

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Lisa Franziska Buchauer, M.Sc. (Physics)
born in Offenbach am Main

Oral examination: _____

Heterogeneous Populations in B Cell Memory Responses and Glioblastoma Growth

Referees:

Prof. Dr. Thomas Höfer
Prof. Dr. Hedda Wardemann

Denn was fängt man an am Jüngsten Tag, wenn die menschlichen Werke gewogen werden, mit drei Abhandlungen über Ameisensäure, und wenn es ihrer dreissig wären?! Andererseits, was weiss man vom Jüngsten Tag, wenn man nicht einmal weiss, was alles bis dahin aus der Ameisensäure werden kann?!

– Robert Musil, *Mann ohne Eigenschaften*

Contents

Zusammenfassung	1
Summary	3
Introduction	5
I. B Cell Immune Response to Complex Antigen.	9
1. Background - B Cell Memory Responses	13
1.1. The Humoral Immune Response	13
1.1.1. Role within the Vertebrate Immune System	13
1.1.2. Development, Characteristics and Function of B Cells	14
1.1.3. Germinal Centre Reactions	17
1.2. Computational Models of the B Cell Immune Response	20
1.2.1. Antibody-Antigen Binding Models	20
1.2.2. Agent-Based Germinal Centre Models	22
1.2.3. Analytical Germinal Centre Models	24
2. A Flexible Computational Model of B Cell Immunity	27
2.1. Motivation	27
2.2. Binding Model Allowing for Different Antigenic Complexities	33
2.3. Germinal Centre-Based Model of B Cell Memory Formation	39
2.A. Supplement: Binding Model Matrices, Pseudocode, Parameters	44
3. Vaccination Scenarios with Various Boundary Conditions	51
3.1. Antigen Complexity and Clonal Diversity	51
3.2. Multi-Shot Immunisation and General Parameter Space	56
3.3. Avalanche Effect and Repertoire Differences	61
4. Case Study: A Malaria Vaccination Trial	67
4.1. Introduction to Trial Protocol and Experimental Analysis	67
4.2. Comparison of Experimental Data and Computational Model	70
5. Discussion - B Cell Memory Responses	75

II. Stem Cell-Driven Glioblastoma Growth	81
6. Background - Cancer Stem Cells in Glioblastoma	85
6.1. Glioblastoma Multiforme	85
6.1.1. Epidemiology and Pathology of Glioblastoma	85
6.1.2. Glioblastoma Mouse Models	86
6.1.3. Cancer Stem Cells in Glioblastoma	88
6.2. Mathematical Models of Cancer	90
6.2.1. Models of Glioblastoma	90
6.2.2. Models of Spatial Aspects of Solid Tumour Growth	92
6.2.3. Models of Cancer Stem Cells	93
6.3. Mouse Models and Experiments Underlying this Work	95
6.3.1. Induction of Glioblastoma Growth	95
6.3.2. Co-Expression of Tlx and GFP	95
6.3.3. Bioluminescence Imaging of Tumour Volume	95
6.3.4. Viral dsRed-Labeling of Proliferating Progeny	96
6.3.5. TAM-Induced Expression of Rosa26-Brainbow Colours	96
6.3.6. TAM-Induced miRNA Knock-Down of Tlx and Switch of Fluorescent Protein Cassettes	97
7. Exponential Tumour Growth and Cellular Migration	99
7.1. Phenomenological Modelling of Tumour Growth Curves	99
7.2. Three-dimensional Tumour Growth Simulation	100
8. Model Development and Calibration	109
8.1. A Mathematical Model of Glioblastoma Growth	109
8.2. Simulation of Population Experiments	111
8.2.1. BLI Growth Curves	111
8.2.2. Decline of proliferation index in the progeny compartment	111
8.2.3. Tumour composition information	112
8.3. Parameter Estimation, Method and Results	114
8.3.1. Short Introduction to Bayesian Parameter Estimation	115
8.3.2. MCMC Sampling of the Posterior Distribution	116
8.3.3. MAP Parameter Estimates and Confidence Intervals	119
9. Model Applications	123
9.1. IdU incorporation study	123
9.2. Tracing of single-cell derived clones	127
9.3. Treatment Approaches	135
9.3.1. Chemotherapy with Temozolomide	135
9.3.2. Induced knockdown of Tlx	138
10. Discussion - Cancer Stem Cells in Glioblastoma	141
Bibliography	147

Zusammenfassung

Heterogenität ist ein Markenzeichen biologischer Systeme auf jedweder Skala. In dieser Arbeit entwickle ich mathematische Methoden, welche die Beschreibung biologischer Heterogenität ermöglichen und untersuche diese im Kontext zweier biomedizinisch relevanter Anwendungsgebiete. Hierbei handelt es sich zum Ersten um das Hervorrufen schützender B-Zell-Immunantworten durch Impfung und zum Zweiten um die Wachstumsdynamik eines aggressiven Gehirntumors.

In der großen Mehrheit der derzeit lizenzierten Impfungen sind Antikörper-Titer starke Korrelate der Immunität. Allerdings widersetzen sich Krankheiten wie Influenza, Tuberkulose und Malaria bis jetzt der Entwicklung effizienter Impfungen und häufig sind die Mechanismen, welