are represented explicitly in the TLT formation domain model. These cells are the haematopoetic T and B lymphocytes, and mesenchymal stromal cells which undergo a differentiation process adopting various distinct phenotypes. Finite state machines can be used to explicitly capture each state that the cells may occupy, the conditions required for transition between states, and any activity that occurs as a result of these transitions (**?**). The unified modelling language (UML) is a widely-used international standard for describing object-oriented systems in engineering, and has more recently found use in describing complex biological systems (**???**). The UML permits expression of finite state machines through the representation of states in rectangular boxes, with arrows delineating transitions that may occur, 'guards' expressed within square brackets that describe the condition that must be true for the transition to occur, and notes preceded by a backslash denote activities that occur as a result of the transition. An entity may have orthogonal states, in which the object exists in multiple states simultaneously. This notation has been applied to a description of T and B lymphocytes, and stromal cells. The process of describing cell entities using finite state machines requires explicit identification of biological parameters, these are described in Table 3.9. A key advantage of using the UML to describe finite state machines is the inclusion of a specification for *hierarchically-nested states*: effectively sets of states inside of states. If state 'B' and 'C' are within state 'A' then B is said to be a substate of A, and A is the superstate of B; collectively the set of current states within A form a *composite state*. This is particularly useful for describing the

Parameter	Model Entity	Value	Description
T Cell Entry Time	Lymphocytes	Day 0	Time point during TLT formation that T cells colonise tissue
T Cell Entry Rate	Lymphocytes	--	Unknown, assumed to be linear
B Cell Entry Time	Lymphocytes	Day 8	Time point during TLT formation that B cells colonise tissue
B Cell Entry Rate	Lymphocytes	--	Unknown, assumed to be linear
T Cell Velocity	Lymphocytes	10.7 µm/s	Mean observed value, from Miller (2003)
B Cell Speed	Lymphocytes	4.16 µm/m	Mean observed value, from Shulman (2014)
CXCR5 Surface Levels	Lymphocytes	1.00E+06	(Approximately on the order of, measured in lab)
CCR7 Surface Levels	Lymphocytes	--	Unknown
CXCR5 Recycling Rate	Lymphocytes	--	Unknown
CCR7 Recycling Rate	Lymphocytes	--	Unknown
S2 CCL19 Expression	Stroma	0 (RQ)	CCL19 Expression by Localised Stromal Precursors (S2) **(Fig 3.3)**
S3 CCL19 Expression	Stroma	800 (RQ)	CCL19 Expression by Localised Stromal Precursors (S3/FRC) **(Fig 3.3)**
S3 CXCL13 Expression	Stroma	~100,000 (RQ)	CXCL13 Expression by Maturing Stromal Precursor (S3/FRC) **(Fig 3.3)**
S3 VLA4 Expression	Stroma	~400,000 (RQ)	VLA4 Expression by Maturing Stromal Precursor (S3/FRC) **(Fig 3.3)**
S4 CCL19 Expression	Stroma	~25 (RQ)	CCL19 Expression by Localised Stromal Precursors (S4/FDC) **(Fig 3.3)**
S4 CXCL13 Expression	Stroma	~200 (RQ)	CXCL13 Expression by Maturing Stromal Precursor (S3/FDC) **(Fig 3.3)**
ICAM1 Expression	Stroma	2200 (MFI)	ICAM1 Expression by stimulated stroma **(Fig 3.4)**
VCAM1 Expression	Stroma	300 (MFI)	VCAM1 Expression by Maturing Stromal Precursor (S3/FDC) **(Fig 3.4)**
S2-S3 T cell Stim Reqs	Stroma	--	Requirements for differentiation from S3-S3, Unknown
S3-S4 B cel Stim Reqs	Stroma	--	Requirements for differentiation from S3-S4, Unknown
Gland Stroma Density	Stroma	--	Characteristics of stroma in submandibular gland, Unknown

Figure 3.9: Table of identified Domain Model parameters.

complexity of stromal cell development succinctly, as the four phenotypes it may adopt and the properties associated with each can be described as four composite states, with the phenotype-specific states within these described as substates of the phenotypic state.

Several abstractions and assumptions have been made in the construction of the cell state machines, removing extraneous detail that is not necessary to describe the dynamics that result in TLT formation. Additional assumptions are necessary when entities identified within the domain model have unknown values associated with them or are themselves unknown, this is particularly the case with this domain model as the aim is to determine in a highly theoretical manner whether two differential feedback loops driven by the secretion of and response to chemokine is sufficient to describe TLT formation generally, without regard to the myriad other cells, cytokines and signalling molecules known to be present in TLT in different pathologies. All such assumptions and abstractions must be clearly stated to produce a transparent domain model in the interest of aiding interpretation and avoiding ambiguity within the model. While these assumptions are documented within the

domain model description, they are also explicitly addressed and justified using goal-structuring notation in Section 1.4.1.3 along with all platform model assumptions and simulation implementation decisions. Together, these diagrams form a coherent argument that the model is fit for purpose.

1. **T and B Lymphocytes**

   For the purposes of describing TLT formation, T and B lymphocyte dynamics are very similar; the principle difference being the chemokine to which they respond to and the effect they have upon stromal cells. T and B lymphocytes both undergo adhesion to stroma and chemotaxis in response to chemokine gradients in essentially the same manner, although B cells respond only to lymphoid chemokine CXCL13, and T cells also to CCL19. In the domain, B lymphocytes can also respond to a lesser degree to CCL19, and T cells to CXCL13 (**?**) although for the purposes of the model, this is assumed not to be the case. The primary role of these additional chemokine responses are to orchestrate the immune response within lymphoid tissue by enabling T-B interactions; as the model is concerned with the establishment of tissue and does not capture the humoral response, we feel it is justified to reduce the system to the simple differential response described herein.

   (a) T Lymphocytes

   Depicted as a UML finite state machine in Figure 3.10, the T cell begins in the initial '*psuedostate*' depicted as a black circle, and immediately assumes several orthogonal states: 'S3 Stromal Stimulation Factor Expression', 'CCR7 expression', and 'LFA-1 Expression'. The cell also must make a choice between two states, either responsiveness to local CCL19 or a state of being in random motion – the former state is adopted if and only if the T cell is expressing sufficient levels of cognate receptor CCR7 and the local chemokine concentration is high enough to stimulate downstream signalling from CCR7 that results in the chemotactic process, otherwise, the cell will adopt the random motion state. In addition,

there is an activity notation stating that the cell is either already present at T_0 or enters at any point after. If the T cell adopts a chemotactic response, then receptor-ligand (CCR7-CCL19) complexes are internalised, the CCL19 is degraded and CCR7 is either recycled back to the cell surface or degraded, marked on the diagram by an activity notation below the chemotaxis state box. Quantitative T-cell motility data available from **?**.

Once a motility state has been adopted and the associated activity (movement, receptor internalisation) executed, there is a possibility that the cell will adopt a state of adhesion, in which it is bound to a stromal cell via receptor LFA-1 and cognate adhesion molecules on the stromal cell (VCAM-1, ICAM-1). The conditions for this to occur are that the T cell is in contact with a stromal cell, the stromal cell expresses adhesion molecules, the T cell expresses adhesion molecule receptors (LFA-1), and the bind between the cells is sufficient. In the domain, whether binding occurs will be a highly complex function including the number of adhesion molecule-receptor complexes that formed due to cell contact, and shear forces on the cell due to surrounding cells, fluid flow dynamics, and the cell's motility. Modelling such complexity on the scale of thousands of cells expressing millions of receptors is both intractable, and not desirable with respect to the model aims. Therefore, a probabilistic function is assumed, described by the guard 'Bind between cells is sufficient'. This probability is assumed to be directly proportional to the level of LFA-1 expression on the T cell and the level of ICAM-1/VCAM-1 adhesion molecule expression by the stromal cell. It should also be noted that there are multiple cognate receptors to stromal cell adhesion molecules beyond LFA-1, including for example, VLA-4. For the sake of simplicity, these receptors are assumed to have the same function and are abstracted into a single 'LFA-1' entity. If the conditions for adhesion are not met,

then the system makes a new determination as to the motility state of the lymphocyte and re-assesses whether the cell is in an adhesion state. If a state of adhesion is adopted, then the cell undergoes prolonged contact with the stromal cell with only localised movement around the cell.

The stromal cell response to T cell contact is captured separately within the stromal cell UML finite state machine, and it is assumed that stromal stimulation can only occur if the T cell adheres to the stromal cell such that 'glancing contact' is insufficient to stimulate differentiation or phenotypic accentuation in the stromal cell. The state of adhesion ends when the bind is no longer sufficient – this, again, is a complex function of biophysical forces and intracellular signalling processes within the domain, and for the purposes of the domain model, it is assumed that there is a probability that the cells will dissassociate at any given moment, which leads the cell back to determination of cell motility, resulting in an iterative loop of motility determination followed by adhesion determination. It should be noted that these transitions are not necessarily a function of time, the determination of motility and adhesion is described as occurring sequentially since the cell cannot undergo chemotactic migration if it is adhered, but these state determinations are not sequential in time; in the domain, motility and adhesion are continuous orthogonal processes. For the sake of simplicity, this is abstracted into the iterative sequential processes of state determination depicted iusing a UML finite-state machine.

(b) B Lymphocytes

Depicted as a UML finite-state machine in Figure 3.11, the B cell shares many similarities with the T cell. The principal difference is the 'S2 Stromal Stimulation Factor Expression' state replacing 'S3 Stromal Stimulation Factor', CXCR5 and CXCL13 replace CCR7 and CCL19 respectively, and cellular entry into the tissue occurs from the 5th day of the

15 day development process only, an abstraction of their recruitment to the ectopic tissue compartment through T cell-induced CXCL13 expression by FRC-like 'S3' lymphocyes. An iterative process of motility state determination and stromal adhesion occurs identical to that of the T cell described above. Quantitative B-cell motility data are available from **?**.

2. **Stromal Cells**

A stromal cell may exist in one of four phenotypes, as determined by the informal 'four-state stromal development model' described in Figure 3.5. The FRC-like and FDC-like phenotypes will be referred to as 'FRCs' and 'FDCs' for the sake of convenience, but it should be noted that also the lymphoid stroma phenotypes found in TLT are very similar to those of secondary lymphoid organs, they develop in a distinct manner and should not be considered to be equivalent in the strictest sense. These phenotypes are sequential developments from mesenchymal tissue-resident fibroblastic stromal progenitors and phenotypic changes are induced in response to the inflammatory chemokine millieu and stimulation from T and B lymphocytes, and are therefore captured within one state machine, despite the significant functional difference between the phenotypes. In addition to the four phenotypes represented as sub-states labelled S1–S4, there are several orthogonal states common to all stromal cells. These states are 'Adhesion Molecule (ICAM-1/VCAM-1) expression, 'TNFR-1 expression', 'IL13R Expression', and finally, if the cell is in the 'S2' state, adopts a state of S2 Stromal Stimulation Factor Receptor Expression, and upon differentiation into the 'S3' state, adopts an 'S3 Stromal Stimulation Factor Receptor Expression' state.

Stromal cells are assumed to be present throughout the ectopic tissue spatial environment in a network-like structure.The model aims to capture the development of distinct phenotypic niches within the network but is not concerned with remodelling or development of the network structure itself beyond T_0, thus the domain model makes the assumption of a static network despite it

being known that network density and structure does remodel in response to lymphocyte colonisation and subsequent lymphoid tissue development, as it is assumed for the sake of the domain model that the TLT self-organisation and development process can occur without stromal remodelling, which is likely to be related to permitting efficient immune responses but is not necessary to sustain the developmental process.

(a) Stromal Progenitor Cell (SPC) – *State 'S1'*

The stromal cell begins in the 'S1' state at T_0. In this state, no additional sub-states exist and the cell only exhibits behaviour manifested through the orthogonal states segregated by swim-lanes (dashed lines) at the bottom of the diagram. These cells differentiate into state 'S2', the localised stromal precursor cell, in response to inflammatory cytokines present due to local inflammation caused through autoimmunity. These cytokines are assumed to be present at T_0 as the model aims to capture the 15 day development process following establishment of chronic inflammation (Figure 3.3 – the cytokines responsible for S1–S2 differentiation are principally TNF-α and IL13).

(b) Localised Stromal Precursor (LSP) – *State 'S2'*

This composite state represents the localised stromal precursor phenotype. Immediately upon adoption of the 'S2' state, the cell adopts a substate of 'T cell chemokine expression'. This reflects that there are multiple chemokines to which T cells respond, although for the purposes of the model these are all subsumed into one abstract chemokine - references in the domain model to CCL19 therefore relate to this abstract notion, and the same is therefore true of CCR7 expression on T cells. An orthogonal substate also exists, the 'Adhesion Molecule Upregulation Substate'. T-cell contact that does not result in differentiation and transition to the FRC phenotype (state 'S3') accentuates the LSP phenotype by increasing adhesion molecule expression and inducing a T-cell chemokine

upregulation state. It is assumed that there is a maximum expression level for CCL19 achievable while the cell is in the 'S2' superstate, and a maximum overall expression level for adhesion molecules: once these are reached then T-cell contact that does not result in differentiation has no further effect on the LSP cell.

The conditions for differentiation to state 'S3' are T-cell contact bound via adhesion molecules (as determined by the T-cell UML finite state machine, and the expression of S2 Stromal Cell Stimulation Factor). A 'sufficient level' of T-cell contact is necessary for S2–S3 differentiation, however in the domain the specific factors expressed by T cells responsible for LSP-FRC differentiation are largely unknown, therefore it must be assumed that there is cumulative contact time from all T cells required for the adoption of the FRC phenotype. It is for this reason that stimulation which does not lead to differentiation accentuates the present phenotype – it results in a more gradual transition over time as the stroma is stimulated as opposed to sudden stepwise changes in expression levels of chemokines.

(c) Fibroblastic Reticular Cell (FRC) – *State 'S3'*

The FRC composite state, often referred to as 'T-cell Stroma' (**?**), provides the microanatomical niche for the T cell zone in secondary and tertiary lymphoid tissues. Once the cell has transitioned into the 'S3' state, it immediately adopts a state of 'B-cell chemokine Expression' and a state of 'T-cell Chemokine Upregulation', the latter increases the cell's maximum expression level of CCL19. Continued T-cell contact does not result in differentiation, but results in continued increases of T-cell and B-cell chemokine expression levels until a maximum threshold for the FRC phenotype is reached. This is expressed in the state machine by an arrow showing a transition from the state of 'T-cell/B-cell Chemokine Upregulation' back onto itself, with an activity of increasing chemokine

expression level demarked alongside the guard requiring T cell contact. If a B cell is in contact with the FRC stromal cell then the cell will either transition into state 'S4' (FDC) or result in an increase in B-cell chemokine expression until the maximum threshold level is reached. It is assumed that B-cell stimulation of an FRC-like cell is likely to preferentially induce increased B-cell chemokine (CXCL13) in order to further attract B cells in a positive feedback loop that increases the opportunity for the stromal cell to differentiate into the FDC phenotype, state 'S4'. This differential response of stroma to T and B lymphocytes is hypothesised to be sufficient to result in the development of B-cell follicles within an FDC network surrounded by a T-cell zone supported by a stromal network with an FDC phenotype.

(d) Follicular Dendritic Cell (FDC) – *State 'S4'*

The FDC composite state, often referred to as 'B-cell Stroma'. Immediately upon adoption of this phenotype, expression of CXCL13 is increased through lymphocyte stimulation and prolonged B-cell signalling will result in down-regulation of T-cell chemokines CCL19/21. The Markov model finite state machine is sufficient to describe changes in factor expression based on Stromal phenotype.

Figure 3.10: UML Finite State Machine Domain Model diagram describing the role of T lymphocytes in TLT formation.

B Cell | TLT Domain Model | State Machine Diagram

/ enter TLT after ~ day 5

j

[Local B cell chemoattractant level
sufficient for migration]

c

Responsive to local B
cell chemokine level

Random Movement

/ lose & replenish CXCR5
/ chemotaxis

/ random walk

j

c   ?else?

j

[B cell in contact with stromal cell
& Cell expresses adhesion molecules
& Bind between cells is sufficient
& Cell adhesion level is sufficient]

State of Adhesion

Prolonged contact with localised
movement around adhesion
molecule expressing cell

[Bind is insufficient | Cell adhesion level is insufficient]

Express S3 Stromal
Stimulation Factor

Express CXCR5
(B Cell Chemokine Receptor)

Express LFA-1
(Adhesion Molecule Receptor)

[exit tissue via blood vessels at periphery]

Figure 3.11: UML Finite State Machine Domain Model diagram describing the role of B lymphocytes in TLT formation.
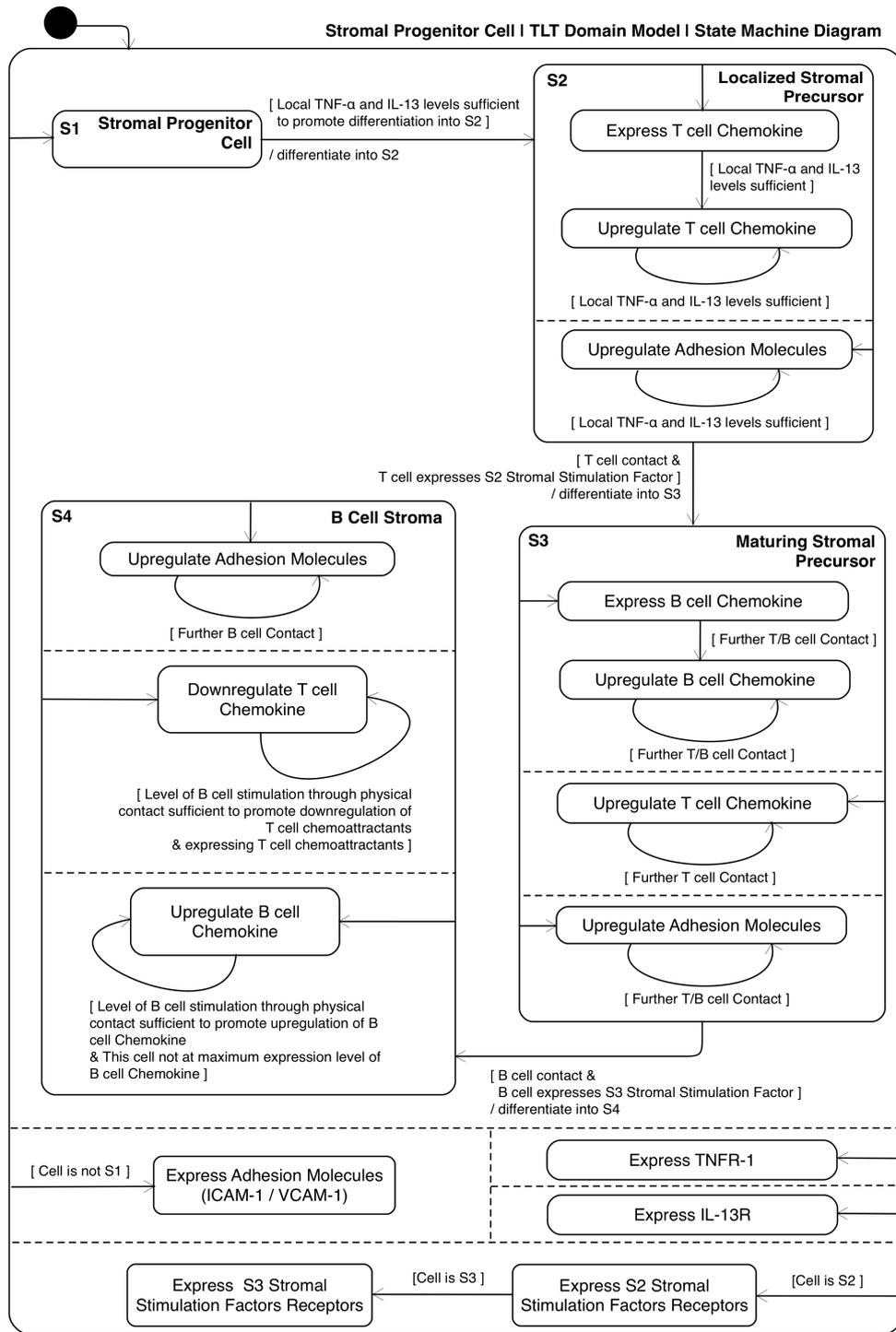
Figure 3.12: UML Finite State Machine Domain Model diagram describing the role of stromal precursor cells in TLT formation, including differentiation into FRC-like and FDC-like phenotypes.

3.2.3.2   Describing Lymphocyte–Stroma Crosstalk

The domain model UML 'Activity Diagram' shown in Figure 3.14, in effect, integrates the state machines for each model entity, and describes the interactions expected to occur and the sequence they are expected to occur in over time. Also contained within the activity diagram are the overall effect of each interaction (for instance, the up- or down- regulation of a particular molecule on a cell surface). Considering the massively parallel nature of biological systems, many entities will be undergoing interactions at each point of the diagram, and it is this nature that allows *emergence* of high-level behaviours from low-level basic molecular interactions.

3.3   Summary

In this chapter a novel domain model of TLT formation has been formulated following the CoSMoS process (**?**), and an overview of the biological domain was presented, providing a rationale for the underlying biology within the model. The TLT formation model contains many layers of complexity, describing phenomena occurring over several spatiotemporal scales. In order to develop a tractable simulation that can capture the inherent complexity, it is necessary to hybridise a variety of modelling techniques to capture the different biological processes responsible for determining model outcomes. Therefore, in Chapter 4, a methodology is presented for the simul-

Expression	Meaning
X → Y	Differentiation of X into Y
⇧ Z	Express Z
⇧⇧ A	Upregulate A
⇩ B	Downregulate B
C ⧓ D	C and D are in contact
E ↯ F	E undergoes chemotaxis due to F
G ⇨ H	G migrates to H
I ↭ J	I adheres to J due to adhesion molecules

Figure 3.13: Key describing the ASCII symbols used in the UML Activity Diagram of Figure 3.14 as notation for common immunological interactions between entities to simplify model communication.

Activity Diagram I Domain Model I Tertiary Lymphoid Tissue Induction
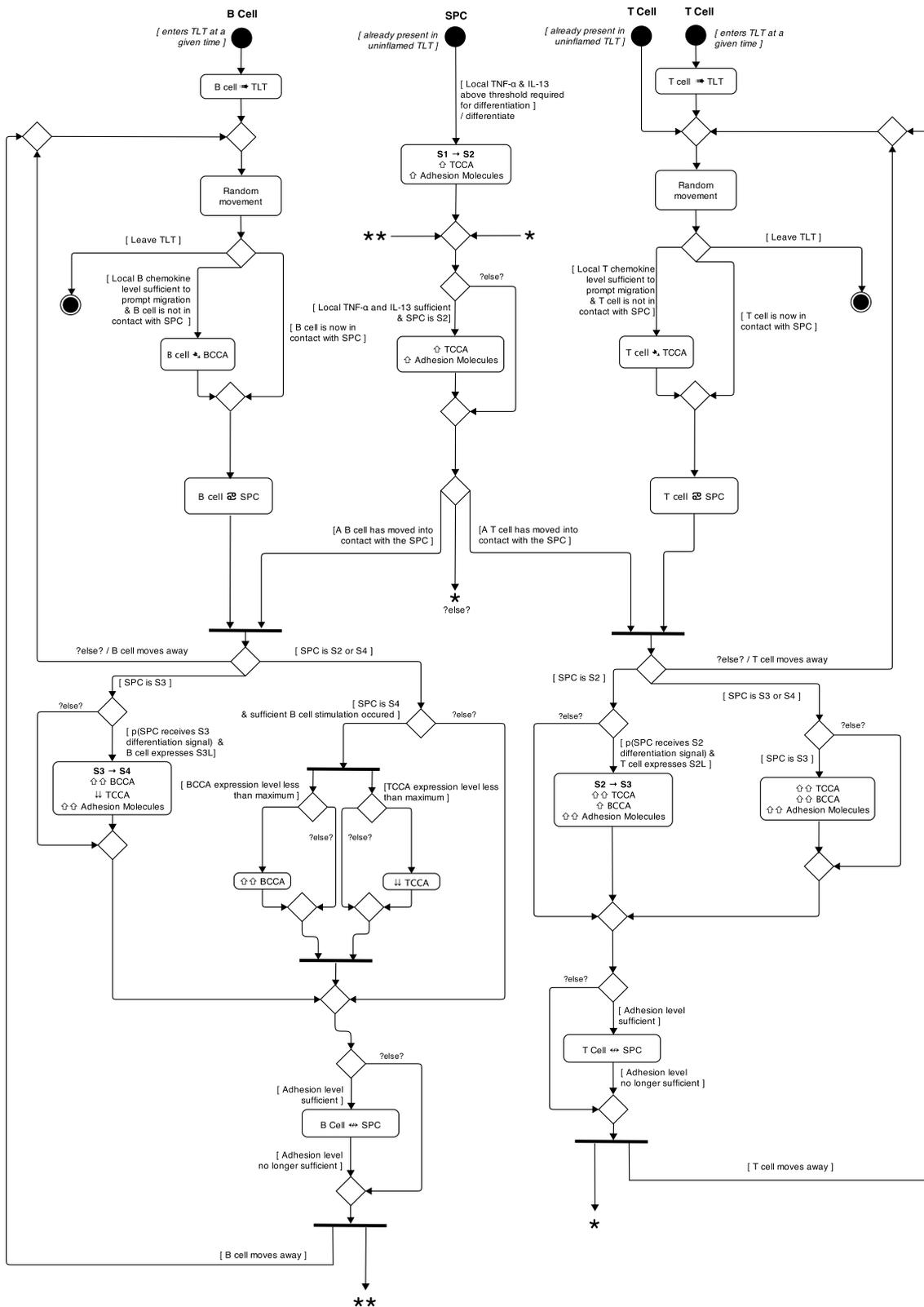
Figure 3.14: UML Activity Diagram describing interactions between stroma, B and T lymphocytes that result in the self-organisation of TLT.

taneous hybridisation of multiple mathematical and computational techniques into a single executable simulation, this is subsequently used to develop a hybrid model of TLT formation that utilises the most appropriate technique for each model entity.

The following chapter describes the development of a platform model and executable simulation derived from the biological domain model described in this chapter. In order to describe each model entity using the most appropriate degree of granularity, it was necessary to develop a framework and notation to describe hybrid models composed of many interconnected 'sub-models'. Following development of the platform model, the veracity of the hypothesis that TLT formation is driven by a differential response of B and T lymphocytes to stroma may be confirmed, candidate intervention strategies can be subsequently evaluated, incorporating them into the simulation to determine their potential therapeutic effect with various dosing regimes.

The MDE paradigm discussed in Section 3.2 is applied in Chapter 6, in which *in vivo* experiments are performed to test predictions of therapeutic efficacy derived from the TLT formation model using machine learning techniques. Specifically, the prediction that the addition of anti-VLA4 post TLT induction will result in improved pathology resolution, that anti-TNF improves this effect if given sufficiently early, but that anti-TNF is ineffective in established disease. The *in silico* experiments concerning biologic therapy interventions are described in Chapter 6. Although the adoption of MDE has only recently started to have an impact on immunology research, it is starting to have a very significant impact on other areas of biology. We propose that the increased accessibility of computational models, high-performance computing resources, the increased familiarity and understanding of simulations as tools to understand immune function and the capacity to apply in silico approaches to identify potential therapeutic approaches and disease biomarkers will accelerate the application of MDE as a methodology understand and target disease resolution. The MDE paradigm as described in Figure 3.1 is broadly applicable to any disease pathology providing cell culture or animal model experimental data are available in

addition to the development of complementary *in silico* models.

Chapter 4

Developing and Applying Hybrid Modelling Techniques to Construct a Simulation Platform of TLT Formation

The key biological processes believed to drive TLT formation were described in the previous chapter, in which a novel model of stromal cell differentiation in response to lymphocyte-stroma crosstalk based on in vivo and in vitro experimentation was presented. In order to develop a functional simulation of TLT formation, the biological model must be 'de-constructed' into its constituent entities: each represented at the required granularity to answer any model research questions, also taking into consideration the nature and granularity of experimental data that are available. When dealing with many separate 'sub-model' entities grounded in various mathematical methodologies, it was necessary to develop a means of describing the model in terms of information flow between each sub-model in order to communicate the model coherently. A general framework for defining and specifying hybridised, multiscale models of complex biological processes is developed.

Firstly, a generalised schema for defining hybrid models is introduced, and applied

to the TLT formation platform model. Each biological entity and its mathematical implementation in terms of developing a software simulation is subsequently described. The hybridisation framework developed is centred upon information flow between sub-models, and feedback loops are observed to emerge between the sub-models themselves. The diagrammatic methods developed for describing integrated hybrid models and simulations clearly and efficiently communicate the platform model architecture. Additionally, the safety-critical systems argumentation method Goal Structuring Notation is applied as the experimental, domain, platform and implementation level, ensuring that the model remains demonstrably fit-for-purpose. The visualisation methods developed in Chapter 2 were applied to the simulation to permit high-throughput histological image analysis to operate on simulated data, utilised for in silico experimentation regarding adhesion and TLT morphology in Chapter 5.

The following sections describe the development of the TLT Platform Model, derived from the domain model detailed in Chapter 3. For a more concise description of the model, please see Appendix A.

4.1 Determining Spatial Dimensionality

In the domain under study, TLT organs are 3D structures that have developed ectopically in tissues in a state of chronic infection; in Sjögren's syndrome, these tissues are principally the salivary glands and tear ducts. The platform model is required to describe the minimum requirements to sustain lymphoid neogenesis within the salivary gland. Each spatial dimension within a system adds an orthogonal axis within which objects within that space may move, enormously increasing an object's degrees-of-freedom - this is often described as the 'curse of dimensionality'. Therefore the computational expense of simulations in three dimensions is significantly higher than in two dimensions. This alone is not sufficient justification for using a 2D model system; it must be demonstrated that a 2D representation is adequate with respect to the domain under study and the questions that are being asked of the model. By abstraction of the complex cellular biophysics involved in cellu-
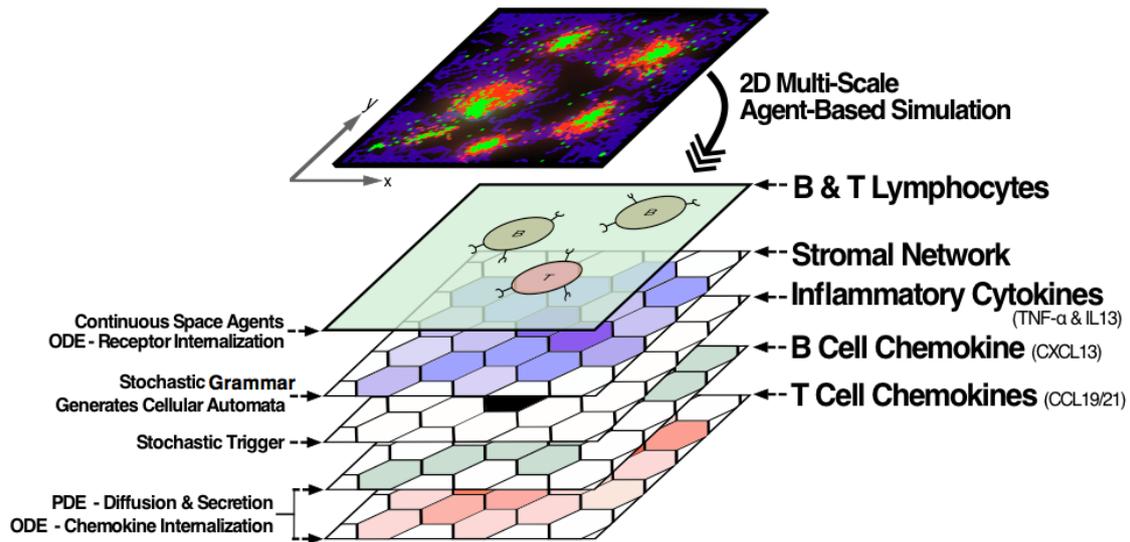
Figure 4.1: Each model entity has been assigned a putative implementation method. This figure serves to introduce each aspect of the overarching model in terms of multiple 'sub-models', and can also serves a introductory process as the Expected Behaviours diagram does when first introducing a Domain Model description. The figure illustrates how each model entity has its own notion of space that map to each other to produce the resulting 2D/Quasi-3D simulation.)

lar movement, contact and collision within tissues, permitting lymphocyte agents to occupy the same spatial locations simultaneously, one can achieve what might be termed a '*quasi-3D*' environment – a truly 3D environment would not permit models of multiple volume-occupying objects to occupy the same spatial locations as this would negate the Pauli exclusion principle (**?**). Due to the flexibility and compressibility of cellular membranes, this is an appropriate abstraction given that our principle concern is self-organisation through lymphocytic stimulation of stroma and resulting differential chemokine expression of stromal cells.

The placing of lymphocytes in a continuous space two dimensional grid which maps to the identically sized CDS (Cellular Data Structure) that stores the stromal network structure and details of its composition in terms of individual cells, generated through a cellular automaton implementation of a generative grammar (detailed further in Section 4.7) lymphocytes will be able to determine contact with co-localised stromal cells and respond to resulting chemokine gradients. Therefore, our platform model and its resulting simulator utilise a '*quasi-3D*' dimensional space

evolving over time, in which the spatial data structures of each sub-model all map homomorphically to each other. This mapping is illustrated in Figure 4.1, which shows a decomposition of the 2D TLT simulation visualisation at the top of the figure into each individual sub-model containing a 2D spatial representation that all map, or 'overlay', to the same spatial coordinates to permit interaction between model entities from different sub-models.

4.2   Developing a Model Hybridisation Framework and Schema

When attempting to describe a mathematical model constructed from multiple 'sub-models', each representing a specific biological entity or process, it can quickly be difficult to understand the model topology and there currently lacks notation to succinctly describe the model structure and organisation. Therefore it is considered that the most informative representation should be centred on information flow between sub-models. Each input/output between the various sub-models can be defined using arrows between sub-models defined within boxes. An example of such a diagram is presented in Figure 4.2.

This example hybrid model diagram includes a 'meta-feedback loop' between sub-models 1 to 4 as information passes between the 4 sub-models. This semi-formal diagrammatic approach to defining complex hybrid models can be easily extended or modified, for example to take multiple compartments into consideration; in this case, one may wish to either represent each compartment of a multi-compartment model in a separate spatial area, with sub-models involved shown within the appropriate area — alternatively, the sub-models could include the compartment(s) to which they operate on and what information is transmitted to the compartments by the sub-models.

Model Hybridisation Schema:
Information Flow Between Component Sub-models
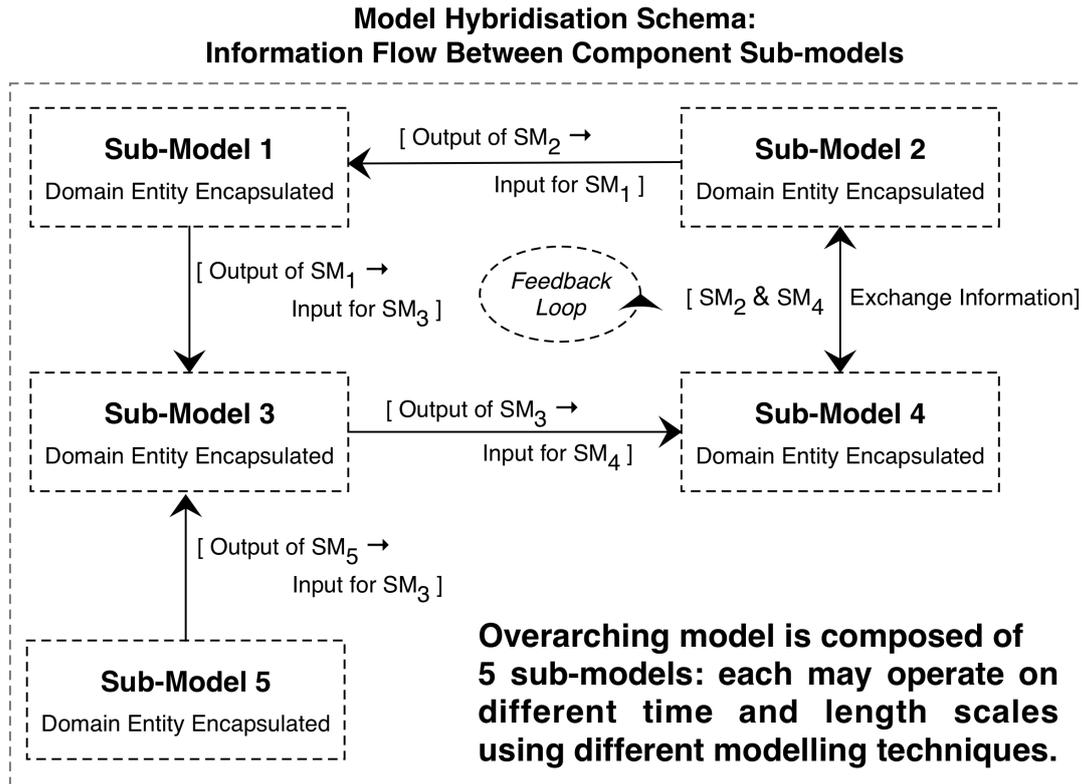
Figure 4.2: Generalised schema for model hybridisation, a useful tool for visualising the relationship between each sub-model and the domain entity that they are intended to represent. Arrows indicate the general direction of information flow within the model. Sub-models may influence other sub-models based on parameters, or receive inputs from other sub-models. Sub-models may also reciprocally influence each other, indicated using a double arrowed line.

4.3   Agent-Based Modelling of Lymphocytes

Lymphocytes need to move freely in a spatial environment such that they can chemotactically migrate due to spatially resolved chemokine gradients, and interact with stroma to stimulate phenotypic and 'state' changes to initiate and maintain the positive feedback loop of stromal cell differentiation, chemokine secretion, and emergent lymphocyte self-organisation. The flexibility afforded by agent-based models in describing objects that can interact with their environment in innumerate ways, such as determining and modulating local chemokine gradients in the presence of other cells also doing the same, makes this approach well-suited to provide a 'bedrock' for the overall model.
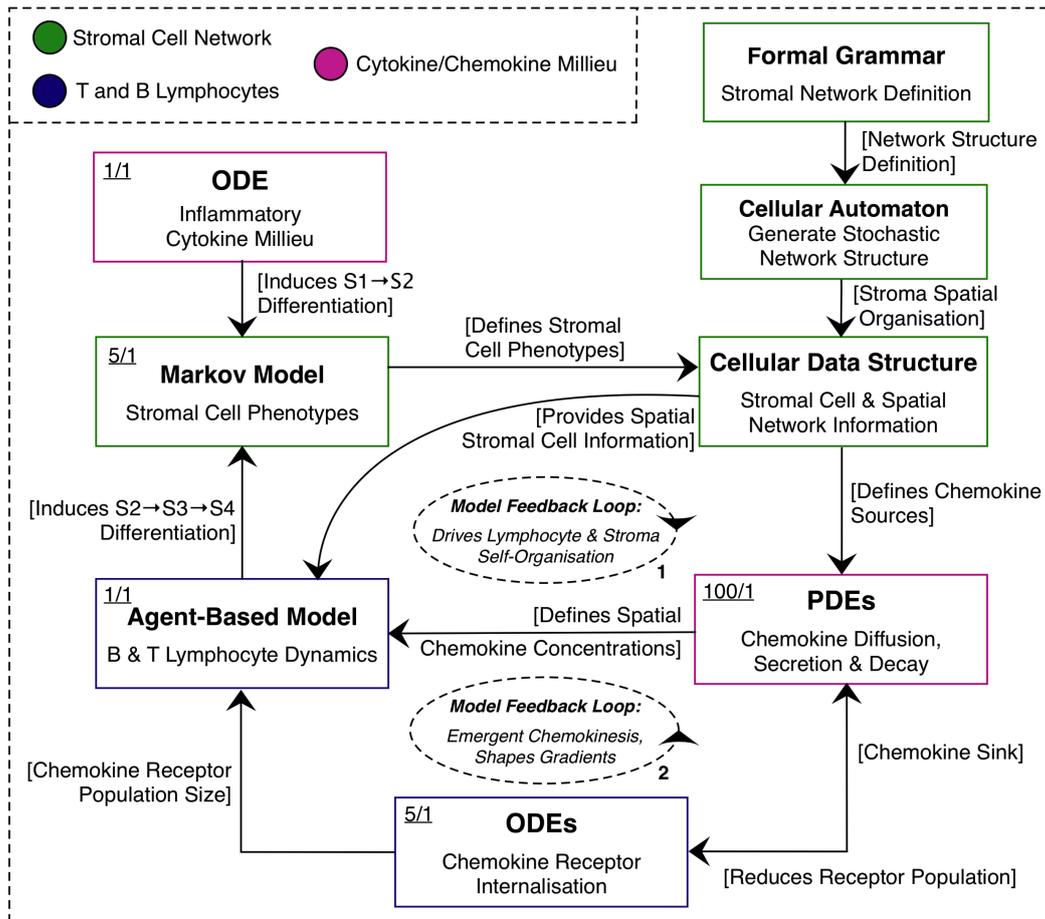
Figure 4.3: Simplified hybridisation scheme demonstrating how each of the seven sub-models that comprise the platform model share information to represent TLT development as a whole. Underlined numbers indicate the number of time-steps that occur within that sub-model with respect to other sub-models, relative to the agent-based model, labelled '1/1'. For example, the PDE iterates 100 times while the remainder of the system is held quasi-static, and as such is labelled '100/1'.

4.3.1 Capturing Cellular Dynamics

This section details how the dynamics for B and T lymphocytes, whose role in stimulating and regulating stromal cell phenotype was detailed in the domain model (Section 3.2.1), can be defined in a manner suitable for implementation in a simulation. This requires a combination of UML, to describe lymphocyte states and conditions required for change in states, and mathematics, to describe how the chemotactic response and adhesion molecules are implemented within the ABM. Internalisation, recycling and decomposition of chemokine receptor-ligand complexes by lymphocytes is captured by an ODE model described in Section 4.5.

Although B and T lymphocytes are distinct phenotypes, of which there are many subsets, they do have much in common – particularly with respect to their role in lymphocyte-stroma crosstalk. For the sake of simplicity, it is assumed that B cells respond only to lymphoid chemokine CXCL13, through expression of CXCR5, and not to CCL19, that is, that B cells do not express CCR7, although it is known not to be the case, and B cells can chemotactically respond to CCL19 gradients (**?**). The research question is concerned with the *minimum requirements* for TLT formation; the hypothesis states that capturing the differential response of inflamed stroma to T and B cells, and the resulting self-organisation of TLT is the result of a differential response of T and B cells to CCL19 and CXCL13 respectively. Therefore, the structure of the finite state machines that describe T and B cells are very similar, as they are in the domain model. The difference between B and T cells in the platform model solely lie in the chemokines they respond to, and the effect that they have upon the stromal cells they stimulate.

4.3.1.1 UML Finite-state Machines describing Lymphocytes

T cells begin either in a tissue-resident state, or begin to enter the simulation at a linear rate, while B cells do not enter the tissue until day 5 of the formation process, each ABM time-step represents 10 minutes of time, and therefore they begin to enter after the 720th simulation time-step at a linear rate. The process

determining the number of lymphocytes entering the simulation space at each time

space is detailed in Section 4.3.1.2. T and B cells both then determine whether to

follow a random walk or to migrate chemotactically against their respective lymphoid

chemokine CCL19 or CXCL13. If the cell's chemokine receptor expression is greater

than a minimum parameter value, the '$minimumReceptorThreshold$' and the local

chemokine gradient in at least one of the six adjacent grid spaces is greater than the

parameter value '$minimumChemokineThreshold$', then the cell enters a chemotactic

state and uses the mathematical process described in Section 4.3.1.3 to determine

the direction of motion to follow, which involves making 10 'sub-movements' in

which an angle is chosen and the cell moves 1/10th of the total distance per time-

step. Alternatively, the cell will randomly select a direction to move, also making 10

smaller sub-movements within the ABM time-step. All other aspects of the model

are held to be quasi-static while these 10 sub-movements take place. The process

of making several smaller movements within one time-step produces a walk with a

closer approximation to continuous motion; as the number of sub-movements taken

approaches infinity, lymphocyte motion becomes continuous. This 'sub-movement'

process was used in **?** for LTi cell motion in a model of pre-natal lymphoid organ

development. For a full mathematical formulation of the chemotaxis model, please

see Section A.5.

Once a T or B cell has determined whether it is chemotactic, it may be in

contact with stroma and enter a state of adhesion via stromal adhesion molecules.

To determine if this occurs, the continuous space grid in which T and B cells move

is discretised with an element width of 10 microns, matching the discrete stromal

network grid (CDS, described previously in Section 4.7.2). If the discrete location of

the lymphocyte is occupied by a stromal cell in the CDS, then they are taken to be in

contact. Further requirements are that the expression of stromal adhesion molecules,

as a value between 0 and 1, is greater than a random number also between 0 and 1.

This equates to the adhesion molecule expression level determined by a stromal cell's
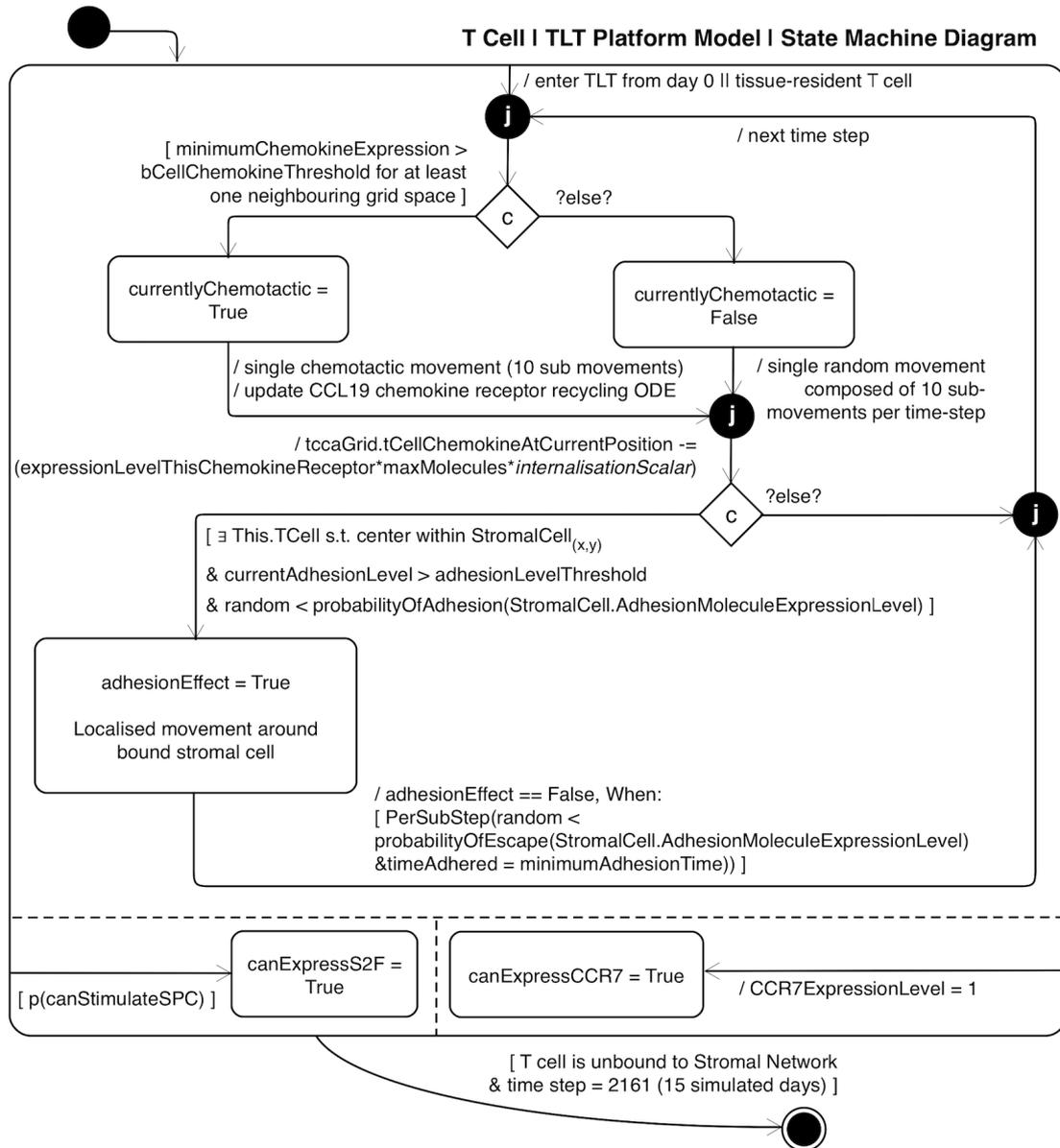
Figure 4.4: UML Finite State Machine Platform Model diagram describing the role of T lymphocytes in TLT formation.

Markov chain emissions being equivalent to a probability that any given lymphocyte will successfully adhere to it. Finally, in order to prevent permanent adhesion to stroma and a loss of motility, there is a maximum time that any one lymphocyte can adhere to a stromal cell in one instance, this is reflected by the guard '*timeAdhered > minimumAdhesionTime*' in Figures 4.4 and 4.5. This is equal to the number of 'sub-movements' that a lymphocyte makes in each time-step, and is incremented one per attempted sub-movement — the function defined as '*perSubStep()*' indicates
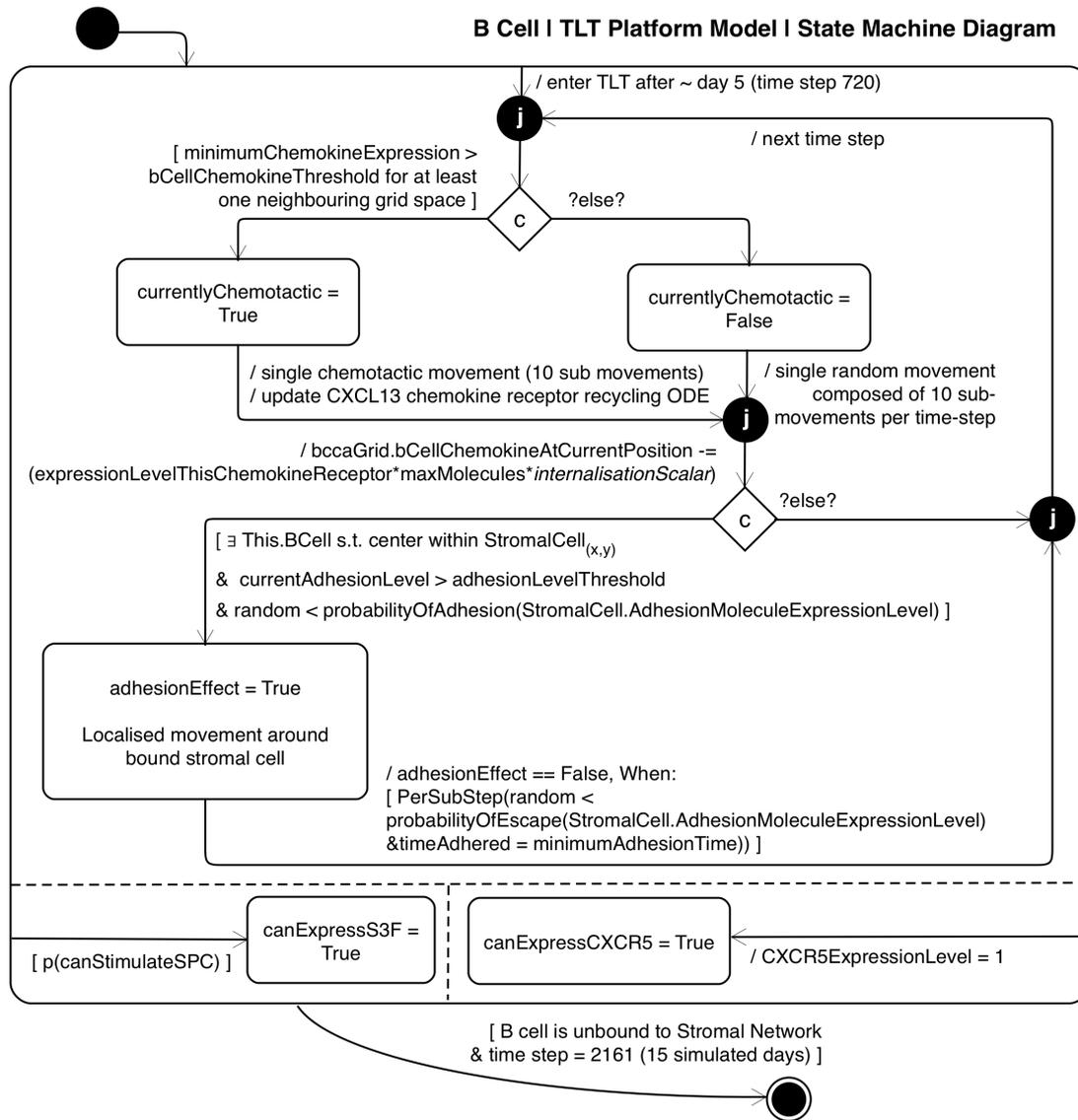
Figure 4.5: UML Finite State Machine Platform Model diagram describing the role of B lymphocytes in TLT formation.

that the probabilistic guard is re-evaluated for each sub-movement made within an ABM time-step, such that adhered lymphocytes will move at between 0–100% of their usual velocity, in increments of 10%. This represents a departure from the model of adhesion molecules used in **?**, in which a state of adhesion resulted in a lower velocity for each sub-movement, rather than evaluation of a binary decision of full-velocity motion or no movement at each interval within the time-step.

A further departure, from both **?** and the TLT domain model in Chapter 3 is the abstraction of adhesion molecule receptor expression in lymphocytes - this

is replaced with the probability of adhesion being determined only by the stromal cell. Independent probabilistic guards on both the stromal cells and lymphocytes is redundant, and executing two probabilistic guards would serve no useful purpose. Therefore, there is no differential expression of adhesion molecules across the lymphocyte population; there is, however, over the stromal cell population. The abstraction of adhesion molecules from lymphocytes represents an implicit assumption that all lymphocytes express adhesion molecule receptors (e.g. VLA-4, LFA-1), and as the lymphocytes in the platform model are not specialised to perform any particular effector function, there is no mechanism through which a differential response might be established – a normal distribution of expression could be included, but it is not clear what benefit this added complexity would bring to the model in light of the research questions.

4.3.1.2 Defining Lymphocyte Tissue Entry Rate

Lymphocyte tissue entry rate is defined as a linear rate of population increase depending on the total lymphocyte population size and the number of lymphocytes to be added to the simulation space during a specified interval. These assumptions are in the absence of data to the contrary, however, it should be noted that the model does not attempt to capture emergent recruitment to the ectopic region of interest, that is, cell exit/entry rates are predefined and assumed to be linear fixed effects.

4.3.1.3 Determining Motility by Modelling the Chemotactic Response

This section describes the process that B and T lymphocytes undergo when in a chemotactic state to determine their direction of motion, and provides a mathematical definition. Note that this section neglects the sequestration of chemokine and loss of chemokine receptors through the internalisation and degradation of chemokine-ligand complexes, as this process is managed by an ODE-based model detailed in Section 4.5.

Lymphocytes measure the chemokine concentration in the six hexagonal grid

elements neighbouring the one in which the lymphocyte exists when its continuous space location is discretised such that the lymphocyte exists within an element that maps directly to the hexagonal grids that contain chemokine concentrations and the stromal network defined within the CDS. The chemokine grid is a 2D matrix of double precision values that lie between 0 and 1, in which '1' represents a saturated 10μ M^2 region of space, and '0' indicates no chemokine is present. Weighted by a scalar multiplication upon the hexagonal element with the highest concentration, the probability of moving in the direction of each of the six neighbouring grid elements is constructed by dividing the concentration of chemokine in each of the neighbours by the total chemokine concentration in all six neighbouring grid elements. From this, a cumulative probability density function is constructed which the lymphocyte randomly samples to determine the neighbouring grid element in whose direction it shall move. From the centre of the hexagonal grid element that the discrete-space lymphocyte maps to, there is a range of $\frac{\pi}{3}$ radians (60°) between the vertices of the shared edge of the neighbouring hexagonal grid element selected by sampling the cumulative distribution function and the element which the lymphocyte spatially maps to. A random angle within this range is selected and the lymphocyte's position co-ordinates are updated on the respective ABM continuous space grid for B or T lymphocytes, placing the lymphocyte at a distance of 1/10th of the intrinsic cell velocity multiplied by the length of the time-step. This is reflective of the 10 'sub-movements' a lymphocyte makes within one ABM time-step in an effort to increase isotropy in cellular motion captured by the TLT formation model, therefore, this process of gradient measuring and direction determination is repeated a maximum of ten times. The number of sub-movements made is dependent upon whether the cell adheres to a stromal cell following each sub-movement, in which case no movement is made – adherence is determined by a probabilistic guard in which the probability depends on the level of adhesion molecules expressed by the stromal cell that maps to the same location of the discrete-location lymphocyte, this process is detailed in Section 4.3.1.1.

This process can be mathematically formulated as follows. When a T or B cell is in the state *'currentlyChemotactic'* as described in Figures 4.4 and 4.5, the random motion is biased towards the direction of the gradient of CCL19 (T Cell chemokine) and CXCL13 (B cell chemokine), respectively. This is achieved by constructing a probability distribution from the chemokine gradient, such that the probability a cell will move in one of the six directions of the underlying hexagonal grid that contains chemokine concentration data is defined in equation 4.1.

$$p(\text{Cell moves up}) = c_{\text{up}}/C$$

$$p(\text{Cell moves down}) = c_{\text{down}}/C$$

$$p(\text{Cell moves upper left}) = c_{\text{upperleft}}/C$$

$$p(\text{Cell moves upper right}) = c_{\text{upperright}}/C \tag{4.1}$$

$$p(\text{Cell moves lower left}) = c_{\text{lowerleft}}/C$$

$$p(\text{Cell moves lower right}) = c_{\text{lowerright}}/C$$

Where c_{up} is the concentration of chemokine in the hexagon above the hexagonal element the cell currently lies in, c_{down} is the concentration of the hexagonal element below, and so on. As lymphocytes exist on their own continuous space grid overlying the discrete hexagonal grid containing chemokine concentration data, once the general direction based on the chemokine concentration has been defined, a specific angle of motion needs to be determined. Each of the six directions contains a range 60 degrees (360/6), thus a random angle is selected between 0 and 60 degrees as described in Figure 4.6.
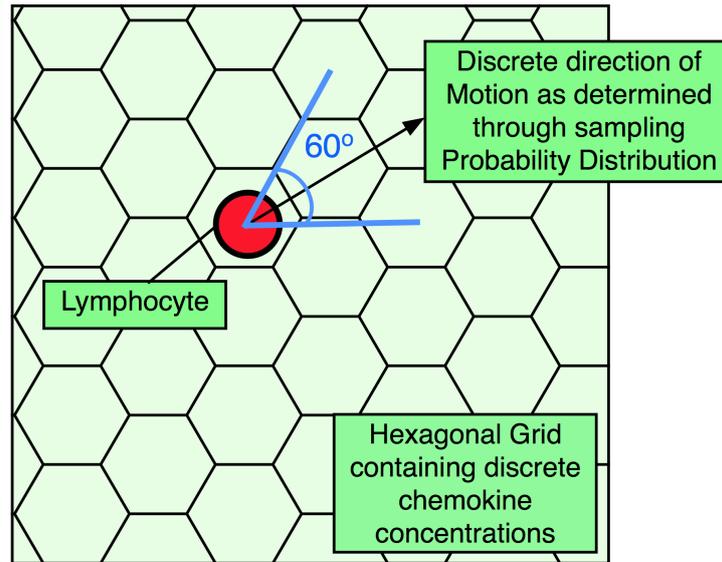
Figure 4.6: Figure illustrating the selection of direction for continuous Lymphocytes following determination of the discrete direction to move based on the hexagonal grid that contains chemokine concentrations. The cell selects at random an angle in the 60 degree range of possible directions once a particular hexagon has been selected through probabilistic sampling according to equation 4.1

4.4 Numerical Modelling of Chemokine Secretion, Diffusion and Decay

Stromal cells stored in the CDS map to a discrete hexagonally-packed square grid. This enables them to produce chemokines which can diffuse in a more isotropic manner as each element has 6 boundaries rather than 4, and also increases the number of degrees of freedom for network growth, permitting stromal networks that are intuitively a better representation than simple squares, retaining connectivity through a grid edge at all times. The stromal network produces the lymphoid chemokines CCL19 and CXCL13, to which T and B cells are chemotactically attracted, respectively. The location of stromal cells within the CDS can be used as a set of point sources for the secretion of chemokine, and therefore a natural starting point for the implementation of a chemokine diffusion model is the 'diffusion equation', a partial differential equation first described by ? that describes the collective motion of many small particles in a material as a result of their random motion. The diffusion

equation is discretised for implementation in a hexagonally-packed 2D square grid and extended to incorporate diffusion from stromal cells and decay of chemokine molecules over time. The hexagonally-packed grid permits chemokine diffusion to overlay with the stromal network, and each stromal cell element to have an associated chemokine concentration, along with unoccupied space between the network. We begin with the 2D heat equation as described by **?**:

$$\frac{\partial \phi_{(x,y,t)}}{\partial t} = D \left(\frac{\partial^2 \phi_{(x,y,t)}}{\partial x^2} + \frac{\partial^2 \phi_{(x,y,t)}}{\partial y^2} \right) \tag{4.2}$$

Where $\phi_{(x,y,t)}$ is the concentration at location (x, y) and at time, t, and D is the homogeneous diffusion co-efficient. Equation 4.2 is then discretised using the two-way finite differences method (described in **?**) as follows (let ΔL be the length of the edge of each square in the discrete grid):

$$
\begin{aligned}
\frac{\phi_{(x,y,t)} - \phi_{(x,y,t-\Delta t)}}{\Delta t} &\approx \frac{\partial \left(\frac{\phi_{(x,y,t)} - \phi_{(x-\Delta x,y,t)}}{\Delta x} \right)}{\partial x} + \frac{\partial \left(\frac{\phi_{(x,y,t)} - \phi_{(x,y-\Delta y,t)}}{\Delta y} \right)}{\partial y} \\
&\approx \frac{\frac{\phi_{(x+\Delta x,y,t)} - \phi_{(x,y,t)}}{\Delta x} - \frac{\phi_{(x,y,t)} - \phi_{(x-\Delta x,y,t)}}{\Delta x}}{\Delta x} + \frac{\frac{\phi_{(x,y+\Delta y,t)} - \phi_{(x,y,t)}}{\Delta y} - \frac{\phi_{(x,y,t)} - \phi_{(x,y-\Delta y,t)}}{\Delta y}}{\Delta y} \\
&= \frac{\phi_{(x+\Delta x,y,t)} + \phi_{(x-\Delta x,y,t)} + \phi_{(x,y+\Delta y,t)} + \phi_{(x,y-\Delta y,t)} - 4\phi_{(x,y,t)}}{\Delta L^2}
\end{aligned}
\tag{4.3}
$$

Equation 4.3 may be more conveniently expressed in terms of the concentration at (x, y) at the next discrete time-step in terms of the present time-step, as follows:

$$\phi_{x,y}(t+1) = \left(\phi_{x,y}(t) - \frac{4D\Delta t \cdot \phi_{x,y}(t)}{\Delta x^2} \right) + \frac{D\Delta t}{\Delta x^2} \sum_{i=1}^{i=4} \phi_i(t) \tag{4.4}$$

Where $\phi_i(t)$ is the value of each of the 4 adjacent edges in the discrete grid at location (x, y) from $\phi_{x,y}(t)$ (i.e. the Von Neumann neighbourhood of $\phi_{x,y}$). Equation 4.4 provides a suitable diffusion equation on a 2D square grid that allows one to iterate over a discrete grid of chemokine values to model their diffusion, and a starting point for our chemokine diffusion model. Some extension to this is necessary to make it suitable for use in the TLT model, including modification to suit

the hexagonally-packed geometry used by the stromal network model, and the incorporation of chemokine secretion by many point sources, and decay over time.

To extend Equation 3 to utilise a hexagonally-packed square grid is trivial, the Von Neumann neighbourhood that is summed should be replaced with a neighbour of the 6 adjacent grid elements in a hexagonally-packed grid, and the negation of chemokine in the expression $4D\Delta t$ needs to be replaced with $\mathbf{6}D\Delta t$ to incorporate the additional edges. This results in equation 4.5 below:

$$\phi_{x,y}(t+1) = (\phi_{x,y}(t) - \frac{6D\Delta t \cdot \phi_{x,y}(t)}{\Delta x^2}) + \frac{D\Delta t}{\Delta x^2} \sum_{i=1}^{i=6} \phi_i(t) \qquad (4.5)$$

Finally this is extended by incorporating all chemokine point sources, and integrating the resulting discretised expression into the hybrid simulation. The final form of the discretised chemokine diffusion equation is given in equation 4.6 A specific point may be a chemokine source if it is co-located with a stromal cell stored in the CDS. If a stromal cell exists in that location, its expression values can be queried from the CDS to determine appropriate diffusion rates. In terms of the software implementation and integration with the ABM, for each hexagonal grid, one checks whether the location is equal to an object of type StromalCell, and if so, extracts the expression value and calculates the resulting diffusion over the previous time-step. In equation 4.6, $\phi_{x,y}(t+1)$ is the chemokine concentration at point (x, y) at the following time-step from $\phi_{x,y}(t)$. Δt is the difference in time between t and $t+1$. D is the diffusion co-efficient, and λ is the decay constant.

$$\phi_{x,y}(t+1) = (((\phi_{x,y}(t) - \frac{6D\Delta t \cdot \phi_{x,y}(t)}{\Delta x^2}) + \frac{D\Delta t}{\Delta x^2} \sum_{i=1}^{i=6} \phi_i(t)) - \phi_i(t)e^{-\lambda\Delta t}) + s \quad (4.6)$$

4.5   Ordinary Differential Equation Model of Chemokine Receptors on Lymphocyte Agents

When chemokine binds to a cognate receptor, the complex may be internalised by the cell in a process of recycling that aims to maintain expression of the receptor during chemotaxis, enabling the cell to continue to effectively identify the chemokine gradient. This process has previously been explored in a simple mathematical model of dark zone-light zone cycling by chemotactic GC B cells (**?**). This ODE has been adapted for use on an individual agent level, discretised such that each B cell updates its receptor expression level at each time step, based on a function of recycling, degradation and synthesis, as illustrated in Fig. 4.7.

Modelling the internalisation and recycling of chemokine receptors CXCR5 (on B cells) and CCR7 (on T cells) allows lymphocytes to undergo desensitisation in response to locally high chemokine concentrations, as the loss of ligand surface expression due to an increased rate of internalisation will reduce lymphocyte sensitivity to chemokines and therefore reduce chemotactic capability. However, once lymphocytes migrate to an area with a lower local chemokine concentration, intracellular receptors will be recycled to the cell surface at a greater rate than they are being internalised, resulting in re-sensitisation and therefore increased chemotactic activity. We can write an ODE expressing the rate of change of receptor expression with respect to time as follows:

$$\frac{dr}{dt} = S - \psi\pi + \sigma, \tag{4.7}$$

where ψ is the number of receptors that have been internalised and degraded, σ represents the quantity recycled - a given proportion of ψ (determined via the parameter 'receptorInternalizationScalarT' and 'receptorInternalizationScalarB' for T and B cells respectively), and S represents the addition of chemokine receptor to the intracellular pool due to synthesis. In equation 4.7, π is a function of the local
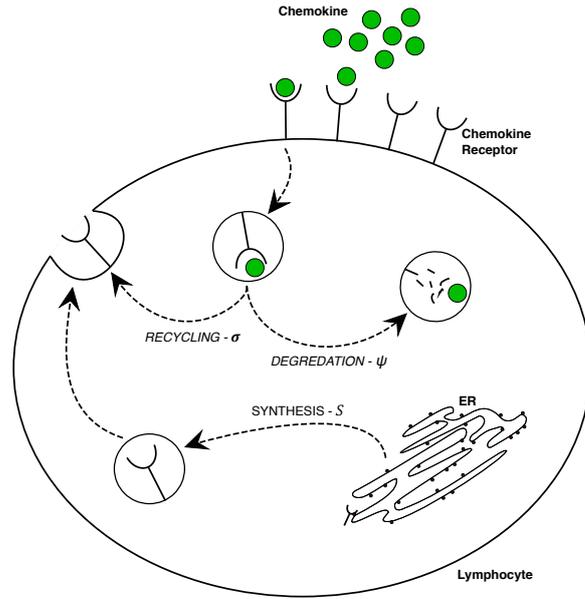
Figure 4.7: Figure illustrating the process of chemokine receptor-ligand internalisation. When a complex is internalised, it is either recycled or degraded. In the platform model, each cell has an intracellular pool of chemokine receptors derived from what has been synthesised or recycled, which is placed back onto the cell surface at a given rate. Internalisation of chemokine also occurs at a given rate defined as a parameter.

chemokine concentration, current receptor expression level, and a scaling variable; it is defined as:

$$\pi = \frac{gnqr(t)}{Q},$$
(4.8)

where q is the local chemokine level, Q is the maximum possible chemokine level, $r(t)$ is the current number of chemokine receptors on the cell surface, n is a scaling parameter to modulate the rate of internalisation by calibration, , and g is a random number in the range [0,1] sampled from a Gaussian distribution using the Mersenne twister (**?**) psuedo-random number generator to incorporate heterogeneity. Combining and discretisation of equations 4.7 and 4.8 yields equation 4.9, to be incorporated into the model for each lymphocyte agent:

$$r(t+1) = \left(r(t) - \psi\frac{gnqr(t)}{Q}\right) + \sigma + S.$$
(4.9)

While equation 4.9 describes the change in chemokine receptor expression levels

on lymphocytes, there must be a corresponding reduction of chemokine availability. The model must account for the internalisation of chemokine-ligand complexes through provision of a means of removing chemokine from the environment to reflect the quantity that has been absorbed and broken down by lymphocytes (**?**).

By operating on the same hexagonally-packed square grid that stores chemokine concentration data, this process effectively couples the internalisation ODE with the diffusion, secretion and decay PDE. This is achieved by determining the total number of chemokine molecules that have been internalised, I, shown in Equation 4.10:

$$I = \psi \frac{nqr(t)}{Q} + \sigma \frac{nqr(t)}{Q} \tag{4.10}$$

Following this, we remove the same quantity from the current chemokine level at the location of the lymphocyte in the hexagonally-packed grid, as shown in Equation 4.11, below. Let c be the change in chemokine level due to diffusion and stromal cell secretion, defined by the PDE in Section 4.4:

$$q(t + 1, x, y) = q(t, x, y) + c - I \tag{4.11}$$

4.6 Integrating the Agent-Based Model with the ODE and PDE Models

The discretised chemokine secretion and diffusion PDE model applied to a hexagonal grid is maps to that used by the stromal network in terms of spatial dimensions and grid size – this ensures chemokine secretion can be easily incorporated into the PDE by adding chemokine to a stromal cell's cognate spatial location on the two chemokine grids (B-cell and T-cell chemokines: CXCL13 and CCL19). The PDE discretisation is a finite-difference approximation of the diffusion equation with additional decay and secretion terms. This permits agents to interact with the grid to determine concentrations, gradients and therefore motion vectors, and also with the agent chemokine receptor ODE to consume chemokine. A similar discretisa-

tion scheme is applied in use for the ODE system to describe cell surface protein expression levels for the chemokine receptors on B and T lymphocytes.

4.7 Modelling the Stromal Network

This section introduces the sub-models used to define stromal cells and the stromal network within the context of the 'overarching' TLT formation model, providing a rationale for each modelling technique used and describing how the components interoperate to provide a functional stromal network model suitably amenable to hybridisation with other components of the TLT model; subsequent sections formally define each of the three sub-models in detail — those of generative grammars, Markov models, and cellular automata.

Stroma has a fundamental role in the formation of TLT; it is responsible for producing adhesion molecules and the lymphoid chemokines CXCL13 and CCL19. Stromal cells form complex interconnections resulting in a network structure that aids lymphocyte migration and organisation (??). Crosstalk between lymphocytes and stromal cells results in changing chemokine expression by stroma, as seen from the data in Figure 3.3 discussed in Chapter 3. Therefore, a method is needed to appropriately describe a spatially resolved network of individual elements, representative of the stromal network structure *in vivo*. The following elements are therefore needed to adequately model the stromal network:

- A stochastic algorithm that generates networks of individual elements with parameters permitting control of network density and element size.

- A method of determining stromal cell phenotype in response to stimulation by lymphocyte agents from the agent-based sub-model.

- A means of implementing the network generation algorithm and storing the result within a data structure amenable to hybridisation with other model types, bringing together spatial network structure and individual stromal cell phenotype.

Each of the elements described above are best-suited to different modelling approaches. Given the choice of a continuous space ABM for T and B lymphocytes, a discretised per-agent ODE describing chemokine sequestration and receptor internalisation, and a discretised PDE that describes chemokine secretion, diffusion and decay, the generated network should occupy space within a grid of the same dimensions of the spatial environment defined for the ABM — a square grid representing $40\mu\text{M}^2$ and utilise a data structure permit interaction with the ABM, ODE and PDE sub-models.

4.7.1 Determining Appropriate Modelling Methodologies for Individual Stromal Cell Network Components

This section describes the decision making processes that led to the adoption of each technique used to model the stromal network and stromal cell development. The existing literature is evaluated to determine the suitability of approaches previously applied to similar problems and justifications are given for the techniques selected. Following this, each of the three models is formally defined and discussed in detail.

Determining Appropriate Methodology for Stromal Network Structure Generation

Previous approaches to modelling stromal networks have focused on generating mathematical networks of edges connecting vertices using random sampling to determine edge location (**???**), these types of approach are inappropriate for the TLT model as they do not occupy a volume in order to permit emergent lymphocyte adherence and crosstalk (for example, in **?**, lymphocytes are confined to the network) and it is not clear whether such approaches would be valid abstractions when applied in 2 dimensions. Other approaches to capturing the stromal network have attempted to accurately capture tissue geometry from imaging data, **?** used confocal microscopy to construct 3D geometric models of lymph node morphology which were then converted to meshed models which, while ideally suited to finite-element

modelling, are not appropriate for defining individual stromal cells and hybridisation with other modelling methodologies used to develop the TLT formation model; finally, there is no suitable means of adapting meshed geometric models for use in a 2D environment.

The possibility was raised in **?** of algorithms that 'grow' stromal networks by beginning with a single node that propagates outward, although it was noted that such algorithms tend to produce several long edges and results in a non-random tree-like structure (**?**) that is not representative of the stromal network, which tends to maintain a small number ($\tilde{3}$) of edges (connections) (**?**) between each vertex (representing the stromal cell). Indeed, it is this branching property of such algorithms that has permitted their use in modelling plant growth (**?**), using formal grammars to define rewriting systems that recursively generate network structures. Therefore, any generative algorithm would need to address the issue of long edges for use in modelling stromal networks. Furthermore, the use of mathematical networks to describe the stromal network, while suitable for investigations of lymphocyte-stroma spatial dynamics in terms of understanding their effect on antigen scanning and cell-cell contact rates (**?**), are less well-suited to modelling a stromal network when the intention is to provide a substrate for lymphocyte self-organisation and stromal cell phenotypic development, in which a volumetric stromal network model is preferable (or, in the case of 2 spatial dimensions, an area-occupying stromal network model). Whilst the geometric models produced by **?** are volumetric, they do not lend themselves well to the hybridisation framework necessary to describe the other elements involved in TLT formation.

The issue of long edges arising from generative algorithms could be mitigated by defining multiple starting locations for vertices and allowing networks to overlap — a sufficiently dense network representative of the mesenchymal reticular tissue in which TLT forms should mitigate the formation of long edges. In **?**, a cellular automaton is used to describe lymphocyte motion along a generated FRC network of edges and vertices. In the TLT model, a generative grammar was defined, in

which the stromal network is randomly 'grown' through implementation in a cellular automaton consisting of a discretised grid of dimensions equal to that of the agent-based model. Instead of a non-volumetric network model consisting of edges connecting vertices, our stromal network occupies the entire area of each grid element within the cellular automaton in which the stromal cell is present; this permits a spatial area to allow for lymphocyte adherence and stimulation, and is suitable for defining point sources for chemokine secretion/diffusion using partial differential equations. The background of formal grammars followed by a complete description of the generative grammar created for stromal network structure definition within the TLT model, is described below (Section 4.7.2). Once the stromal network generation algorithm has been applied within a cellular automaton, the resulting data structure can be used as an interface between the stromal cell network and other components of the TLT model.

Determining Appropriate Methodology for Modelling Stromal Cell Differentiation

A Markov model using Monte Carlo techniques is well suited to model the current and possible future phenotype of stromal cells in response to lymphocyte stimulation, and lends itself well to hybridisation with other techniques. A discrete Markov chain is a system in which a random process governs transitions between a finite (or 'countably infinite') number of states, usually evolving over time. A probability exists that when in its present state, the process will change to another state, and these are defined for each possible state change. A process is 'Markovian' if its future state does not depend on the past, but only the processes present state (i.e. a *memoryless* process). A Monte Carlo process is a random sampling event to determine whether or which state transition takes place. Ordinarily, the Markov chain model would be simulated for n steps, as it evolves over time, and then the distribution of how much time spent in each state analysed to determine the Markov chain's 'steady state' — alternatively, Markov chains can be used to make predictions about the future

state of the system after a given interval; however in this application, each stromal cell is associated with one Markov chain that evolves as a function of contact with lymphocytes within an ABM, and to analyse the result we assess the final state of each Markov chain within each time-step (as a stromal cell may be stimulated by lymphocytes many times in one time-step, and thus many Monte Carlo samples of the Markov chain can occur).

The Monte Carlo process to determine transition is determined by contact with lymphocytes, which means in terms of the platform model that a lymphocyte within the ABM is in the same location in the simulation space as a stromal cell entity defined in the CDS, and that the lymphocyte is adhered to the stromal cell; this is in contrast to 'traditional' Markov chains — often called DTMC, a discrete-time Markov chain, which usually operate over a time dimension. While Markov chains operating over other dimensions do exist, the stromal cell Markov chain model is unique in terms of the stochastic non-linearity of the step function, at any given moment it is impossible to know when a lymphocyte will trigger a further Monte Carlo process. Markov chains have been used in dimensions other than time previously, mainly in the form of the *Monte Carlo Markov chain* class of algorithms that utilise Markovian properties to perform an array of tasks such as evaluating complex integral equations (**?**) and computing large hierarchical models in Bayesian statistics (**?**). There are few examples of models using the structure of a DTMC without 'time' as the dimension over which it evolves. One example however, is in the 'PageRank' algorithm used by the search engine developed by Google Inc. to determine the relevance of web-pages, (**?**), although this essentially is just utilising the power method for finding the largest eigenvalue (**?**). Markov chain models are suitable because they allow definition of a finite state of specific entities, and transition through probabilities in a manner common in evaluating cell-cell interactions *in silico* (**???**), and Markov model 'emissions', or external changes induced by the Markov chain model, are very well suited to defining specific phenotype with regard to exactly how much CCL19, CXCL13 and adhesion molecule expression is

occurring for each cell. Given that the biology of lymphocyte-stroma crosstalk is largely unknown in terms of the molecular signals on B and T cells responsible for the differential chemokine response, we can assume that in contact with a lymphocyte, a stromal cell has a defined, unknown, probability of differentiating. Since up/down-regulation of FRCs and FDCs does occur in response to lymphocyte stimulation beyond large phenotypic changes such as FRC to FDC differentiation, if a Markov chain model remains in the same state, its phenotype may be accentuated - for example, B cell stimulation of an FDC will result in a small upregulation of CXCL13 and downregulation of CCL19 — within defined limits. The transition probabilities, and the maximum/minimum expression limits of the stromal cell for each phenotype, are abstract platform parameter values that represent and allow us to model the unknown factors driving lymphoid stroma development. The Markov chain is formally defined in Section 4.7.3.

Integrating the Stromal Network and Stromal Cell Models

The three sub-models (generative grammar, cellular automaton, and Markov chain) that collectively describe the stromal network in the TLT model, are shown in Figure 4.8, illustrating how they can provide a suitable stromal network model when in combination that can be used within the context of a multiscale hybrid model. The generative grammar provides the instruction set that defines the stromal network structure, this is implemented using a cellular automaton within which the network is 'grown' prior to T_0 (the point in the model or simulation at which time 'begins'). Once the network generation process within the cellular automaton is complete, the network structure and individual cell information is stored within the matrix data structure used by the cellular automaton — this is referred to as the cellular data structure (CDS) . Each stromal cell within the CDS has a Monte Carlo Markov chain associated with it, such that every time a lymphocyte is in contact with and adhered to a stromal cell (as defined by the agent rules within the ABM) a Monte Carlo process is triggered on the associated Markov chain, which determines whether

the stromal cell changes 'state' (representative of stromal progenitor, localised stromal precursor, FRC or FDC phenotypes), or otherwise modifies expression levels of chemokines (known as Markov chain 'emissions', a concept borrowed from hidden Markov models typically used to infer unknown states, but useful for defining stromal phenotype beyond the simple 'four-state model' defined in the domain model). There is some precedent in integrating Markov chain models with cellular automata, 'probabilistic cellular automata' is an extension of cellular automata, also known as 'locally interacting Markov chains', in which the state of a collection of entities is updated according to a simple, uniform rule throughout the automaton space (**??**), this uniform rule is essentially a discrete-time Markov chain operating over the cellular automaton. In our case, there are many Markov chains operating within the data structure remaining from the application of a cellular automaton to implement the formal grammar that stochastically defines the spatial location of the Markov chains within the CDS. Changes in the stromal cell phenotype are updated within the CDS, which is used to determine chemokine sources for secretion and diffusion, which subsequently leads to lymphocyte migration, self-organisation, and eventually, TLT formation. Once the network is generated and time has begun within the simulation, the CDS effectively acts as an interface between the collection of Markov chains each associated with a stromal cell entity, and all the TLT formation sub-models which interact with them.
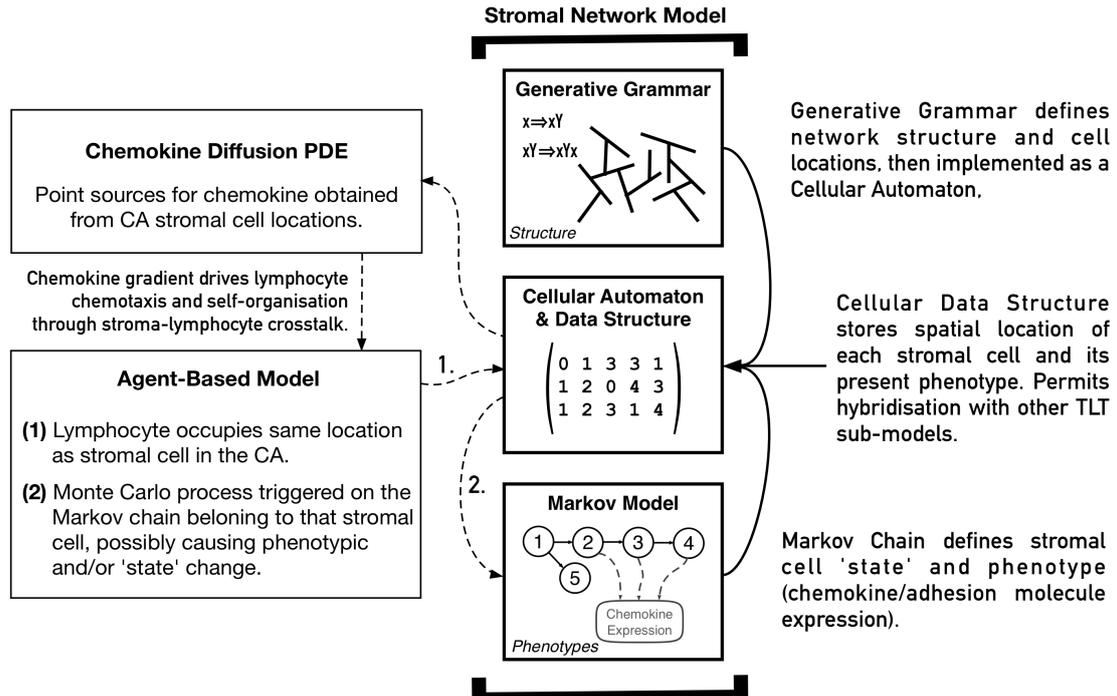
Figure 4.8: Schematic illustrating how the three model components that constitute the stromal network are connected, and how lymphocytes within the agent-based network migrate toward and stimulate stromal cells. Note that the chemokine receptor recycling ODE described in Section 4.5 also influences lymphocyte behaviour and chemokine concentrations.

4.7.2 Stromal Network Generation using Generative Grammar

Introducing Formal Grammars

A formal grammar falls within the framework of formal language theory, and can be defined as a set of production rules for producing strings within a given language; formal grammars were developed in the context of the linguistic theory of *generative grammar* – so called due to the intention that such grammars would be capable of producing every possible 'utterance' within a language (**?**) that is considered grammatically correct. The production rules describe how strings may be formed from the defined alphabet of the language which the grammar describes. The formal grammar itself describes the creation of strings through recursive replacing of 'non-terminal' characters within a string according to a defined set of production rules,

beginning with an initial 'axiom', and are therefore a form of string rewriting system. The strings can have 'meaning' applied to them, enabling interpretation by software parsers to utilise formal grammars within a wide variety of domains, extending far beyond their linguistic origins, this is due in large part to the combination of formal grammars and an information-theoretic approach within the 'Shannon tradition' of information and coding theory (**?**). Within the life sciences, perhaps the most common application is their use in many modern RNA structure prediction algorithms (**???**) through stochastic sequence generation that may then be evaluated for correctness.

Through the assignment of operations to each letter in the grammar's alphabet, strings generated by a given grammar may be interpreted as an instruction set for the production of geometric structures – this principle led to the development of 'L-systems' or Lindenmayer systems (**?**) used to describe the growth of trees and other branching processes (**?**).

Grammars are generally classified according to the nature of their production rules; fundamentally there are three classes of formal grammar: a *context-sensitive* grammar has production rules that are dependent not only on the symbol undergoing rewriting, but those adjacent to it – that is, the production rules are dependent on the context of the string; conversely, a *context-free* grammar has no such restrictions, however each production rule must replace only one character. Far more flexibility is afforded by *unrestricted* grammars, in which production rules may operate upon any set of non-terminal characters, although this comes at the cost of increased difficulty of creating parsers capable of interpreting resulting strings, and the non-trivial complexity of mathematical analysis of the grammar. Grammars may be augmented with additional properties; for instance, grammars may be *stochastic*, containing probabilistic production rules, or *parametric* in which case characters are associated with defined parameter values (in which case, the tuple of the character and its associated parameter are termed a *module*). A full description of the mathematical foundations of formal grammar theory is beyond the scope of this thesis,

however the reader is directed to **?** for a comprehensive overview and to **?** for a review of the historical literature following their development by **?**.

Defining the Stromal Network Generative Grammar

The grammar used to develop stochastic network-like structures of stromal cells for use in our hybrid multi-scale model of TLT formation is described in Figure 4.9. This grammar is both stochastic, such that each generated sequence is different even with identical parameters, initial axiom and rule application, and parametric, allowing values to be assigned to the grammar, permitting modulation of stromal cell size, network density and connectivity.

Stochastic Parametric Unrestricted Grammar for Stromal Network

Generation on a 2D Hexagonally-Packed Square Grid

Axiom: c

Alphabet: $C\,Q\,D\,P\,F\,L\,R\,\{\,\}\,[\,]$

Constants: c q f l r

Parameters: B N n

Rules:

1. $\{\,(\,[\,(\,c\,d\,)^{\,n}\,]\,q\,)^{B}\,\}^{N} \Rightarrow (P\,\{\,(\,[\,(\,c\,d\,)^{n}\,]q\,)^{B}\,\}\,)^{N}$

2. $(\,cd\,)^{\,n}C \Rightarrow \{\,(\,[\,(\,cd\,)^{n}\,]\,q\,)^{\,B}\,\}\,q\,C$

3. $c \Rightarrow c\,d\,C$

4. $d \Rightarrow F\,\|\,L\,\|\,R$

Further stipulations:

- These rules are in order of precedence, the first applicable rule must be applied.
- During parsing, the spatial grid must be treated as toroidal and later elements placed should overwrite existing elements.
- The generative process must continue until:

$$\frac{Nn}{2} = \sum_{i=1}^{i=\mathrm{len}(s)} \delta(\,\{\,...\,\}\,,s_i),$$

where s_i is the ith element of the generated sentence, s.

Figure 4.9: Complete definition of the stochastic parametric unrestricted grammar that defines every possible stromal network structure.

Lower-case letters in generated strings are *non-terminal* characters or sets thereof, and may be re-written according to the production rules, while upper-case letters represent *terminal* characters that are immutable and cannot be re-written. The parameters $\boldsymbol{n}$, $\boldsymbol{B}$, and $\boldsymbol{N}$ form modules with alphabet characters *cd*, *([(cD)^n]q)*, and $\{([(cD)^n]q)^B\}$ respectively. It is important to note that these modules utilise an abuse of notation to indicate that a parameter operates on several characters in a manner similar to algebraic notation with standard brackets – '*(*' and '*)*' – indi-

cating where expansion should occur, whilst square and curly brackets are reserved as characters and form part of the defined language. For example, cd^n, where $n = 3$ is equivalent to writing '*cdcdcd*'. Furthermore, the parameter-character modules are nested, such that the module of parameter $\boldsymbol{B}$ contains parameter $\boldsymbol{n}$, and parameter $\boldsymbol{N}$ contains both parameters $\boldsymbol{B}$ and $\boldsymbol{n}$ within its associated module. Rule 4 is stochastic, such that $d \Rightarrow F \parallel L \parallel R$ should be read as 'd is to be randomly replaced with F, L, or R with equal probability'. These three outcomes correspond to growth of the network Forward, Left, or Right, respectively. The four rules are stated in an order of precedence, such that the first rule that applies to a given structure *must* be applied, and this continues until termination because the pre-defined number of networks defined by parameter $\boldsymbol{N}$ have been created. The grid is treated as toroidal to maintain stromal network connectivity and prevent multiple 'short branches being produced at the grid edges. For the purposes of chemokine diffusion from stromal cells, however, the grid space is treated as non-toroidal as described in Section 4.4. The following list defines the meaning mapped to each of the letters in the generative grammar alphabet defined in Figure 4.9:

- C — Place stromal cell in grid location.

- Q — Select random element from the preceding set of $\boldsymbol{n}$ elements.

- D — Branch out in an undetermined direction.

- R — Branching direction for current stromal cell is to the right.

- F — Branching direction for current stromal cell is forward.

- L — Branching direction for current stromal cell is to the left.

- P — Randomly select new (x,y) co-ordinates for the new sub-network.

- { — Begin new individual branch (equivalent to start of one 'stromal cell').

- } — End current branch (equivalent to end of one 'stromal cell').

- [— Begin new network of $\boldsymbol{B}$ branches.

-] — End current network of ***B*** branches.

Figure 4.10 illustrates the 'growth' process represented by the grammar upon implementation. The generative grammar is intended to operate over hexagonally packed grid toroidal about the X and Y axes, initially the resulting 'sentence' should be generated, and this should subsequently be parsed as the network structure is created (in practice, however, the grammar can be algorithmically defined in a more computationally efficient manner such that the network is directly generated using a series of recursive loops). Rule 3 produces strings of the form '*cd...cd*', this creates a branch of individual grid elements that when combined may be classified as one stromal cell. Rule 2 subsequently rewrites this by replacing each branch with a new structure containing both the previous structure and a new branch. Rule 1 is responsible for creating numerous branching networks that overlap each other to avoid the issue of unnatural long-edges typically produced by generative grammar-type algorithms. The order of precedence is defined such that the 'highest-level' operation possible is performed first, with the lowest level being the addition of a new grid element to a stromal cell and the highest level being the addition of a new network of stromal cells to the grid environment. This process is further illustrated in Figure 4.11 which demonstrates an expansion from the initial axiom of a stromal network structure (left), alongside one of the many possible structures this sentence could produce depending upon the stochastic choice imparted by rule four of the grammar (right).

The grammar described in Figure 4.9 was developed manually by tracing over micrographs of stromal networks and attempting to define informal rules for occupying squares in a hexagonally-packed square grid, prior to formalisation of these rules in the form of the grammar described herein. Figure 4.9 provides a schematic illustration of the process formally defined in Figure 4.9, and Figure 4.11 subsequently provides an example of the 'sentence generation' process and how the 'sentence' produce describes the spatial organisation of the abstract stromal network model. The important characteristics that the resulting structures must have were that each
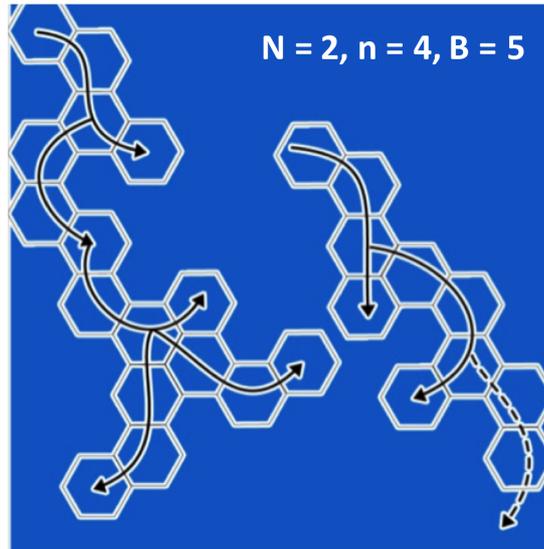
Figure 4.10: Illustrative demonstration of the network generation process imparted by the grammar of Figure 4.9 when applied to a hexagonally-packed grid with parameters defined as $B = 5$, $N = 2$ and $n = 4$. Extremely simple parameter values have been chosen in these examples for the purposes of communicating how the grammar operates.

generated network must be unique through stochastic generation, yet consistent with each other for a fixed set of parameter values (e.g. maintain a similar overall density through the parameter values N and n) such that they are structurally *locally unique* to prevent simulation artefacts resulting from aberrant nuances in a specific network structure. Therefore, every single simulation run is conducted using a unique, non-repeating network structure. It is difficult to formally define the maximum number of possible structures this grammar is capable of producing, however it may be considered infinite for all practical purposes providing that care is used in selecting the random seed used for the pseudo-random number generator for each simulation execution.

B = 2, N = 2, n = 3
1. C
2. cdC
3. cdcdC
4. cdcdcdC = cd³C
5. {([(cd)³]q)²}qC
6. {([(cd)³]q)²}qcdC
7. {([(cd)³]q)²}qcdcdC
8. {([(cd)³]q)²}qcd³C
9. {([(cd)³]q)²}²
10. (P{([(cd³)]q)²})²

'Sentence' produced:
P{[cdcdcd]q[cdcdcd]}qP{[cdcdcd]q[cdcdcd]}P
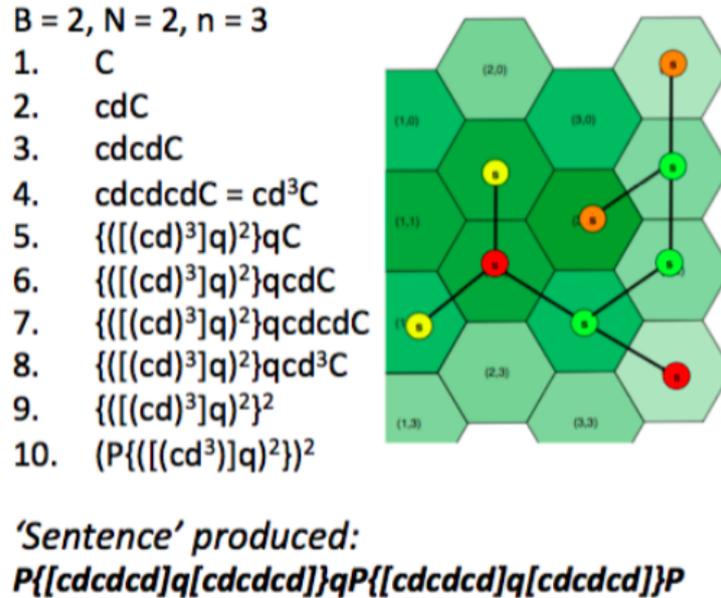
Figure 4.11: Demonstrative implementation of the stromal network generative grammar, resulting in a 'sentence' that provides instructions for spatially resolving the network on a hexagonally-packed grid within a cellular automata.

4.7.3 Markov Model to Describe Stromal Phenotypes and Differentiation

This section defines the Markov chain for a stromal cell, including its emissions (expression of adhesion molecules and lymphoid chemokines), and this is then developed into a finite state machine using the UML in a manner amenable to hybridisation with agent-based models, as the system evolves as a function of lymphocyte contact — which itself is a complex non-linear stochastic function of time — and therefore does not lend itself well to elementary analytical treatment. While emissions and emission probabilities are a property of hidden Markov models, it is conducive to adopt their use in the stromal cell Markov model to represent phenotype with greater granularity than the 'four-state stroma model' discussed in Chapter 3, permitting precise expression levels of CXCL13, CCL19 and adhesion molecules. The justification for using a Markov chain model to describe stromal cell phenotypes within the stromal network was introduced in Section 4.7.1. The Markov chain describing one stromal cell, using finite state machine notation (?) with square boxes to define
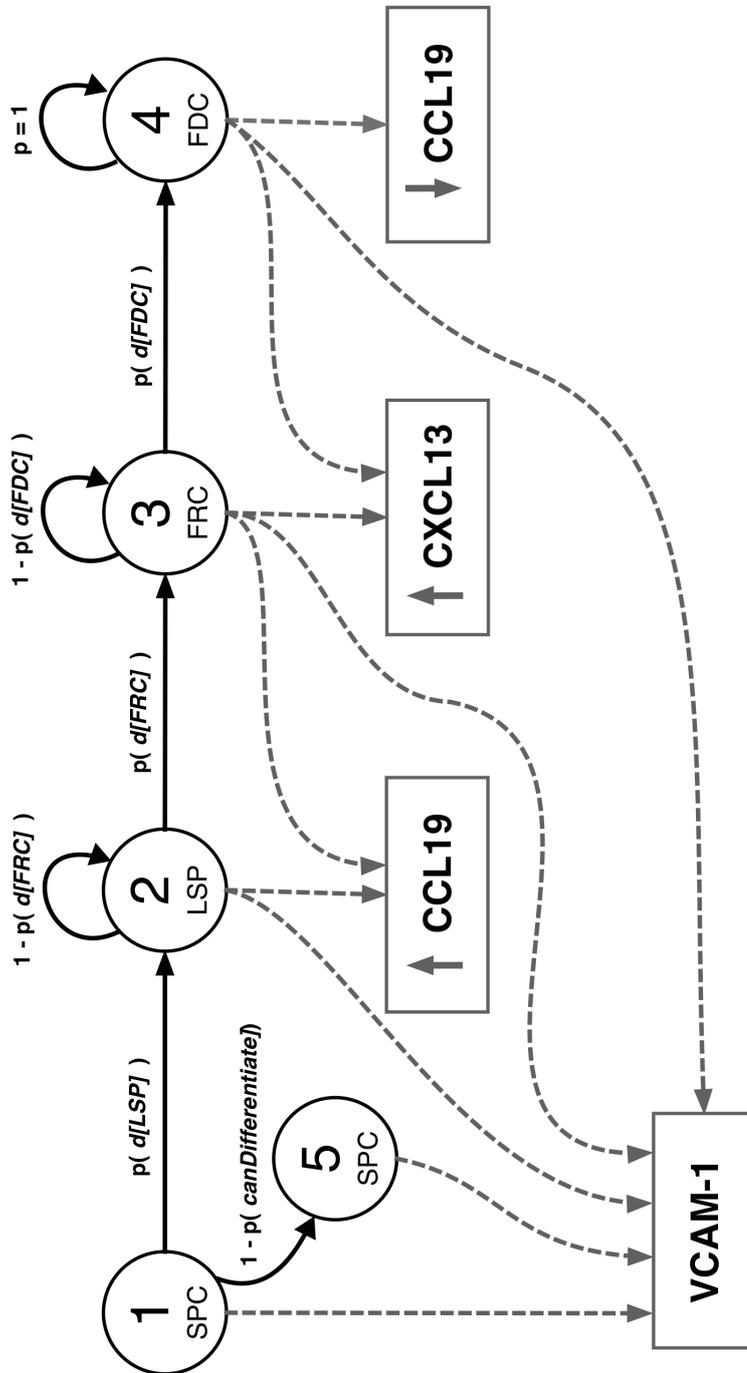
Figure 4.12: Markov chain describing transitions between stromal phenotypes, and the 'emissions' each state is responsible for; these correspond to the soluble and membrane-bound proteins produced by each phenotype. The Markov chain is hybridised with an agent-based model and Monte Carlo probability sampling to determine when state transitions occur. The emissions in this model are parameterised using the *in vitro* data described in 3.3

emissions, is shown in Figure 4.12.

Parameter values for the Markov chain are provided in Table 4.1, these represent the probability of a given stromal cell phenotype differentiating into the next type described by the chain, which itself is derived from the four stage differentiation domain (biological) model described in 3.5. It is necessary to use a probabilistic approach in which Monte Carlo sampling determines stromal cell differentiation as the differentiation signals presented by B and T cells are presently not known. Variation of these probabilities change the rate at which stromal cells develop, as does variation in the number of B and T cells present in the simulation as defined by the initial conditions. The presented parameter values were calibrated such that, with a total of 8000 T cells and 8000 B cells, the differentiation of stromal cells into FRC and FDC phenotypes allows the self-organisation process to emerge over the 15 day period observed in the acute TLT murine model described in Chapter 3. Given the theoretical and unidentifiable nature of this model, it is not viable to present ranges for these parameter values, especially considering the argument from **?** that parameter measurements must be impractically complete and precise to constrain model behaviours in complex systems models.

It is useful to additionally define the stromal cell Markov chain and its emissions as a UML finite state machine (**?**) in the manner used to define T and B lymphocyte behaviour within the agent-based model in Section 4.3, as this permits tight integration of lymphocytes and stromal cells within a single simulation during the software implementation phrase, otherwise referred to as the 'Simulation Platform' in CoSMoS parlance.

Parameter Name	Value	Hybridised With	Monte Carlo Method Triggered By	Event Time-Scale
p(d[LSP])	$I_0 e^{-\lambda \Delta T}$	CA	Stromal CA evaluating ODE	10 Minutes
p(d[FRC])	0.003	ABM + CA	T-Cell adhered to Stroma	2 Minutes
p(d[FDC])	0.000001	ABM + CA	B-Cell adhered to Stroma	2 Minutes
1 - p(canDifferentiate)	0.2	N/A	Initiated at T=0	Once

Table 4.1: Table defining the stromal differentiation Markov model probabilities and the sub-model with which each is hybridised. Interactions with the ABM and CA dictate when transitions will be evaluated using Monte Carlo methods for each stromal cell within the simulation model. The numerical probabilities provided were determined through calibration.
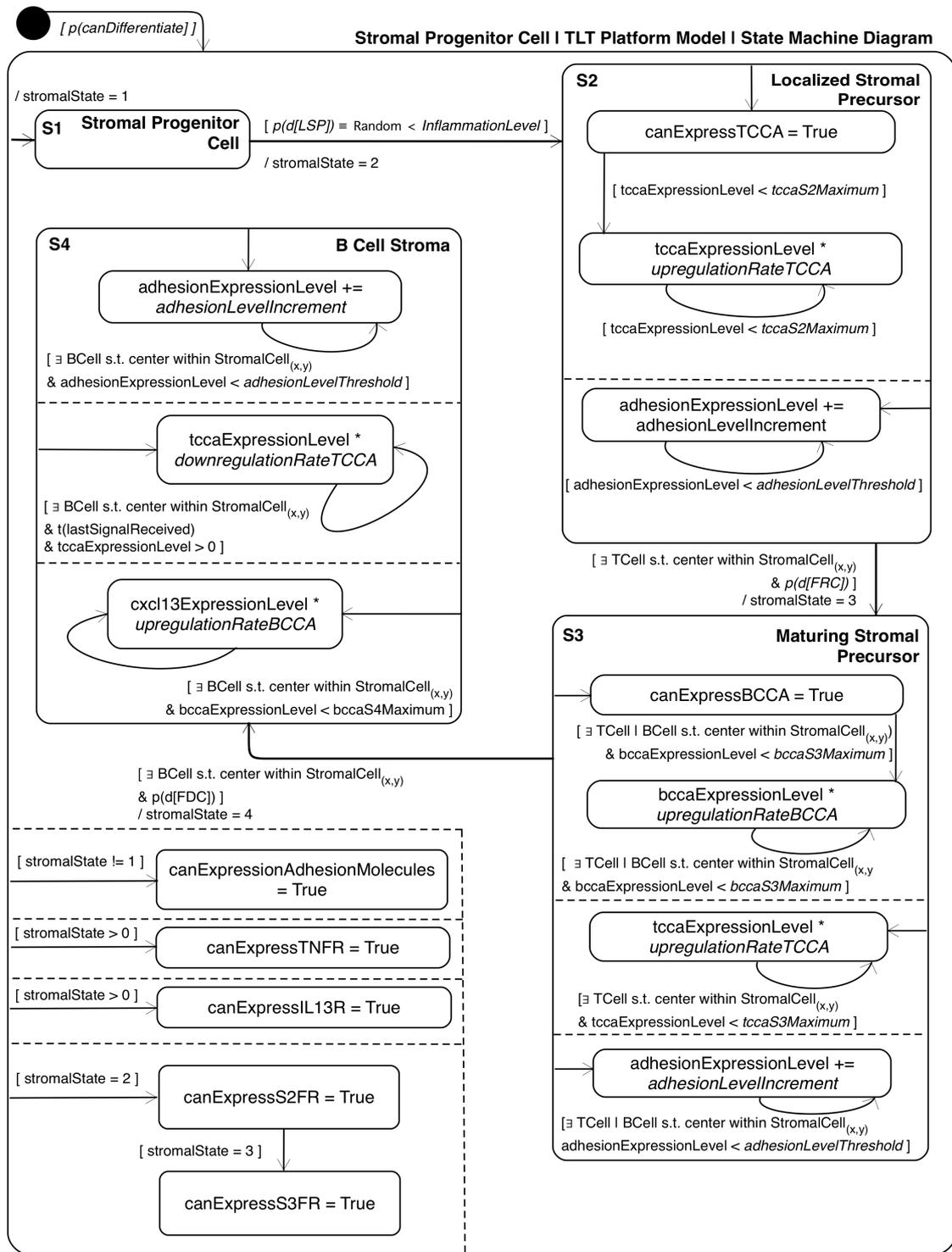
Figure 4.13: UML Finite State Machine Platform Model diagram describing the role of stromal precursor cells in TLT formation, including differentiation into FRC-like and FDC-like phenotypes.

4.8   Deriving Model Outputs Using the Hybrid Modelling Schema

This section explores a minor extension to the hybrid modelling schema introduced in Section 4.2, permitting defining model outputs clearly in terms of sub-models, the data extracted from them, and their combination to produce outputs that may be used to evaluate model responses. Model outputs or responses can be defined in terms of the individual component models that comprise the resulting hybrid agent-based simulation. Figure 4.14 provides a generalised schema illustrating this processes, in which high-dimensional model outputs are specified using visual notation that relates them to multiple specific model components.

Figure 4.15 provides an example of how high dimensional data such as that provided by flow cytometry can be derived from models parameterised with low-dimensional data. Outputs approximating imaging techniques, such as immunohistochemistry confocal microscopy, are defined in terms of the stromal Markov chain,

Schema for Generation of Model Outputs & Visualisations from Hybrid Models
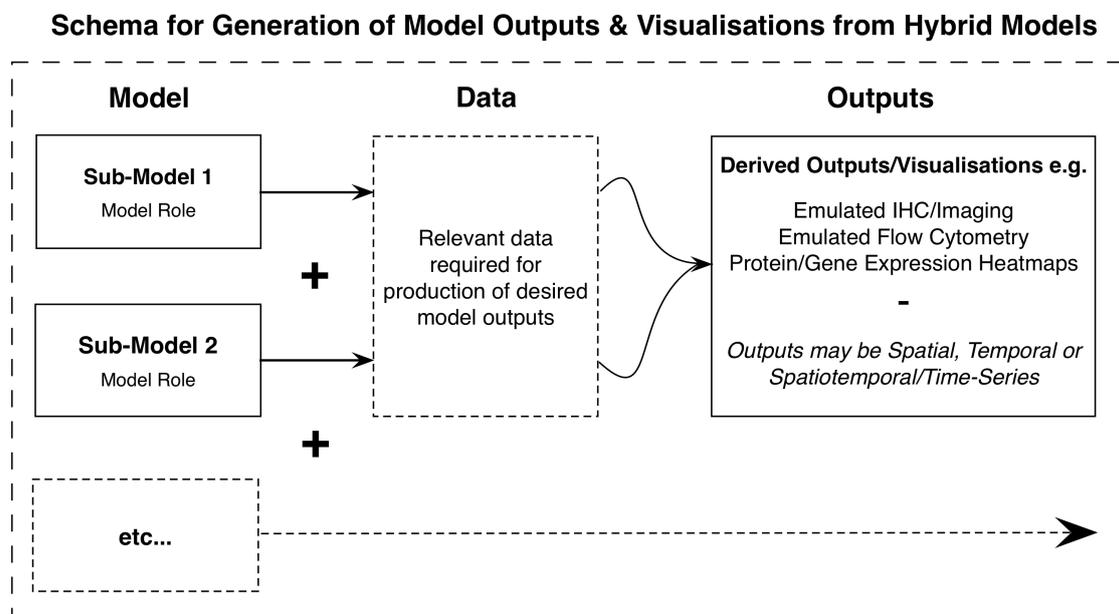
Figure 4.14: Generalised schema for defining the overarching multi-dimensional model outputs and visualisations with respect to combined sub-models integrated into an executable simulation, producing novel, useful simulation outputs, including those analogous to techniques in experimental biology, as discussed in Chapter 2.
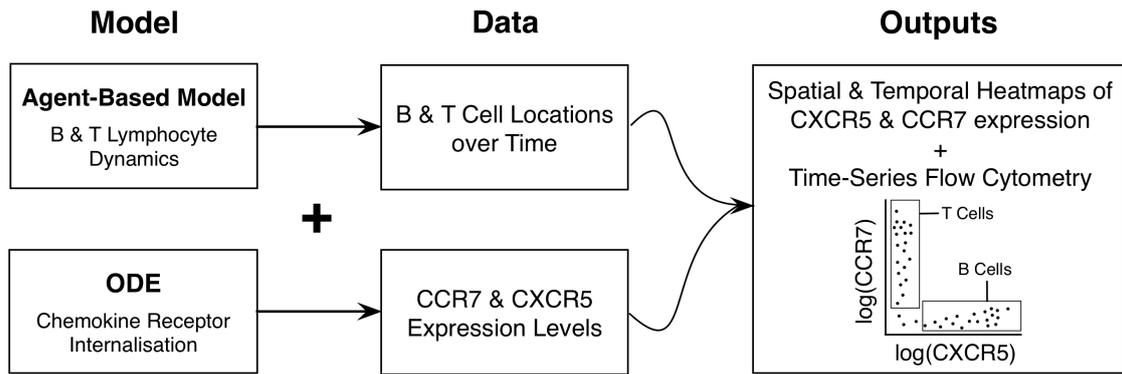
Figure 4.15: Scheme for combining the agent-based sub-model and receptor recycling ODE to produce emulated flow cytometry and IHC/imaging outputs.

the 'cellular data structure' and the generative grammar that defines the generation process for the stromal network structure, as presented in Figure 4.16. Understanding the spatial distribution of chemokines over time is possible by combining the chemokine diffusion and receptor internalisation models, as shown in Figure 4.17.

The model outputs defined here are based on the methodologies developed in Chapter 2 using the Peyer's patch development simulation, PPSim, as a case study. Those in Figure 4.16 form the basis for understanding model responses to therapeutic intervention in Chapter 6, and those in Figure 4.17 are used to explore co-localisation of chemokines and lymphocytes during calibration of model parameters to fit the temporal development pattern outlined in the TLT domain model.

In addition to the possible outputs discussed above, high-throughput image analysis is applied to graphical simulation outputs (introduced in Section 5.1.2) using software designed for use with microscopy images. Additionally, high dimensional data such as cell counts, particularly of the four stromal phenotypes within the model, lend themselves well to a variety of high-dimensional image analysis methodologies. These are introduced in Chapter 5
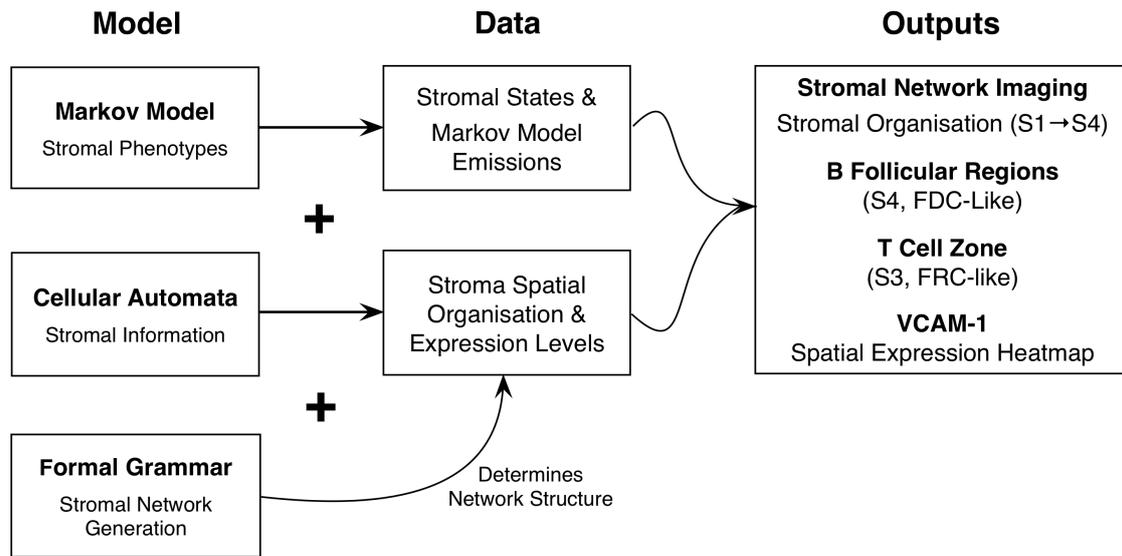
Figure 4.16: Scheme for combining the Markov model of stromal cell differentiation and the CA that stores the phenotype data for each individual stromal cell and the spatial organisation of the network as defined by the stochastic formal grammar. This permits outputs of images illustrating the spatial distribution of stromal cell type and heat-maps of VCAM-1 adhesion molecule expression within the spatial compartment.
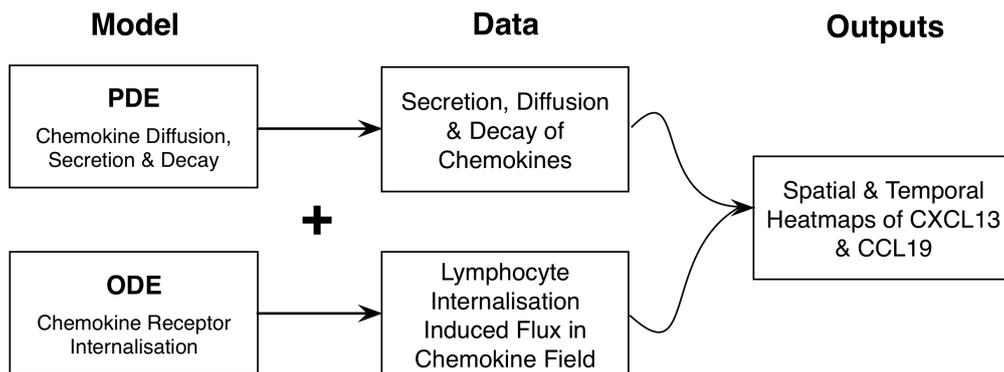
Figure 4.17: Scheme for combining the PDE responsible for describing the secretion, diffusion and decay of chemokines with the chemokine receptor recycling ODE. This permits production of spatial heat-maps of CCL19 and CXCL13 chemokine distribution, and temporal heat-maps illustrating the change in total concentration over time.

4.8.1 Model Outputs

Both qualitative and quantitative outputs are of interest from this model. Given the somewhat theoretical foundations for the model, based largely on hypotheses derived from human *in vitro* mesenchymal stem cell models of stroma and an *in vivo* murine model of Sjörgen's syndrome, it is not possible to define a quantitative level of precision with respect to any other model of the disease. Therefore, the principal output of interest is whether the model as constructed, implementing the TLT formation theory as described in Chapter 3, results in the formation of structures in a self-organising manner consistent with those found *in vivo*. This is fundamentally a qualitative, or semi-quantitative, output because of the enormous heterogeneity of TLT observed in the domain, and the lack of quantification of the extent of this heterogeneity.

Using high-throughput image analysis, the variation in the morphology of TLT within the model can be assessed as parameters are peturbed. This is explored for modulating the level of adhesion molecule expression in Chapter 5. Further to this, by incorporating potential and previously trialled therapeutic interventions into the model and observing the relative change in the population of stromal phenotypes, insight can be derived into the potential therapeutic benefits of the intervention. The key output in this sense is the change in the population size of FDC-like (S4) stromal cells. These cells are required to support lymphoid follicles and germinal centre reactions known to correlate with disease activity in Sjörgen's syndrome. This may therefore be used as a surrogate for disease activity for the purposes of evaluating therapeutics. This is explored further in Chapter 6.

4.9 Argument-driven Model Validation

This section explores the argumentation structure for the TLT model, this contains 4 key strategies identified for arguing the fitness-for-purpose of the model. GSN provides a useful way to present the evidence used to construct the model and

simulation, including all assumptions that have been made and justifications thereof. The strategies used to argue fitness-for-purpose of a model developed using the CoSMoS process are best provided in this author's opinion through decomposition of the argumentation structure into the Domain, Domain Model, Platform Model and Implementation – one must argue that the experimental data are appropriate for the domain model (Strategy 1.1), that the domain model captures the relevant biology (Strategy 2.1), that the platform model appropriately abstracts the domain model and preserves emergent phenomena (Strategy 3.1), and that the simulation is correctly implemented (Strategy 4.1).

Strategy 1.1: Experimental Data Sources

Strategy 1.1 (Figure 4.19) aims to argue that the experimental data used in the development of the domain model are appropriate and explore how these data were integrated into a model. This currently contains two sub-claims: Claim 1.1.1 states that relative expression levels of chemokines produced by stroma *in vitro* are known. This is evidenced in **?** [1] (Evidence 1.1.1.1) and has the assumption (1.1.1.1) that *in vitro* data may be used for our *in silico* model of an *in vivo* process. Claim 1.1.2 states that Adipose Derived Stem Cells (ADSCs) treated with inflammatory cytokines are a suitable model of lymphoid stromal cells. This is evidenced from observations *in vitro* that ADSC-derived stromal cells adopt a "lymphoid" phenotype when cultured with inflammatory cytokines.

Strategy 1.2: Domain Model

Strategy 1.2 (Figure 4.20) makes a case that the Domain Model encapsulates all relevant aspects of the domain within its experimental scope. Currently this is evidenced by the observation that IL13 KO mice do not develop TLT structures, along with the data discussed in Strategy 1.1.

Strategy 1.3: Platform-Independent Model Strategy 1.3 (Figure 4.21) is comprised of 4 claims. These are 1) the model unambiguously formalises all possible cellular interactions within the model scope, 2) that lymphocyte movement within

---

[1]Experiments shown here from **?** were performed by Bridget Glaysher

the tissue can be modelled as a random walk (essentially Brownian motion), 3) that chemokine diffusion can be modelled using the heat equation and finally 4) that the stromal network generated by the L-System is an appropriate abstraction of stroma. Various pieces of evidence from literature, and justifications for implementation decisions are then included. Note that a white diamond on a box indicates that it has not yet been developed into a full argument.

Strategy 1.4: Simulation Implementation Strategy 1.4 (Figure 4.22) is concerned with the actual implementation - whether space and time are modelled appropriately, and whether it meets its aim of determining the minimum requirements for TLT formation.

Figure 4.18: The complete GSN Argumentation making a case that the simulation is fit for purpose and an adequate representation of the biology within the intended experimental scope. This figure is included for illustrative purposes only, larger versions of the 4 strategies are presented below.
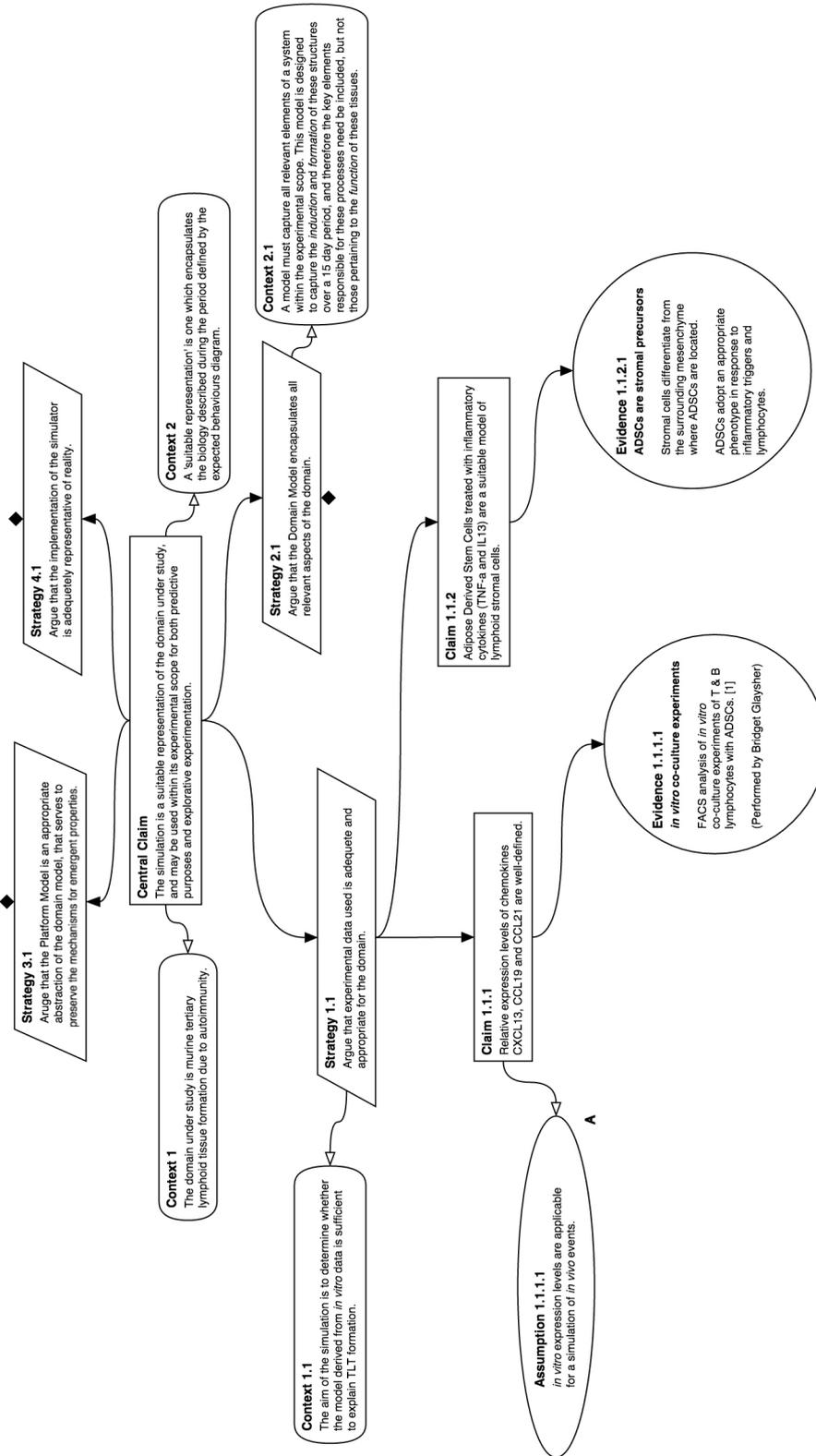
Figure 4.19: GSN Argumentation Structure – Strategy 1.1
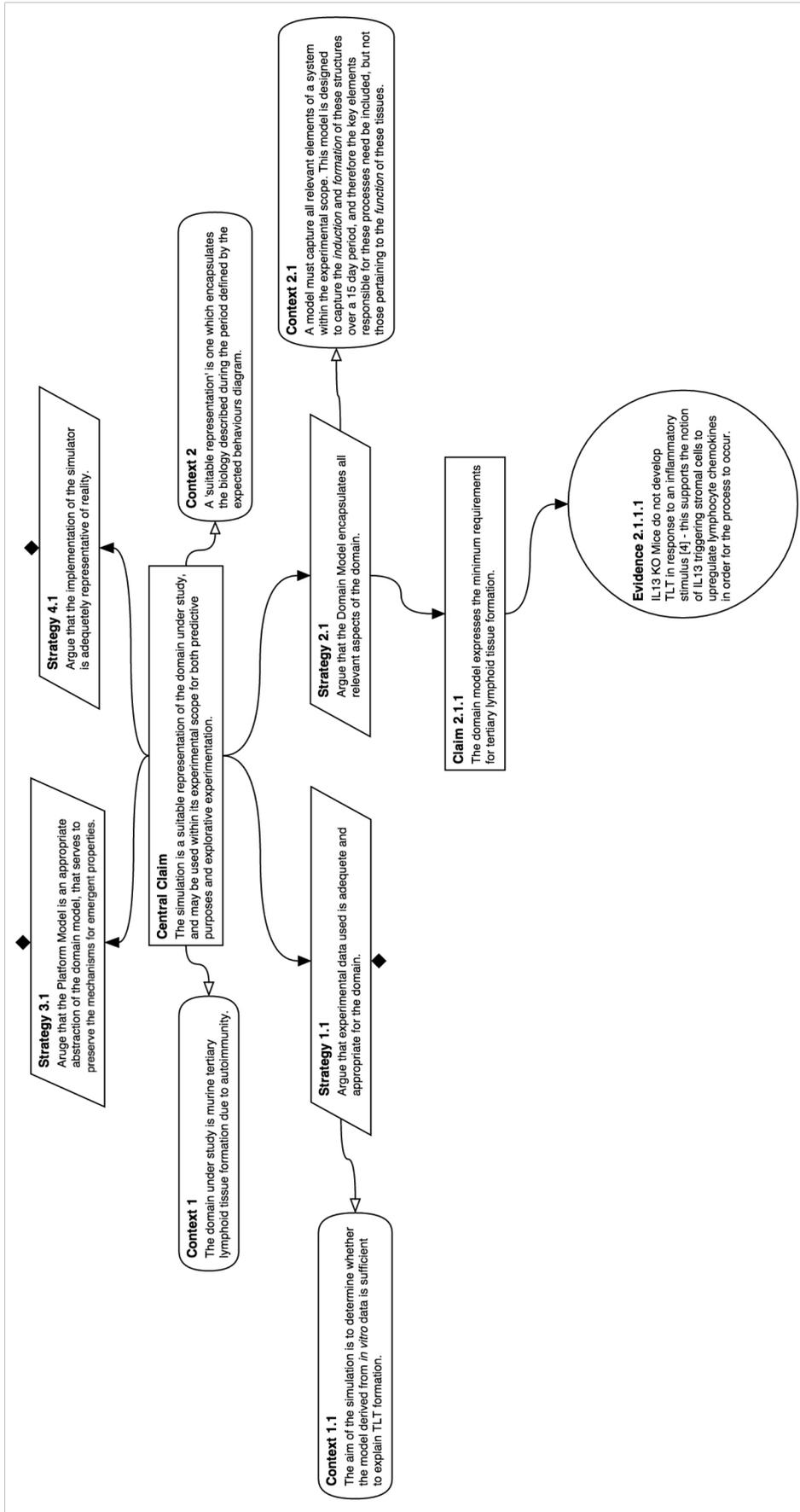
Figure 4.20: GSN Argumentation Structure – Strategy 1.2
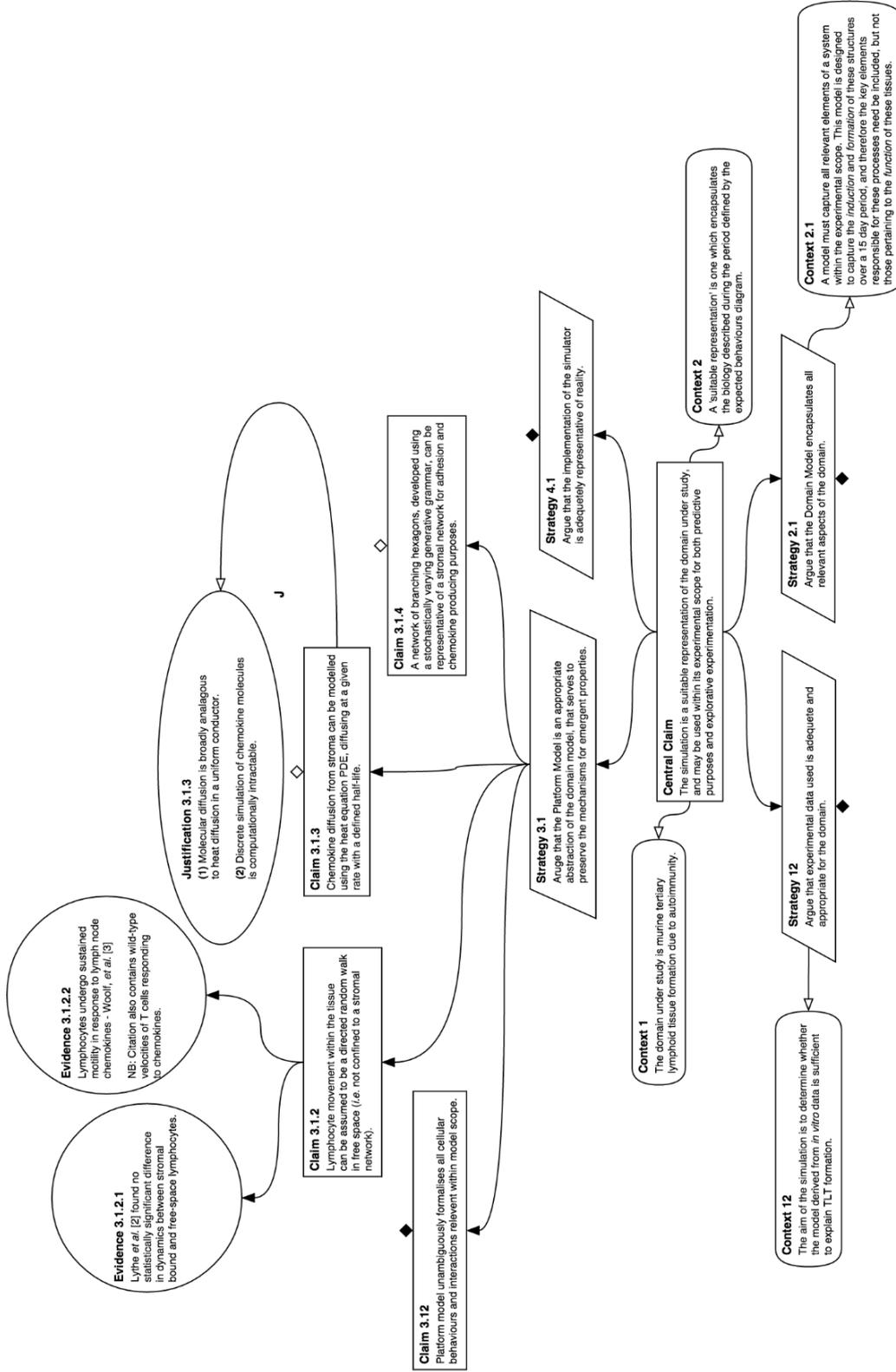
Figure 4.21: GSN Argumentation Structure - Strategy 1.3
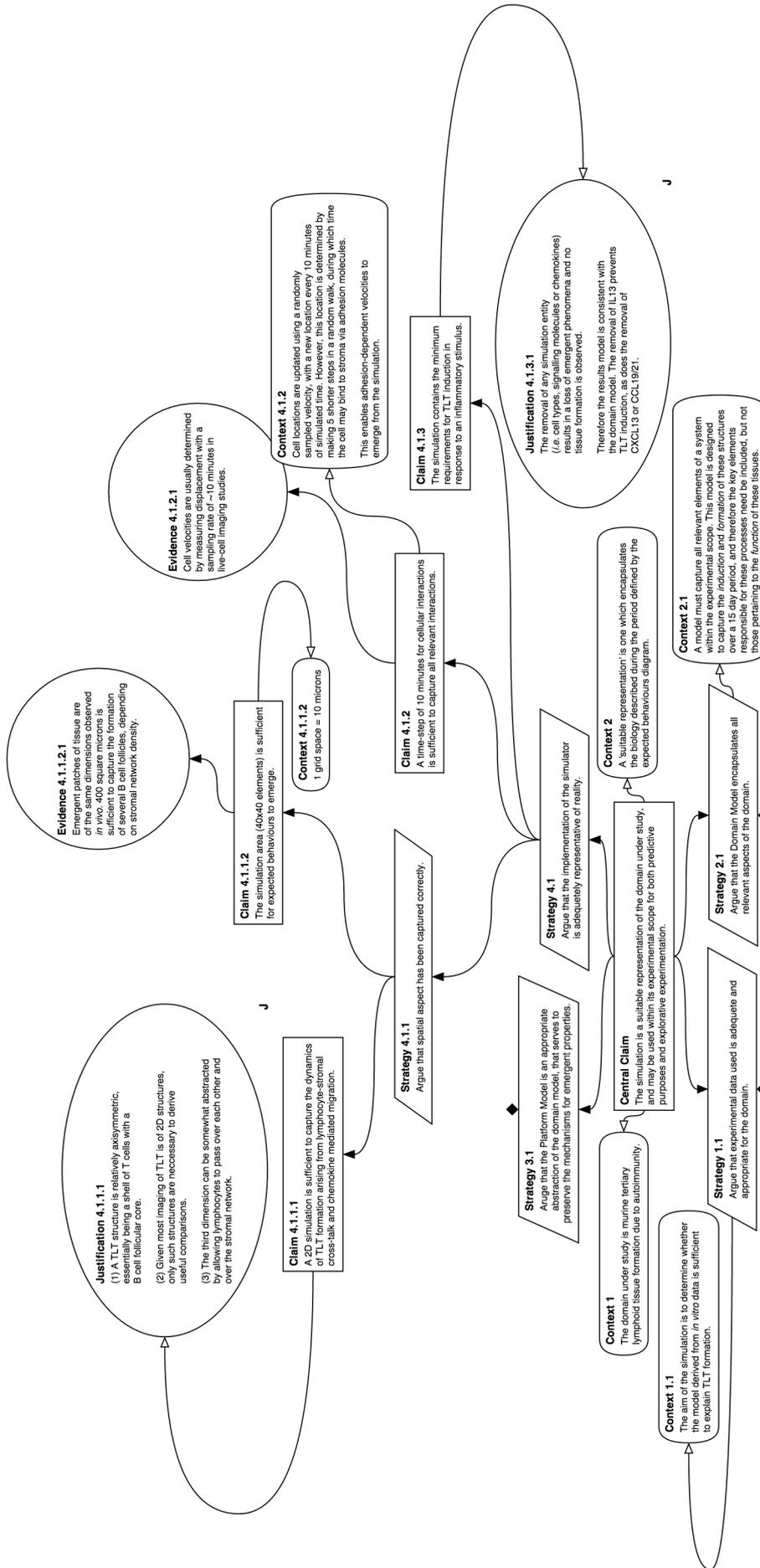
Figure 4.22: GSN Argumentation Structure – Strategy 1.4

4.10   Developing the Simulation Platform

The software simulator, '*NeoSim*' was developed in Java using the MASON library **?** to provide a scheduler and environment for the agent-based aspects of the model. *NeoSim* will be released and made available following publication of this thesis on the York Computational Immunology Website (`www.york.ac.uk/computational-immunology`). The simulation 'class diagram', showing the software classes and their relationships (e.g. inheritance) is shown in Figure 'ref-fig:class. A full description of the model as implemented is described in Appendix A.

A screen-shot of the resulting software simulation, with a graphical user interface (GUI) for real-time model interrogation and visualisation is shown in Figure 4.24. A typical run in the console takes approximately 5-6 minutes using one of eight cores on a 2.2 GHz Intel Core i7 laptop with 16 GB of RAM. With the visualisation updating each of the 2161 time-steps (each representing 10 minutes of simulated time), a typical run takes about 12 minutes to complete. This disparity is because the simulation is still restricted to a single processing core but must also handle the graphical overhead as well as running the simulation logic and writing outputs.

Figure 4.23: NeoSim UML Class Diagram showing TLT formation model implementation as a software simulator in Java using the MASON agent-based modelling library.
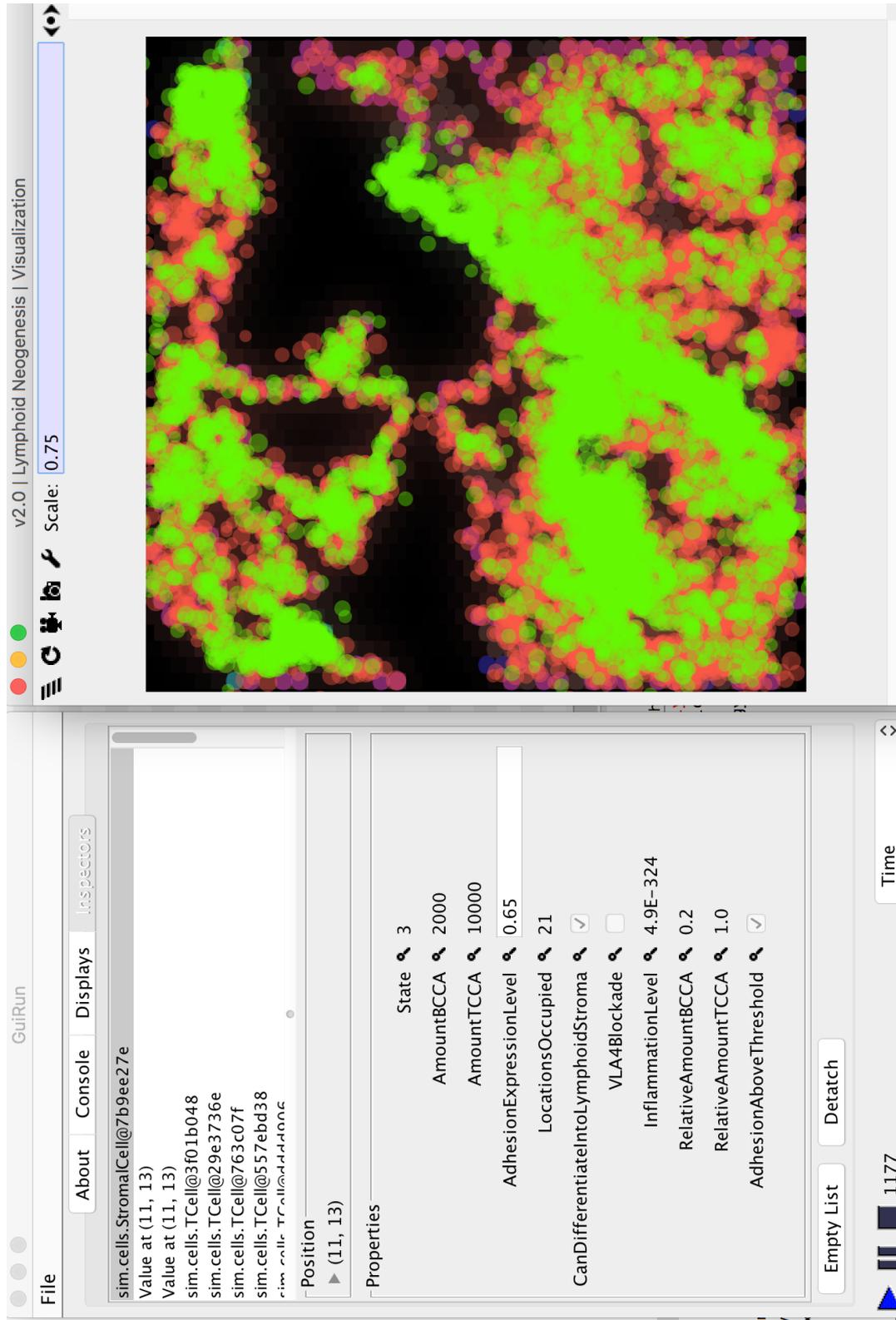
Figure 4.24: NeoSim software screenshot. The GUI elements are implemented using the MASON simulation toolkit derived from the Swing GUI library.

4.11 Summary

This chapter introduced a platform model describing TLT formation in Sjögren's syndrome, providing mathematical and computational descriptions of each biological entity within the model. Through the development of a model hybridisation schema, the model could be effectively described in such a manner that each entity is described using the most appropriate modelling technique prior to re-integration into a single simulation tool that has been named NeoSim. Implementation details of the NeoSim software tool have been provided.

NeoSim integrates agent-based modelling with Markov chains, a discretised PDE model describing chemokine secretion and diffusion, and an ODE model describing the recycling of chemokine receptors on lymphocytes. A generative grammar is described that produces stochastic stromal network-like structures on a discrete hexagonal grid. The generative grammar provides a key platform for lymphocyte-stroma crosstalk, which lies at the heart of the hypothesis presented in Chapter 3 that lymphocytes induce differential chemokine secretion profiles in stromal cells and that this results in their self-organisation into TLT structures.

The following chapter performs analysis of the model and simulation, determining whether the theoretical model outlined in Chapter 3, when implemented in the manner described in this chapter, produces structures that are consistent with TLT as observed *in vivo*. Further to this, the roles of adhesion molecules in modulating TLT morphology *in silico* are explored, and some consideration to simulation sensitivity and robustness is given. Baseline parameter values for the model presented in this chapter are provided in Chapter 5, Section 5.2. A complete list of parameters, values, initial conditions and values used for various experimental conditions is provided in Appendix B.

Chapter 5

TLT Model and Simulation Analysis

The previous chapter described the platform model and implementation specification for TLT formation. The simulation can now be calibrated against experimental data, utilised to test our initial hypotheses regarding formation mechanisms, evaluate the effect of biologics and other therapeutic interventions, and perform other explorative experimentation. We present the simulation results at key time-points compared to TLT formation observed in the in vivo *model, and apply IHC emulation with high-throughput image analysis to determine how perturbing the model affects TLT morphology. The effect of changing the rate at which stroma and lymphocytes upregulate adhesion molecules and their cognate receptors is explored against TLT morphology (aggregation factor), area, number of follicles formed using CellProfiler (Broad Institute) for high-throughput analysis of emulated confocal IHC micrographs.*

We apply Kohonen networks and other machine learning approaches to simulation outputs, and explore the possibility of using the Kohonen network 'Unified Distance Matrix' variance to determine the appropriate number of repeats required to adequately ameliorate stochastic effects arising from aleatory uncertainty.

5.1 Confirming TLT Formation Hypothesis Veracity

A key test of our TLT formation hypothesis is whether the model conforms to expectations from *in vivo* data. This is explored in Section 5.1.1 to determine if the model is consistent with TLT formation.

5.1.1 Calibration Against Expected Behaviours

To illustrate conformity to biological observations, a representative simulation run visualisation, applying the techniques developed during the Peyer's patch case study discussed in Section 2.2.2, is compared with the key observable phenomena determined from the murine model in the 'expected behaviours' diagram described in Section 3.2.2. Figure 5.1 clearly shows that the calibrated simulation reproduces the behaviour over time of TLT formation. The 'observable phenomena' described in Figure 5.1 are domain-expert descriptions of the acute TLT formation process informed from a murine model, and the images are taken from a typical simulation run, illustrate that the simulation developed from the theoretical computational model described in Chapter 4 produces outputs that are consistent with these descriptors.

For the purposes of this comparison, the stromal cell distribution over the simulation space was limited so that only one follicle is likely to emerge. Therefore Section 5.1.4 looks at a large sample of simulation runs analysed using high-throughput image analysis software, to further validate the hypothesis veracity by assessment of high-dimensional results model data using Kohonen (self-organising) maps.
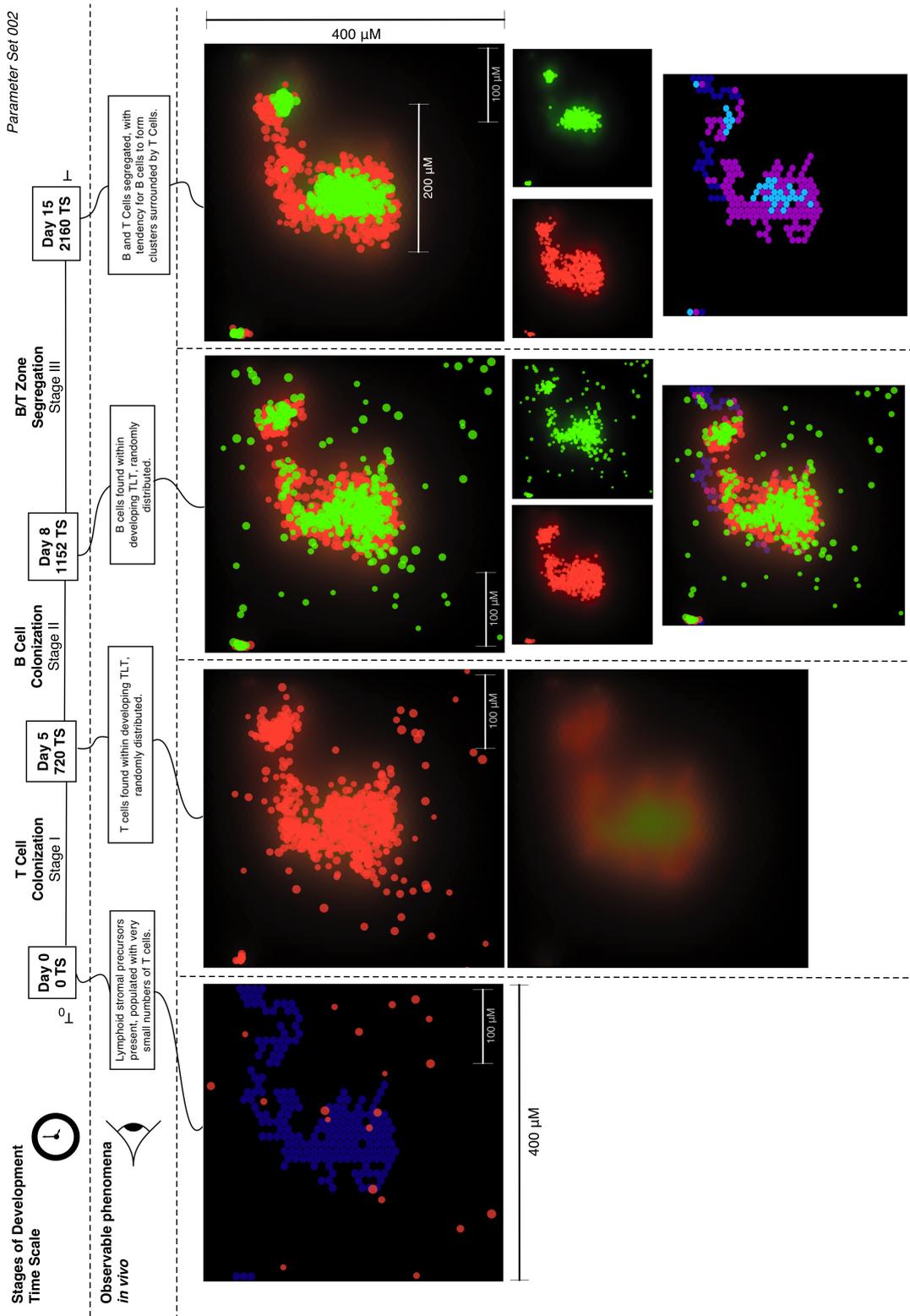
Figure 5.1: Baseline simulation output vs. expected behaviours diagram for TLT formation. This figure demonstrates that tissue development in the simulation outputs at key observable phenomena time-points matches well what is expected and observed experimentally (e.g. in the inducible TLT murine model described by **?**). Note that all data linked to the observed *in vivo* descriptions are simulation outputs from NeoSim.

5.1.2 *In Silico* Image Analysis

As discussed in Section 2.1, emulating experimental techniques can be useful beyond simulation analysis, visualisations may be helpful for communicating results models to those in fields with little exposure to mathematical and computational modelling, and to provide insights into how a simulation result equates to the domain under study. Figure 5.2 presents a confocal micrograph of a TLO compared to a baseline simulation result that has been processed using the techniques described in Section 2.2.2. The principal output, as discussed in Section 4.8.1, is whether the simulation implementation of the theory of TLT formation developed in Chapter 3 and formalised as a mathematical and computational model in Chapter 4, is whether structures form over 15 days of simulated time that are qualitatively similar to those observed *in vivo*. The restriction to qualitative or semi-quantitative outputs in this sense is due to the lack of existing data describing the variation of TLT *in vivo*. However, there are clear criteria established in the model expected behaviours diagram from Figure 3.8 under 'expected behaviours'. These were compared to observed simulation behaviours in Figure 5.1. In terms of developed TLT, the key observables of interest are that dense B-cell follicles form, supported by FDC-like stroma, surrounded by T cells randomly distributed over the FRC-like stromal network. This can clearly be seen to be the case in Figure 5.2 where a typical simulation run is shown against an IHC micrograph of murine TLT. The simulation output here has been processed using the methodology described in Section 2.2.2.1.

The principal research question driving this model is determining the veracity of our hypothesis regarding the minimum requirements to sustain TLT development. Figure 5.3 shows how multi-dimensional TLT morphology data are obtained through analysis with CellProfiler (software developed by the Broad Institute, MA, USA. **?**) comprising TLO and lymphoid follicle number, area, compactness, perimeter and solidity. Figure 5.4 shows the U-matrix for 100 baseline simulation runs of TLT formation, trained on these morphology metrics.
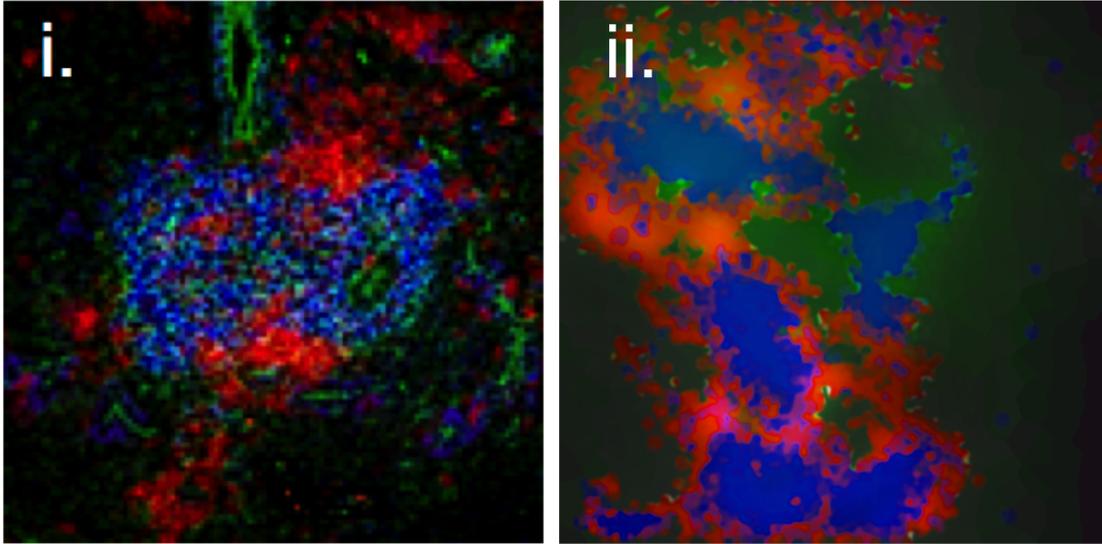
Figure 5.2: **i.** Actual confocal microscopy image of a fluorescent antibody stained murine TLO in a salivary gland. **ii.**Emulated immunohistochemistry micrograph from a simulation at 15 days post-TLT induction. This process is based on techniques discussed in Chapter 2, this figure is analogous to Figure 2.13 for the Peyer's patch simulation case study using the methodology described in Section 2.2.2.1. Note that this figure compares a 3D section of murine submandibular tissue against a 2D simulation, and as such less T cells are visible surrounding the *in vivo* structure.
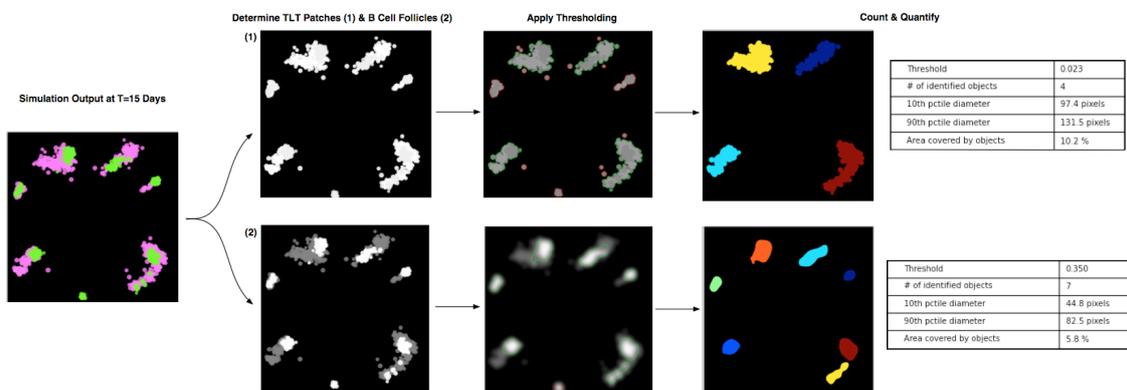
Figure 5.3: Example Cell Profiler analysis of the TLT simulation, segregating the TLOs and lymphoid follicles and measuring many parameters including area, perimeter, population size and compactness.

5.1.3 Introducing Kohonen maps as a means of simulation results interpretation

Kohonen networks, otherwise known as self-organising maps (SOM), are a form of perceptron-based neural network model introduced in (**?**) that utilises unsupervised

learning to produce a low-dimensional representation of a higher dimensional input
dataset which preserves the underlying topological properties of the data, permit-
ting clustering of data samples on a 2D map which approximates the state space of
the input, and visualisation of the distribution of individual output measures over
the state space. Unsupervised learning refers to machine learning algorithms that
process unlabelled data sources, such that they do not know, for example, whether
any particular 'input vector' is a control sample or a test sample, this permits identi-
fication of hidden topological data structures. A Kohonen network is constituted of
an array of nodes which constitute an abstract representation of a biological neuron,
each associated with a location co-ordinate in the map space and a weight vector
with dimensionality equal to that of the input data vectors. The map itself has the
form of a discretised grid, these may be rectangular or hexagonal, although only
hexagonal maps are used for the purposes of this thesis.

The process of mapping input vectors to the SOM is termed vector quantization,
a simplified description is provided here; for the interested reader, a full description
of this process and variations thereof may be found in **?**. The weight vectors for each
node are initialised, usually randomly or linearly. An input vector is selected, and
the node with the shortest Euclidean distance (most similar) weight vector identified.
The weight vectors for all nodes within the neighbourhood of the identified most
similar node are adjusted toward the input vector. The neighbourhood is defined
as the six surrounding hexagonal elements for a hexagonal network (either the Von
Neumann or Moore neighbourhoods of 4 or 8 nodes may be used with rectangular
grids). This process is repeated over all input vectors or 'training data', and applied
iteratively until sufficient convergence is achieved between the Kohonen network
topology and the topological structure of the set of input vectors. It is important to
note that the discrete spatial representation of the nodes is non-linear such that the
Euclidean distance between adjacent nodes is different for each node; two apparently
similar nodes may be separated by a large distance. This property permits the use
of unified-distance matrices, in which the colour or shading of each node is mapped

to its 'codebook vector' (Euclidean distance over the node), this is a useful means of visualising clustering of input vectors when mapped to the SOM: if the level of grey for each node maps to the codebook vector, then large boundaries separating clusters will be visible as dark lines that separate lighter-coloured regions. This property is exploited in Section 5.3 to propose a novel means of quantifying and ameliorating aleatory uncertainty in stochastic simulations.

The Kohonen networks presented in the following sections and in Chapter 6 were created using the open source python package and front-end Orange (version 2.4, University of Ljubljana). Full details of how the package implements Kohonen networks can be found at the Orange library documentation for Self-Organising Maps (available at `https://docs.orange.biolab.si/2/reference/rst/Orange.projection.som.html`). The formal mathematical derivation of Kohonen networks from which they were implemented in this package is presented in **?**, and the source code from which the Kohonen networks were generated is available at the Orange version 2 Git repository (`https://github.com/biolab/orange`)

Visualising Simulation Outputs Over Parameter Perturbation

The usual motivation for use of Kohonen networks is as an unsupervised classifier, such that the network is 'trained' with a sufficient number of input vectors (in our case, individual simulation results), building the map through competitive learning, this then permitting classification of unknown vectors to clusters of known samples based on their similarity to the weight vectors of nodes. Thus, the process is divided into modes: 'training' and 'mapping'. However, for the purposes of evaluating the effect of simulation parameter perturbation and therapeutic intervention with *in silico* disease models, the motivation for utilising Kohonen networks is for their capacity to provide visualisations of efficacy surrogate outputs over the trained map of high-dimensional simulation results. Such maps crucially permit quantification of the level of stratification between specific simulation sample sets, visualisation of the differences in topological structure of high-dimensional inputs, and simultaneously,

visualisation of output values of interest such as surrogates of disease activity. A map visualising specific model outputs over the nodes and input vectors is known as a 'component plane', as individual values of one component from the input vectors are mapped to the network. In combination with mapping input vectors, this presents a powerful tool for evaluating the effect of perturbing a parameter while taking into consideration the topological structure arising both due to parameter perturbation and stochastic (aleatory) effects.

5.1.4 Stochastic Variation in TLO Development

The lack of clustering occurring in Figure 5.4 illustrated by the mostly white colouring of the hexagonal elements indicates that the results observed are not due to aleatory uncertainty. The plotted circles on the U-matrix indicate the average number of TLOs that formed in a simulation, and the radius of the circle indicates the number of simulation runs that fall into that location within the U-matrix. It can be seen that the vast majority of results show one to three TLO structures after 15 days, with very few outlying results in which more disperse lymphoid aggregates developed. Means of interpreting features of Kohonen networks are provided herein as necessary where they are utilised, however the unfamiliar reader is referred to ? for a useful and more complete summary of their interpretation generally. Note that the grey-scale in each node (hexagon) is proportional to the codebook vector distance (similarity metric, described above in Section 5.1.3) across that node, and therefore the apparent regularity in spacing on the following figure does not suggest all samples are equally similar. There are clear outliers in the bottom left of the figure, and it can be seen that the Kohonen network has stratified samples very clearly in proportion to the number of TLT structures formed, as identified using CellProfiler.
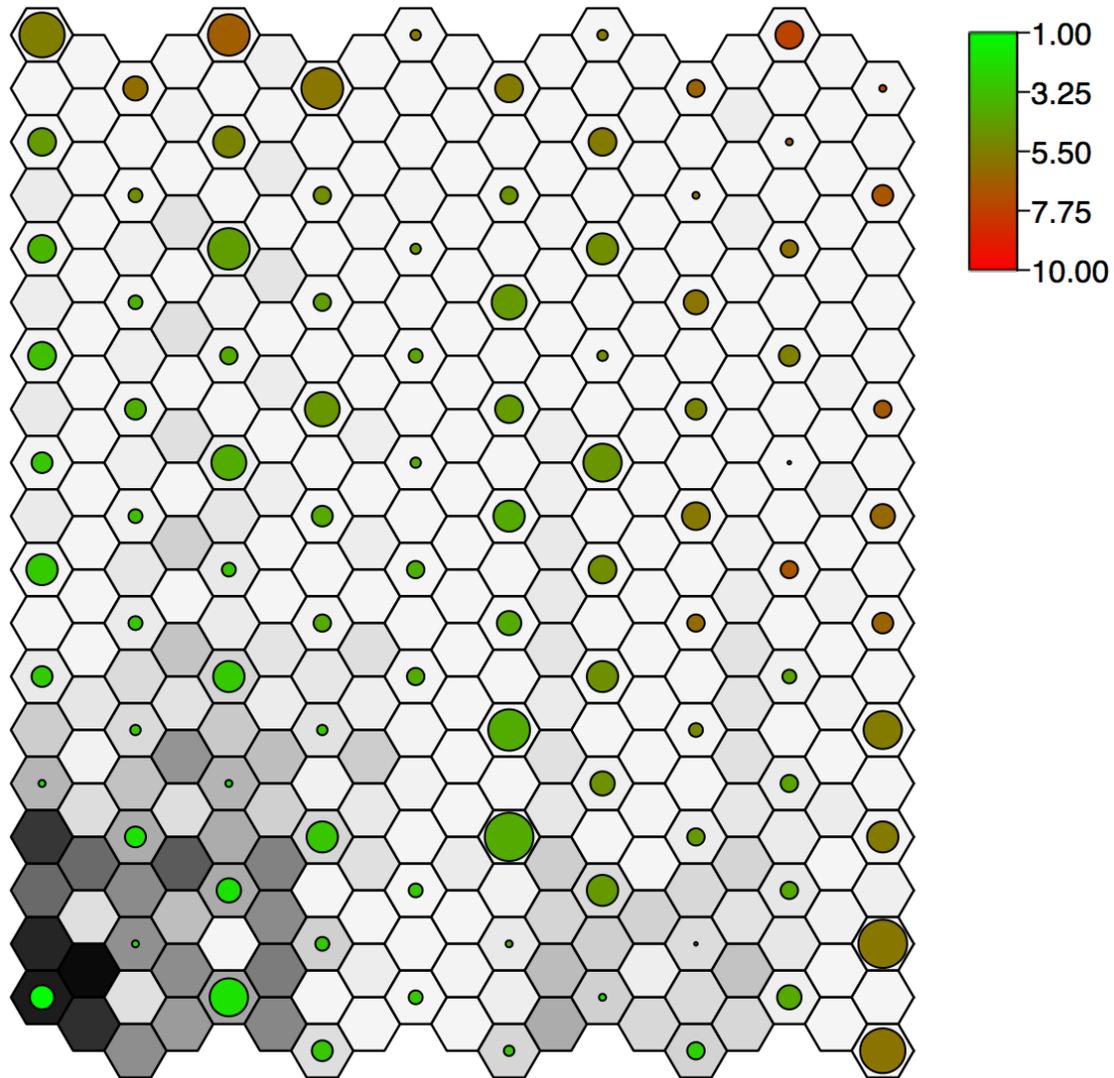
Figure 5.4: U-Matrix for 100 simulation runs of the TLT model, with shading from green to red indicating the number of TLOs that formed.

5.2   Platform Model Parameter Values

This section provides a complete list of all platform model parameters, a description of their purpose, the baseline value and initial conditions used, and an indication of the source of each parameter value. Those that are marked 'Calibrated to Expected Behaviours' were perturbed by hand such that simulated TLT development conformed to the 15 day process outlined above in Section 5.1.1.

Parameter (NeoSim Name)	Model Entity	Description	Value	Determined
tCellBegin	Simulation Initial Conditions	Agent time-step at which T cells begin to enter tissue compartment	0	*In Vivo - Bombardieri (2012)*
tCellCease	Simulation Initial Conditions	Agent time-step at which T cells cease to enter tissue compartment	720	*In Vivo - Bombardieri (2012)*
tCellInputPopulationSize	Simulation Initial Conditions	Number of T cells to be added to simulation in total, via linear increment between tCellBegin and tellCease.	8000	Arbitrary Input
bCellBegin	Simulation Initial Conditions	Agent time step at which B cells begin to enter tissue compartment	720	*In Vivo - Bombardieri (2012)*
bCellCease	Simulation Initial Conditions	Agent time step at which B cells cease to enter tissue compartment	1152	*In Vivo - Bombardieri (2012)*
bCellInputPopulationSize	Simulation Initial Conditions	Number of B cells to be added to simulation in total, via linear increment between bCellBegin and bCellCease.	8000	Arbitrary Input
elementsPerCell	Simulation Initial Conditions	(Generative Grammar, '*B*') Number of hexagonal grid elements to place per branch stromal cell	24	Calibrated by hand via confocal IHC imaging
fractionOfGridToCover	Simulation Initial Conditions	(Generative Grammar) To determine total number of Stromal Cells to place in GG; expressed as product of total grid area to hold stromal cells	1.4	Calibrated by hand via confocal IHC imaging
beginNewNetwork	Simulation Initial Conditions	(Generative Grammar, '*N*'), number of connected stromal cells to place before new branch, must be multiple of sqrt(n)	6--12 (12)	Calibrated by hand via confocal IHC imaging
beginNewCell	Simulation Initial Conditions	(Generative Grammar, 'n') Number of connected branches defined as 1 stromal cell.	4--24 (12)	Calibrated by hand via confocal IHC imaging
initialCCL19Expression	Simulation Initial Conditions	Inflammatory trigger for T-Cell Chemotaxis	6	Fig 3.3 *In Vitro* Data

Figure 5.5: Table of 'System-wide' parameters including those defining model initial conditions.

Parameter (NeoSim Name)	Model Entity	Description	Value	Determined
maxChemokinePerElementB	CXCL13	Maximum concentration of chemokine in each discrete element of the CXCL13 hexagonal grid.	30000	Dummy parameter
diffusionCoefficientB	CXCL13	Diffusion Coefficient, D, for Chemokine PDE (CXCL13)	4E-14	Calibrated to Expected Behaviours*
bccaDecayConstant	CXCL13	Decay Constant, Lamda, for Chemokine PDE (CXCL13)	0.0005	Calibrated to Expected Behaviours
sufficientBChemokineThreshold	CXCL13	Minimum concentration of chemokine to which a B cell will respond chemotactically	0.02	Arbitrary low value required, Prevents chemotaxis with negligable presence of chemokine
maxChemokinePerElementT	CCL19	Maximum concentration of chemokine in each discrete element of the CCL19 hexagonal grid.	30000	Dummy parameter
diffusionCoefficientT	CCL19	Diffusion Coefficient, D, for Chemokine PDE (CCL19) -- Einstein-Stokes would be 0.000000000158.	4E-14	Calibrated to Expected Behaviours*
tccaDecayConstant	CCL19	Decay Constant, Lamda, for Chemokine PDE (CCL19).	0.0005	Calibrated to Expected Behaviours
sufficientTChemokineThreshold	CCL19	Minimum concentration of chemokine to which a B cell will respond chemotactically	0.02	Arbitrary low value required, Prevents chemotaxis with negligable presence of chemokine

Figure 5.6: Table of Chemokine model-related parameters. The chemokine diffusion parameter value was initially determined using the Einstein-stokes relationship, but this was found to be substantially higher than suitable for model expected behaviour, and the values had to be reduced by 2 orders of magnitude. *Recent evidence from the lab (unpublished) has confirmed that the Einstein-Stokes relationship is not applicable for chemokine diffusion *in vivo* due to protein interactions with the stromal network.

Parameter (NeoSim Name)	Model Entity	Description	Value	Determined
decayConstantInflammation	Stromal Cell	Decay constant for capacity of stromal progenitors to differentiate into a lymphoid stroma phenotype (lower value = more lymphoid stroma)	0.001	Calibrated to Expected Behaviours development timeline
probabilityS1S2PerTimeStep	Stromal Cell	Probability that any S1 becomes S2 each time step.	0.05	Calibrated to Expected Behaviours
inflammationLevel	Stromal Cell	*a priori* Probability that any S1 can become S2. This abstracts presence of inflammatory cytokines.	0.95	Calibrated to Expected Behaviours
bccaS3Maximum	Stromal Cell	Maximum number of cxcl13 molecules secreted per time step when S3 (Feature scaled).	0.2	Fig 3.3 *In Vitro* Data
tccaS4Maximum	Stromal Cell	Maximum number of tcca molecules secreted per time step when S4 (Feature scaled).	0.3	Fig 3.3 *In Vitro* Data
tccaS3Maximum	Stromal Cell	Maximum number of tcca molecules secreted per time step when S3 (Feature scaled).	1	Fig 3.3 *In Vitro* Data
tccaS2Maximum	Stromal Cell	Maximum number of tcca molecules secreted per time step when S3 (Feature scaled).	0.05	Fig 3.3 *In Vitro* Data
upregulationRateBCCA	Stromal Cell	CXCL13 upregulation rate stroma is in contact with lymphocyte (Feature scaled).	1.01	Rate assumed to be Linear
upregulationRateTCCA	Stromal Cell	CCL19 upregulation rate when stroma is in contact with lymphocyte (Feature scaled).	1.01	Rate assumed to be Linear
downregulationRateBCCA	Stromal Cell	CXCL13 down-regulation rate when stroma is in contact with lymphocyte (Feature scaled).	0.99	Rate assumed to be Linear
downregulationRateTCCA	Stromal Cell	CCL19 down-regulation rate when stroma is in contact with lymphocyte (Feature scaled).	0.99	Rate assumed to be Linear
probabilityS1S2	Stromal Cell	Probability that environmental factors cause S1 stroma to differentiate into S2 (Feature scaled).	1	Calibrated to Expected Behaviours development timeline
probabilityS2S3	Stromal Cell	Probability that environmental factors cause S2 stroma to differentiate into S3 (Feature scaled).	0.0004	Calibrated to Expected Behaviours development timeline
probabilityS3S4	Stromal Cell	Probability that environmental factors cause S3 stroma to differentiate into S4 (Feature scaled).	0.000005	Calibrated to Expected Behaviours development timeline
bccaMaxMoleculesPerStep	Stromal Cell	Number of absolute CXCL13 molecules secreted per time step per hexagonal element.	10000	Only for outputting absolute values, ODE model is relative & feature scaled.
tccaMaxMoleculesPerStep	Stromal Cell	Number of absolute ccl19/21 molecules secreted per time step per hexagonal element.	10000	Only for outputting absolute values, ODE model is relative & feature scaled.
initialAdhesionLevel	Stromal Cell	Percentage of maximum adhesion molecule expression at T0	0	Assumed that mesenchymal progenitors do not express adhesion factors prior to stimulation by lymphocytes
adhesionLevelIncrement	Stromal Cell	Upregulation increment of adhesion factor expression following lymphocyte stimulation.	0.01	Calibrated to Expected Behaviours development timeline
adhesionLevelThreshold	Stromal Cell	Threshold on adhesionLevel above which adhesion is possible.	0.65	Calibrated to Expected Behaviours development timeline

Figure 5.7: Table of Stromal Cell-related parameters.

Parameter (NeoSim Name)	Model Entity	Description	Value	Determined
initialPopulationSizeBCell	B Cell	Number of B cells present at T_0.	0	*In Vivo - Bombardieri (2012)*
chemokineConsumptionB	B Cell	True if B Cell Receptor Recycling ODE Model Enabled.	TRUE	(Receptor recycling ODE On/Off flag)
synthesisRateBCCA	B Cell	Specifies linear rate of synthesis of B-cell chemokine receptor.	3	Calibrated to Expected Behaviours
bChemotaxisWeighting	B Cell	Option to weight probability of B-cell motion direction of highest B-cell chemokine value by this factor. Abstraction of cellular persistence.	1.2	Calibrated to Expected Behaviours
maxBCCARMolecules	B Cell	Maximum number of B-cell chemokine receptor molecules on surface.	1000000000	Not used in model logic (but measured approximately in the lab)
minRadiusB	B Cell	Minimum B cell radius as proportion of one hexagon (10um^2).	0.625	Measured from *In Vivo* Micrographs
maxRadiusB	B Cell	Minimum B cell radius as proportion of one hexagon (10um^2).	0.375	Measured from *In Vivo* Micrographs
receptorInternalizationScalarB	B Cell	Value in range [0,1] defines the proportion of chemokines are internalised on average (steady state).	0.0001	Calibrated to Expected Behaviours
receptorRecyclingRateB	B Cell	Rate at which CXCL13 receptors are recycled back to cell surface.	0.8	Calibrated to Expected Behaviours
maximumRecyclingCacheB	B Cell	Maximum size of intracellular pool (range [0,1]).	0.75	Calibrated to Expected Behaviours
receptorProportionRecycledB	B Cell	Proportion of CXCL13 receptors internalised that are recycled.	0.5	Calibrated to Expected Behaviours
minimumAdhesionTimeB	B Cell	Minimum amount of time (in 1/10th time-steps) a stable B cell-Stroma bond can form for.	5	Value from *Alden (2012)*
initialChemokineReceptorExpressionB	B Cell	Expression of CXCL13 by B Cell at T_0 (Range [0,1])	1	Assumed lymphocytes begin with maximum expression of chemokine receptors
initialPopulationSizeTCell	T Cell	Number of T cells present at T_0.	50	Estimated number of tissue resident T cells prior to inflammatory signal
chemokineConsumptionT	T Cell	True if T-cell Receptor Recycling ODE Model Enabled.	TRUE	Flag to enable/disable internalisation ODE
synthesisRateTCCA	T Cell	Specifies linear rate of synthesis of T-cell chemokine receptor.	3	Calibrated to Expected Behaviours
tChemotaxisWeighting	T Cell	Option to weight probability of T-cell motion direction of highest B chemokine value by this factor. Abstraction of cellular persistence.	1.2	Calibrated to Expected Behaviours
maxTCCARMolecules	T Cell	Maximum number of B cell chemokine receptor molecules on surface.	1000000000	Not used in model logic (but measured approximately in the lab)
minRadiusT	T Cell	Minimum B cell radius as proportion of one hexagon (10um^2).	0.625	Measured from *In Vivo* Micrographs
maxRadiusT	T Cell	Minimum B cell radius as proportion of one hexagon (10um^2).	0.375	Measured from *In Vivo* Micrographs
receptorInternalizationScalarT	T Cell	Value in range [0,1] defines the proportion of chemokines are internalised on average (steady state).	0.0001	Calibrated to Expected Behaviours
receptorRecyclingRateT	T Cell	Rate at which CXCL13 receptors are recycled back to cell surface.	0.8	Calibrated to Expected Behaviours
maximumRecyclingCacheT	T Cell	Maximum size of intracellular pool (range [0,1]).	0.7	Calibrated to Expected Behaviours
receptorProportionRecycledT	T Cell	Proportion of CXCL13 receptors internalised that are recycled.	0.5	Calibrated to Expected Behaviours
minimumAdhesionTimeT	T Cell	Minimum amount of time (in 1/10th time-steps) a stable B cell-Stroma bond can form for.	5	Value from *Alden (2012)*
initialChemokineReceptorExpressionT	T Cell	Expression of CXCL13 by B Cell at T_0 (Range [0,1])	1	Assumed lymphocytes begin with maximum expression of chemokine receptors

Figure 5.8: Table of Lymphocyte-related parameters.

5.3   Quantifying Aleatory Uncertainty using Kohonen Networks

Section 1.4.1.3 discussed a method by which the aleatory uncertainty introduced by stochastic simulation events may be ameliorated, such that observed effects may be attributable to a change in parameter values rather than random fluctuations. These techniques were either computationally intensive and limited in scope with regard to results measured, or assumed a normal, or other known, distribution. This can be an issue if model outputs are not known to be independent random variables, in which case the central limit theorem would ensure normality (**?**), or otherwise if no known distribution can be identified for an output. The distribution of outputs in the TLT formation simulation are unknown, thus alternative means of ensuring results are reliable have been sought. Kohonen maps may be considered a non-linear generalisation of principal component analysis (PCA). The data analysis work-flow, whereby baseline samples of 10, 25, 50 and 100 simulations are evaluated using SOMs is shown in Figure 5.9. The reduction in clustering and increasing uniformity of samples as the number of simulations, n, increases, is demonstrative of a reduction in aleatory uncertainty. The reduction in the variation of grey in the hexagonal nodes indicates increasing uniformity of codebook vectors, and an increase in similarity between each simulation run.

Kohonen maps are based on iteratively stratifying results into a similarity neighbourhood within a 2D grid using a perceptron network utilising competitive learning. These prove invaluable in Chapter 6 in identifying clusters when therapeutic interventions are applied to simulation results. However, the U-Matrix, or unified distance matrix, wherein the 'codebook vector' Euclidean distance between simulation results which broadly describes their similarity while preserving overall data structure topology, is displayed as a grey-scale level on a 2D map such that clusters may be identified with samples that are far apart. This can be used to identify the point at which the number of samples of the same parameter value is sufficiently
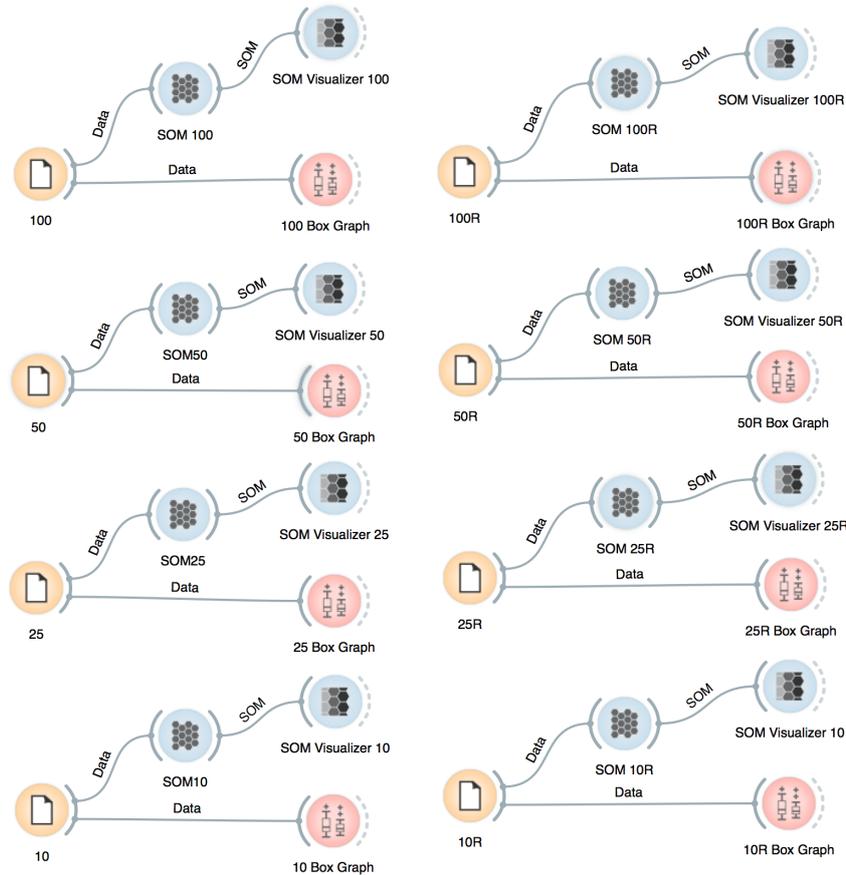
Figure 5.9: Pipeline for Aleatory Analysis of 10, 25, 50 and 100 simulation samples at baseline using the Orange toolkit in the Python programming language. The inputs are files containing either absolute and relative stromal phenotype data (left), or relative data alone (right), the outputs of this analysis are a SOM visualisation tool or a box-plot graph generator for attributes within the dataset. For the Kohonen map, linear initialisation (opposed to standard random initialisation) is used to ensure comparability between samples, and 1000 iterations are applied for each sample set, with a Gaussian neighbourhood, hexagonal topology and initial radius of 10 and final radius of 1 in a 10x10 grid.

high that aleatory uncertainty can be considered ameliorated. Figure 5.10 shows the U-matrix for 10, 25, 50 and 100 simulation runs using identical baseline parameter values.

A key benefit of this approach is that the user can sequentially increase the number of samples taken one at a time until the U-matrix reaches pre-specified acceptable levels, and rather than functioning on just a few outputs, considers the entirety of the multidimensional results model simultaneously. The principal nu-
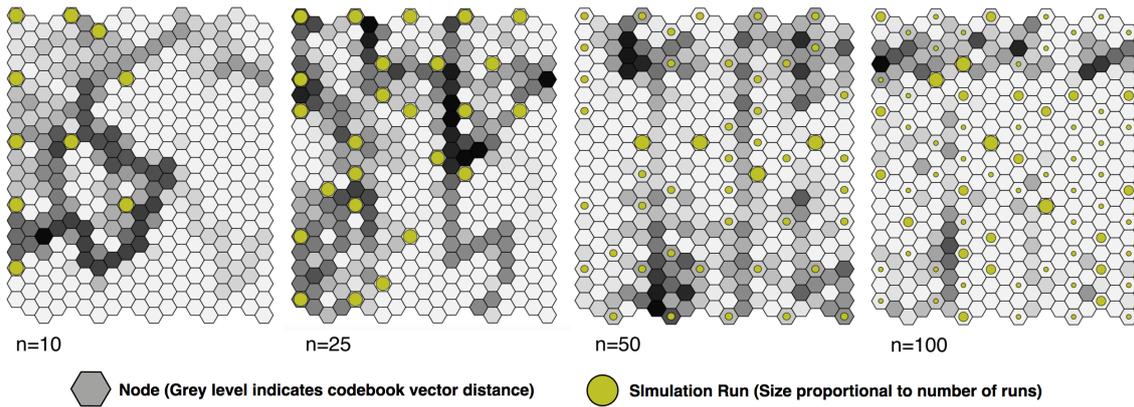
Figure 5.10: Simulation results model Kohonen map unified codebook vector distance matrix (U-matrix) for n=10,25,50,100. It can be seen significant clustering occurs in the n=10 sample with respect to the n=100 sample.

merical outputs of the TLT formation simulation are the population sizes of the four stromal phenotypes described by the Markov Chain Monte Carlo model in Section 4.7.3. These are expressed as absolute population sizes based on the number of elements of the hexagonally-packed stromal network grid elements occupied, however the precise size of this varies between simulations. Therefore, one could conduct an aleatory analysis on the absolute values, or normalise the values by measuring relative percentages of occupied area. We note that normalising the data prior to analysis results in a significant improvement in reducing codebook vector distances between sets of identical simulation samples, as illustrated in Figure 5.11, in which there is a significant codebook vector distance between certain subsets despite identical simulation parameters when using both absolute and normalised data – this effect is not present with only non-normalised data, presumably due to stochastic variation in the number of stromal cells in each simulation run. Therefore, when interpreting aleatory uncertainty, an important consideration is whether the apparent aleatory uncertainty is due to stochasticity-induced chaotic effects that impact simulation dynamics, or an artefact of differing absolute quantities between identical samples that may be mitigated through a simple normalisation process. Therefore, it is proposed that U-matrix aleatory uncertainty analysis is performed using data normalised across simulation runs at baseline to determine the appro-

priate number of runs to mitigate stochastic effects, and on a per experiment basis, non-normalised data should be filtered from the multidimensional datasets prior to U-matrix generation for determination of aleatory effects. Further development in this area would lead to a robust means of understanding the role of stochasticity and aleatory uncertainty within complex systems ABMs, rather than the common approach of 'mitigating' aleatory uncertainty by simply measuring the medians of simulation outputs over many runs. I argue that visualising the heterogeneous response of the system outputs is important, as the probabilistic guards implemented in ABMs are an abstraction of the stochastic events within the biological domain that leads to the large variance we see in the biological domain, in everything from individuals' responses to drugs to the enormous variety of life. Kohonen networks, or other machine learning methods for dimensionality reduction that can reduce and stratify a higher dimensional manifold into a 2D plot over which model outputs can be overlaid, provide a vital means of investigating this uncertainty, and, in the context of QSP, can provide a means of predicting *a priori* the likely rate of efficacy of therapeutics or expected heterogeneity within biological systems generally.

Normalising outputs over many identical simulation samples eliminates aleatory influences of differing population sizes when comparing different parameter samples, yet the resulting distribution remains unchanged – normalisation is a linear rescaling of the population sizes. A box-plot of absolute FDC population size and the percentage of all stroma with an FDC phenotype is shown in Figure 5.12; the dots represent values from individual simulation runs, the effect of normalisation can be seen on the modified distribution of individual samples, however there is no significant difference in the overall distribution in terms of relative mean, medians and quartiles (the distribution difference tends to zero as the limit of the number of samples is taken to infinity).

Absolute Stromal Phenotype Values   Normalised Stromal Phenotype Values
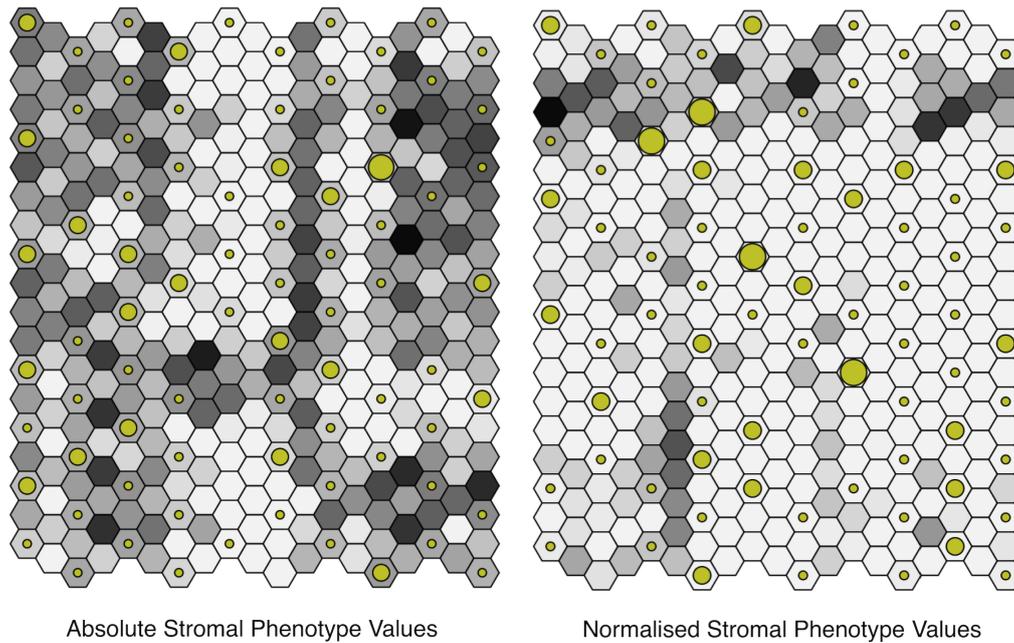
Figure 5.11: U-Matrix for 100 simulation values using both absolute and normalised multidimensional data (left) and normalised data (right) such as ratios and percentages only. It is clear that the absolute value data introduces considerable noise into the simulation results due to differences in individual simulations.

5.4 The Potent Role of Adhesion Molecules in Regulating TLO Morphology

The role of adhesion molecules in TLT induction was explored by varying the 'Adhesion Molecule Up-regulation Rate' parameter from very low to high values, this defines the rate of increase in VCAM-1 and ICAM-1 expression in response to lymphocyte stimulation. Three simulation output measures were used in this experiment, these were the median number of TLT patches observed within the simulation space, the total area covered by lymphocytes (TLT area), and a new metric termed 'TLT Aggregation Factor'. The TLT aggregation factor is a metric defined by normalising the total area of the simulation space ($400\mu m^2$) occupied by TLT, such that the lowest observed area occupied by TLT is defined as 1 and the most highly is defined as 0; given that there is no exit mechanism for lymphocytes during the simulation, those with the lowest occupied areas represent the densest aggregated
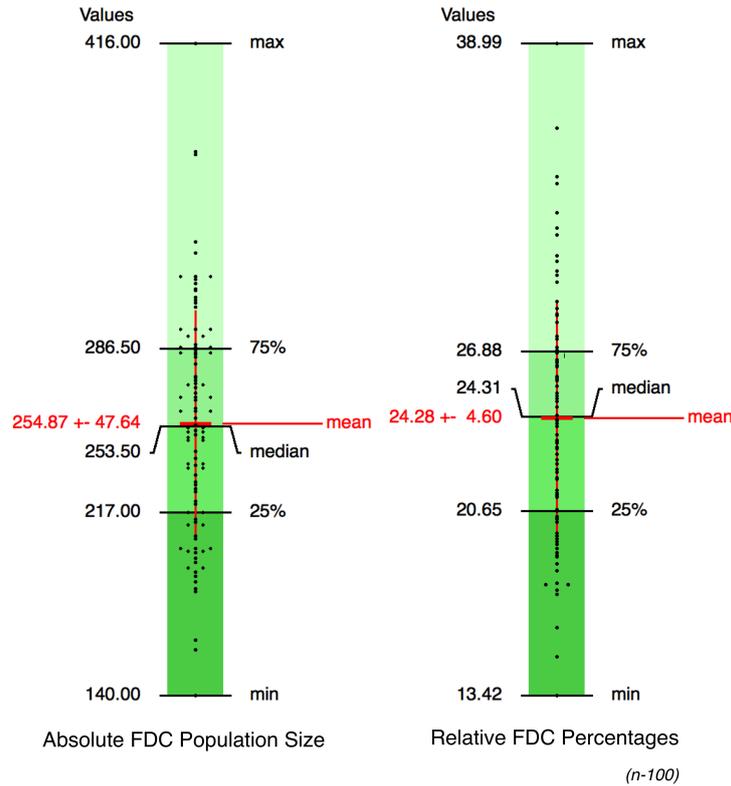
Figure 5.12: Comparison of absolute FDC stromal cell grid elements (left) vs percentage of stroma with FDC phenotype, derived from 100 baseline samples. The distributions are identical, however the Kohonen map U-matrix data shown in Figures 5.10 and 5.11 are markedly different when using only relative data.

lymphoid tissue.

Figure 5.13 shows a box and whisker plot illustrating the change in TLT aggregation factor as the rate of adhesion molecule up-regulation increases. There is a clear relationship, showing an exponential increase in aggregation as the adhesion molecule expression level increases (note the x-axis is non-linear). From this, it is reasonable to conclude that lymphocytes do indeed require stromal cell expression of adhesion molecules to properly self-organise. If lymphocytes cannot easily adhere to the stromal network, then lymphocyte-stroma crosstalk is significantly reduced, preventing development of lymphoid stroma phenotypes and induction of the chemokine feedback loops responsible for driving organisation. Significance in Figure 5.13 was confirmed using One-way ANOVA (Analysis of Variance) in combination with Tukey's range test, confirming that the rise in aggregation is statistically significant until the plateau following X=0.001, the analysis summary table is shown
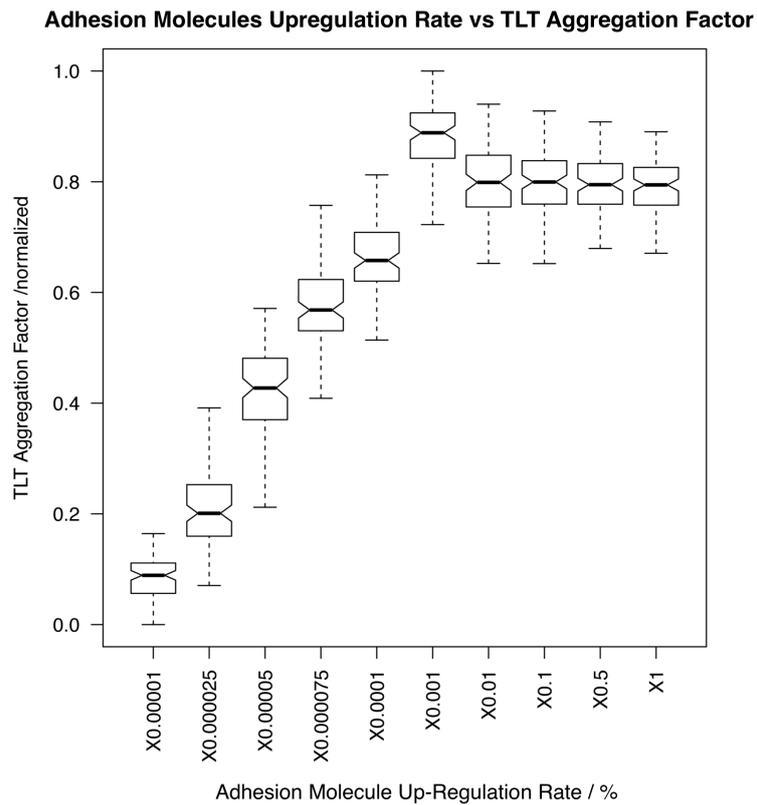
Figure 5.13: Boxplot of Adhesion Molecule Upregulation (X-axis) Rate vs TLT Aggregation Factor. There is a rapid, non-linear increase in observed aggregation as the adhesion molecule up-regulation rate is increased. All differences are statistically significant between each other (adjusted $p < 2e16$) from X=0.0001 to 0.001), however values 0.01 to 1 are not statistically significant between themselves. This indicates that adhesion molecule expression has a significant effect until a threshold is reached, at which point the degree of tissue formation plateaus. *(n=100)*.

	Df	Sum Sq	Mean Sq	F Value	Pr (>F)	
Ind	9	68.04	7.559	1788	<2e16	***
Residuals	990	4.19	0.004	-	-	

Table 5.1: Tukey multiple comparisons of means 95% family-wise confidence level for Figure 5.13
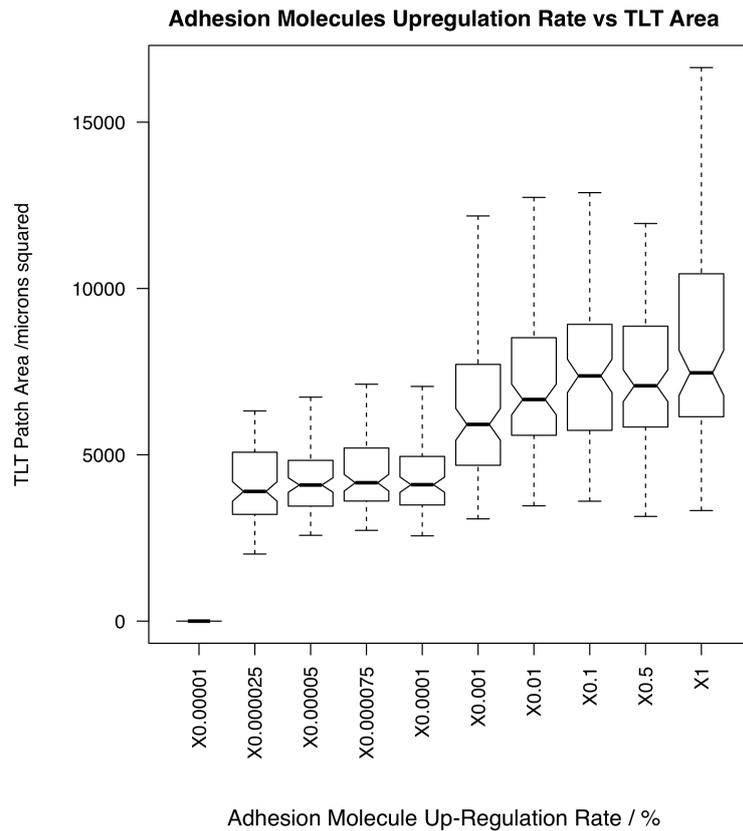
Figure 5.14: Box-plot of Adhesion Molecule Upregulation Rate vs TLOs formed per 40x40 micron simulation space, illustrating the change in the number of discrete TLOs or tertiary lymphoid tissue aggregates forming as the rate of adhesion molecule up-regulation increases. n=100.

in Table 5.1. X in these plots refers to 'Adhesion Molecule Upregulation Rate', called 'adhesionLevelIncrement' in the stroma parameters table in Figure 5.7.

Figure 5.14 demonstrates the change in tissue surface area vs adhesion molecule up-regulation rate. This graph indicates that there is a minimum requirement for the

	Df	Sum Sq	Mean Sq	F Value	Pr (>F)	
Ind	9	6.648e9	738675885	70.33	<2e16	***
Residuals	990	1.040e+10	10502671	-	-	

Table 5.2: Tukey multiple comparisons of means 95% family-wise confidence level for Figure 5.14.

Parameter Value	p-value (adj)	Significance
X1-X0.001	1	None
X0.5-X0.01	0.999	None
X0.1-X0.01	0.999	None
X1-X0.1	0.999	None
X0.1-X0.001	0.999	None
X0.0001-X0.000075	0.999	None
X0.5-X0.1	0.998	None
X1-X0.01	0.997	None
X0.01-X0.001	0.984	None
X1-X0.5	0.965	None
X0.0001-X0.00005	0.955	None
X0.5-X0.001	0.898	None
X0.5-X0.000075	0.878	None
X0.01-X0.000075	0.657	None
X0.000075-X0.00005	0.555	None
X0.5-X0.0001	0.389	None
X0.1-X0.000075	0.358	None
X0.01-X0.0001	0.181	None
X1-X0.000075	0.144	None
X0.001-X0.000075	7.721E-2	None
X0.1-X0.0001	5.875E-2	None
X0.00005-X0.000025	3.809E-2	*
X1-X0.0001	1.47837E-2	*
X0.5-X0.00005	1.250E-2	*
X0.001-X0.0001	6.217E-3	*
X0.01-X0.00005	2.963E-3	**
X0.1-X0.00005	4.827E-4	**
X0.0001-X0.000025	2.524E-4	**
X1-X0.00005	6.490E-5	***
X0.001-X0.00005	1.970E-5	***
X0.000075-X0.000025	5.599E-6	*****
X0.000025-X0.00001	0	*****
X0.00005-X0.00001	0	*****
X0.000075-X0.00001	0	*****
X0.0001-X0.00001	0	*****
X0.001-X0.00001	0	*****
X0.01-X0.00001	0	*****
X0.1-X0.00001	0	*****
X0.5-X0.00001	0	*****
X1-X0.00001	0	*****
X0.001-X0.000025	0	*****
X0.01-X0.000025	0	*****
X0.1-X0.000025	0	*****
X0.5-X0.000025	0	*****
X1-X0.000025	0	*****

Table 5.3: Adjusted p-values for change in TLO area as a function of Adhesion Molecule up-regulation rate from Figure 5.14. * = p<0.05, ** = p<40.005, *** = p<0.0005, **** = p<0.00005.

level of adhesion molecules expressed in order to support TLT formation, although it appears there is not a strong dependency on the expression level in terms of tissue area, the adjusted p-values calculated from a Tukey test and a one-way ANOVA do indicate a large degree of significance between different parameter samples when considering their entire distribution. Although *statistical* significance is demonstrated, it is unclear whether this translates to *biological* significance, given the consistent median and quartiles across the adhesion molecule up-regulation rates evaluated. A second threshold around a X = 0.01, which enables significantly larger TLT patches to form, demonstrated by the large increase in variance for the six highest sampled parameter values (further illustrated by the data in Table 5.3, adjusted p-values at zero for many parameters).

	Df	Sum Sq	Mean Sq	F Value	Pr (>F)	
Ind	9	2166	240.72	110.7	<2e16	***
Residuals	990	2154	2.18	-	-	

Table 5.4: Tukey multiple comparisons of means 95% family-wise confidence level for Figure 5.15
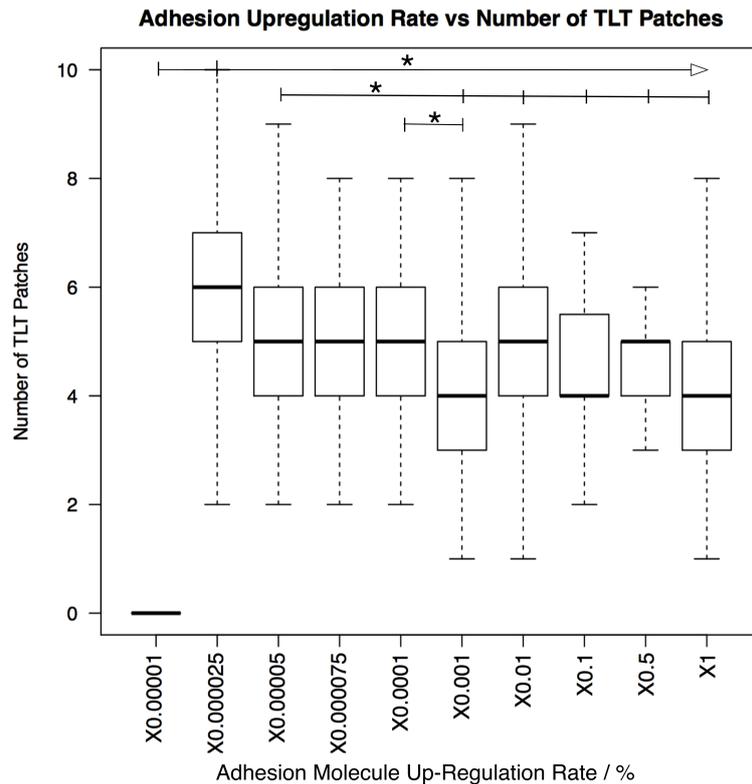
Figure 5.15: Box-plot of 'Adhesion Molecule Upregulation Rate' vs TLOs formed per 40x40 micron simulation space, illustrating the change in the number of discrete TLOs or tertiary lymphoid tissue aggregates forming as the rate of adhesion molecule up-regulation increases. The lines above the boxes indicate statistical significance, where ticks above a box indicate $p<0.05$ with respect to the line start point. All parameters are significant with respect to X=0.0001 and X=0.00025. n=100. X, 'Adhesion Molecule Upregulation Rate' is the linear rate at which adhesion molecule expression is upregulated on stromal cells following lymphocyte contact.

Figure 5.15 provides additional evidence that there exists a minimum threshold of adhesion molecule expression by stroma required for TLT induction, but beyond that, there is no correlation with the number of discrete TLOs formed, and as seen in Figure 5.14, there is a weak correlation with median TLT tissue area.

5.5   Model Sensitivity

A complete sensitivity analysis of NeoSim has not been completed at this stage, but is currently under-way for further publication of model results. However, at this stage the author can provide qualitative commentary on observed sensitivity of the model. In terms of the principal output of spatial structure formed at Day 15, the model is highly insensitive to parameter perturbation. This is believed to be representative of the domain, in which autoimmune processes are known to be extremely robust and resilient to change through external influences such as therapeutic interventions. Even highly implausible changes to the system, such as allowing B cells to enter from Day 0 and restricting T cells to enter from Day 5, thus reversing what is observed *in vivo*, does not prevent the self-organisation process, it merely extends the time taken to form by 1 to 2 days of simulated time. The section above explores the change in model response as the rate of adhesion molecule expression changes are modulated, and it can be seen that changes over an order of magnitude are required to see notable changes in TLT structure, and the vast majority of parameterisations still result in structures that are qualitatively similar to TLT structures as described in Chapter 3

The exception to this observed sensitivity lies with the receptor recycling ODE. The receptor internalisation rate and recycling rate are the most sensitive parameter within the model, and together have a profound effect on the stability of chemokine gradients that form as a result of lymphocyte–stroma crosstalk. Further investigation is necessary to constrain these values, and will likely require additional *in vivo* experimentation to assist in appropriate parameterisation. Anecdotally, discussions with other modellers that have incorporated receptor recycling have observed similar levels of sensitivity. The lack of quantified values within the domain makes this aspect of the model non-trivial to make robust.

5.6   Summary

This chapter determined that the theory of TLT formation in Sjörgen's syndrome developed in Chapter 3, and developed into a mathematical and computational model in Chapter 4, is consistent with the 15 day formation process observed *in vivo*. The application of IHC microscopy emulation further demonstrated the similarity between structures that form in NeoSim are qualitatively similar to *in vivo* sections of TLT tissue in a murine model of Sjörgen's syndrome. Further work on sensitivity and robustness analyses are on-going to provide additional support the evidence-based and argument-driven validation of NeoSim provided in this thesis, and this will appear in a future publication.

Furthermore, adhesion molecules were identified as having a potent role in the regulation of TLT morphology. This may present an interesting and novel avenue for investigation of a means of disrupting the formation process in a manner useful for therapeutic intervention. The following chapter is therefore concerned with testing existing therapeutics *in silico* to determine the usefulness of the model in determining therapeutic efficacy, which subsequently permits investigation of a novel therapeutic approach in treating Sjörgen's syndrome, the anti-VLA4 antibody natalizumab.

Chapter 6

Evaluating Therapeutic Intervention in TLT Formation *in silico*

We now evaluate the efficacy of three candidate biologic therapies in silico, *allowing determination of optimal dosing and combination therapies. Furthermore, we make the case for applying model-driven experimentation combining our TLT formation simulation with experiments to identify mechanisms driving lymphoid tissue formation and function, and we discuss potential applications of this experimental paradigm to identify novel therapeutic targets for TLT pathology. This leads us to propose that anti-VLA4 therapy is a suitable candidate worthy of further investigation* in vivo.

The use of Kohonen networks (or 'Self Organising Maps'/SOMs) permits projection of component planes providing a surrogate for efficacy, in particular by quantifying the change in FDC population size over different experimental groups. Simulation results are projected into a Unified Distance Matrix, enabling unsupervised clustering and visualisation of how subtle changes in the stromal population phenotype distribution resulting from biologic interventions can be stratified by the variance this causes in the reduced dimensionality projection represented by the U-Matrix. Justifi-

cation of our predictions with respect to efficacy are supported by various component plane projections quantifying relationships between efficacy, intervention, timing and dosage.

Anti-TNFα, Anti-VLA4 and baminercept interventions were examined as candidate therapies for treatment of TLT pathology in Sjörgen's syndrome, and the combination therapy of anti-VLA4 and anti-TNFα are also included. Clinical trials failed to find any significant benefit from anti-TNF therapy, and this also reflected in our in silico *evaluation (discussed in Section 6.4.1). The following table (Figure 6.1) describes each intervention examined, the relative dose range and the time-point at which the intervention is introduced into the simulation. Intervention at day 5 in our acute 15 day* in silico *model provides a limited model for prediction of therapies that remain very effective in well-established disease.*

Intervention *(n=10)*	Dose	Time-Point
Baseline	NA	NA
Baminercept	Low	Immediate
Baminercept	High	Immediate
Anti-TNF	NA	Immediate
Anti-TNF	NA	Day 5
Anti-VLA4 & Anti-TNF	Medium & NA	Immediate
Anti-VLA4 & Anti-TNF	Medium & NA	Day 5
Anti-VLA4	Medium	Immediate
Anti-VLA4	Medium	Day 5
Anti-VLA4	High	Immediate
Anti-VLA4	Very High	Immediate
Anti-VLA4	Very High	Day 5

Figure 6.1: List of therapeutic interventions included in QSP-Machine Learning driven analysis of simulation results. Precise parameterisation data are provided in Figure 6.2 below.

Experiments using a murine model of TLT induction and anti-VLA4 are ongoing in an attempt to provide experimental evidence of its suitability as a candidate therapeutic. Our data also suggest that anti-TNFα therapy in addition to anti-VLA4 may be beneficial in those who receive prompt diagnosis — established disease is likely

to be resistant to anti-TNF therapy as chronic autoimmunity establishes and TLT becomes self-sustaining through positive feedback.

6.1   Extending Existing Simulation Platforms

We have previously discussed the issue of extending a complex systems model to explore the role of particular pathways (**?**). The changes made to accommodate any intervention, and all assumptions and abstractions made during implementation of the intervention must be explicitly declared. Unlike the modifications we made to ARTIMMUS in **?** and **?** in which novel pathways were added to the model, in this instance we focused upon the manipulation of existing signalling pathways to determine the effect of various biologic therapies. The modelling methodology used to implement each of the biologic therapies evaluated are described in the following sections. Section 6.1.1 describes implementation of a TNF blockade within NeoSim, Section 6.1.2 describes the incorporation of baminercept, and Section 6.1.3 describes how VLA4 blockade (i.e. natalizumab) is incorporated into the model and simulation.

6.1.1   Anti-TNF Intervention Methodology

Inflammation is represented in the model as a probability that each stromal cell becomes inflamed/activated at any given time step, $p(\text{canDifferentiate})$ as defined in the finite-state machine derivation of the stromal cell Markov chain in Figure 4.13. As can be seen in the probabilistic guard described in Figure 4.13, a Monte Carlo sampling of the value of I_t determines at each time-step whether each stromal cell is capable of differentiation into lymphoid phenotypes S2-S4. The probability value is determined by Equation 6.1 below, originally introduced as a Markov chain parameter in Table 4.1:

$$p(\text{canDifferentiate}) = I_t = I_0 e^{-\lambda \Delta T}, \tag{6.1}$$

where I_t is the inflammation level at time t, λ is the decay constant and ΔT is the time-step length. The initial conditions for this equation are defined by the parameters in Appendix B, Figure B.2. Inflammatory factors trigger the self-sustaining and self-organising lymphocyte swarm resulting in TLT formation. Incorporating anti-TNF into the model requires that an additional two parameters be defined, anti-TNF dose and the time-point at which it is introduced, collectively termed α_t, determining the degree to which the probability value is reduced in response to anti-TNF signalling. Thus Equation 6.1 is extended to become Equation 6.2:

$$I_t = (I_0 e^{-\lambda \Delta T})(1 - \alpha_t). \tag{6.2}$$

By exploring results with immediate and delayed interventions, and a range of doses, we can begin to determine precise potential treatment regimes. Anti-TNF has been evaluated against placebo in clinical trials with no significant improvement in the treatment group.

6.1.2 Baminercept (LT-β Receptor Fusion Protein) Intervention Methodology

Baminercept is a fusion protein lymphotoxin-beta receptor (LT-βR), lymphotoxin is known to have an important role in the initiation of lymphoid neogenesis alongside TNF-α and IL-13. As it is available as a pharmaceutical for use in humans, Baminercept is an attractive candidate therapy. Baminercept reached Phase II clinical trials (U.S. Clinical Trial Identifier 'NCT01552681', `https://clinicaltrials.gov/ct2/show/NCT01552681`) prior to termination due to lack of efficacy, therefore we can predict *a priori* that Baminercept will not have a significant effect on FDC population size, GC formation and general capacity to resolve TLT pathology.

6.1.3 Anti-VLA4 Intervention Methodology

VLA-4 is expressed by lymphocytes and enables them to adhere to stromal cell adhesion molecules. By modulating the Monte Carlo-sampled parameter value, it is possible to emulate a VLA-4 blockade by reducing expression levels. Lymphocytes that are not bound to stroma cannot stimulate it to differentiate, and therefore it is plausible that reducing stroma-lymphocyte cross-talk may help reduce acute TLT pathology and as such anti-VLA4 was chosen as our third candidate intervention.

6.1.4 Combination and Time-Delayed Interventions

Combination therapies can be simulated by enabling multiple interventions simultaneously. By incorporating a dose start time, delayed interventions are possible, i.e. only beginning the therapeutic intervention after 5 days of untreated disease pathology.

6.1.5 Model Parameterisation for Interventions

The following figure (6.2) provides a list of the additional parameters created to model the role of anti-TNF and anti-VLA4, and the values used that correspond to the qualitative descriptors of dose used in Figure 6.1. Note that implementation of anti-LTβ (baminercept) does not depend on new parameters, but modulates the existing parameter, '*inflammationLevel*'. Simulation initial conditions and parameterisation remain at the baseline levels as defined in Appendix B for all interventions.

Parameter	Description	Value	Comment
antiVLA4Level	Dose of anti-VLA4	0.8	Medium Dose (feature scaled units)
antiVLA4Level	Dose of anti-VLA4	0.96	High Dose (feature scaled units)
antiVLA4Level	Dose of anti-VLA4	0.99	Very High Dose (feature scaled units)
antiVLA4timepoint	Time-step to introduce anti-VLA4	0	Day 0
antiVLA4timepoint	Time-step to introduce anti-VLA4	720 (10 min steps)	Day 5
antiTNFLevel	Dose of anti-TNF	0.95	(Feature scaled units)
antiTNFtimepoint	Time-step to introduce anti-TNF	0	Day 0
antiTNFtimepoint	Time-step to introduce anti-TNF	720	Day 5
inflammationLevel	Existing parameter modulated as surrogate for baminercept	50	Day 0
inflammationLevel	Existing parameter modulated as surrogate for baminercept	75	Day 0

Figure 6.2: Table of therapeutic intervention-related parameters for biologics anti-TNF, anti-VLA4 (natalizumab), and anti-LTβ (baminercept)

6.2   Clustering and Stratification using Kohonen Networks and the Unified Distance Matrix

The Kohonen map in Figure 6.3 shows all simulation results described in the table from Figure 6.1 following dimensionality reduction with 1) all available data and 2) only normalised stromal ratios utilised. The area of the pie chart drawn within each node is directly proportional to the number of simulation results placed at that node. Simulation results with the same colours (and therefore interventions) will tend to naturally cluster together as they share latent properties. Note that the key metric for interpreting simulation responses across different interventions is observation of simulation runs from each intervention stratifying into separate regions within the Kohonen network, along with observation of surrogate measures of disease severity and stromal cell populations within those regions.
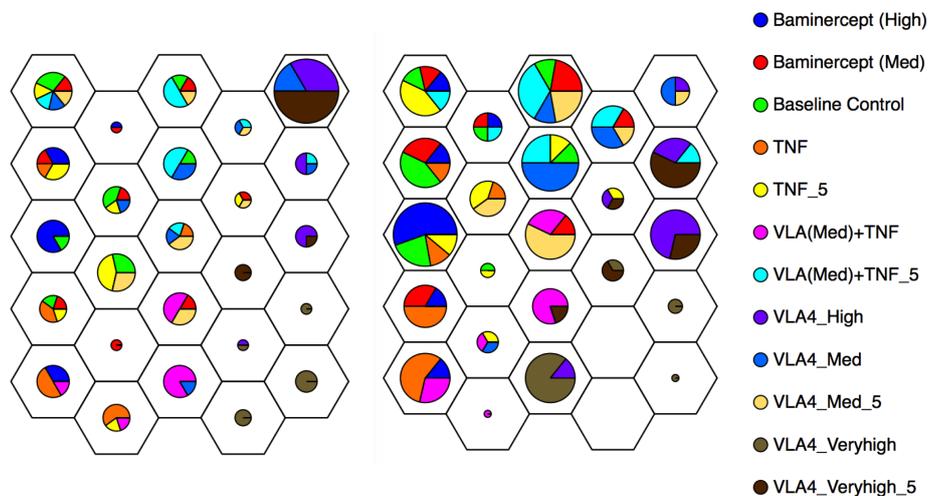
Figure 6.3: Kohonen Map (Self-organising Map) showing all simulated interventions clustered into a 64 node input layer arranged as an 8x8 hexagonal grid using 500 iterations and a Gaussian neighbourhood, repeated with all stromal phenotypic data (Left) and only relative data, such as FRC-to-FDC ratio (Right). Clustering is more efficient when data are restricted to relative attributes, as including the additional parameters such as absolute cell numbers for each phenotype (in which the total varies in every simulation run) requires more parameters associated with each experiment to be effectively clustered by the artificial neural network. The data and neural network applied in these two maps are identical.

6.2.1   Model Outputs Observed

Model outputs observed in this chapter are the absolute and relative population sizes of S1 (stromal progenitor), S2 (localised precursor), S3 (FRC-like), and S4 (FDC-like) stromal cells. The principal output providing a surrogate disease activity is the population of FDC-like cells, as these are required to support B-cell follicles and germinal centre reactions. Therefore a reduction in the population size of FDC-like cells is indicative of lower disease severity. These data are acquired by interrogating the CDS data structure developed in Section 4.1 and described in Figure 4.8, the contents of which are output every 12 hours of simulated time.

6.3   Summary of Interventions and Observed Effects

The Unified Distance Matrix, as introduced in Section 5.3, is calculated for the set of all simulation runs for each therapeutic intervention (Figure 6.4). In Figure 6.4, light grey depicts more closely spaced node vectors and dark grey indicate more widely separated node vectors (that is, distance on the map is not the 'codebook vector' distance: the distance between nodes (hexagonal elements) is based on the similarity within their underlying data). Thus, darker regions separating lighter areas may therefore be seen as 'extended' boundaries between plotted data points, implying greater separation if a set of samples is segregated by a dark grey line on the hexagonal grid. In this sense, Kohonen maps derived from Kohonen networks are essentially a non-linear generalisation of principal component analysis. The samples generally well cluster into groups based on the intervention applied, despite being an unsupervised analysis with relatively low sample size of n=10.

The mean difference in final stromal phenotype distribution between each of the intervention strategies can be visualised using a horizontal bar graph, as illustrated in Figure 6.5. It can clearly be seen that only high VLA-4 blockade can cause a
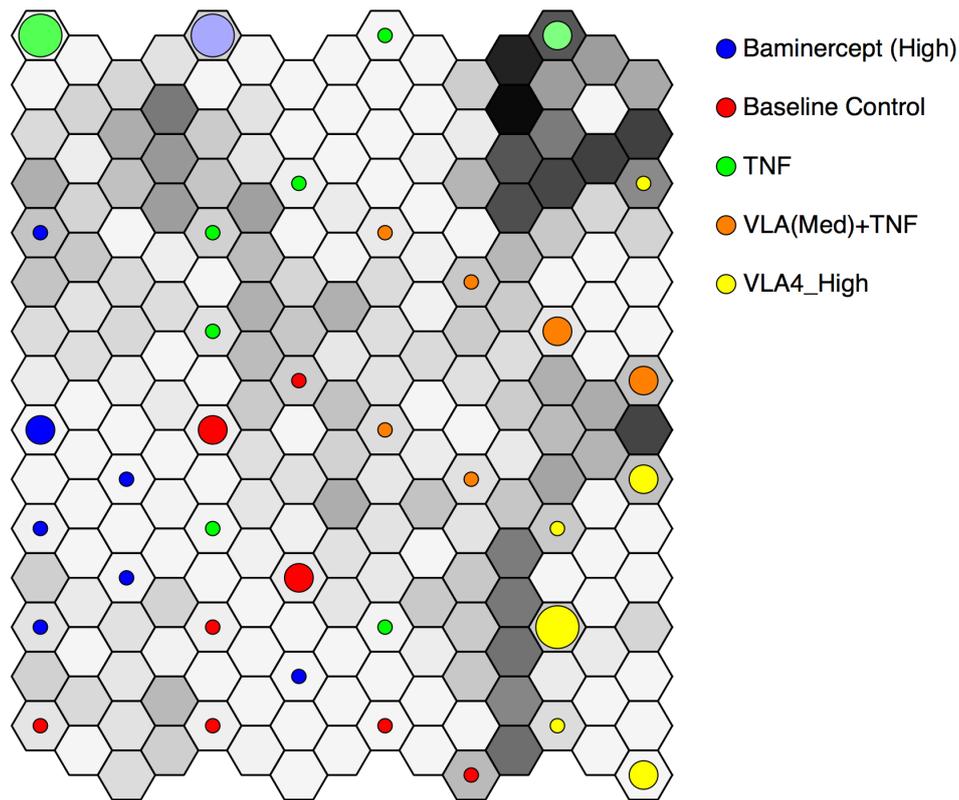
Figure 6.4: Kohonen Map U-Matrix for all therapies, combination therapies and delayed interventions simulated. The grey-scale tone map represents the Euclidean distance between each simulation sample.

significant reduction in FDC population size. Lower-dose VLA-4 blockade appears to increase the ratio of FRC to FDC populations, whereas high doses significantly increases the population of localised stromal precursors that have not adopted an inflammatory lymphoid phenotype (the LSP population, or state 'S2'). TNF blockade reduces the FRC population size and the overall level of inflammation, however anti-TNF therapy did not significantly affect FDC population size.

The ratio of FDC to FRC stromal cell population sizes could be a useful surrogate measure of efficacy, as any decrease in this ratio would suggest an increase in FRCs over FDCs – as is seen in VLA-4 blockade (Figure 6.5). To examine the change in this ratio and the ratio of inflamed, activated stromal cells, a bar graph for each of the three biologic therapies can be drawn examining the relative FDC and FRC percentage population sizes (Figure 6.6). The most evident difference from baseline
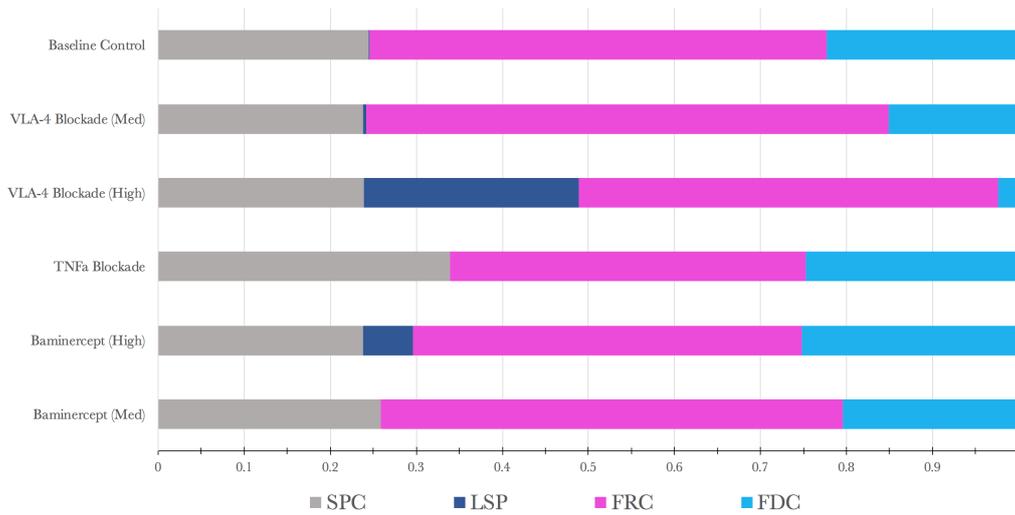
Figure 6.5: Horizontal cumulative bar graph showing the stromal cell phenotype (Stromal Progenitor Cell, Localised Stromal Precursor, Fibroblastic Reticular Cell, Follicular Dendritic Cell) distribution for each of five interventions vs baseline control simulation.

outcomes is observed with the VLA-4 blockade, showing a large reduction in mean FDC population size, a modest decrease in FRC population size and a significant increase in the FRC to FDC ratio (together, this is indicative within the domain under study of reduced GC activity and therefore disease severity, the proxy for which in our simulation is the change in relative FDC population size).
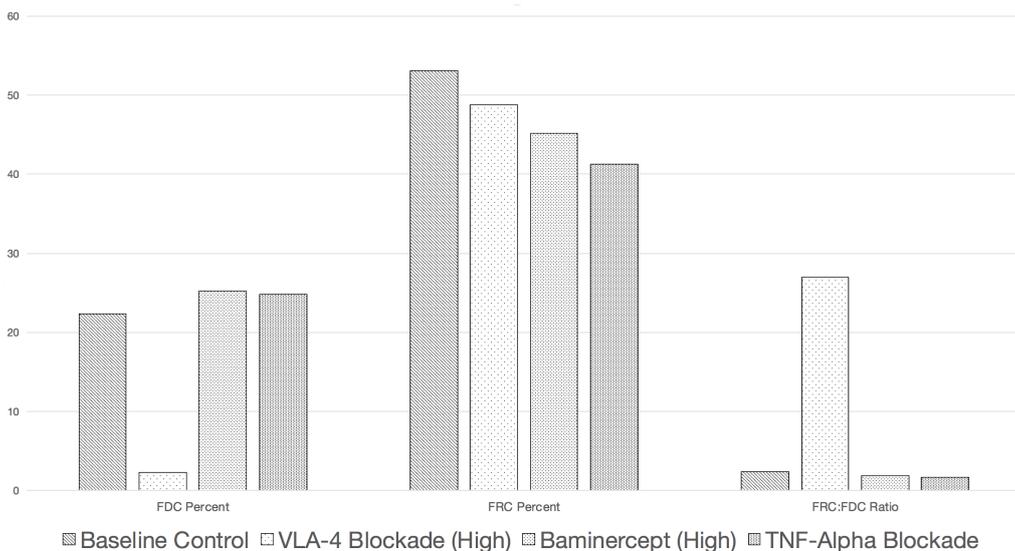
Figure 6.6: Bar graph of FDC/FRC stroma percentages and ratio for TNFα, VLA-4 and Baminercept biologic therapeutics evaluated *in silico*.

One platform model definition of inflammation is the percentage of stromal cells that have begun to develop a differentiated lymphoid phenotype, as this number is directly proportional to initial inflammation level/local TNF-α concentration – described as an exponential decay ODE, each stromal cell in the network sequentially evaluates whether it has adopted an inflammatory phenotype according to a stochastic guard every 2 minute time-step. A bar chart was plotted using absolute and normalised (relative) values for inflammation based on this definition, shown in Figure 6.7. Such abstract notions of inflammation can be useful to identify effect magnitudes and key changes in observables within abstract simulations, and provide indications of what may be suitable for further exploration. For example, later investigations reveal that TNF blockade is rather ineffective, particularly compared to VLA-4 blockade, despite the latter's negligible effect on inflammation in the model when viewed from this perspective.
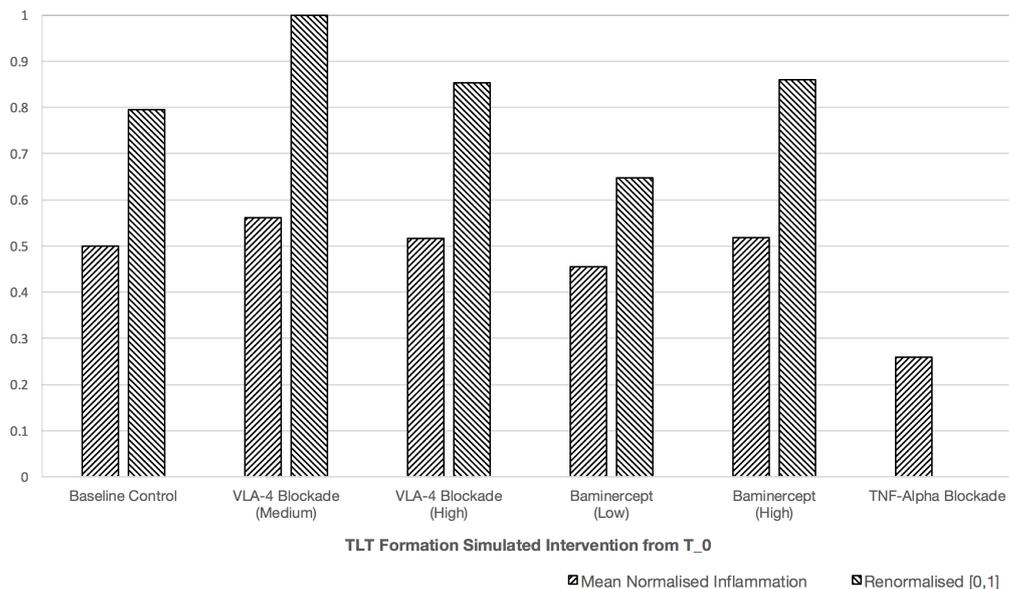
Figure 6.7: Bar graph of mean (re)normalised inflammation for 5 interventions *in silico* vs baseline control simulation. Inflammation is defined here as the means of the normalised ratio (across all experiments) of lymphoid stroma (LSP, FRC, FDC) to SPC (Stromal Progenitor Cells). The renormalisation is a standard rescaling of these values after the mean per experiment has been computed, hence the renormalised inflammation for TNF blockade is 0 because it was lowest. *n=10 per sample*

Feature-mapping by projecting component planes on clustered Kohonen maps

produced using artificial neural networks permits an assessment of the magnitude and relative effect size of different therapeutic interventions of single and dual biologic therapies. In simple terms, if simulation results including interventions are well separated by some definition, and there is a clear relationship between the test groups and the feature chosen for mapping in the component plane – such as relative FDC population size as a surrogate measure of clinical efficacy, it should be very apparent by plotting feature maps comparing all of the interventions against baseline and each other. Interpreting these data can present difficulty or a learning curve initially due to the densely concentrated information spread over many charts. There is a clear correlation between unsupervised clusters, the applied intervention and the value of the projected component - percentage of stroma with an FDC phenotype at the end of each simulation run. In the case of Figure 6.8, there is a clear relationship between anti-VLA4 therapeutic intervention and lower FDC population sizes - demonstrated by the overlaid heat-map illustrating the dose-dependent role of anti-VLA4 in decreasing the FDC population size. It is worth noting that while we know the absolute change in phenotype distribution is relatively small (from Figure 6.6), it appears to be critical enough to prevent the formation of densely aggregated TLT structures – that is, a relatively small deviation from homeostasis is responsible for the resulting dysregulation and autoimmune processes.

The Kohonen map shown in Figure 6.9 shows the same data points projected against FRC population size, with a very approximately inverse relationship to the FDC heat-map in Figure 6.8. Baminercept appears to notably reduce FRC population size, whilst – as observed from Figure 6.6, anti-VLA4 slightly increases FRC population size at the expense of the FDC population size, or alternatively put, biases lymphoid stromal develop to preferentially adopt an FRC-like stroma, or 'T-cell stroma', phenotype.

When the simulation results are projected against the entire percentage of stroma that had adopted a lymphoid phenotype at the simulation end – one definition of
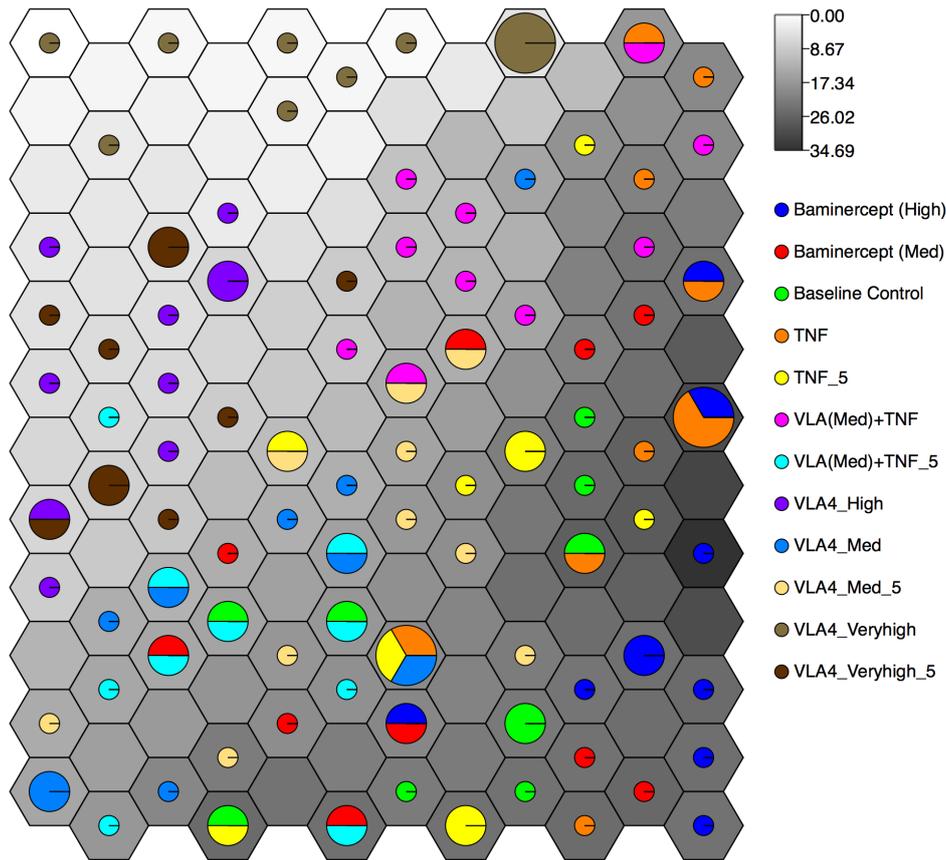
Figure 6.8: Kohonen Feature Map (Self-organising feature map) of all simulated interventions clustered and visualised against the mean percentage of stroma per run with an FDC phenotype at T_{END}. An unsupervised learning Kohonen network with random map initialization, a 20x20 hexagonal grid and 1000 iterations produces a generalised non-linear equivalent to principal component analysis (PCA) with projected output data (FDC population percent) as a grey-scale level. This is an excellent technique for visualising high-dimensional data in low dimensions while preserving the toplogy of the underlying data through use of competitive, neighbourhood learning rather than traditional machine learning 'error correction' methodologies, *n=10 per sample.*

inflammation within the context of this model, we note that there is not much variance between samples, however there is a notable drop with the combination therapy intervention in which both anti-VLA4 and anti-TNF are given together.

A randomly initialised network of 10x10 nodes is used to create an FDC-like stroma feature map in Figure 6.11 that clearly highlights the superior capacity of anti-VLA4 treated samples to resist the development of FDC-like stromal networks and the associated autoimmune response.
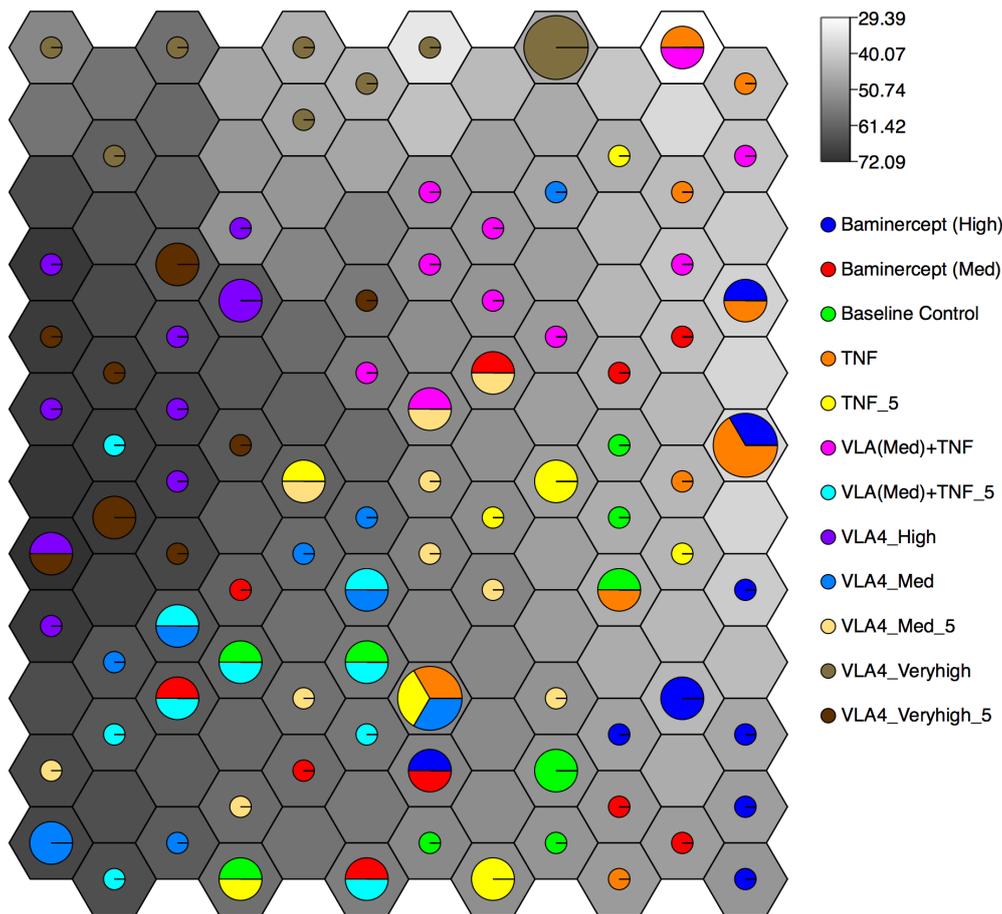
Figure 6.9: Kohonen Feature Map (Self-organising feature map) of all simulated interventions clustered and visualised against the mean FRC population size of stroma with an FRC phenotype at T_{END} (population size in units of 'stromal cell elements': individual occupied hexagonal grid elements, of which up to 24 elements stochastically constitute one stromal cell, within the Cellular Data Structure described in Section 4.7.2.
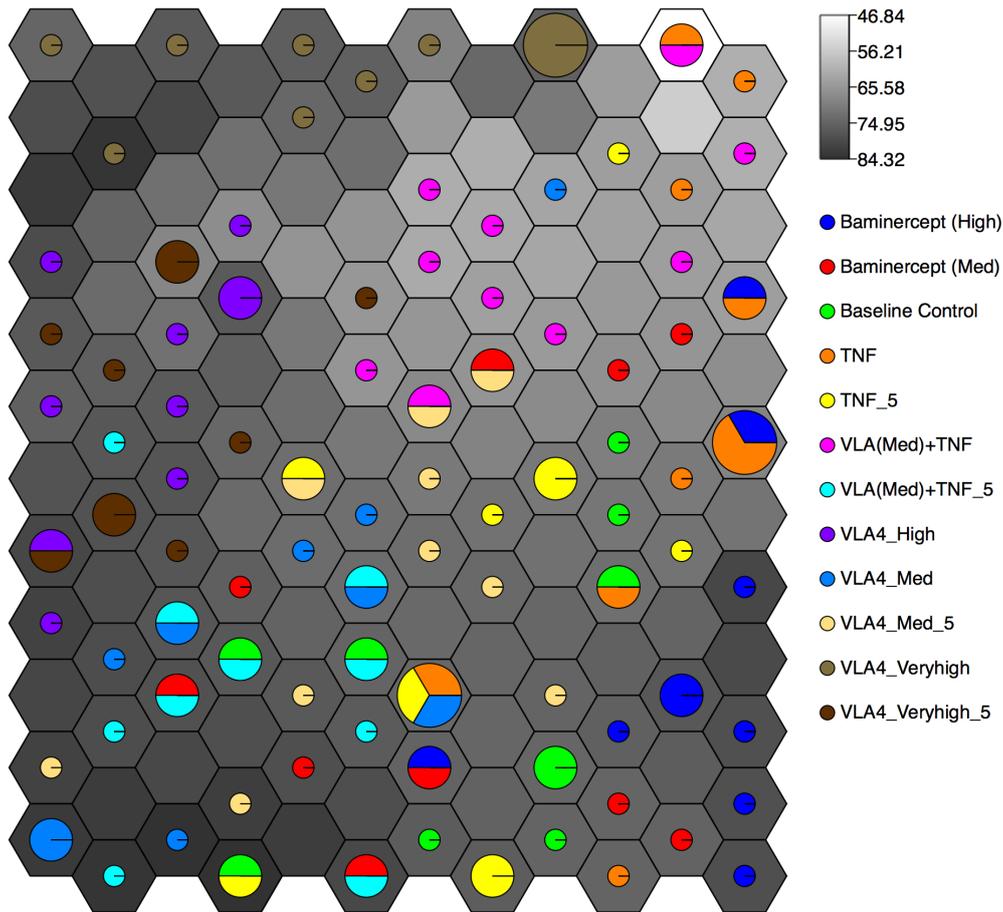
Figure 6.10: Kohonen Feature Map (Self-organising feature map) of all simulated interventions clustered and visualised against the percentage of stroma that has adopted an inflamed lymphoid phenotype per run by T_{END}.
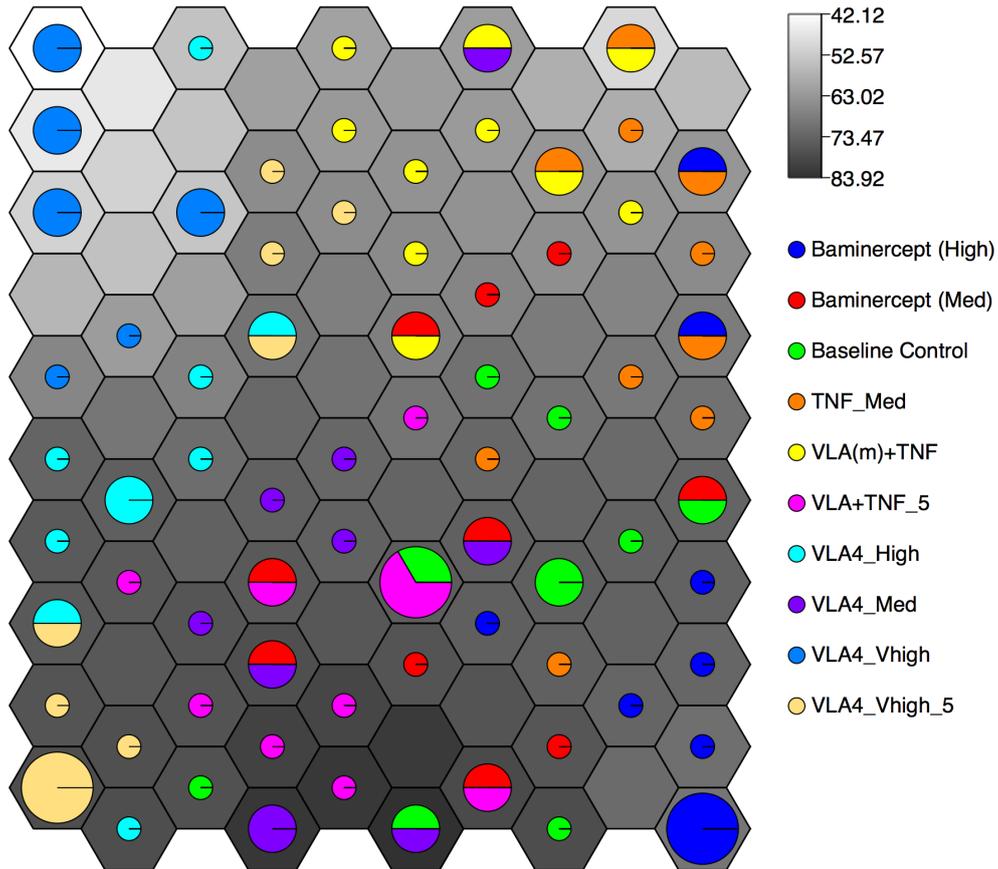
Figure 6.11: Kohonen Feature Map (Self-organising feature map) of all simulated interventions clustered and visualised against the mean percentage of stroma per run with any lymphoid phenotype (LPC, FRC, FDC) at T_{END}. *n=10 per sample.*

6.4   Evaluating Efficacy through Feature Mapping using Unsupervised Machine Learning

To better understand how these approaches can be used to understand potential treatment strategies, it will be helpful to look at a smaller number of sample types such that the plots are easier to interpret, therefore we will now examine Anti-TNF and Anti-VLA4 against the 'control sample' – baseline simulation results without any therapeutic intervention – and determine the relative predicted effect and efficacy.

6.4.1   Efficacy of TNF-Blockade in Reducing TLT Formation and Associated Pathology

This section evaluates the anti-TNF therapeutic intervention with respect to the simulation baseline (no therapeutic intervention). Firstly, we calculate the Unified Distance Matrix codebook values for the control and experimental group, shown in Figure 6.12. Clear boundaries between the baseline simulation samples and those that include the anti-TNF therapy indicate that there is not sufficient aleatory uncertainty to prevent the use of feature mapping of component planes in determining efficacy from Kohonen maps.

Feature mapping the FDC-like and FRC-like stromal phenotype percentages results in the maps shown in Figures 6.13 and 6.14 respectively. There is no statistically significant relationship between FDC-like population size in anti-TNF treated samples as compared to baseline results. Anti-TNF therapy results in a modest decline in FRC relative population size compared to baseline. This might suggest that anti-TNF is a more suitable therapeutic in T-cell mediated autoimmunity, but has negligible effect on FDC-like population and therefore the presence of B cell follicles that could support GC reactions and the autoantibody response.
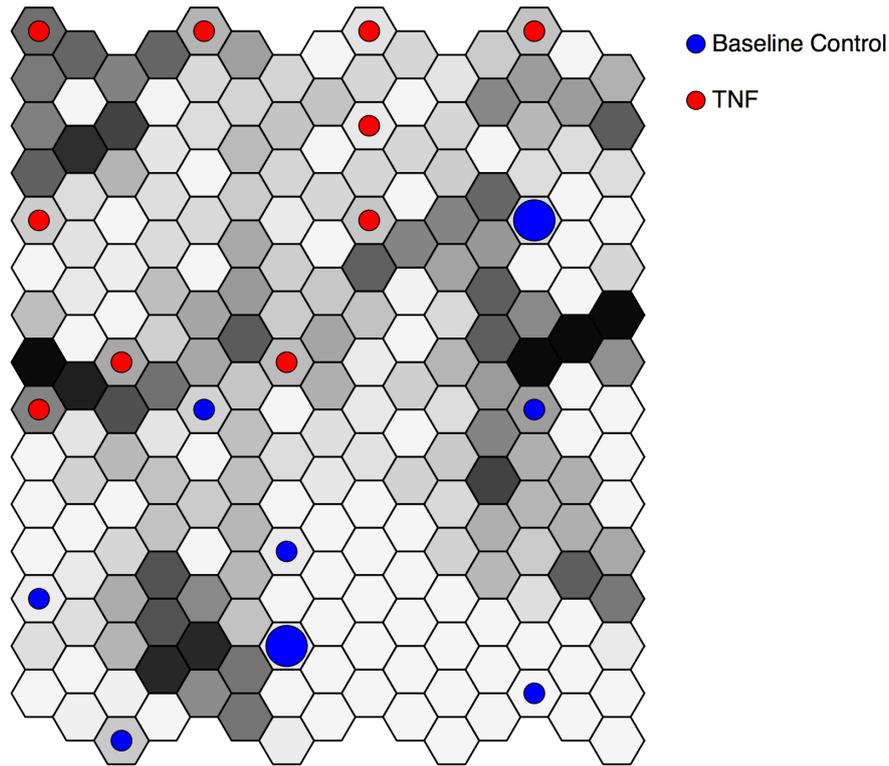
Figure 6.12: TNF blockade vs baseline U-Matrix: clear boundaries between the TNF blockade and the baseline control datasets indicate these samples can be used to reliably predict efficacy of TNF blockade in TLT formation.

The TNF blockade has resulted in a notably lower overall percentage of inflamed stromal cells, however those that do remain preferentially adopt FDC-like phenotypes – this suggests that the tissue can still support substantial B-cell follicles with GC reactions, despite the observed lower inflammation (also seen in Figure 6.7). This can be seen in the inflammation feature plot of Figure 6.16 and the previously defined FDC-FRC ratio metric of GC activity, and its inverse.

Figure 6.13: Kohonen Feature Map (Self-organising feature map) of TNF blockade vs baseline control, clustered and visualised against the mean percentage of stroma per run with an FDC-like phenotype at T_{END}. A greater percentage of stromal area has adopted an FDC-like phenotype during the TNF blockade compared to the baseline control. *n=10 per sample*
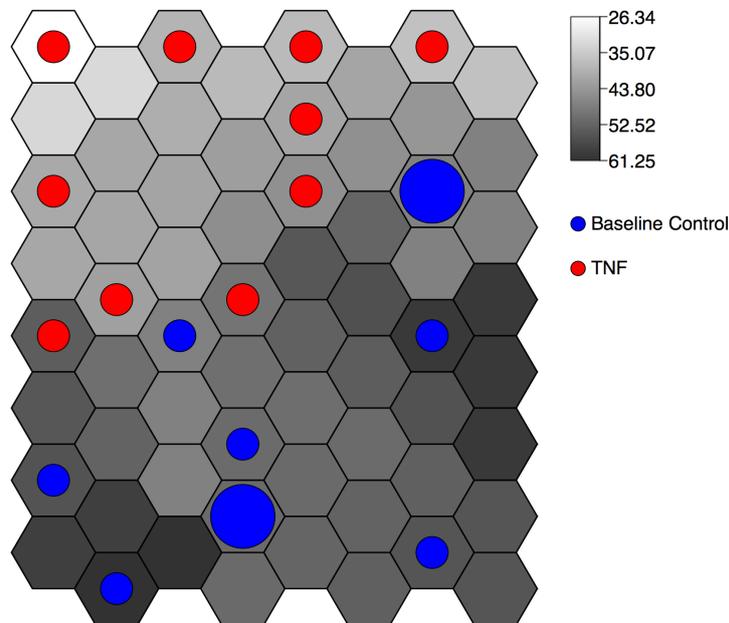
Figure 6.14: Kohonen Feature Map (Self-organising feature map) of TNF blockade vs baseline control, clustered and visualised against the mean percentage of stroma per run with an FRC phenotype at T_{END}. A notably lower percentage of stromal area has adopted an FRC phenotype during the TNF blockade compared to the baseline control. *n=10 per sample*
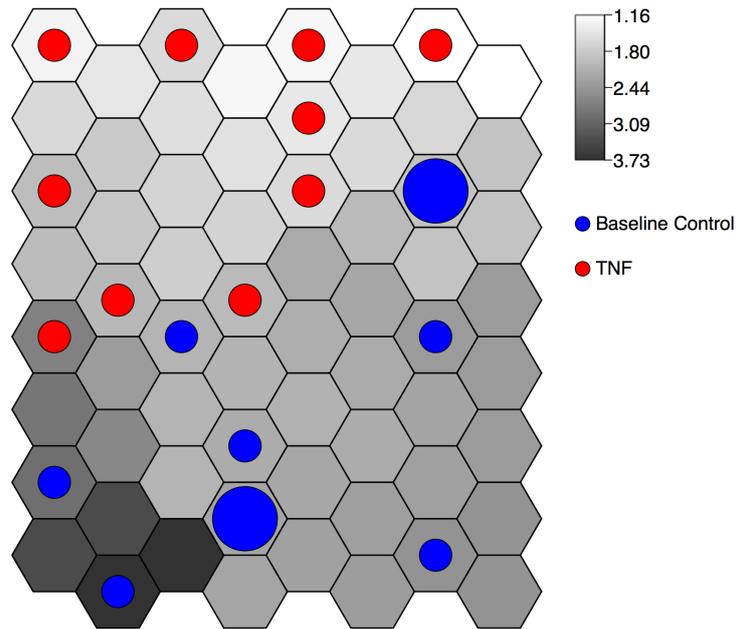
Figure 6.15: Kohonen Feature Map (Self-organising feature map) of TNF blockade vs baseline control, clustered and visualised against the mean percentage of stroma per run with an FRC phenotype at T_{END}. *n=10*
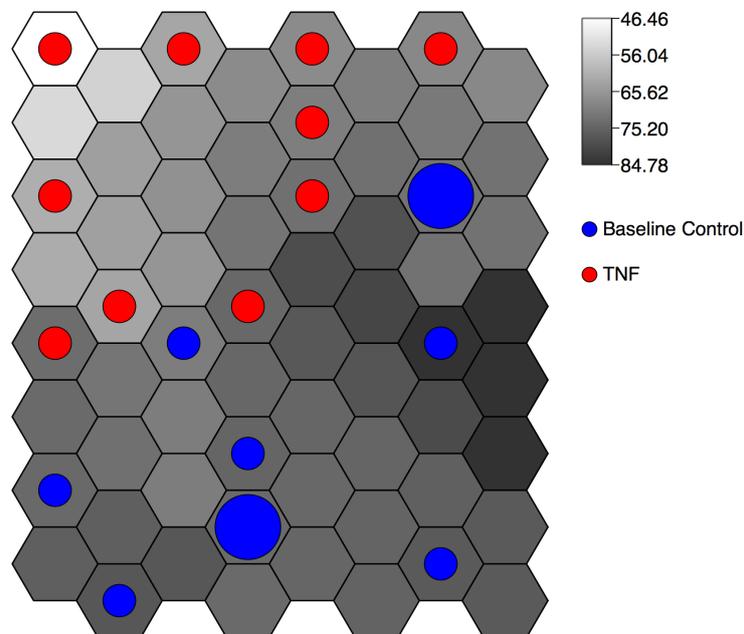
Figure 6.16: Kohonen Feature Map (Self-organising feature map) of TNF blockade vs baseline control, clustered and visualised against the mean percentage of stroma per run with an inflamed, lymphoid phenotype at T_{END}. *n=10*

6.4.2 Efficacy of TNF-Blockade in Reducing TLT Formation and Associated Pathology

This section evaluates the anti-VLA4 therapeutic intervention with respect to the simulation baseline (no therapeutic intervention). We can re-calculate the codebook vectors and ensure proper clustering of intervention and baseline samples. Feature mapping of the percentage of stroma with A) a lymphoid phenotype and B) either an FRC or FDC-like phenotype. The clear dose dependent clustering and relationship to the inflammation surrogate measures utilised provides further evidence of efficacy of anti-VLA4 treatment in Sjögren's syndrome within the context of this *in silico* model. There have been no clinical trials of the use of anti-VLA4 in Sjögren's syndrome, however we believe the mechanistic and *in silico* evidence arising from the unsupervised neural (Kohonen) network analysis is sufficiently high to warrant further studies of this biologic *in vivo* within the context of murine models.

The difference in absolute stromal population size can be seen in the boxplot in Figure 6.17. The absolute stromal population size falls to a minute fraction of the baseline median, despite maintaining the same distribution of stromal phenotype population sizes. The Kohonen feature map illustrating the relationship between Anti-VLA4 dose received and stromal inflammation level, shown in Figure 6.18, clearly shows a dose-response relationship for the three dose values evaluated ('antiVLA4Level' = 0.80, 0.96 and 0.99) in which higher anti-VLA4 doses result in further reduction of FDC-like cells, and inflamed (FRC-like and FDC-like) stroma generally. With three doses, it is not possible to determine a dose-response curve or accurately assess the non-linearity of the dose-response effect, although the author notes anecdotally that only high values of the parameter 'antiVLA4Level' had a large impact on FDC level. It should also noted that the relationship between the disease severity metric (FDC-like stroma population size) and actual disease severity is also unknown, which means a dose-response curve plotted with these metrics would be of limited relevance.
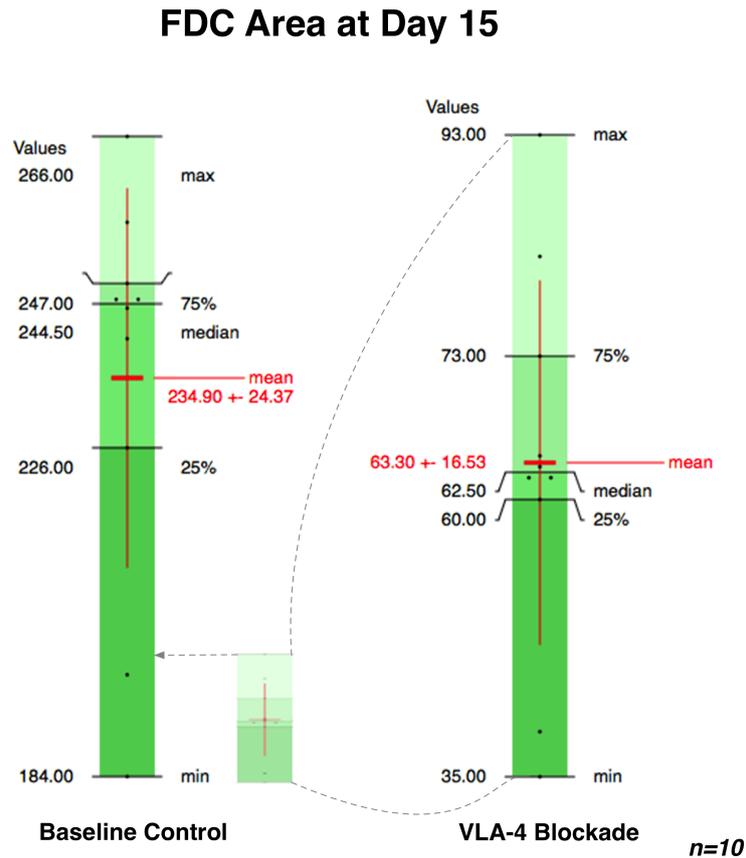
Figure 6.17: Boxplot of FDC-like distribution over multiple simulation runs comparing baseline behaviour to strong VLA-4 blockade. FDC-like stromal cells are greatly reduced during VLA-4 blockade, with a long-tailed distribution and significantly reduced median FDC-like stroma differentiation at T_{END}.
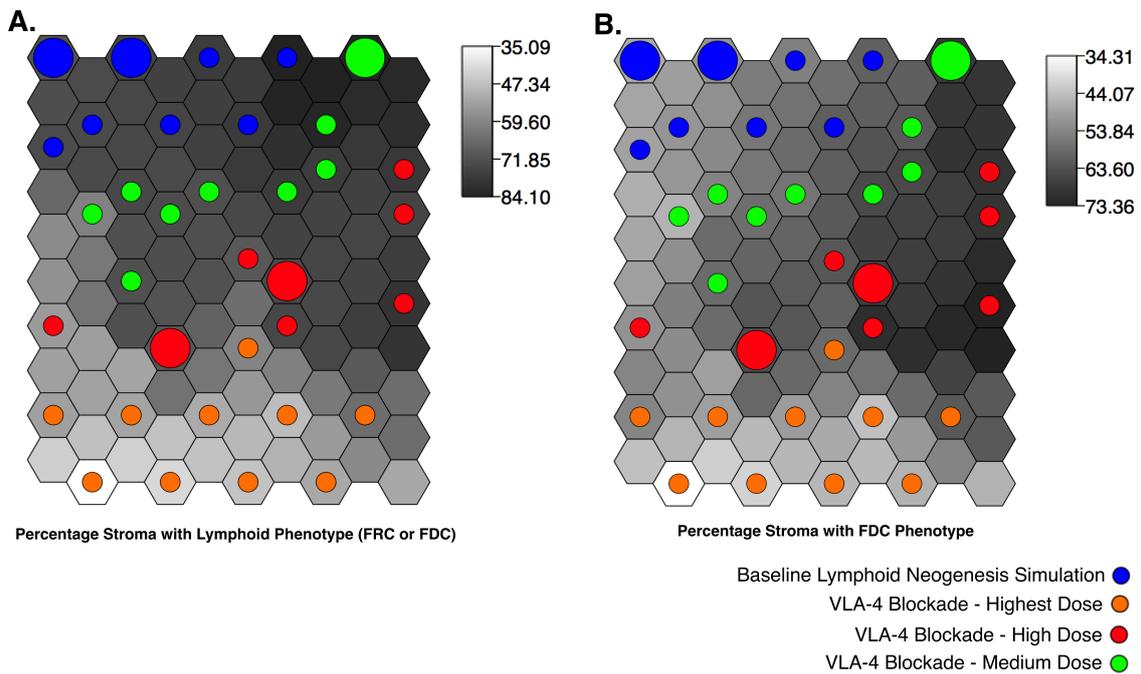
Figure 6.18: Kohonen network feature map showing the percentage lymphoid stroma (FRC or FDC-like) and percentage FDC-like stroma component planes for VLA-4 blockade interventions of three doses. *n=10*

6.4.3   Importance of Timing in Therapy Efficacy for Anti-TNF and Anti-VLA4

In order to evaluate the importance of timing when receiving anti-VLA4, Anti-TNF or a combination therapy, the unified distance matrix is calculated for samples with the intervention effective from T_0 and from 5 days into the development process, shown in Figure 6.19. Kohonen feature maps were created illustrating changes in the FDC-like and FRC-like population sizes respectively in Figures 6.20 and 6.21. Given the role of TNF-α in triggering the lymphoid neogenesis process, and the systems' capacity for self-organisation, one may predict that delayed interventions of anti-TNF serve no useful purpose in this context as tissue will already be established. Anti-TNF was observed to have no effect as a delayed intervention. The likely response to a delayed anti-VLA4 therapy is less straightforward as adhesion continues to hold an important role in TLT pathology even when inflammatory markers are not significant (in the context of TLT pathology in Sjögren's syndrome. This may not be the case in TLT pathologies in primarily T-cell mediated autoimmune disorders such as multiple sclerosis or rheumatoid arthritis due to the supposed lesser dependence on B cells and germinal centre reactions producing autoantibodies in mediating tissue damage.
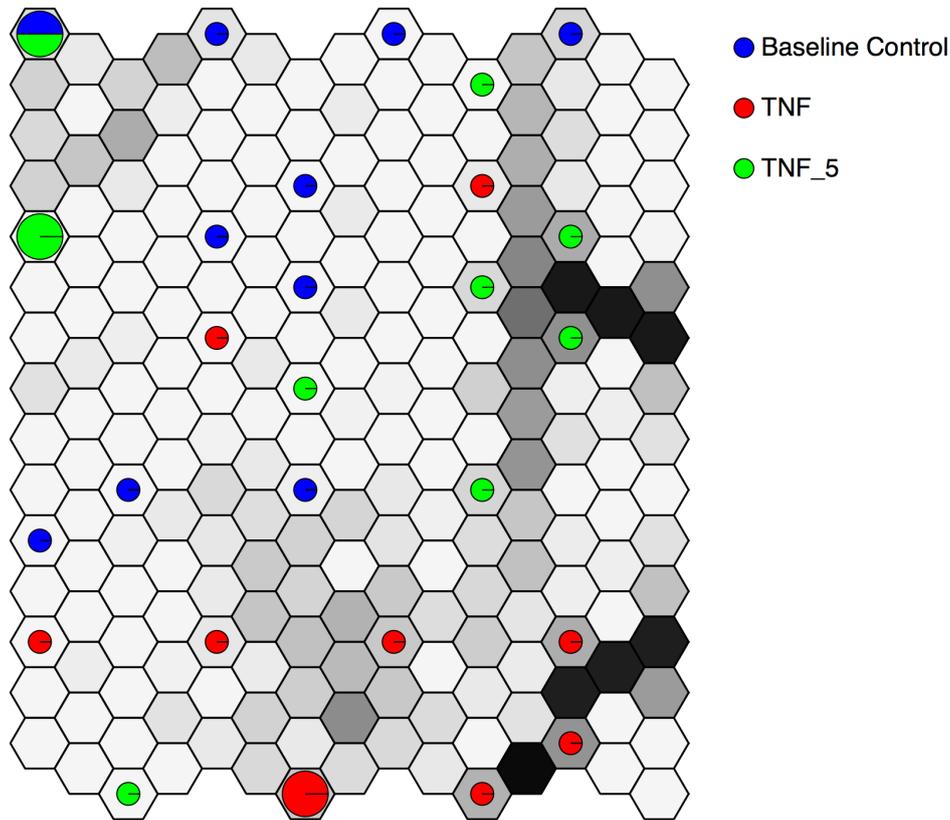
Figure 6.19: TNF blockade at Days 0 and 5 vs Baseline U-Matrix: there is largely clear stratification of TNF blockade at day 0 and the baseline control, however the map cannot clearly segregate simulation runs in the data between the baseline simulation and a TNF blockade introduced at Day 5 of the TLT development process. That there is no significant difference between these (indeed several runs occupy the same node with zero codebook distance) is indicative that blocking inflammatory signals after inducement of the tissue cannot effectively reverse it. This is likely to be the reason that results of anti-TNFa and baminercept therapies have been disappointing (**??**), early intervention is key. The U-Matrix, however, represents differences between all parameters of multi-dimensional data; there may be some mechanistic or clinical effect that can be identified using feature maps with surrogate outcomes as the component plane (see Figures 6.20 and 6.21.
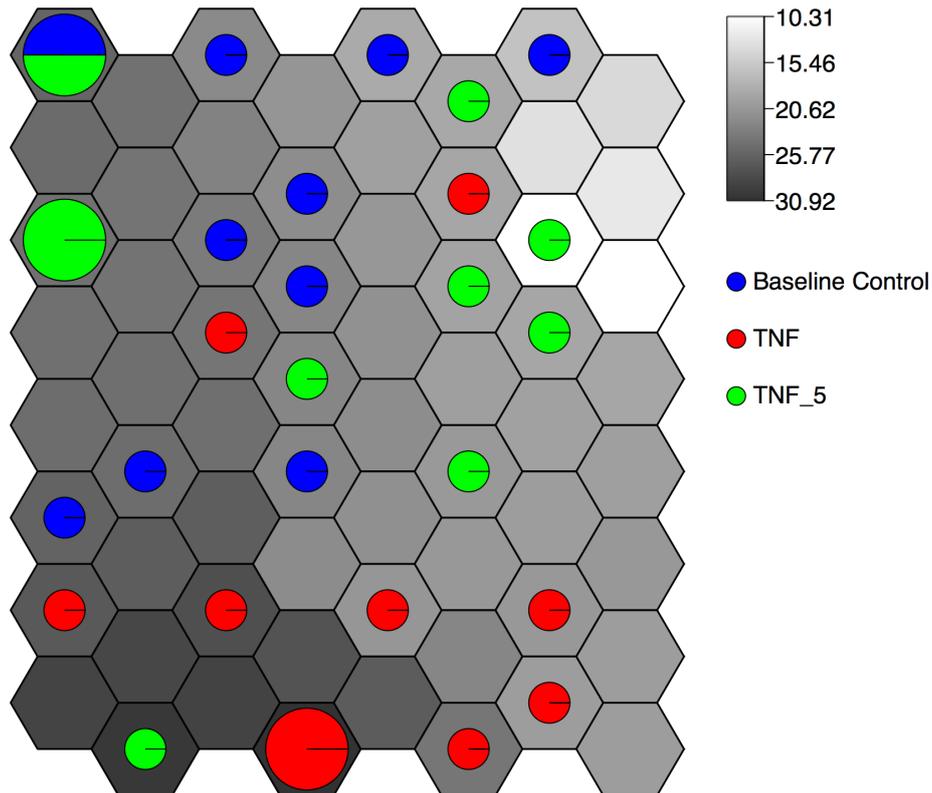
Figure 6.20: Kohonen Feature Map (Self-organising feature map) of TNF blockade at days 0 and 5 vs baseline control, clustered and visualised against the mean percentage of stroma per run with an FDC-like phenotype at T_{END}. A greater percentage of stromal area has adopted an FDC-like phenotype during the TNF blockade at day 0 compared to the baseline control, and there does seem to be a limited separation for a day 5 intervention, although it is much closer to baseline FDC-like levels than if the blockade is initiated 5 days earlier. *n=10.*
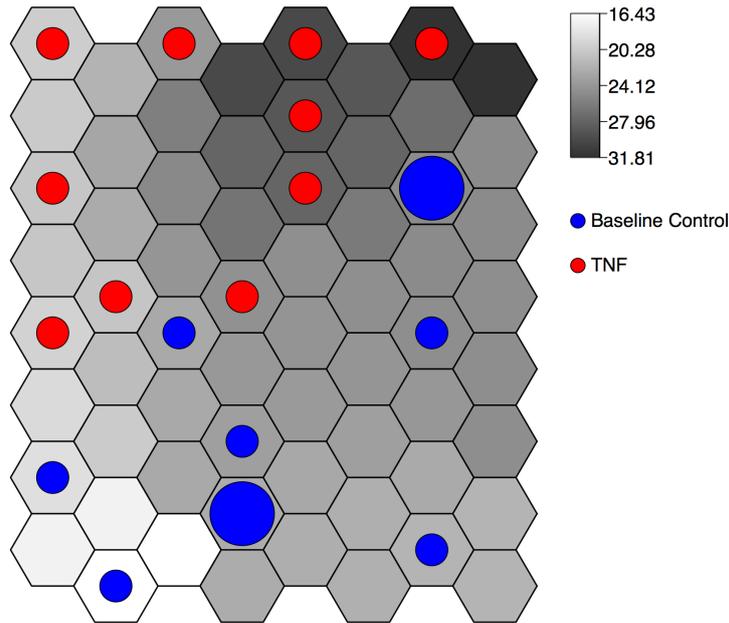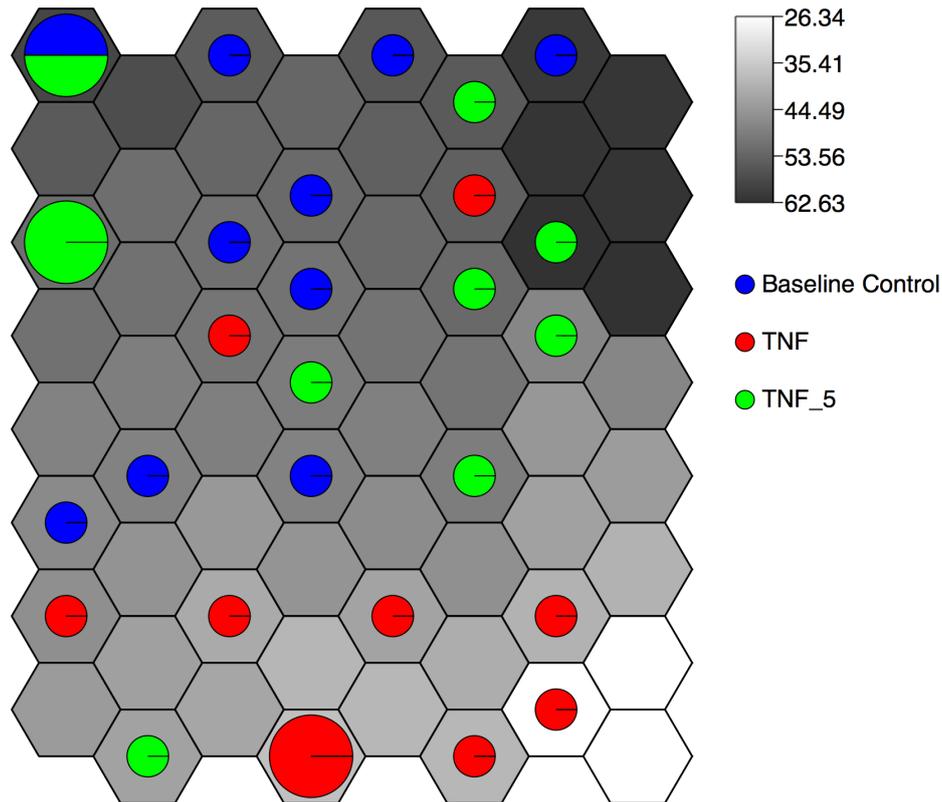
Figure 6.21: Kohonen Feature Map (Self-organising feature map) of TNF blockade at days 0 and 5 vs baseline control, clustered and visualised against the mean percentage of stroma per run with an FRC phenotype at T_{END}. A greater percentage of stromal area has adopted an FRC phenotype during baseline control than compared to the TNF blockade at day 0, however there is significant spread in FRC values for interventions initiated at day 5 with no meaningful distinction against the baseline control values. *n=10.*

6.5 Evaluating Efficacy of VLA-4 Blockade *In Vivo*

Following generation of the prediction that anti-VLA4 will reduce the population of FDC-like stromal cells, and thus reduce the capacity for the tissue to support B-cell follicles and germinal centre reactions, we collaborated with an external lab to evaluate the efficacy of VLA-4 blockade in the murine model of inducible submandibular TLT. It can be seen in Figure 6.22 shows the exocrine gland of a mouse treated at day 8 of the 15 day TLT formation process with A) phosphate buffer solution (PBS, the control group) and B) anti-VLA4 antibody. It can clearly be seen that there is major reduction in CD19+ cells (B cells) in the active sample as compared to the control. Further *in vivo* experimentation is in progress to provide additional validation of the model prediction, the results presented in this section should be treated as preliminary. The intervention procedure is illustrated at the top of Figure 6.22, with IHC micrograph of the murine salivary gland 15 days after pathology induction, showing control (A) and intervention with anti-VLA4 at day 8 (B).

For comparative purposes, the NeoSim model output showing FRC-like (pink) and FDC-like (teal) stromal cells at day 15, following anti-VLA4 treatment (*antiVLA4Level* = 0.99 and *antiVLA4TimePoint* = 1152 time steps) is presented in Figure 6.23. There is notable reduction in FDC-like stroma, and therefore a proportional reduction in the presence of B-cell follicles.

A qualitative assessment of these *in vivo* and *in silico* results suggests that the VLA-4 blockade is more successful *in vivo* than the *in silico* model predicts when the intervention occurs at day 8. Reasons for this could be related to the pharmacokinetic–pharmacodynamic properties of the antibody used, and also due to the lack of plasticity in the stromal cell differentiation model employed in NeoSim. That is, an FDC-like cell cannot 'regress' into an FRC-like phenotype under any circumstances in NeoSim, although it is possible that FDC-like stroma adopts an FRC-like phenotype if insufficient stimulation is received. There is presently insufficient
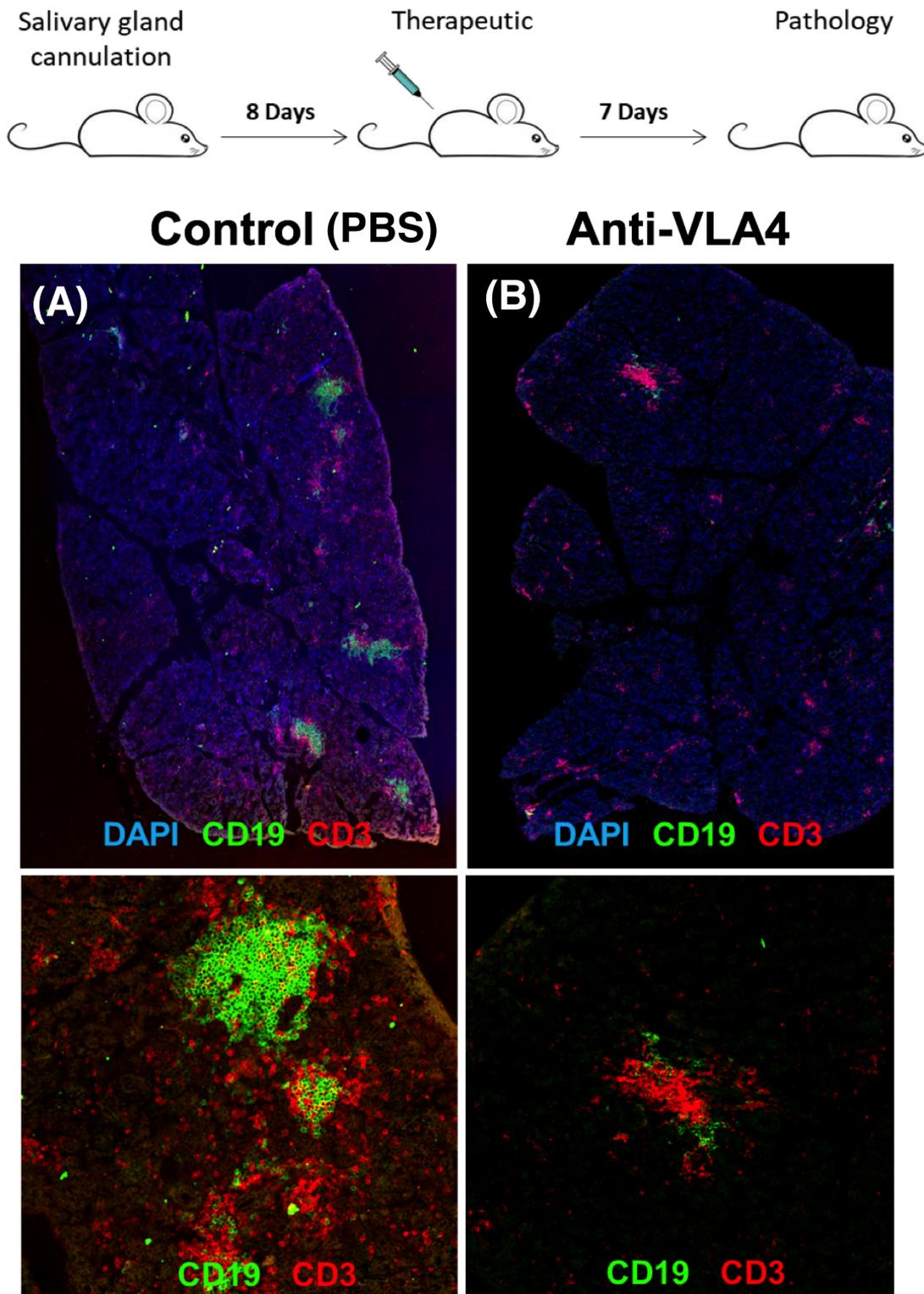
Figure 6.22: Murine model of Sjögren's syndrome treated on day 8 of 15 with (A) control, phosphate buffer solution (PBS), and (B) anti-VLA4 antibody, demonstrating efficacy of VLA-4 blockade in reducing lymphoid follicles (Data from **?**, experimentation performed by Christopher Buckley *et al*). CD19+ (green) are B cells, CD3+ (red) are T cells, DAPI (purple) stains all nucleic cells.
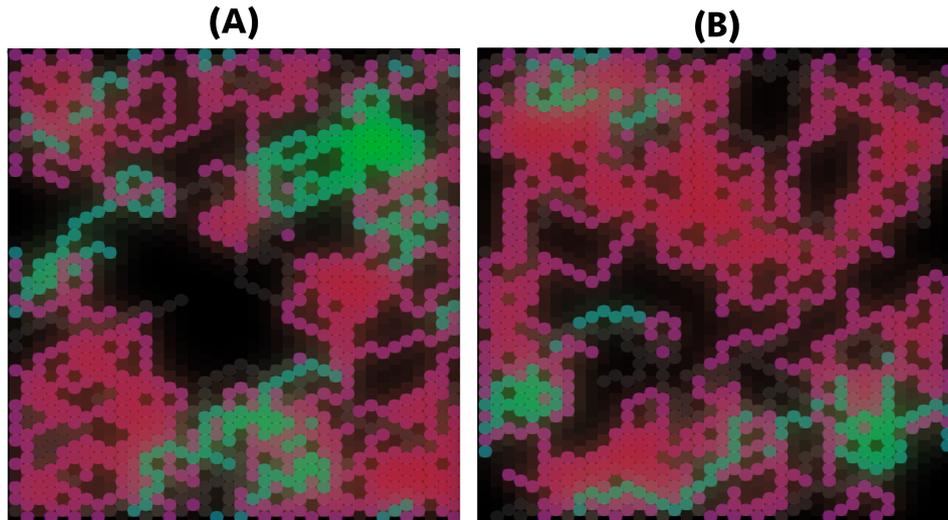
Figure 6.23: NeoSim run treated on day 8 of 15 with (A) anti-VLA4 (0.96), and (B) at baseline, demonstrating efficacy of anti-VLA4 in reducing lymphoid follicles. Red is S3/FRC-like Stroma, Green is S4/FDC-like Stroma.

evidence in the domain to determine whether this is the case.

6.6   Summary and Discussion

The best outcome does appear to result from the simultaneous and early combination of anti-TNF and anti-VLA4 therapies, however we lack safety information with regards to toxicology and complex non-linear effects when using multiple biologic therapies together, which could potentially result in severe immunosuppression. Anti-VLA4 alone is used in the treatment of irritible bowel disease (as natalizumab) and as such is known to have a reasonable safety profile. Anti-VLA4 has distinguished itself as a candidate therapy for the treatment of TLT pathology in Sjögren's syndrome, a model prediction with preliminary *in vivo* confirmation. Anti-VLA4 appears to result in a reduction in the relative population size of FDC-like stromal cells and an increase in FRC-like stromal cell population size. The use of Kohonen networks, U-Matrices and Feature Mapping/Component Plane Projection are found to be valuable tools in the analysis of higher-dimensional simulation result data, based on the highly preliminary analysis contained within this chapter using relatively small sample sizes.

Kohonen networks proved to be a valuable means of visualising the change in efficacy surrogate as dosage and intervention time-points were varied. Kohonen networks have received little attention in biomedical science or pharmacology, however in **?** they were used to visualise latent structures in pharmaceutical preparations, and the authors found that latent properties 'were clearly visualized by self-organising feature maps', and that Kohonen networks successfully clustered variables of interest into similar groups.

Future developments with regard to this technique include identifying latent classes of individuals based on their similarity, permitting clustering and classification based on the unique stochastic pathology of the model system, identifying latent properties such as disease severity through the use of disease surrogates. The use of Kohonen networks may offer an improvement over traditional means of evaluating the extent of the role of aleatory uncertainty has on system dynamics. Larger sample sizes, permitted through the availability of increased computational resources, will offer improved statistical power and potentially more nuanced clustering and sample stratification using U-Matrices.

Preliminary *in vivo* results validating the prediction that VLA-4 blockade is a potentially viable therapeutic intervention in Sjögren's syndrome, indicate that anti-VLA4 is indeed capable of significantly reducing the presence of B-cell follicular structures and has the potential to reduce disease activity. Further *in vivo* investigation is needed to assess the extent to which this occurs as compared to other biologic therapies recently identified as potentially ameliorating Sjögren's syndrome, such as rituximab (anti-CD20) (**?**).

Chapter 7

Discussion

The thesis has presented the development of 'NeoSim' by the present author. NeoSim is a hybridised complex systems model and simulation that successfully reproduces lymphocyte and stromal dynamics over a 15 day induction and tissue formation process. NeoSim provides a useful tool in quantitative systems pharmacology for the evaluation of various therapeutics for the treatment of TLT pathology and associated immune disorders: NeoSim comprises eight 'sub-models' including ordinary and partial differential equations, Markov chains, formal grammar, and other modelling techniques. The model is made comprehensible through use of the hybridisation schema notation developed in Chapter 4, and was successfully integrated into a software simulation tool. NeoSim has qualitatively reproduced clinical trial results with regard to anti-TNF and baminercept treatment, and provides strong theoretical evidence in support of anti-VLA4 biologics as an effective therapeutic strategy to disrupt the formation process. Furthermore, in vivo experimentation has validated that anti-VLA4 does reduce inflammatory activity and TLT formation in an acute murine model of Sjögren's syndrome, warranting further investigation due to the lack of current therapies in treating this disease.

This thesis began with the development of methods for emulating techniques in experimental biology, and these were applied in a case study to a pre-existing agent-based simulation of Peyer's patch formation. Following this, the focus has been upon TLT formation and potential means of disrupting this process. Therefore a biological

domain model describing the state of the art, best understanding of TLT formation was compiled, using the UML and additional modelling tools, in collaboration with domain experts and opinion leaders; this model was then translated into a Platform-Independent model with a formal specification and Simulation Platform written in Java. Finally, three candidate biologic therapeutics were evaluated for their efficacy in reducing FDC stroma population size with various dosing regimes and several predictions regarding TLT pathology and its treatments have been made.

A common rhetoric is that models ought to be made as simple as possible, but no simpler. This is not necessarily true of modelling methodology itself: it is important that the implementation of a model and its means of analysis are sufficiently complex to enable us to answer questions that are asked of the model. By determining the most appropriate representation for each aspect of the model individually, we can ensure that all abstractions made are for the sake of simplifying the biological domain to improve model usefulness, and not merely due to constraints placed on the system by the chosen modelling methodology. For instance, the model presented in Chapters 3 and 4 is very simple in some sense – it comprises three cell types, two diffusive chemokines, and a handful of intercellular signals. However, the implementation combines agent-based modelling with Markov chains, cellular automata, generative grammars, and ordinary/partial differential equations.

The development of a 'hybrid modelling paradigm' that formalised the hybridisation of different computational and mathematical modelling techniques enables abstraction for the purpose of increasing domain simplicity to improving model outcomes, and reduce the need to utilise abstraction as a means of increasing simulation tractability and feasibility at the cost of granularity at scales of interest and the capacity of models to answer the questions we wish to ask of them.

5 principle thesis aims were outlined in Section 1.5, these are re-produced below:

1. *Derive a hypothesis-driven model of tertiary lymphoid tissue (TLT) formation during autoimmune disease for use as a case study in applying the result of*

aims 2 and 3 below.

2. *Development of a schema and novel paradigm for developing a highly integrated hybrid multiscale model.*

3. *Development of improved methods for visualising simulation outputs that are analogous to those utilised in experimental biology.*

4. *Determine the veracity of the model hypothesis describing the minimum requirements for TLT formation.*

5. *Utilize the TLT formation model to evaluate the efficacy of therapeutic interventions for Sjögren's syndrome in silico.*

Chapters 3, 4 and 5 taken together satisfy aim 1 through the production of a software TLT simulator (termed 'NeoSim'). Chapter 4 satisfies aim 2 in providing clear descriptions of complex hybrid models. Aim 3 was met by the case study in Chapter 2 on the emulation of experimental techniques including flow cytometry and gene expression analysis. Aim 4 is met by virtue of the development of a functional simulation of TLT formation that confirms the veracity of our minimum requirements hypothesis. Aim 5 is satisfied by the substance of Chapter 6, in which five intervention strategies with three biologic therapeutics were evaluated, and one of those candidates, anti-VLA4, stood out as a particularly promising therapy with respect to efficacy in resolving TLT formation. This prediction was validated through preliminary in vivo investigation of anti-VLA4 in a murine model of Sjögren's syndrome.

Well-defined multi-scale models will have great capacity for capturing complex disease processes, ensuring every element of the model is appropriately captured without compromise due to constraints in the modelling methodology employed. This has important applications in developing models that can aid in understanding pathophysiology, and furthermore, evaluate the efficacy of therapeutic intervention strategies – both known and those predicted by the model. There is therefore great potential with the hybrid model paradigm to improve and accelerate the drug development process. This aspect is explored in Chapter 6, which replicates the results found in existing

Figure 7.1: This figure illustrates advancements made in applying the CoSMoS framework as shown in Fig 1.13 to modelling complex biological systems, incorporating large-scale model hybridisation and experimental technique emulation/visualisation. The framework consists of the iterative process of Domain-Platform-Simulation-Results modelling.

clinical trials, and proposes untested therapeutics that show a significant improvement over current therapies in Sjögren's syndrome through modulation of tertiary lymphoid tissue morphology and structure.

7.1 Summary of Contributions

This section briefly summarises the contributions made to the field of complex systems modelling in immunology and quantitative systems pharmacology as a result of the endeavours described within this thesis.

The CoSMoS modelling framework and supporting tools were presented in Figure 1.13. This process has been extended to include the additional visualisation, analysis and hybridisation tools and techniques that have been developed during

the production of NeoSim and this thesis; this is presented in Figure 7.1. The ovals show tools and techniques used to support the utilisation of this framework during biological model development: *i)* describing biological systems using the unified modelling language (**?**), *ii)* describing the evidence base and providing an argumentation structure that the model is fit for purpose using GSN (**?**) with *Artoo* (**?**), *iii)* statistical analysis of simulation results using the SPARTAN package (**?**) in R, *iv)* ODE/Kinetic model parameter sampling using ASPASIA (Dyson, et al. unpublished), *v)* model hybridisation through deconstruction of the model into component entities and reconstruction using the most applicable technique for that component in isolation (Chapter 4), *vi)* defining and developing model outputs using the hybridisation schema to combine data structures (Section 4.8), *vii)* emulating experimental techniques within the simulation platform and performing bioinformatics analysis of model outputs at the results model process stage (Chapter 2 and **?**), and *vii)* applying Kohonen networks and other deep learning strategies to evaluate complex simulation results and assess the effect of interventions such as efficacy evaluation for therapeutic agents. Contributions resulting from work contained in this thesis are highlighted in red in Figure 7.1.

7.1.1 Visualisation and Machine Learning in Bottom-Up Simulation Analysis

Models and simulations of complex biological systems are becoming increasingly sophisticated, and with the generation of ever more 'omics' data, the capacity for simulations to develop and test hypothesis *in silico* is ever-growing. However, the growth of computational resources and data availability has significantly outpaced the development of modelling methodologies and analysis techniques that are capable of maximally leveraging these resources. We present strategies for combining models that capture phenomena on different scales and at various levels of granularity into integrated 'overarching' models that can make effective use of computational resources. Furthermore, by taking inspiration from experimental biology, to develop

techniques for effective interrogation of the vastly complex 'results models' that modern simulations produce, and to improve the communicability of *in silico* models across multidisciplinary teams of computer scientists, engineers, biologists and clinicians, such that their results can achieve maximum impact. Visualisation is an important aid for both model analysis, and model communication – the importance of the latter cannot be understated – biological models presented using arcane statistical analyses and obtuse methodologies are unlikely to be fully understood by opinion leaders outside of computational modelling fields (i.e. computer science), increasing the likelihood that important results and observations may go unnoticed, Chapter 2 therefore proposes a standard for models being *effectively communicable* as well as being demonstrably fit-for-purpose and addressing relevant, well-defined research questions.

The utilisation of machine learning applied to high-dimensional biological data including sources such as emulated flow cytometry and gene expression analysis permits a more formally grounded link between the results model and the domain model. By transforming simulation results into a structure 'homomorphic' to those used in experimental biology. Due to the large number of simulation runs and data points involved in these types of stochastic simulation, interpreting FACS or gene expression data additionally requires the integration of data science techniques once the results have been generated in order to fully interpret the model. Kohonen networks have presented themselves as a viable means of stratifying simulation results according intervention-induced therapy, utilising Composite Planes to map clustered simulation data to simulation outputs (e.g. FDC population size change as a surrogate for GC reactions and level of auto-Ab response).

7.1.2 Tertiary Lymphoid Tissues: Formation and Therapeutic Interventions

Our hypothesis stated that TLT formation occurs as a 'self-organising' emergent structure in Sjögren's syndrome following priming of mesenchymal stroma due to

the presence inflammatory cytokines IL-13 and TNFα, mediated via two differential feedback loops of T and B chemokine expression induction, lymphocyte migration and further stimulation by T and B cells, causing development of lymphoid stromal phenotypes from tissue-resident progenitors. This was determined to be consistent insomuch that the cellular dynamics described in Chapter 3, when implemented as an executable simulation as described in Chapter 4, reproduced the 15-day development process. We consider this to be a powerful means of testing the veracity of hypotheses that cannot currently be tested with the best available technology. In addition to veracity confirmation, and reproduction of the negative results in clinical trials of efficacy of baminercept and anti-TNF in treating Sjögren's syndrome, predictions regarding unknown aspects of the domain can be tested experimentally using *in vivo* disease models — such as the efficacy of anti-VLA4 therapy in resolving TLT pathology based on *in silico* experimentation, and for which *in vivo* experiments are currently underway to evaluate this.

Several inferences and predictions can be made based on the *in silico* therapeutic interventions applied to NeoSim. Anti-VLA4 presents itself as a potentially viable candidate for the treatment of TLT pathology in Sjögren's syndrome, showing a significantly greater reduction in FDC population size by reducing lymphocyte-stroma cross-talk, which we hypothesise restricts the capacity of the pathophysiological system to develop a functional auto-antibody response and therefore cause tissue damage. Neither baminercept or anti-TNF presented themselves as viable candidates due to lack of evidence of efficacy, a finding supported by clinical trials. We found that delayed treatment of anti-VLA4 can still provide a significant benefit, especially at higher doses. There is a strong dose-response relationship in terms of the relationship between anti-VLA4 dose and the resulting reduction in FDC-like stroma. It was found that adding anti-TNF to anti-VLA4 therapy *in silico* provides a minor additional benefit, most likely by suppressing the total number of stromal cells available to adopt a lymphoid phenotype. Toxicology studies may be necessary to determine the safety profile of anti-VLA4 biologics at higher doses, and in

particular when they are used in combination with anti-TNF, as potent immuno-suppression with clinical manifestation of symptoms related to immunodeficiency would be a significant possibility.Anti-TNF therapy provides no meaningful benefit in isolation unless at unrealistically high doses, given at the same time as TLT induction. No reduction in FDC population was found in delayed interventions for anti-TNF.

NeoSim is constructed from a theoretical model of TLT formation, as described in Chapter 3, which itself is derived from limited qPCR and flow cytometry analysis of *in vitro* human mesenchymal stem cells stimulated with inflammatory cytokines and lymphocytes. NeoSim therefore effectively demonstrates that the veracity of theoretical models of pathophysiological processes can be determined with unidentifiable models that utilise 'sloppy' parameters. The *in vivo* confirmation of the prediction that anti-VLA4 ablates B-cell follicular structures additionally demonstrates that such models can also have significant predictive capacity.

7.1.3   Model Hybridisation: Extending the Limitations of Simulation

Presently, limited model hybridisation has taken place in which agent-based models and differential equation systems have been integrated explicitly (**?**), however this has been on a smaller scale and without any formalised means of describing the interconnections between each 'sub-model' that comprise the overarching model. With the hybrid models described in **?**, **?**, and **?**, for example, hybridisation is applied along compartmental lines - that is, a model may contain multiple tissue compartments, and each may be modelled using a different mathematical technique – whereas in the hybridisation framework proposed within this thesis, hybridisation occurs based on the 'de-composition' of the domain model into its constituent components separately described using the most appropriate technique given our research questions. As the complexity of hybridisation increases, maintaining the communicability of models will require the establishment of such a formalised

methodology. The motivation for the work underlying this thesis is therefore fundamentally to ensure that model development and analysis techniques keep pace with improvements in data availability and computational resources, and that one can continue to ask complex questions of complex systems, by enabling multi-scale hybrid model development that utilises an arbitrarily large number of methodologies to construct the model, while maintaining the ability to communicate the model platform in terms of each individual model and its connections to others.
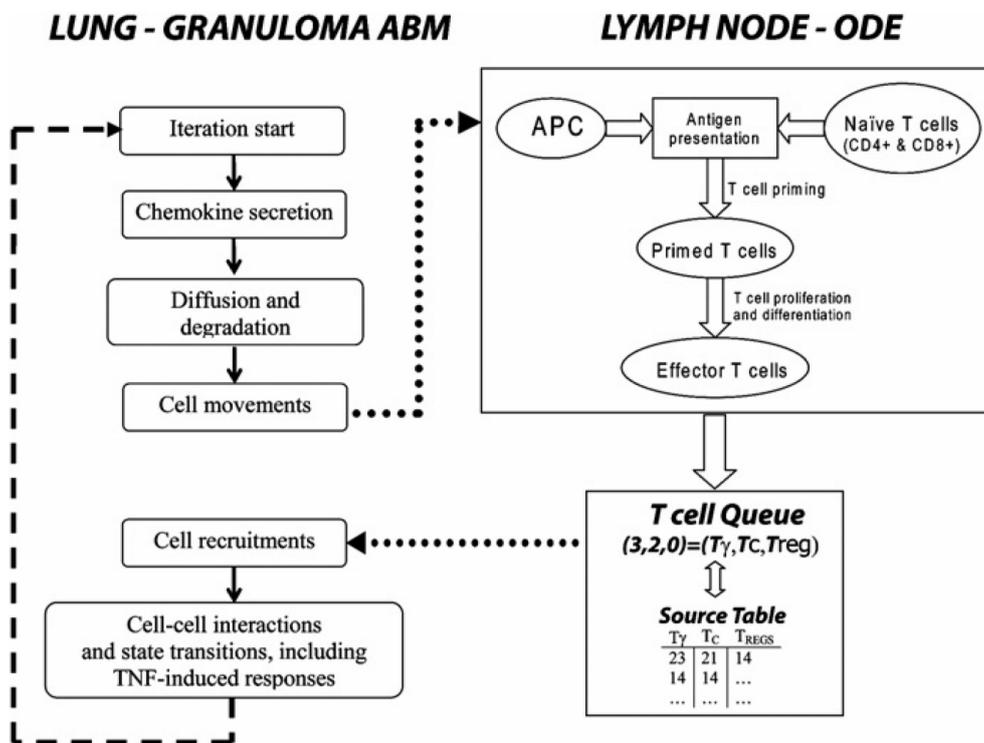
Figure 7.2: Figure describing hybridisation in the tuberculosis model described in **?**. Whilst the diagram used here successfully communicates the two spatial compartments and the mathematics that connect them, it does not clearly delineate separate models in terms of data flowing between the ABM and ODE-system. Figure taken from **?**.

This thesis addressed the need for a flexible schema that permits the hybridisation of multiple biological 'sub-models' constructed using various modelling techniques, which may address phenomena occurring on different time and length scales, into an integrated model and executable simulation. Furthermore, we provide strate-

Model Hybridisation Schema: Granuloma Formation from Marino et al. (2011a)
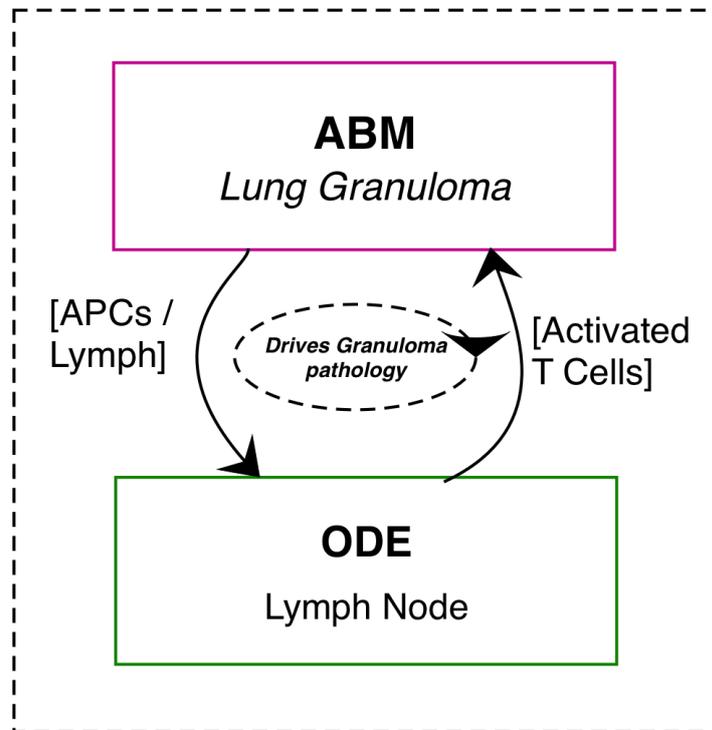
Figure 7.3: Figure describing hybridisation in the tuberculosis model described in **?**. This diagram is much simpler as ABM specifics can be described elsewhere using the UML but communicates the same process of positive feedback driving disease pathology.

gies for the development of more sophisticated results models with strong links to the domain, through the production of novel simulation outputs that are grounded in — but go beyond — the methodologies of experimental biology. Through the combination of novel hybridised, multiscale models and results analogous to experimental methods, the ultimate aim is to derive a robust framework for the development of *in silico* models of highly complex biological phenomena and disease processes that address processes occurring from the molecular to the tissue level; such models then present themselves as useful scientific tools in medicine and pharmacology for the evaluation of intervention strategies against complex disease processes.

7.1.4 Generalisation of Methodologies Developed

This thesis has only been concerned with the application of the methodologies discussed within the biological domain. The quantification and visualisation techniques developed in Chapter 2 are only applicable to models ultimately grounded in results from experimental biology or biomedical sciences. However, the concept of emulating experimental methods from the domain-under-study to improve model communicability and to derive further insights into model dynamics, should extend to other domains. For example, molecular dynamics simulations could theoretically be extended to produce outputs utilised in physical chemistry such as nuclear magnetic resonance and crystallography. Such considerations are beyond the scope of this thesis.

Regarding the hybrid modelling paradigm developed in Chapter 4, this approach to modelling easily generalises in a domain-agnostic manner to any system sufficiently complex to warrant the integration of varied mathematical and computational techniques in order to achieve a fit-for-purpose description of the domain. Indeed, the generalised schema introduced in Figure 4.2 (Section sec:hybridparadigm) is domain-agnostic and applicable to any domain for which such complex-systems modelling is deemed an appropriate tool.

NeoSim itself is specifically intended to describe the formation process of TLT in Sjögren's syndrome, with the intention of determining the veracity of the theory of TLT formation derived in Chapter 3, and to investigate potential avenues for therapeutic intervention that can disrupt this formation process as explored in Chapter 6. NeoSim is designed to describe the process in a theoretical sense, and not to generalise to specific patients. Such *in silico* experimental capacity would be dependent upon data availability across a large cohort of patients on a scale that is presently not feasible. The methodologies applied to develop NeoSim, however, generalise to any pathophysiological process for which consistent theories and hypotheses can be derived, even if data available for parameterisation are relatively sparse.

7.2   Future Developments and Further Experimentation

The NeoSim software implementation will be made publicly available at the York Computational Immunology Lab website at `www.york.ac.uk/ycil`. Two key future developments, and a natural progression from the work described within this thesis, are software tools related to the modelling of therapeutic intervention strategy efficacy and to the production of data structures and outputs conducive to analysis from a bioinformatics, or 'top-down' perspective. Firstly, as illustrated in Figure 7.4, a proposed, domain-specific software tool-kit that can utilise reflection to determine the molecular processes included within a model, then produce and evaluate therapeutic interventions that provide appropriate agonistic or antagonistic responses. An informal proposed schema is presented in Figure 7.4
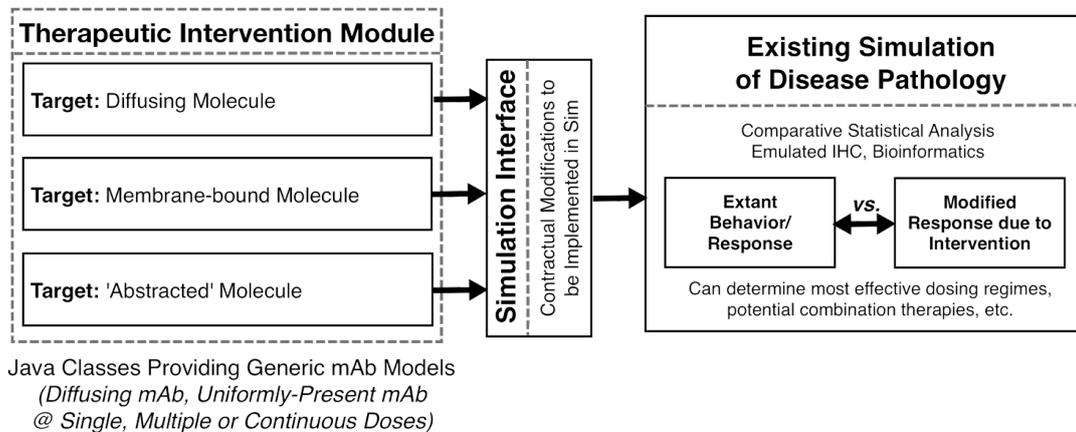
Figure 7.4: Proposed software module for automated incorporation of intervention strategies into hybrid complex systems-based models of different diseases under study. Automated incorporation of intervention strategies into hybrid complex systems-based models and simulations of any pathology of interest.

Secondly, we propose that the collection of methods and scripts utilised in Chapter 2 be developed and refined into a broad toolkit that permits *in situ* visualisation and emulation of experimental techniques. Outputs from this would be useful in providing a bridge between the abstract statistical evaluation of effect magnitudes

measured by Spartan, and the 'wet-laboratory' investigation that led to 'top-down' high-dimensionality data collection. This is depicted in Figure 7.5.
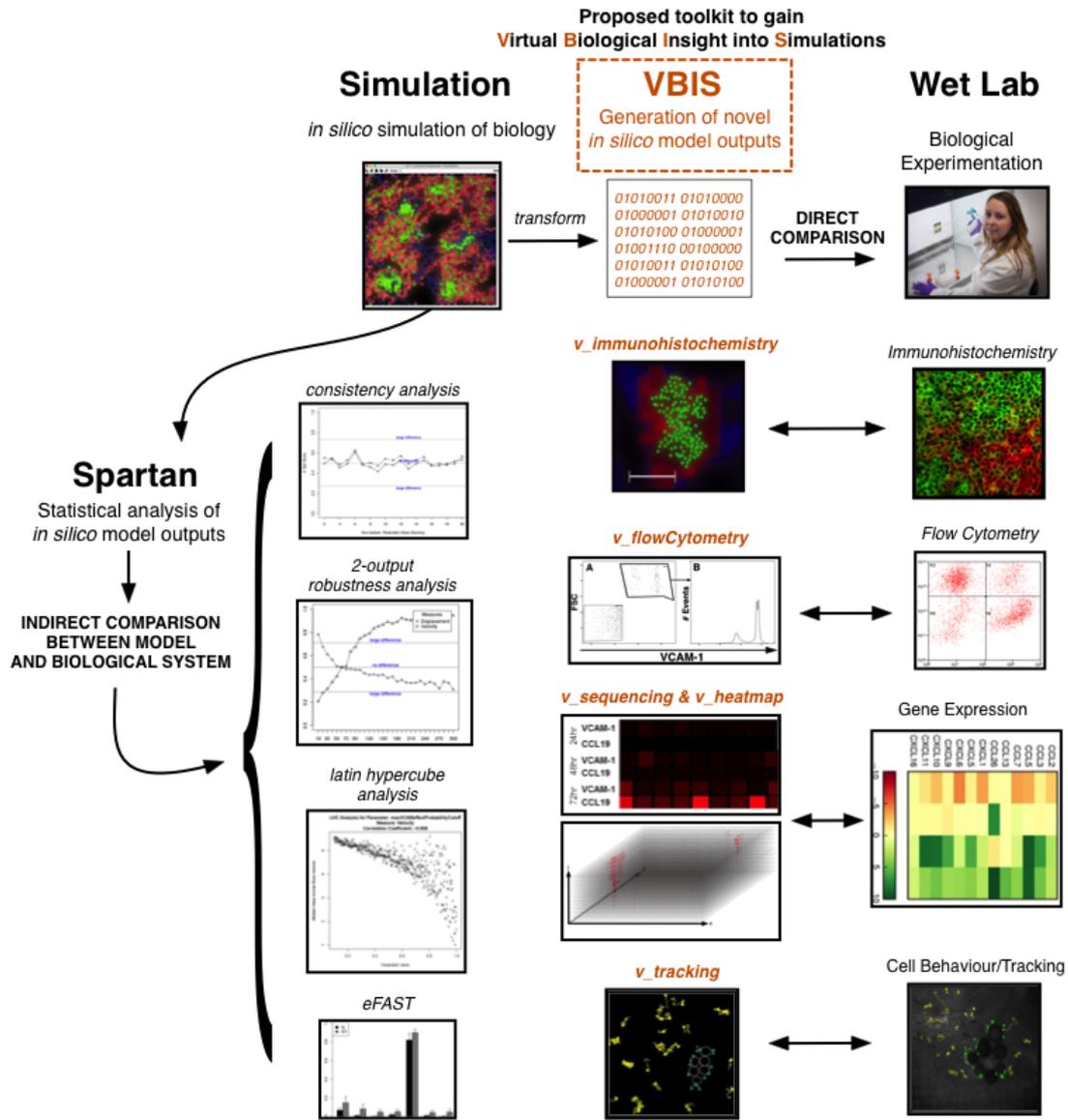
Figure 7.5: Proposed software toolkit for automated bioinformatic data structure generation and high-throughput image analysis. Note that parameterisation of the 'Simulation' track of this figure is dependent upon the specific modelling methodology and data availability for the domain-under-study.

Appendices

Appendix A

Model Summary

This section contains a concise summary of *NeoSim*, a Platform Model describing the minimum requirements for TLT formation in Sjögren's syndrome. The specifics of each sub-model, including the relative time-scales they operate on, are described in each model's section herein.

A.1   Hybrid Model Summary

This section summarises the hybrid nature of the TLT formation platform model developed in Chapter 4, showing how model components are integrated at a schematic level, and the integrated class diagram of the resulting software simulation.
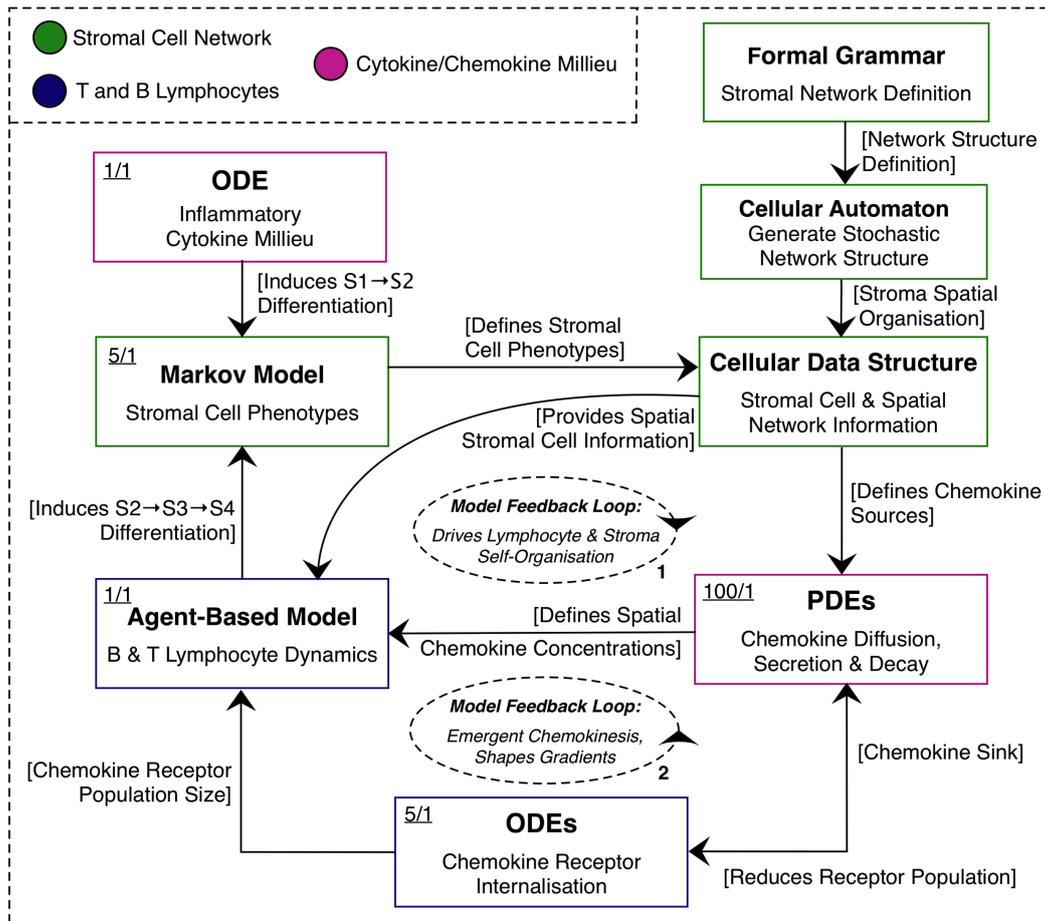
Figure A.1: Simplified hybridisation scheme demonstrating how each of the seven sub-models that comprise the platform model share information to represent TLT development as a whole. Underlined numbers indicate the number of time-steps that occur within that sub-model with respect to other sub-models, relative to the agent-based model, labelled '1/1'. For example, the PDE iterates 100 times while the remainder of the system is held quasi-static, and as such is labelled '100/1'.
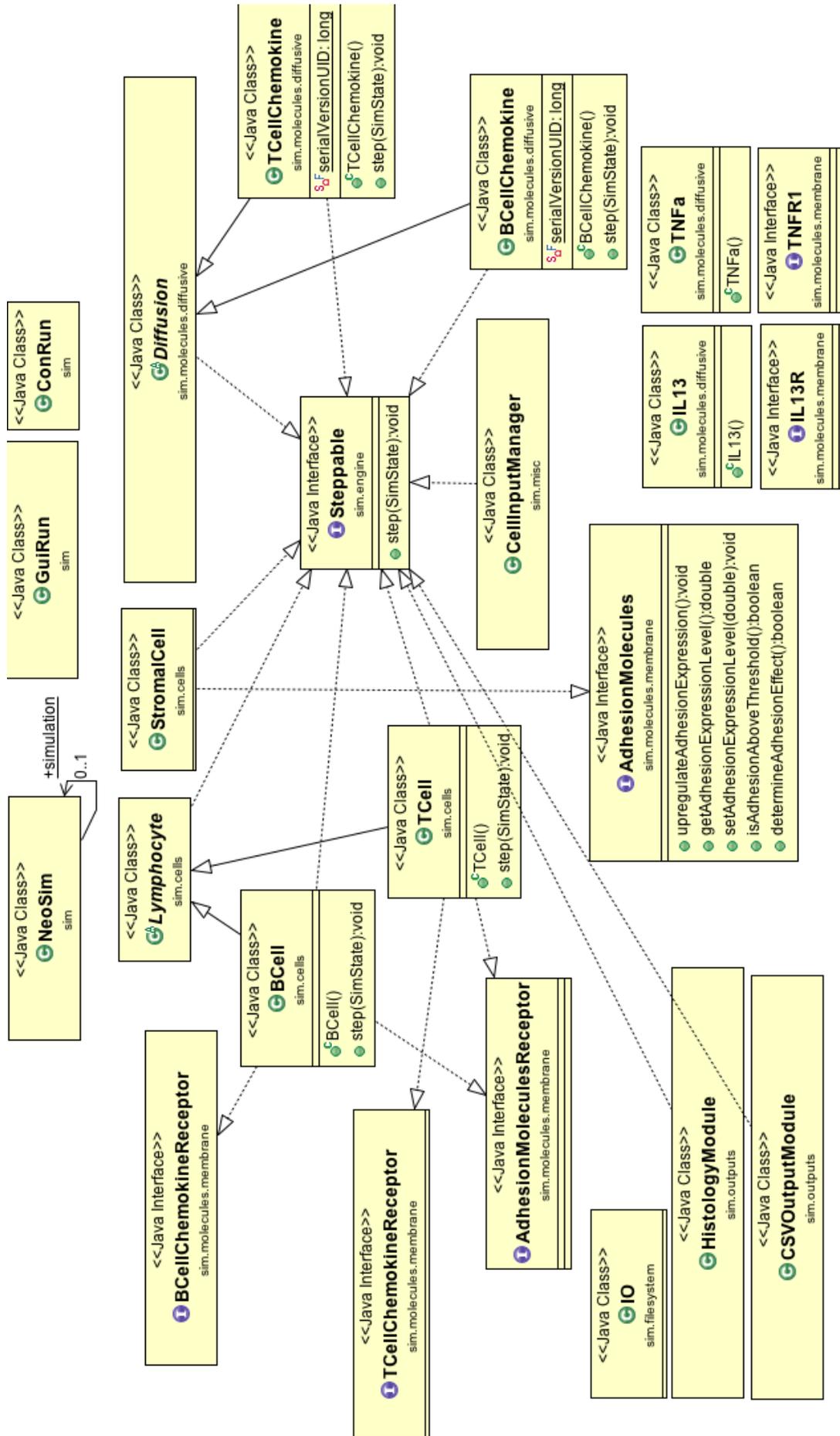
Figure A.2: NeoSim UML Class Diagram showing TLT formation model implementation as a software simulator in Java.

A.2 TLT Platform Model State Machines

These UML state machine diagrams describe the behaviour of T cells, B cells and Stroma. For further exposition on the T and B cell state machines the reader is referred to their full description in Section 4.3.1.1. For the stromal cell, which is implemented in the simulator within the context of an agent-based model, but formally defined as a Markov model, the reader is referred to section 4.7.3.

Figure A.3: UML Finite State Machine Platform Model diagram describing the role of T lymphocytes in TLT formation.

B Cell | TLT Platform Model | State Machine Diagram

/ enter TLT after ~ day 5 (time step 720)

/ next time step

[minimumChemokineExpression >
bCellChemokineThreshold for at least
one neighbouring grid space]

?else?

currentlyChemotactic =
True

currentlyChemotactic =
False

/ single chemotactic movement (10 sub movements)
/ update CXCL13 chemokine receptor recycling ODE

/ single random movement
composed of 10 sub-
movements per time-step

/ bccaGrid.bCellChemokineAtCurrentPosition -=
(expressionLevelThisChemokineReceptor*maxMolecules*$internalisationScalar$)

?else?

[∃ This.BCell s.t. center within StromalCell$_{(x,y)}$

& currentAdhesionLevel > adhesionLevelThreshold

& random < probabilityOfAdhesion(StromalCell.AdhesionMoleculeExpressionLevel)]

adhesionEffect = True

Localised movement around
bound stromal cell

/ adhesionEffect == False, When:
[PerSubStep(random <
probabilityOfEscape(StromalCell.AdhesionMoleculeExpressionLevel)
&timeAdhered = minimumAdhesionTime))]

canExpressS3F =
True

canExpressCXCR5 = True

/ CXCR5ExpressionLevel = 1

[p(canStimulateSPC)]

[B cell is unbound to Stromal Network
& time step = 2161 (15 simulated days)]

Figure A.4: UML Finite State Machine Platform Model diagram describing the role
of B lymphocytes in TLT formation.
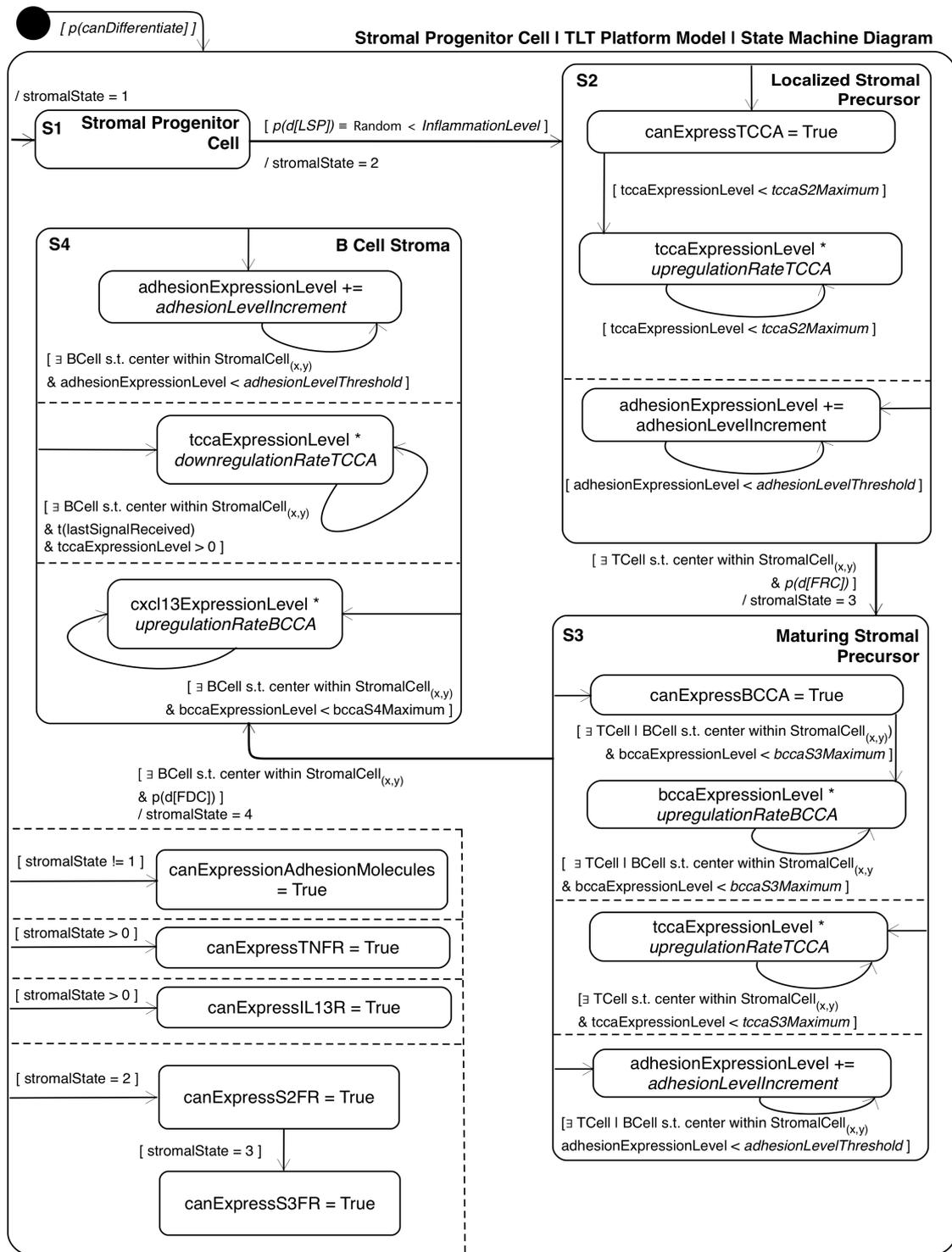
Figure A.5: UML Finite State Machine Platform Model diagram describing the role of stromal precursor cells in TLT formation, including differentiation into FRC-like and FDC-like phenotypes.

A.2.1   Stromal Cell Markov Model Description

The Markov Model defining the Platform description of Stromal Cells is illustrated in A.6

The table below (A.1) defines the parameters used in this Markov Model.

Parameter Name	Value	Hybridised With	Monte Carlo Method Triggered By	Event Time-Scale
p(d[LSP])	$I_0 e^{-\lambda \Delta T}$	CA	Stromal CA evaluating ODE	10 Minutes
p(d[FRC])	0.003	ABM + CA	T-Cell adhered to Stroma	2 Minutes
p(d[FDC])	0.000001	ABM + CA	B-Cell adhered to Stroma	2 Minutes
1 - p(canDifferentiate)	0.2	N/A	Initiated at T=0	Once

Table A.1: Table defining the stromal differentiation Markov model probabilities and the sub-model with which each is hybridised. Interactions with the ABM and CA dictate when transitions will be evaluated using Monte Carlo methods for each stromal cell within the simulation model. The numerical probabilities provided were determined through calibration.
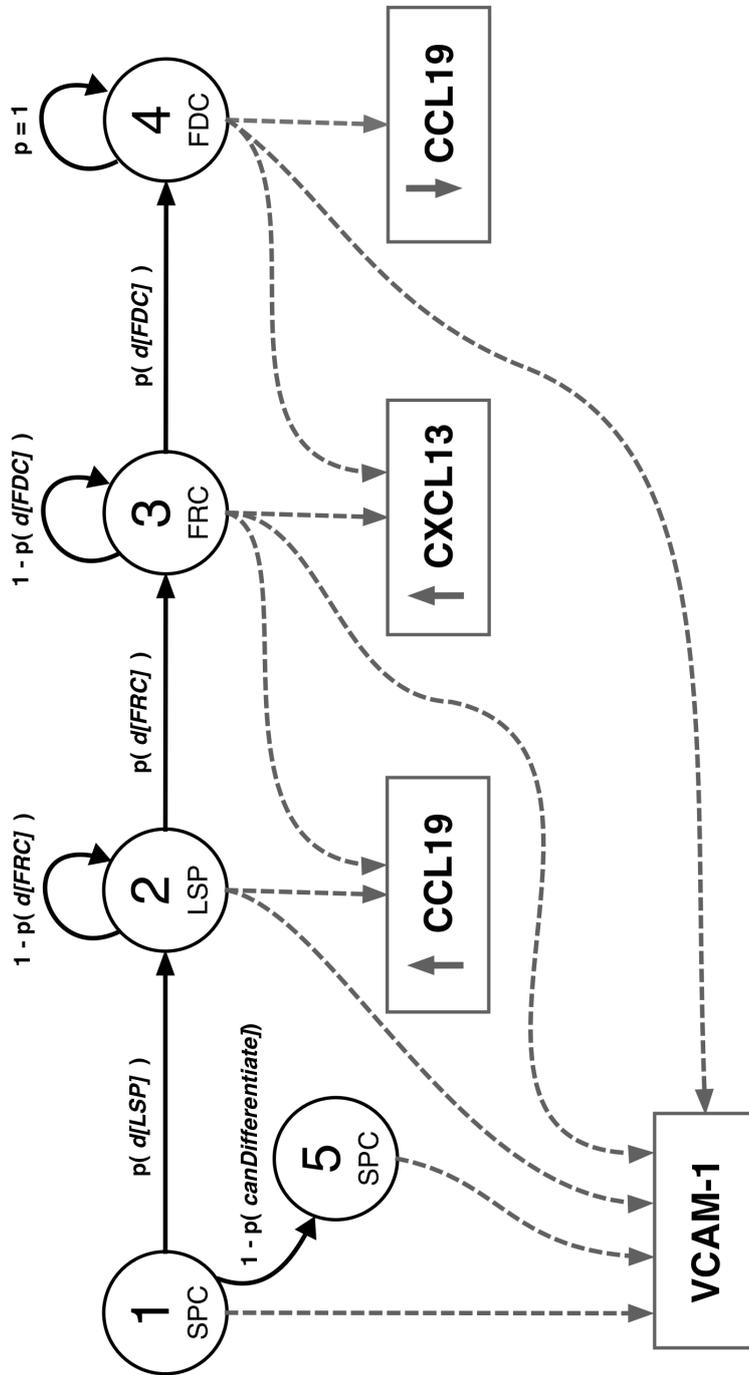
Figure A.6: Markov chain describing transitions between stromal phenotypes, and the 'emissions' each state is responsible for; these correspond to the soluble and membrane-bound proteins produced by each phenotype. The Markov chain is hybridised with an agent-based model and Monte Carlo probability sampling to determine when state transitions occur.

A.3   Chemokine Diffusion

Discretised chemokine diffusion equation (equation A.1 as implemented for T-cell chemokines (CCL19) and B-cell chemokines (CXCL13). The diffusion of chemokines operates on a different time-scale to the T cell, B cell and Stromal cell agents, the rest of the model is treated as quasi-static while the diffusion equation is iterated 100 times, to account for the approximately two orders of magnitude difference between molecular and cellular motion:

$$\phi_{x,y}(t+1) = (((\phi_{x,y}(t) - \frac{6D\Delta t \cdot \phi_{x,y}(t)}{\Delta x^2}) + \frac{D\Delta t}{\Delta x^2} \sum_{i=1}^{i=6} \phi_i(t)) - \phi_i(t)e^{-\lambda\Delta t}) + s \quad (A.1)$$

$\phi_{x,y}(t+1)$ is the chemokine concentration at point (x,y) at the following time-step from $\phi_{x,y}(t)$. Δt is the difference in time between t and $t+1$. D is the diffusion co-efficient, and λ is the decay constant. The secretion rate of new chemokines by stromal cells is defined as s, the value is dependent on the number of contacts each stromal cell has had with lymphocytes and is defined in the state machine diagrams in Section A.2, defined as cxcl13ExpressionLevel and tccaExpressionLevel for B cell and T cell chemokines respectively. This discretised equation operates on a 40x40 hexagonally-packed square grid, the '6' thus refers to the six boundaries each hexagon shares with its neighbours. The grid is non-toroidal, and chemokines that diffuse beyond the grid are removed from the simulation.

A.4   Lymphocyte Chemokine Receptor Internalisation

B and T cells express receptors for CXCL13 and CCL19, termed CXCR5 and CCR7 respectively. When chemokines bind to these receptors, they cause the lymphocytes to move in the direction of the chemokine gradient, and the receptor-ligand (CXCL13-CXCR5 and CCL19-CCR7) is internalised into the cell. This section

describes the discretised equations implemented on T and B cells to model and simulate this phenomenon. For the ODE derivation and detailed exposition the reader is referred to Section 4.5.

Equation A.2 describes the change in chemokine receptor expression levels on lymphocytes. $r(t)$ is the cell-surface expression level of chemokine at time t. ψ is the number of receptors that have been internalised and degraded, σ is the receptor quantity recycled, s represents the addition of chemokine receptors to the intracellular pool due to the cell's production of new receptors. Receptor values are feature-scaled (normalised to the range [0,1]) where 0 is no receptors and 1 is the maximum possible number of receptors. This permits parameterisation in the absence of biological data measuring precise cell-surface numbers of receptors. n is a scaling parameter to modulate the rate of internalisation by calibration to observed phenomena *in vivo* due to lack of data regarding the rate of recycling, and g is a random number in the range [0,1] sampled from a Gaussian distribution using the Mersenne twister (**?**) psuedo-random number generator. q is the local chemokine level, equal to $\phi_{x,y}(t)$ at the location (x, y) of the T or B cell, calculated from the chemokine diffusion equation described in Section A.3.

$$r(t + 1) = \left(r(t) - \psi \frac{gnqr(t)}{Q} \right) + \sigma + S \qquad (A.2)$$

Internalised chemokine needs to be removed from the environment. Equation A.1 handles the secretion, diffusion and decay of chemokine, but not loss due to internalisation. This is instead achieved by determining the total number of chemokine molecules that have been internalised, I, shown in Equation A.3:

$$I = \psi \frac{gnqr(t)}{Q} + \sigma \frac{gnqr(t)}{Q} \qquad (A.3)$$

This quantity is then removed from the hexagonally-packed square grid by performing the calculation described in equation A.4 below. Let c be the change in chemokine level due to diffusion and stromal cell secretion, defined by the PDE

in Section 4.4. This effectively couples the receptor internalisation and recycling ODE-derived model with the chemokine secretion, diffusion and decay PDE-derived model described above in Section A.3

$$q(t + 1, x, y) = q(t, x, y) + c - I \tag{A.4}$$

A.5   Chemotaxis Model

When a T or B cell is in the state *'currentlyChemotactic'* as described in Figures A.3 and A.4, the random motion is biased towards the direction of the gradient of CCL19 (T Cell chemokine) and CXCL13 (B cell chemokine), respectively. This is achieved by constructing a probability distribution from the chemokine gradient, such that the probability a cell will move in one of the six directions of the underlying hexagonal grid that contains chemokine concentration data is defined in equation A.5

$$
\begin{aligned}
&p(\text{Cell moves up}) = c_{\text{up}}/C \\
&p(\text{Cell moves down}) = c_{\text{down}}/C \\
&p(\text{Cell moves upper left}) = c_{\text{upperleft}}/C \\
&p(\text{Cell moves upper right}) = c_{\text{upperright}}/C \\
&p(\text{Cell moves lower left}) = c_{\text{lowerleft}}/C \\
&p(\text{Cell moves lower right}) = c_{\text{lowerright}}/C
\end{aligned}
\tag{A.5}
$$

Where c_{up} is the concentration of chemokine in the hexagon above the hexagonal element the cell currently lies in, c_{down} is the concentration of the hexagonal element below, and so on. As lymphocytes exist on their own, continuous space grid, once the general direction based on the chemokine concentration has been defined, a specific angle of motion needs to be determined. Each of the six directions contains a range 60 degrees (360/6), thus a random angle is selected between 0 and 60 degrees as described in Figure A.7
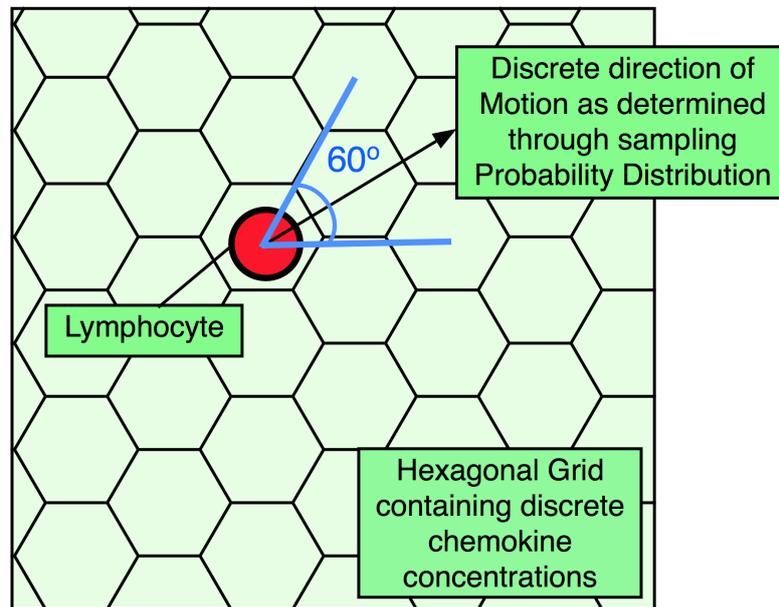
Figure A.7: Figure illustrating the selection of direction for continuous Lymphocytes following determination of the discrete direction to move based on the hexagonal grid that contains chemokine concentrations. The lymphocyte model entity selects at random an angle in the 60 degree range of possible directions once a particular hexagon has been selected through probabilistic sampling according to equation 4.1

Appendix B

Model Parameters

This appendix details the Domain Model and Platform Model parameters, their values, and sources.

B.1   Domain Model

This section summarises the parameters contained within the domain model.

Parameter	Model Entity	Value	Description
T Cell Entry Time	Lymphocytes	Day 0	Time point during TLT formation that T cells colonise tissue
T Cell Entry Rate	Lymphocytes	--	Unknown, assumed to be linear
B Cell Entry Time	Lymphocytes	Day 8	Time point during TLT formation that B cells colonise tissue
B Cell Entry Rate	Lymphocytes	--	Unknown, assumed to be linear
T Cell Velocity	Lymphocytes	10.7 µm/s	Mean observed value, from Miller (2003)
B Cell Speed	Lymphocytes	4.16 µm/m	Mean observed value, from Shulman (2014)
CXCR5 Surface Levels	Lymphocytes	1.00E+06	(Approximately on the order of, measured in lab)
CCR7 Surface Levels	Lymphocytes	--	Unknown
CXCR5 Recycling Rate	Lymphocytes	--	Unknown
CCR7 Recycling Rate	Lymphocytes	--	Unknown
S2 CCL19 Expression	Stroma	0 (RQ)	CCL19 Expression by Localised Stromal Precursors (S2) **(Fig 3.3)**
S3 CCL19 Expression	Stroma	800 (RQ)	CCL19 Expression by Localised Stromal Precursors (S3/FRC) **(Fig 3.3)**
S3 CXCL13 Expression	Stroma	~100,000 (RQ)	CXCL13 Expression by Maturing Stromal Precursor (S3/FRC) **(Fig 3.3)**
S3 VLA4 Expression	Stroma	~400,000 (RQ)	VLA4 Expression by Maturing Stromal Precursor (S3/FRC) **(Fig 3.3)**
S4 CCL19 Expression	Stroma	~25 (RQ)	CCL19 Expression by Localised Stromal Precursors (S4/FDC) **(Fig 3.3)**
S4 CXCL13 Expression	Stroma	~200 (RQ)	CXCL13 Expression by Maturing Stromal Precursor (S3/FDC) **(Fig 3.3)**
ICAM1 Expression	Stroma	2200 (MFI)	ICAM1 Expression by stimulated stroma **(Fig 3.4)**
VCAM1 Expression	Stroma	300 (MFI)	VCAM1 Expression by Maturing Stromal Precursor (S3/FDC) **(Fig 3.4)**
S2-S3 T cell Stim Reqs	Stroma	--	Requirements for differentiation from S3-S3, Unknown
S3-S4 B cel Stim Reqs	Stroma	--	Requirements for differentiation from S3-S4, Unknown
Gland Stroma Density	Stroma	--	Characteristics of stroma in submandibular gland, Unknown

Figure B.1: Table of identified Domain Model parameters.

B.2 Platform Model

Parameter (NeoSim Name)	Model Entity	Description	Value	Determined
tCellBegin	Simulation Initial Conditions	Agent time-step at which T cells begin to enter tissue compartment	0	*In Vivo - Bombardieri (2012)*
tCellCease	Simulation Initial Conditions	Agent time-step at which T cells cease to enter tissue compartment	720	*In Vivo - Bombardieri (2012)*
tCellInputPopulationSize	Simulation Initial Conditions	Number of T cells to be added to simulation in total, via linear increment between tCellBegin and tellCease.	8000	Arbitrary Input
bCellBegin	Simulation Initial Conditions	Agent time step at which B cells begin to enter tissue compartment	720	*In Vivo - Bombardieri (2012)*
bCellCease	Simulation Initial Conditions	Agent time step at which B cells cease to enter tissue compartment	1152	*In Vivo - Bombardieri (2012)*
bCellInputPopulationSize	Simulation Initial Conditions	Number of B cells to be added to simulation in total, via linear increment between bCellBegin and bCellCease.	8000	Arbitrary Input
elementsPerCell	Simulation Initial Conditions	(Generative Grammar, 'B') Number of hexagonal grid elements to place per branch stromal cell	24	Calibrated by hand via confocal IHC imaging
fractionOfGridToCover	Simulation Initial Conditions	(Generative Grammar) To determine total number of Stromal Cells to place in GG; expressed as product of total grid area to hold stromal cells	1.4	Calibrated by hand via confocal IHC imaging
beginNewNetwork	Simulation Initial Conditions	(Generative Grammar, 'N'), number of connected stromal cells to place before new branch, must be multiple of sqrt(n)	6--12 (12)	Calibrated by hand via confocal IHC imaging
beginNewCell	Simulation Initial Conditions	(Generative Grammar, 'n') Number of connected branches defined as 1 stromal cell.	4--24 (12)	Calibrated by hand via confocal IHC imaging
initialCCL19Expression	Simulation Initial Conditions	Inflammatory trigger for T-Cell Chemotaxis	6	Fig 3.3 *In Vitro* Data

Figure B.2: Table of 'System-wide' parameters including those defining model initial conditions.

Parameter (NeoSim Name)	Model Entity	Description	Value	Determined
maxChemokinePerElementB	CXCL13	Maximum concentration of chemokine in each discrete element of the CXCL13 hexagonal grid.	30000	Dummy parameter
diffusionCoefficientB	CXCL13	Diffusion Coefficient, D, for Chemokine PDE (CXCL13)	4E-14	Calibrated to Expected Behaviours*
bccaDecayConstant	CXCL13	Decay Constant, Lamda, for Chemokine PDE (CXCL13)	0.0005	Calibrated to Expected Behaviours
sufficientBChemokineThreshold	CXCL13	Minimum concentration of chemokine to which a B cell will respond chemotactically	0.02	Arbitrary low value required, Prevents chemotaxis with negligable presence of chemokine
maxChemokinePerElementT	CCL19	Maximum concentration of chemokine in each discrete element of the CCL19 hexagonal grid.	30000	Dummy parameter
diffusionCoefficientT	CCL19	Diffusion Coefficient, D, for Chemokine PDE (CCL19) -- Einstein-Stokes would be 0.000000000158.	4E-14	Calibrated to Expected Behaviours*
tccaDecayConstant	CCL19	Decay Constant, Lamda, for Chemokine PDE (CCL19).	0.0005	Calibrated to Expected Behaviours
sufficientTChemokineThreshold	CCL19	Minimum concentration of chemokine to which a B cell will respond chemotactically	0.02	Arbitrary low value required, Prevents chemotaxis with negligable presence of chemokine

Figure B.3: Table of chemokine model-related parameters.

Parameter (NeoSim Name)	Model Entity	Description	Value	Determined
decayConstantInflammation	Stromal Cell	Decay constant for capacity of stromal progenitors to differentiate into a lymphoid stroma phenotype (lower value = more lymphoid stroma)	0.001	Calibrated to Expected Behaviours development timeline
probabilityS1S2PerTimeStep	Stromal Cell	Probability that any S1 becomes S2 each time step.	0.05	Calibrated to Expected Behaviours
inflammationLevel	Stromal Cell	*a priori* Probability that any S1 can become S2. This abstracts presence of inflammatory cytokines.	0.95	Calibrated to Expected Behaviours
bccaS3Maximum	Stromal Cell	Maximum number of cxcl13 molecules secreted per time step when S3 (Feature scaled).	0.2	Fig 3.3 *In Vitro* Data
tccaS4Maximum	Stromal Cell	Maximum number of tcca molecules secreted per time step when S4 (Feature scaled).	0.3	Fig 3.3 *In Vitro* Data
tccaS3Maximum	Stromal Cell	Maximum number of tcca molecules secreted per time step when S3 (Feature scaled).	1	Fig 3.3 *In Vitro* Data
tccaS2Maximum	Stromal Cell	Maximum number of tcca molecules secreted per time step when S3 (Feature scaled).	0.05	Fig 3.3 *In Vitro* Data
upregulationRateBCCA	Stromal Cell	CXCL13 upregulation rate stroma is in contact with lymphocyte (Feature scaled).	1.01	Rate assumed to be Linear
upregulationRateTCCA	Stromal Cell	CCL19 upregulation rate when stroma is in contact with lymphocyte (Feature scaled).	1.01	Rate assumed to be Linear
downregulationRateBCCA	Stromal Cell	CXCL13 down-regulation rate when stroma is in contact with lymphocyte (Feature scaled).	0.99	Rate assumed to be Linear
downregulationRateTCCA	Stromal Cell	CCL19 down-regulation rate when stroma is in contact with lymphocyte (Feature scaled).	0.99	Rate assumed to be Linear
probabilityS1S2	Stromal Cell	Probability that environmental factors cause S1 stroma to differentiate into S2 (Feature scaled).	1	Calibrated to Expected Behaviours development timeline
probabilityS2S3	Stromal Cell	Probability that environmental factors cause S2 stroma to differentiate into S3 (Feature scaled).	0.0004	Calibrated to Expected Behaviours development timeline
probabilityS3S4	Stromal Cell	Probability that environmental factors cause S3 stroma to differentiate into S4 (Feature scaled).	0.000005	Calibrated to Expected Behaviours development timeline
bccaMaxMoleculesPerStep	Stromal Cell	Number of absolute CXCL13 molecules secreted per time step per hexagonal element.	10000	Only for outputting absolute values, ODE model is relative & feature scaled.
tccaMaxMoleculesPerStep	Stromal Cell	Number of absolute ccl19/21 molecules secreted per time step per hexagonal element.	10000	Only for outputting absolute values, ODE model is relative & feature scaled.
initialAdhesionLevel	Stromal Cell	Percentage of maximum adhesion molecule expression at T0	0	Assumed that mesenchymal progenitors do not express adhesion factors prior to stimulation by lymphocytes
adhesionLevelIncrement	Stromal Cell	Upregulation increment of adhesion factor expression following lymphocyte stimulation.	0.01	Calibrated to Expected Behaviours development timeline
adhesionLevelThreshold	Stromal Cell	Threshold on adhesionLevel above which adhesion is possible.	0.65	Calibrated to Expected Behaviours development timeline

Figure B.4: Table of stromal Cell-related parameters.

Parameter (NeoSim Name)	Model Entity	Description	Value	Determined
initialPopulationSizeBCell	B Cell	Number of B cells present at T_0.	0	*In Vivo - Bombardieri (2012)*
chemokineConsumptionB	B Cell	True if B Cell Receptor Recycling ODE Model Enabled.	TRUE	(Receptor recycling ODE On/Off flag)
synthesisRateBCCA	B Cell	Specifies linear rate of synthesis of B-cell chemokine receptor.	3	Calibrated to Expected Behaviours
bChemotaxisWeighting	B Cell	Option to weight probability of B-cell motion direction of highest B-cell chemokine value by this factor. Abstraction of cellular persistence.	1.2	Calibrated to Expected Behaviours
maxBCCARMolecules	B Cell	Maximum number of B-cell chemokine receptor molecules on surface.	1000000000	Not used in model logic (but measured approximately in the lab)
minRadiusB	B Cell	Minimum B cell radius as proportion of one hexagon (10um^2).	0.625	Measured from *In Vivo* Micrographs
maxRadiusB	B Cell	Minimum B cell radius as proportion of one hexagon (10um^2).	0.375	Measured from *In Vivo* Micrographs
receptorInternalizationScalarB	B Cell	Value in range [0,1] defines the proportion of chemokines are internalised on average (steady state).	0.0001	Calibrated to Expected Behaviours
receptorRecyclingRateB	B Cell	Rate at which CXCL13 receptors are recycled back to cell surface.	0.8	Calibrated to Expected Behaviours
maximumRecyclingCacheB	B Cell	Maximum size of intracellular pool (range [0,1]).	0.75	Calibrated to Expected Behaviours
receptorProportionRecycledB	B Cell	Proportion of CXCL13 receptors internalised that are recycled.	0.5	Calibrated to Expected Behaviours
minimumAdhesionTimeB	B Cell	Minimum amount of time (in 1/10th time-steps) a stable B cell-Stroma bond can form for.	5	Value from *Alden (2012)*
initialChemokineReceptorExpressionB	B Cell	Expression of CXCL13 by B Cell at T_0 (Range [0,1])	1	Assumed lymphocytes begin with maximum expression of chemokine receptors
initialPopulationSizeTCell	T Cell	Number of T cells present at T_0.	50	Estimated number of tissue resident T cells prior to inflammatory signal
chemokineConsumptionT	T Cell	True if T-cell Receptor Recycling ODE Model Enabled.	TRUE	Flag to enable/disable internalisation ODE
synthesisRateTCCA	T Cell	Specifies linear rate of synthesis of T-cell chemokine receptor.	3	Calibrated to Expected Behaviours
tChemotaxisWeighting	T Cell	Option to weight probability of T-cell motion direction of highest B chemokine value by this factor. Abstraction of cellular persistence.	1.2	Calibrated to Expected Behaviours
maxTCCARMolecules	T Cell	Maximum number of B cell chemokine receptor molecules on surface.	1000000000	Not used in model logic (but measured approximately in the lab)
minRadiusT	T Cell	Minimum B cell radius as proportion of one hexagon (10um^2).	0.625	Measured from *In Vivo* Micrographs
maxRadiusT	T Cell	Minimum B cell radius as proportion of one hexagon (10um^2).	0.375	Measured from *In Vivo* Micrographs
receptorInternalizationScalarT	T Cell	Value in range [0,1] defines the proportion of chemokines are internalised on average (steady state).	0.0001	Calibrated to Expected Behaviours
receptorRecyclingRateT	T Cell	Rate at which CXCL13 receptors are recycled back to cell surface.	0.8	Calibrated to Expected Behaviours
maximumRecyclingCacheT	T Cell	Maximum size of intracellular pool (range [0,1]).	0.7	Calibrated to Expected Behaviours
receptorProportionRecycledT	T Cell	Proportion of CXCL13 receptors internalised that are recycled.	0.5	Calibrated to Expected Behaviours
minimumAdhesionTimeT	T Cell	Minimum amount of time (in 1/10th time-steps) a stable B cell-Stroma bond can form for.	5	Value from *Alden (2012)*
initialChemokineReceptorExpressionT	T Cell	Expression of CXCL13 by B Cell at T_0 (Range [0,1])	1	Assumed lymphocytes begin with maximum expression of chemokine receptors

Figure B.5: Table of lymphocyte-related parameters.

B.2.1   Therapeutic Intervention Parameters

Parameter	Description	Value	Comment
antiVLA4Level	Dose of anti-VLA4	0.8	Medium Dose (feature scaled units)
antiVLA4Level	Dose of anti-VLA4	0.96	High Dose (feature scaled units)
antiVLA4Level	Dose of anti-VLA4	0.99	Very High Dose (feature scaled units)
antiVLA4timepoint	Time-step to introduce anti-VLA4	0	Day 0
antiVLA4timepoint	Time-step to introduce anti-VLA4	720 (10 min steps)	Day 5
antiTNFLevel	Dose of anti-TNF	0.95	(Feature scaled units)
antiTNFtimepoint	Time-step to introduce anti-TNF	0	Day 0
antiTNFtimepoint	Time-step to introduce anti-TNF	720	Day 5
inflammationLevel	Existing parameter modulated as surrogate for baminercept	50	Day 0
inflammationLevel	Existing parameter modulated as surrogate for baminercept	75	Day 0

Figure B.6: Table of therapeutic intervention-related parameters for biologics anti-TNF, anti-VLA4 (natalizumab), and anti-LTβ (baminercept).

References

K. Alden. *Simulation and statistical techniques to explore lymphoid tissue organogenesis.* PhD thesis, University of York, 2012a.

K. Alden. Simulation and statistical techniques to explore lymphoid tissue organogenesis. 2012b.

K. Alden, P. Andrews, J. Timmis, H. Veiga-Fernandes, and M. Coles. Towards argument-driven validation of an in silico model of immune tissue organogenesis. *Artificial Immune Systems*, pages 66–70, 2011.

K. Alden, M. Read, J. Timmis, P. Andrews, H. Veiga-Fernandes, and M. Coles. spartan: A comprehensive tool for understanding uncertainty in simulations of biological systems. *Submitted to PLoS Computational Biology*, 2012a.

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

K. Alden, M. Read, J. Timmis, P. S. Andrews, H. Veiga-Fernandes, and M. Coles. Spartan: a comprehensive tool for understanding uncertainty in simulations of biological systems. *PLoS Comput Biol*, 9(2):e1002916, 2013.

K. Alden, P. S. Andrews, F. A. Polack, H. Veiga-Fernandes, M. C. Coles, and J. Timmis. Using argument notation to engineer biological simulations with increased confidence. *Journal of The Royal Society Interface*, 12(104):20141059, 2015.

K. Alden, J. Timmis, P. Andrews, H. Veiga-Fernandes, and M. Coles. Extending and applying spartan to perform temporal sensitivity analyses for predicting changes in influential biological pathways in computational models. 2016.

C. D. Allen, T. Okada, H. L. Tang, and J. G. Cyster. Imaging of germinal center selection events during affinity maturation. *Science*, 315(5811):528–531, 2007.

S. Allerheiligen. Next-generation model-based drug discovery and development: quantitative and systems pharmacology. *Clinical Pharmacology & Therapeutics*, 88(1), 2010.

S. Allerheiligen. Impact of modeling and simulation: myth or fact? *Clinical pharmacology and therapeutics*, 96(4):413–415, 2014.

F. Aloisi and R. Pujol-Borrell. Lymphoid neogenesis in chronic inflammatory diseases. *Nature Reviews Immunology*, 6(3):205–217, 2006.

P. W. Anderson et al. More is different. *Science*, 177(4047):393–396, 1972.

R. Anderson, R. May, and S. Gupta. Non-linear phenomena in hostparasite interactions. *Parasitology*, 99(S1):S59–S79, 1989.

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

P. S. Andrews, F. A. Polack, A. T. Sampson, S. Stepney, and J. Timmis. The cosmos process version 0.1: A process for the modelling and simulation of complex systems. *University of York Technical Report YCS–2010*, 453, 2010.

S. T. Arron, R. M. Ribeiro, A. Gettie, R. Bohm, J. Blanchard, J. Yu, A. S. Perelson, D. D. Ho, and L. Zhang. Impact of thymectomy on the peripheral t cell pool in rhesus macaques before and after infection with simian immunodeficiency virus. *European journal of immunology*, 35(1):46–55, 2005.

M. Ashyraliyev, Y. Fomekong-Nanfack, J. A. Kaandorp, and J. G. Blom. Systems biology: parameter estimation for biochemical models. *Febs Journal*, 276(4):886–902, 2009.

S. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57. Springer Science & Business Media, 2007.

E. Astorri, M. Bombardieri, S. Gabba, M. Peakman, P. Pozzilli, and C. Pitzalis. Evolution of ectopic lymphoid neogenesis and in situ autoantibody production in autoimmune nonobese diabetic mice: cellular and molecular characterization of tertiary lymphoid structures in pancreatic islets. *The Journal of Immunology*, 185 (6):3359–3368, 2010.

G. S. Ayton, W. G. Noid, and G. A. Voth. Multiscale modeling of biomolecular systems: in serial and in parallel. *Current opinion in structural biology*, 17(2): 192–198, 2007.

A. N. Baer, G. Noaiseh, A. Parke, A. Coca, T. Utset, M. C. Genovese, D. J. Wallace, J. Mcnamara, K. Boyle, L. Keyes-elstein, et al. The clinical efficacy and safety of baminercept, a lymphotoxin-beta receptor fusion protein, in primary sjögren's syndrome: Results from a randomized, double-blind, placebo-controlled phase ii trial. *Arthritis & Rheumatology*, 67:3844–3845, 2015.

R. Bagabir, R. Byers, I. Chaudhry, W. Müller, R. Paus, and A. Bayat. Site-specific immunophenotyping of keloid disease demonstrates immune upregulation and the presence of lymphoid aggregates. *British Journal of Dermatology*, 167(5):1053–1066, 2012.

M. Bajénoff, J. G. Egen, L. Y. Koo, J. P. Laugier, F. Brau, N. Glaichenhaus, and R. N. Germain. Stromal cell networks regulate lymphocyte entry, migration, and territoriality in lymph nodes. *Immunity*, 25(6):989–1001, 2006.

S. Banerjee and M. E. Moses. A hybrid agent based and differential equation model of body size effects on pathogen replication and immune system response. In *Artificial Immune Systems*, pages 14–18. Springer, 2009.

J. Bard and I. Lauder. How well does turing's theory of morphogenesis work? *Journal of Theoretical Biology*, 45(2):501–531, 1974.

A. L. Barnes, P. G. Genever, S. Rimmer, and M. C. Coles. Collagen–poly (n-isopropylacrylamide) hydrogels with tunable properties. *Biomacromolecules*, 17 (3):723–734, 2016.

J. B. Beltman, A. F. Marée, and R. J. de Boer. Analysing immune cell migration. *Nature Reviews Immunology*, 9(11):789–798, 2009.

H. Bersini, D. Klatzmann, A. Six, and V. Thomas-Vaslin. State-transition diagrams for biologists. *PloS one*, 7(7):e41165, 2012.

T. Beyer and M. Meyer-Hermann. Cell transmembrane receptors determine tissue pattern stability. *Physical review letters*, 101(14):148102, 2008a.

T. Beyer and M. Meyer-Hermann. Mechanisms of organogenesis of primary lymphoid follicles. *International immunology*, 20(4):615–623, 2008b.

C. C. Bleul, J. L. Schultze, and T. A. Springer. B lymphocyte chemotaxis regulated in association with microanatomic localization, differentiation state, and b cell receptor engagement. *The Journal of experimental medicine*, 187(5):753–762, 1998.

M. Blum. Independent unbiased coin flips from a correlated biased sourcea finite state markov chain. *Combinatorica*, 6(2):97–108, 1986.

M. Bombardieri and C. Pitzalis. Ectopic lymphoid neogenesis and lymphoid chemokines in sjogren's syndrome: at the interplay between chronic inflammation, autoimmunity and lymphomagenesis. *Current pharmaceutical biotechnology*, 2012.

M. Bombardieri, F. Barone, D. Lucchesi, S. Nayar, W. B. van den Berg, G. Proctor, C. D. Buckley, and C. Pitzalis. Inducible tertiary lymphoid structures, autoimmunity, and exocrine dysfunction in a novel model of salivary gland inflammation in c57bl/6 mice. *The Journal of Immunology*, 189(7):3767–3776, 2012.

P. L. Bonate. Clinical trial simulation in drug development. *Pharmaceutical research*, 17(3):252–256, 2000.

G. P. Breschet. *Le systeme lymphatique considere sous les rapports anatomique, physiologique et pathologique.* Chez J. B. Balliere, Librarie de L'academie Royale de Medecine, 1836.

J. L. Browning. Inhibition of the lymphotoxin pathway as a therapy for autoimmune disease. *Immunological reviews*, 223(1):202–220, 2008.

F. J. Bruggeman and H. V. Westerhoff. The nature of systems biology. *Trends in microbiology*, 15(1):45–50, 2007.

J. Butler, A. Sawtell, S. Jarrett, J. Cosgrove, R. Leigh, J. Timmis, and M. Coles. Imaging immunity in lymph nodes: Past, present and future. In *Biophysics of Infection*, pages 329–346. Springer International Publishing, 2016a.

J. A. Butler, K. Alden, H. V. Fernandes, J. Timmis, and M. Coles. Novel approaches to the visualization and quantification of biological simulations by emulating experimental techniques. In *Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems*, pages 614–621, 2014.

J. A. Butler, J. Cosgrove, K. Alden, J. Timmis, and M. C. Coles. Model-driven experimentation: A new approach to understand mechanisms of tertiary lymphoid tissue formation, function, and therapeutic resolution. *Frontiers in Immunology*, 7, 2016b.

J. A. Butler, S. Nayar, B. Glaysher, J. Timmis, C. Buckley, F. Barone, and M. Coles. Targeting stroma in immune mediated inflammatory disease: Unexpected roles for il4r and vla4 in tlt formation & maintenance. In *Proc. 4th International workshop on Clinical Therapeutic Tolerance*. Newcastle University, 2017.

R. E. Callard and A. J. Yates. Immunology and mathematics: crossing the divide. *Immunology*, 115(1):21–33, 2005.

D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. Newman, and S. H. Strogatz. Are randomly grown graphs really random? *Physical Review E*, 64(4):041902, 2001.

B. Calvez and G. Hutzler. Automatic tuning of agent-based models using genetic algorithms. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*, pages 41–57. Springer, 2005.

J. J. Campbell, K. E. Murphy, E. J. Kunkel, C. E. Brightling, D. Soler, Z. Shen, J. Boisvert, H. B. Greenberg, M. A. Vierra, S. B. Goodman, et al. Ccr7 expression and memory t cell diversity in humans. *The Journal of Immunology*, 166(2):877–884, 2001.

C. Canevet, S. Gilmore, J. Hillston, M. Prowse, and P. Stevens. Performance modelling with uml and stochastic process algebras. *IEE Proceedings: Computers and Digital Techniques*, 150(2):107–120, 2003.

A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, et al. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology*, 7(10):R100, 2006.

D. M. Carragher, J. Rangel-Moreno, and T. D. Randall. Ectopic lymphoid tissues and local immunity. In *Seminars in immunology*, volume 20, pages 26–42. Elsevier, 2008.

S. Carsons. A review and update of sjogren's syndrome: manifestations, diagnosis, and treatment. *The American journal of managed care*, 7(14 Suppl):S433–43, 2001.

M. Casamayor-Pallejà, P. Mondière, C. Verschelde, C. Bella, and T. Defrance. Bcr ligation reprograms b cells for migration to the t zone and b-cell follicle sequentially. *Blood*, 99(6):1913–1921, 2002.

D. M. Catron, A. A. Itano, K. A. Pape, D. L. Mueller, and M. K. Jenkins. Visualizing the first 50 hr of the primary immune response to a soluble antigen. *Immunity*, 21(3):341–347, 2004.

L. Cervantes-Barragan, R. Züst, F. Weber, M. Spiegel, K. S. Lang, S. Akira, V. Thiel, and B. Ludewig. Control of coronavirus infection through plasmacytoid dendritic-cell–derived type i interferon. *Blood*, 109(3):1131–1137, 2007.

S. D. Chakravarty, G. Zhu, M. C. Tsai, V. P. Mohan, S. Marino, D. E. Kirschner, L. Huang, J. Flynn, and J. Chan. Tumor necrosis factor blockade in chronic murine tuberculosis enhances granulomatous inflammation and disorganizes granulomas in the lungs. *Infection and immunity*, 76(3):916–926, 2008.

C. Chan, M. Billard, S. A. Ramirez, H. Schmidl, E. Monson, and T. B. Kepler. A model for migratory b cell oscillations from receptor down-regulation induced by external chemokine fields. *Bulletin of mathematical biology*, 75(1):185–205, 2013.

B. Chapman and J. Chang. Biopython: Python tools for computational biology. *ACM Sigbio Newsletter*, 20(2):15–19, 2000.

N. Chomsky. Logical structures in language. *American Documentation (pre-1986)*, 8(4):284, 1957.

T. P. Conrads, M. Zhou, E. F. P. III, L. Liotta, and T. D. Veenstra. Cancer diagnosis using proteomic patterns. *Expert review of molecular diagnostics*, 3(4):411–420, 2003.

J. Cornes. Number, size, and distribution of peyer's patches in the human small intestine: Part i the development of peyer's patches. *Gut*, 6(3):225, 1965.

S. Cose, C. Brammer, K. M. Khanna, D. Masopust, and L. Lefrançois. Evidence that a significant number of naive t cells enter non-lymphoid organs as part of a normal migratory pathway. *European journal of immunology*, 36(6):1423–1433, 2006.

J. Cosgrove, J. Butler, K. Alden, M. Read, V. Kumar, L. Cucurull-Sanchez, J. Timmis, and M. Coles. Agent-based modeling in systems pharmacology. *CPT: pharmacometrics & systems pharmacology*, 4(11):615–629, 2015.

E. Crivellato, L. Travan, and D. Ribatti. The hippocratic treatise on glands: the first document on lymphoid tissue and lymph nodes. *Leukemia*, 21(4):591–592, 2007.

T. Cupedo, F. Lund, V. Ngo, T. Randall, W. Jansen, M. Greuter, R. de Waal-Malefyt, G. Kraal, J. Cyster, and R. Mebius. Initiation of cellular organization in lymph nodes is regulated by non-b cell-derived signals and is not dependent on cxc chemokine ligand 13. *The Journal of Immunology*, 173(8):4889–4896, 2004.

J. Z. Dalgaard, A. J. Klar, M. J. Moser, W. R. Holley, A. Chatterjee, and I. S. Mian. Statistical modeling and analysis of the laglidadg family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the hnh family. *Nucleic acids research*, 25(22):4626–4638, 1997.

D. Dan, C. Mueller, K. Chen, and J. A. Glazier. Solving the advection-diffusion equations in biological contexts using the cellular potts model. *Physical Review E*, 72(4):041909, 2005.

J. Dechanet, P. Merville, I. Durand, J. Banchereau, and P. Miossec. The ability of synoviocytes to support terminal differentiation of activated b cells may explain plasma cell accumulation in rheumatoid synovium. *Journal of Clinical Investigation*, 95(2):456, 1995.

S. Dhib-Jalbut and D. V. Kalvakolanu. Microglia and necroptosis: The culprits of neuronal cell death in multiple sclerosis. *Cytokine*, 76(2):583–584, 2015.

M.-C. Dieu-Nosjean, N. A. Giraldo, H. Kaplon, C. Germain, W. H. Fridman, and C. Sautès-Fridman. Tertiary lymphoid structures, drivers of the anti-tumor responses in human cancers. *Immunological reviews*, 271(1):260–275, 2016.

M.-L. Dion, R.-P. Sékaly, and R. Cheynier. Estimating thymic function through quantification of t-cell receptor excision circles. *Immunological Tolerance: Methods and Protocols*, pages 197–213, 2007.

C. B. Do, D. A. Woods, and S. Batzoglou. Contrafold: Rna secondary structure prediction without physics-based models. *Bioinformatics*, 22(14):e90–e98, 2006.

R. Dobrushin, V. Kriukov, and A. Toom. *Stochastic cellular systems: ergodicity, memory, morphogenesis.* Manchester University Press, 1990.

R. L. Dobrushin, V. Kriukov, A. L. Toom, et al. Locally interacting systems and their application in biology. In *School-Seminar on Markov Interaction Processes in Biology (1976: Pushchino, Russia)*. Springer-Verlag, 1978.

G. M. Donovan and G. Lythe. T-cell movement on the reticular network. *Journal of theoretical biology*, 295:59–67, 2012.

G. M. Donovan and G. Lythe. T cell and reticular network co-dependence in hiv infection. *Journal of theoretical biology*, 395:211–220, 2016.

D. C. Douek, R. D. McFarland, P. H. Keiser, E. A. Gage, J. M. Massey, B. F. Haynes, M. A. Polis, A. T. Haase, M. B. Feinberg, J. L. Sullivan&num, et al. Changes in thymic function with age and during the treatment of hiv infection. *Nature*, 396(6712):690–695, 1998.

D. L. Drayton, S. Liao, R. H. Mounzer, and N. H. Ruddle. Lymphoid organ development: from ontogeny to neogenesis. *Nature immunology*, 7(4):344–353, 2006.

B. E. Dutilh and R. J. De Boer. Decline in excision circles requires homeostatic renewal or homeostatic death of naive t cells. *Journal of theoretical biology*, 224 (3):351–358, 2003.

M. Fallahi-Sichani, M. El-Kebir, S. Marino, D. E. Kirschner, and J. J. Linderman. Multiscale computational modeling reveals a critical role for tnf-α receptor 1 dynamics in tuberculosis granuloma formation. *The Journal of Immunology*, 186(6): 3472–3483, 2011.

S. S. Fanchiang, R. Cojocaru, M. Othman, R. Khanna, M. J. Brooks, T. Smith, X. Tang, I. Maricic, A. Swaroop, and V. Kumar. Global expression profiling of peripheral qa-1–restricted cd8$\alpha\alpha$+ tcr$\alpha\beta$+ regulatory t cells reveals innate-like features: Implications for immune-regulatory repertoire. *Human immunology*, 73 (3):214–222, 2012.

C. Fellbaum, J. Sträter, and M.-L. Hansmann. Follicular dendritic cells in extranodal non-hodgkin lymphomas of malt and non-malt type. *Virchows Archiv A*, 423(5): 335–341, 1993.

A. Fick. Ueber diffusion. *Pogg. Ann. Phys. Chem.*, 170(4. Reihe 94):59–86, 1855.

M. T. Figge, A. Garin, M. Gunzer, M. Kosco-Vilbois, K.-M. Toellner, and M. Meyer-Hermann. Deriving a germinal center lymphocyte migration model from two-photon data. *The Journal of experimental medicine*, 205(13):3019–3029, 2008.

A. J. Flugge, J. Timmis, P. Andrews, J. Moore, and P. Kaye. Modelling and simulation of granuloma formation in visceral leishmaniasis. In *Evolutionary Computation, 2009. CEC'09. IEEE Congress on*, pages 3052–3059. IEEE, 2009.

M. Franceschet. Pagerank: Standing on the shoulders of giants. *Communications of the ACM*, 54(6):92–101, 2011.

P. Friedl and B. Weigelin. Interstitial leukocyte migration and immune function. *Nature immunology*, 9(9):960–969, 2008.

Y.-X. Fu and D. D. Chaplin. Development and maturation of secondary lymphoid tissues. *Annual review of immunology*, 17(1):399–433, 1999.

S. Fukuyama, T. Hiroi, Y. Yokota, P. D. Rennert, M. Yanagita, N. Kinoshita, S. Terawaki, T. Shikina, M. Yamamoto, Y. Kurono, et al. Initiation of nalt organogenesis is independent of the il-7r, ltβr, and nik signaling pathways but requires the id2 gene and cd3- cd4+ cd45+ cells. *Immunity*, 17(1):31–40, 2002.

K. G. Gadkar, J. Varner, and F. J. Doyle III. Model identification of signal transduction networks from data using a state regulator problem. *Systems Biology*, 2 (1):17–30, 2005.

P. Garnett, S. Stepney, and O. Leyser. Towards an executable model of auxin transport canalisation. *CoSMoS*, pages 63–91, 2008.

J. Geginat, F. Sallusto, and A. Lanzavecchia. Cytokine-driven proliferation and differentiation of human naive, central memory, and effector memory cd4+ t cells. *The Journal of experimental medicine*, 194(12):1711–1720, 2001.

E. Genth. Cachexia due to sjogren's syndrome. *Deutsche medizinische Wochenschrift (1946)*, 123(49):1499–1499, 1998.

R. N. Germain. The art of the probable: system control in the adaptive immune system. *Science*, 293(5528):240–245, 2001.

R. N. Germain, M. Meier-Schellersheim, A. Nita-Lazar, and I. D. Fraser. Systems biology in immunology–a computational modeling perspective. *Annual review of immunology*, 29:527, 2011.

T. Ghetiu, F. A. Polac, and J. Bown. Argument-driven validation of computer simulations-a necessity, rather than an option. In *Advances in System Testing and Validation Lifecycle (VALID), 2010 Second International Conference on*, pages 1–4. IEEE, 2010.

G. N. Gilbert. *Agent-based models*. Number 153. Sage, 2008.

S. Ginsburg. *The Mathematical Theory of Context Free Languages.[Mit Fig.]*. McGraw-Hill Book Company, 1966.

Y. Goh and D. Sethi. Submandibular gland excision: a five-year review. *The Journal of Laryngology & Otology*, 112(03):269–273, 1998.

C. Golgi and M. F. G. White. Hemolymph glands in domestic animals.

F. Graner and J. A. Glazier. Simulation of biological cell sorting using a two-dimensional extended potts model. *Physical review letters*, 69(13):2013, 1992.

F. Graw and R. R. Regoes. Influence of the fibroblastic reticular network on cell-cell interactions in lymphoid organs. *PLoS Comput Biol*, 8(3):e1002436, 2012.

R. B. Greaves, M. Read, J. Timmis, P. S. Andrews, J. A. Butler, B.-O. Gerckens, and V. Kumar. In silico investigation of novel biological pathways: the role of cd200 in regulation of t cell priming in experimental autoimmune encephalomyelitis. *Biosystems*, 112(2):107–121, 2013.

A. Gregorio, C. Gambini, V. Gerloni, A. Parafioriti, M. Sormani, S. Gregorio, G. De Marco, F. Rossi, A. Martini, and M. Gattorno. Lymphoid neogenesis in juvenile idiopathic arthritis correlates with ana positivity and plasma cells infiltration. *Rheumatology*, 46(2):308–313, 2007.

V. Grimm, U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, et al. A standard protocol for describing individual-based and agent-based models. *Ecological modelling*, 198(1):115–126, 2006.

V. Grimm, U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback. The odd protocol: a review and first update. *Ecological modelling*, 221(23):2760–2768, 2010.

G. L. Gulland. The development of lymphatic glands. *The Journal of Pathology and Bacteriology*, 2(4):447–485, 1894.

R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*, 3(10):e189, 2007.

D. Guy-Grand, O. Azogui, S. Celli, S. Darche, M. C. Nussenzweig, P. Kourilsky, and P. Vassalli. Extrathymic t cell lymphopoiesis ontogeny and contribution to gut intraepithelial lymphocytes in athymic and euthymic mice. *The Journal of experimental medicine*, 197(3):333–341, 2003.

A. Hansen, P. E. Lipsky, and T. Dörner. B cells in sjögren's syndrome: indications for disturbed selection and differentiation in ectopic lymphoid tissue. *Arthritis research & therapy*, 9(4):218, 2007.

D. Harel. A turing-like test for biological modeling. *Nature biotechnology*, 23(4): 495–496, 2005.

T. H. Harris, E. J. Banigan, D. A. Christian, C. Konradt, E. D. T. Wojno, K. Norose, E. H. Wilson, B. John, W. Weninger, A. D. Luster, et al. Generalized lévy walks and the role of chemokines in migration of effector cd8+ t cells. *Nature*, 486(7404): 545–548, 2012.

A. Hatzakis, G. Touloumi, R. Karanicolas, A. Karafoulidou, T. Mandalaki, C. Anastassopoulou, L. Zhang, J. J. Goedert, D. D. Ho, and L. G. Kostrikis. Effect of recent thymic emigrants on progression of hiv-1 disease. *The Lancet*, 355(9204): 599–604, 2000.

M. D. Hazenberg, S. A. Otto, J. W. C. Stuart, M. C. Verschuren, J. C. Borleffs, C. A. Boucher, R. A. Coutinho, J. M. Lange, T. F. R. de Wit, A. Tsegaye, et al. Increased cell division but not thymic dysfunction rapidly affects the t-cell receptor excision circle content of the naive t cell population in hiv-1 infection. *Nature medicine*, 6(9):1036–1042, 2000.

M. D. Hazenberg, J. A. Borghans, R. J. de Boer, and F. Miedema. Thymic output: a bad trec record. *Nature immunology*, 4(2):97–99, 2003.

A. Heyden, H. Lin, and D. G. Truhlar. Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations. *The Journal of Physical Chemistry B*, 111(9):2231–2241, 2007.

W. Hoeffding, H. Robbins, et al. The central limit theorem for dependent random variables. *Duke math. J*, 15(3):773–780, 1948.

E. Huang, S. H. Cheng, H. Dressman, J. Pittman, M. H. Tsou, C. F. Horng, A. Bild, E. S. Iversen, M. Liao, C. M. Chen, et al. Gene expression predictors of breast cancer outcomes. *The Lancet*, 361(9369):1590–1596, 2003.

C. A. Hunt and G. E. Ropella. Moving beyond in silico tools to in silico science in support of drug development research. *Drug Development Research*, 72(2): 153–161, 2011.

C. A. Hunt, R. C. Kennedy, S. H. Kim, and G. E. Ropella. Agent-based modeling: a systematic assessment of use cases and requirements for enhancing pharmaceutical research and development productivity. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 5(4):461–480, 2013.

D. J. Irvine, A. N. Stachowiak, and Y. Hori. Lymphoid tissue engineering: Invoking lymphoid tissue neogenesis in immunotherapy and models of immunity. In *Seminars in immunology*, volume 20, pages 137–146. Elsevier, 2008.

W. Jones. Hippocrates i. loeb classical library, 1923.

T. Junt, E. Scandella, and B. Ludewig. Form follows function: lymphoid tissue microarchitecture in antimicrobial immune defence. *Nature Reviews Immunology*, 8(10):764–775, 2008.

A. Kamal and M. Khamashta. The efficacy of novel b cell biologics as the future of sle treatment: a review. *Autoimmunity reviews*, 13(11):1094–1101, 2014.

M. Karplus. Development of multiscale models for complex chemical systems: from h+ h2 to biomolecules (nobel lecture). *Angewandte Chemie International Edition*, 53(38):9992–10005, 2014.

S. Kaski, J. Nikkilä, and T. Kohonen. Methods for interpreting a self-organized map in data analysis. In *In Proc. 6th European Symposium on Artificial Neural Networks (ESANN98). D-Facto, Brugfes*. Citeseer, 1998.

Y. Kato, H. Seki, and T. Kasami. Rna pseudoknotted structure prediction using stochastic multiple context-free grammar. *IPSJ Digital Courier*, 2:655–664, 2006.

T. Kelly and R. Weaver. The goal structuring notation–a safety argument notation. In *Proceedings of the dependable systems and networks 2004 workshop on assurance cases*. Citeseer, 2004.

T. P. Kelly. *Arguing safety: a systematic approach to managing safety cases*. University of York, 1999.

T. B. Kepler and A. S. Perelson. Cyclic re-entry of germinal center b cells and the efficiency of affinity maturation. *Immunology today*, 14(8):412–415, 1993.

C. Keşmir and R. J. De Boer. A mathematical model on germinal center kinetics and termination. *The Journal of Immunology*, 163(5):2463–2469, 1999.

C. Keşmir and R. J. De Boer. A spatial model of germinal center reactions: cellular adhesion based sorting of b cells results in efficient affinity maturation. *Journal of theoretical biology*, 222(1):9–22, 2003.

S. Kikuchi, Y. Onuki, A. Yasuda, Y. Hayashi, and K. Takayama. Latent structure analysis in pharmaceutical formulations using kohonen's self-organizing map and a bayesian network. *Journal of pharmaceutical sciences*, 100(3):964–975, 2011.

D. Kirschner and S. Marino. Mycobacterium tuberculosis as viewed through a computer. *Trends in microbiology*, 13(5):206–211, 2005.

A. Kislitsyn, R. Savinkov, M. Novkovic, L. Onder, and G. Bocharov. Computational approach to 3d modeling of the lymph node geometry. *Computation*, 3(2):222–234, 2015.

R. Knight, P. Maxwell, A. Birmingham, J. Carnes, J. G. Caporaso, B. C. Easton, M. Eaton, M. Hamady, H. Lindsay, Z. Liu, et al. Pycogent: a toolkit for making sense from sequence. *Genome Biol*, 8(8):R171, 2007.

T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1):59–69, 1982.

T. Kohonen, M. Schroeder, T. Huang, and S.-O. Maps. Springer-verlag new york. *Inc., Secaucus, NJ*, 43, 2001.

I. Kola and J. Landis. Can the pharmaceutical industry reduce attrition rates? *Nature reviews Drug discovery*, 3(8):711–716, 2004.

M. Komai-Koma, A. Donachie, and P. Wilkinson. Antigen-specific chemotaxis of b cells. *Immunology*, 91(4):579–585, 1997.

F.-k. Kong, H. C. Chen-lo, A. Six, R. D. Hockett, and M. D. Cooper. T cell receptor gene deletion circles identify recent thymic emigrants in the peripheral t cell pool. *Proceedings of the National Academy of Sciences*, 96(4):1536–1540, 1999.

N. Koning, L. Bö, R. M. Hoek, and I. Huitinga. Downregulation of macrophage inhibitory molecules in multiple sclerosis lesions. *Annals of neurology*, 62(5):504–514, 2007.

A. Kratz, A. Campos-Neto, M. S. Hanson, and N. H. Ruddle. Chronic inflammation caused by lymphotoxin is lymphoid neogenesis. *The Journal of experimental medicine*, 183(4):1461–1472, 1996.

R. Lalonde, K. Kowalski, M. Hutmacher, W. Ewy, D. Nichols, P. Milligan, B. Corrigan, P. Lockwood, S. Marshall, L. Benincosa, et al. Model-based drug development. *Clinical Pharmacology & Therapeutics*, 82(1):21–32, 2007.

T. Lämmermann and M. Sixt. The microanatomy of t-cell responses. *Immunological reviews*, 221(1):26–43, 2008.

M. R. Lamprecht, D. M. Sabatini, A. E. Carpenter, et al. Cellprofiler: free, versatile software for automated biological image analysis. *Biotechniques*, 42(1):71, 2007.

C. Larochelle, T. Uphaus, A. Prat, and F. Zipp. Secondary progression in multiple sclerosis: Neuronal exhaustion or distinct pathology? *Trends in neurosciences*, 39(5):325–339, 2016.

H. Lassmann, J. Van Horssen, and D. Mahad. Progressive multiple sclerosis: pathology and pathogenesis. *Nature Reviews Neurology*, 8(11):647–656, 2012.

L. Le Pottier, V. Devauchelle, A. Fautrel, C. Daridon, A. Saraux, P. Youinou, and J.-O. Pers. Ectopic germinal centers are rare in sjögrens syndrome salivary glands and do not exclude autoreactive b cells. *The Journal of Immunology*, 182(6):3540–3547, 2009.

L. Lee. Learning of context-free languages: A survey of the literature. *Techn. Rep. TR-12-96, Harvard University*, 1996.

M. Leslie. Immunity goes local. *Science*, 352(6281):21–23, 2016.

F. T. Lewis. The first lymph glands in rabbit and human embryos. *The Anatomical Record*, 3(6):341–353, 1909.

J. Lian and A. D. Luster. Chemokine-guided cell positioning in the lymph node orchestrates the generation of adaptive immune responses. *Current opinion in cell biology*, 36:1–6, 2015.

A. Lindenmayer. Mathematical models for cellular interactions in development ii. simple and branching filaments with two-sided inputs. *Journal of theoretical biology*, 18(3):300–315, 1968.

G. Y. Liu, P. J. Fairchild, R. M. Smith, J. R. Prowle, D. Kioussis, and D. C. Wraith. Low avidity recognition of self-antigen by t cells permits escape from central tolerance. *Immunity*, 3(4):407–415, 1995.

F. Loder, B. Mutschler, R. J. Ray, C. J. Paige, P. Sideras, R. Torres, M. C. Lamers, and R. Carsetti. B cell development in the spleen takes place in discrete steps and is determined by the quality of b cell receptor–derived signals. *The Journal of experimental medicine*, 190(1):75–90, 1999.

G. Luh and W. Cheng. Non-linear system identification using an artificial immune system. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 215(6):569–585, 2001.

S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, and G. Balan. Mason: A multiagent simulation environment. *Simulation*, 81(7):517–527, 2005.

P. K. Maini. The impact of turing's work on pattern formation in biology. *Mathematics today*, 40(4):140–141, 2004.

A. Mantovani, R. Bonecchi, and M. Locati. Tuning inflammation and immunity by chemokine sequestration: decoys and more. *Nature Reviews Immunology*, 6(12): 907–918, 2006.

K. Margaris and R. A. Black. Modelling the lymphatic system: challenges and opportunities. *Journal of The Royal Society Interface*, 9(69):601–612, 2012.

S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology*, 254(1):178–196, 2008.

S. Marino, M. El-Kebir, and D. Kirschner. A hybrid multi-compartment model of granuloma formation and t cell priming in tuberculosis. *Journal of theoretical biology*, 280(1):50–62, 2011a.

S. Marino, J. J. Linderman, and D. E. Kirschner. A multifaceted approach to modeling the immune response in tuberculosis. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 3(4):479–489, 2011b.

W. L. Marmelzat. Galen on glands and spongy flesh: With special reference to an ancient mechanico–biophysical theory for superficial lymph node function. *Journal of the history of medicine and allied sciences*, 46(4):419–439, 1991.

F. Maseras and K. Morokuma. Imomm: A new integrated ab initio+ molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *Journal of Computational Chemistry*, 16(9):1170–1179, 1995.

M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30, 1998.

P. McGonigle and B. Ruggeri. Animal models of human disease: challenges in enabling translation. *Biochemical pharmacology*, 87(1):162–171, 2014.

G. F. Medley, N. A. Lindop, W. J. Edmunds, and D. J. Nokes. Hepatitis-b virus endemicity: heterogeneity, catastrophic dynamics and control. *Nature medicine*, 7(5):619–624, 2001.

D. Meier, C. Bornmann, S. Chappaz, S. Schmutz, L. A. Otten, R. Ceredig, H. Acha-Orbea, and D. Finke. Ectopic lymphoid-organ development occurs through interleukin 7-mediated enhanced survival of lymphoid-tissue-inducer cells. *Immunity*, 26(5):643–654, 2007.

J. M. Meijer, P. Meiners, A. Vissink, F. Spijkervet, W. Abdulahad, N. Kamminga, E. Brouwer, C. G. Kallenberg, and H. Bootsma. Effectiveness of rituximab treatment in primary sjögren's syndrome: a randomized, double-blind, placebo-controlled trial. *Arthritis & Rheumatology*, 62(4):960–968, 2010.

M. Meyer-Hermann and T. Beyer. Models of cell migration. *Reviews in Cell Biology and Molecular Medicine*, 2012.

M. Meyer-Hermann, E. Mohr, N. Pelletier, Y. Zhang, G. D. Victora, and K.-M. Toellner. A theory of germinal center b cell selection, division, and exit. *Cell reports*, 2(1):162–174, 2012.

M. E. Meyer-Hermann and P. K. Maini. Cutting edge: back to one-way germinal centers. *The Journal of Immunology*, 174(5):2489–2493, 2005a.

M. E. Meyer-Hermann and P. K. Maini. Interpreting two-photon imaging data of lymphocyte motility. *Physical Review E*, 71(6):061912, 2005b.

H. Miao, X. Xia, A. S. Perelson, and H. Wu. On identifiability of nonlinear ode models and applications in viral dynamics. *SIAM review*, 53(1):3–39, 2011.

M. J. Miller, S. H. Wei, I. Parker, and M. D. Cahalan. Two-photon imaging of lymphocyte motility and antigen response in intact lymph node. *Science*, 296 (5574):1869–1873, 2002.

M. J. Miller, S. H. Wei, M. D. Cahalan, and I. Parker. Autonomous t cell trafficking examined in vivo with intravital two-photon microscopy. *Proceedings of the National Academy of Sciences*, 100(5):2604–2609, 2003.

P. Milligan, M. Brown, B. Marchant, S. Martin, P. Van Der Graaf, N. Benson, G. Nucci, D. Nichols, R. Boyd, J. Mandema, et al. Model-based drug development: a rational approach to efficiently accelerate drug development. *Clinical Pharmacology & Therapeutics*, 93(6), 2013.

S. Mittal, M. Revell, F. Barone, D. L. Hardie, G. S. Matharu, A. J. Davenport, R. A. Martin, M. Grant, F. Mosselmans, P. Pynsent, et al. Lymphoid aggregates that resemble tertiary lymphoid organs define a specific pathological subset in metal-on-metal hip replacements. *PLoS One*, 8(5):e63470, 2013.

C. Molina-Paris, M. Castro, G. Lythe, and R. Ribeiro. Mathematics in modern immunology. *Interface Focus*, 2016.

T. R. Mosmann and S. Sad. The expanding universe of t-cell subsets: Th1, th2 and more. *Immunology today*, 17(3):138–146, 1996.

P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *science*, 328 (5980):876–878, 2010.

S. N. Mueller and R. N. Germain. Stromal cell contributions to the homeostasis and functionality of the immune system. *Nature Reviews Immunology*, 9(9):618–629, 2009.

G. Müller, U. E. Höpken, and M. Lipp. The impact of ccr7 and cxcr5 on lymphoid organ development and systemic immunity. *Immunological reviews*, 195(1):117–135, 2003.

P. Murdin. Pauli exclusion principle. *Encyclopedia of Astronomy and Astrophysics*, 1:4896, 2000.

A. J. Myers, S. Marino, D. E. Kirschner, and J. L. Flynn. Inoculation dose of mycobacterium tuberculosis does not influence priming of t cell responses in lymph nodes. *The Journal of Immunology*, 190(9):4707–4716, 2013.

N. F. Neel, E. Schutyser, J. Sai, G.-H. Fan, and A. Richmond. Chemokine receptor internalization and intracellular trafficking. *Cytokine & growth factor reviews*, 16 (6):637–658, 2005.

K. Neyt, F. Perros, C. H. GeurtsvanKessel, H. Hammad, and B. N. Lambrecht. Tertiary lymphoid organs in infection and autoimmunity. *Trends in immunology*, 33(6):297–305, 2012.

S. O. Nielsen, R. E. Bulo, P. B. Moore, and B. Ensing. Recent progress in adaptive multiscale molecular dynamics simulations of soft matter. *Physical Chemistry Chemical Physics*, 12(39):12401–12414, 2010.

M. S. Nixon and A. S. Aguado. *Feature extraction & image processing for computer vision*.

M. A. Nowak. Theory is available light. *Current Biology*, 14(11):R406–R407, 2004.

S. Nzula, J. Going, and D. Stott. Antigen-driven clonal proliferation, somatic hypermutation, and selection of b lymphocytes infiltrating human ductal breast carcinomas. *Cancer research*, 63(12):3275, 2003.

P. L. Ogra, D. T. Karzon, F. Righthand, and M. MacGillivray. Immunoglobulin response in serum and secretions after immunization with live and inactivated poliovaccine and natural infection. *New England journal of medicine*, 279(17): 893–900, 1968.

H. Olson, G. Betton, D. Robinson, K. Thomas, A. Monro, G. Kolaja, P. Lilly, J. Sanders, G. Sipes, W. Bracken, et al. Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regulatory Toxicology and Pharmacology*, 32 (1):56–67, 2000.

W. Ouyang, S. Rutz, N. K. Crellin, P. A. Valdez, and S. G. Hymowitz. Regulation and functions of the il-10 family of cytokines in inflammation and disease. *Annual review of immunology*, 29:71–109, 2011.

M. Owen and A. Friedenstein. Stromal stem cells: marrow-derived osteogenic precursors. *Cell and molecular biology of vertebrate hard tissues*, 136:42–60, 1988.

R. Pabst and I. Gehrke. Is the bronchus-associated lymphoid tissue (balt) an integral structure of the lung in normal mammals, including humans? *American journal of respiratory cell and molecular biology*, 3(2):131–135, 1990.

L. Page. United states patent 6,285,999. *USPTO Patent Full-Text and Image Database*, 2001.

M. Pagel and A. Meade. Bayesian analysis of correlated evolution of discrete characters by reversible-jump markov chain monte carlo. *The American Naturalist*, 167(6):808–825, 2006.

A. Patel, N. Harker, L. Moreira-Santos, M. Ferreira, K. Alden, J. Timmis, K. Foster, A. Garefalaki, P. Pachnis, P. Andrews, et al. Differential ret signaling pathways drive development of the enteric lymphoid and nervous systems. *Science signaling*, 5(235):ra55–ra55, 2012.

F. Pereira. Formal grammar and information theory: together again? *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 358(1769):1239–1253, 2000.

M. Peterson and M. Riggs. Fda advisory meeting clinical pharmacology review utilizes a quantitative systems pharmacology (qsp) model: a watershed moment? *CPT: pharmacometrics & systems pharmacology*, 4(3):189–192, 2015.

M. W. Pfaffl. A new mathematical model for relative quantification in real-time rt–pcr. *Nucleic acids research*, 29(9):e45–e45, 2001.

E. M. Pimenta and B. J. Barnes. Role of tertiary lymphoid structures (tls) in anti-tumor immunity: potential tumor-induced cytokines/chemokines that regulate tls formation in epithelial-derived cancers. *Cancers*, 6(2):969–997, 2014.

L. Plitnick and D. Herzyk. *Nonclinical development of novel biologics, biosimilars, vaccines and specialty biologics.* Academic Press, 2013.

F. A. Polack, A. Droop, P. Garnett, T. Ghetiu, and S. Stepney. Simulation validation: exploring the suitability of a simulation of cell division and differentiation in the prostate. In *Proceedings of the 2011 Workshop on Complex Systems Modelling and Simulation, Paris, France*, pages 113–133, 2011.

S. Raimondeau and D. Vlachos. Recent developments on multiscale, hierarchical modeling of chemical reactors. *Chemical Engineering Journal*, 90(1):3–23, 2002.

M. Ramos-Casals, A. G. Tzioufas, J. H. Stone, A. Sisó, and X. Bosch. Treatment of primary sjögren syndrome: a systematic review. *Jama*, 304(4):452–460, 2010.

D. H. Raulet, R. D. Garman, H. Saito, and S. Tonegawa. Developmental regulation of t-cell receptor gene expression. *Nature*, 1985.

M. Read, P. S. Andrews, J. Timmis, and V. Kumar. Techniques for grounding agent-based simulations in the real domain: a case study in experimental autoimmune encephalomyelitis. *Mathematical and Computer Modelling of Dynamical Systems*, 18(1):67–86, 2012.

M. Read, P. S. Andrews, J. Timmis, R. A. Williams, R. B. Greaves, H. Sheng, M. Coles, and V. Kumar. Determining disease intervention strategies using spatially resolved simulations. *PloS one*, 8(11):e80506, 2013.

M. Read, P. S. Andrews, J. Timmis, and V. Kumar. Modelling biological behaviours with the unified modelling language: an immunological case study and critique. *Journal of The Royal Society Interface*, 11(99):20140704, 2014.

M. N. Read. Statistical and modelling techniques to build confidence in the investigation of immunology through agent-based simulation. *PhD Thesis*, 2011.

M. N. Read, K. Alden, L. M. Rose, and J. Timmis. Automated multi-objective calibration of biological agent-based simulations. *Journal of The Royal Society Interface*, 13(122):20160543, 2016.

S. Reagan-Shaw, M. Nihal, and N. Ahmad. Dose translation from animal to human studies revisited. *The FASEB Journal*, 22(3):659–661, 2008.

G. Rozenberg and A. Salomaa. *Lindenmayer systems: Impacts on theoretical computer science, computer graphics, and developmental biology*. Springer Science & Business Media, 2012.

F. R. Sabin. The development of the lymphatic nodes in the pig and their relation to the lymph hearts. *American Journal of Anatomy*, 4(3):355–389, 1905.

Y. Sakakibara, M. Brown, R. Hughey, I. S. Mian, K. Sjölander, R. C. Underwood, and D. Haussler. Stochastic context-free grammers for trna modeling. *Nucleic acids research*, 22(23):5112–5120, 1994.

S. Salomonsson, M. V. Jonsson, K. Skarstein, K. A. Brokstad, P. Hjelmström, M. Wahren-Herlenius, and R. Jonsson. Cellular basis of ectopic germinal center formation and autoantibody production in the target organ of patients with sjögren's syndrome. *Arthritis & Rheumatism*, 48(11):3187–3201, 2003.

R. G. Sargent. Simulation model verification and validation. In *Proceedings of the 23rd conference on Winter simulation*, pages 37–47. IEEE Computer Society, 1991.

B. J. Schmidt, J. A. Papin, and C. J. Musante. Mechanistic systems modeling to guide drug discovery and development. *Drug discovery today*, 18(3):116–127, 2013.

H. M. Senn and W. Thiel. Qm/mm methods for biomolecular systems. *Angewandte Chemie International Edition*, 48(7):1198–1229, 2009.

L. B. Sheiner. Learning versus confirming in clinical drug development. *Clinical Pharmacology & Therapeutics*, 61(3):275–291, 1997.

J. D. Shields, I. C. Kourtis, A. A. Tomei, J. M. Roberts, and M. A. Swartz. Induction of lymphoidlike stroma and immune escape by tumors that express the chemokine ccl21. *Science Signaling*, 328(5979):749, 2010.

N. Shomer, J. Fox, A. Juedes, and N. Ruddle. Helicobacter-induced chronic active lymphoid aggregates have characteristics of tertiary lymphoid tissue. *Infection and immunity*, 71(6):3572–3577, 2003.

Z. Shulman, A. D. Gitlin, J. S. Weinstein, B. Lainez, E. Esplugues, R. A. Flavell, J. E. Craft, and M. C. Nussenzweig. Dynamic signaling by t follicular helper cells during germinal center b cell selection. *Science*, 345(6200):1058–1062, 2014.

N. L. Sicotte and R. R. Voskuhl. Onset of multiple sclerosis associated with anti-tnf therapy. *Neurology*, 57(10):1885–1888, 2001.

G. D. Smith. *Numerical solution of partial differential equations: finite difference methods*. Oxford university press, 1985.

A. Soni, J. Jenkins, and S. Sundaram. Determination of critical network interactions: an augmented boolean pseudo-dynamics approach. *Systems Biology, IET*, 2(2):55–63, 2008.

P. K. Sorger, S. R. Allerheiligen, D. R. Abernethy, R. B. Altman, K. L. Brouwer, A. Califano, D. Z. DArgenio, R. Iyengar, W. J. Jusko, R. Lalonde, et al. Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms. In *An NIH white paper by the QSP workshop group*, volume 48. NIH Bethesda, 2011.

J. Spidlen, W. Moore, D. Parks, M. Goldberg, C. Bray, P. Bierre, P. Gorombey, B. Hyun, M. Hubbard, S. Lange, et al. Data file standard for flow cytometry, version fcs 3.1. *Cytometry Part A*, 77(1):97–100, 2010.

J. D. Storey, W. Xiao, J. T. Leek, R. G. Tompkins, and R. W. Davis. Significance analysis of time course microarray experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 102(36):12837–12842, 2005.

J. K. Tan and T. Watanabe. Artificial engineering of secondary lymphoid organs. *Advances in immunology*, 105:131–157, 2010.

H. Tang, Y. Wang, L. K. Chlewicki, Y. Zhang, J. Guo, W. Liang, J. Wang, X. Wang, and Y.-X. Fu. Facilitating t cell infiltration in tumor microenvironment overcomes resistance to pd-l1 blockade. *Cancer cell*, 29(3):285–296, 2016.

J. H. Teacher and W. Hunter. *Catalogue of the Anatomical and Pathological Preparations of Dr. William Hunter in the Hunterian Museum, University of Glasgow*. James MacLehose, 1900.

E. Theander, L. Vasaitis, E. Baecklund, G. Nordmark, G. Warfvinge, R. Liedholm, K. Brokstad, R. Jonsson, and M. V. Jonsson. Lymphoid organisation in labial salivary gland biopsies is a possible predictor for the development of malignant lymphoma in primary sjögren's syndrome. *Annals of the rheumatic diseases*, 70 (8):1363–1368, 2011.

A. M. Thornton and E. M. Shevach. Suppressor effector function of cd4+ cd25+ immunoregulatory t cells is antigen nonspecific. *The Journal of Immunology*, 164 (1):183–190, 2000.

A. A. Tomei, S. Siegert, M. R. Britschgi, S. A. Luther, and M. A. Swartz. Fluid flow regulates stromal cell organization and ccl21 expression in a tissue-engineered lymph node microenvironment. *The Journal of Immunology*, 183(7):4273–4283, 2009.

Z. Touma, J. Sayad, and T. Arayssi. Successful treatment of sjogren's syndrome with rituximab. *Scandinavian journal of rheumatology*, 35(4):323–325, 2006.

A. V. Tumanov, S. I. Grivennikov, A. N. Shakhov, S. A. Rybtsov, E. P. Koroleva, J. Takeda, S. A. Nedospasov, and D. V. Kuprash. Dissecting the role of lymphotoxin in lymphoid organs by conditional targeting. *Immunological reviews*, 195 (1):106–116, 2003.

A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 237(641):37–72, 1952.

M. H. Ulvmar, K. Werth, A. Braun, P. Kelay, E. Hub, K. Eller, L. Chan, B. Lucas, I. Novitzky-Basso, K. Nakamura, et al. The atypical chemokine receptor ccrl1 shapes functional ccl21 gradients in lymph nodes. *Nature immunology*, 15(7): 623–630, 2014.

T. Ushiki, O. Ohtani, and K. Abe. Scanning electron microscopic studies of reticular framework in the rat mesenteric lymph node. *The Anatomical Record*, 241(1):113–122, 1995.

S. A. van de Pavert and R. E. Mebius. New insights into the development of lymphoid tissues. *Nature Reviews Immunology*, 10(9):664–674, 2010.

P. H. Verburg, B. Eickhout, and H. van Meijl. A multi-scale, multi-model approach for analyzing the future dynamics of european land use. *The Annals of Regional Science*, 42(1):57–77, 2008.

P. Vicini and P. Graaf. Systems pharmacology for drug discovery and development: paradigm shift or flash in the pan? *Clinical Pharmacology & Therapeutics*, 93(5): 379–381, 2013.

S. Visser, E. Manolis, M. Danhof, and T. Kerbusch. Modeling and simulation at the interface of nonclinical and early clinical drug development. *CPT: pharmacometrics & systems pharmacology*, 2(2):1–3, 2013.

S. Visser, D. Alwis, T. Kerbusch, J. Stone, and S. Allerheiligen. Implementation of quantitative and systems pharmacology in large pharma. *CPT: pharmacometrics & systems pharmacology*, 3(10):1–10, 2014.

A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology*, 103(2):227–249, 1976.

A. S. Warthin. The normal histology of the human hemolymph glands. *American Journal of Anatomy*, 1(1):63–79, 1901.

S. H. Wei, I. Parker, M. J. Miller, and M. D. Cahalan. A stochastic view of lymphocyte motility and trafficking within the lymph node. *Immunological reviews*, 195(1):136–159, 2003.

W. Weninger, H. Carlsen, M. Goodarzi, F. Moazed, M. Crowley, E. Baekkevold, L. Cavanagh, and U. von Andrian. Naive t cell recruitment to nonlymphoid tissues: a role for endothelium-expressed cc chemokine ligand 21 in autoimmune disease and lymphoid neogenesis. *The Journal of Immunology*, 170(9):4638–4648, 2003.

P. L. Williams et al. *Gray's anatomy*, volume 378. Churchill livingstone Edinburgh, 1989.

R. A. Williams, R. Greaves, M. Read, J. Timmis, P. S. Andrews, and V. Kumar. In silico investigation into dendritic cell regulation of cd8treg mediated killing of th1 cells in murine experimental autoimmune encephalomyelitis. *BMC bioinformatics*, 14(Suppl 6):S9, 2013.

D. S. Wishart. Current progress in computational metabolomics. *Briefings in Bioinformatics*, 8(5):279–293, 2007.

C. Witt, S. Raychaudhuri, and A. K. Chakraborty. Movies, measurement, and modeling the three ms of mechanistic immunology. *The Journal of experimental medicine*, 201(4):501–504, 2005.

J. Wu and A. Bottaro. Three-dimensional peripheral lymphoid organ cell cultures, Oct. 3 2003. US Patent App. 10/679,184.

J. M. Yabu and F. Vincenti. Novel immunosuppression: small molecules and biologics. In *Seminars in nephrology*, volume 27, pages 479–486. Elsevier, 2007.

P. Ye and D. E. Kirschner. Measuring emigration of human thymocytes by t-cell receptor excision circles. *Critical Reviews in Immunology*, 22(5-6), 2002a.

P. Ye and D. E. Kirschner. Reevaluation of t cell receptor excision circles as a measure of human recent thymic emigrants. *The Journal of Immunology*, 168 (10):4968–4979, 2002b.

L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *Systems, Man and Cybernetics, IEEE Transactions on*, (1): 28–44, 1973.

D. E. Zak, G. E. Gonye, J. S. Schwaber, and F. J. Doyle. Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: insights from an identifiability analysis of an in silico network. *Genome research*, 13(11):2396–2405, 2003.

L. Zhang, S. R. Lewin, M. Markowitz, H.-H. Lin, E. Skulsky, R. Karanicolas, Y. He, X. Jin, S. Tuttleton, M. Vesanen, et al. Measuring recent thymic emigrants in blood of normal and hiv-1–infected individuals before and after effective therapy. *The Journal of experimental medicine*, 190(5):725–732, 1999.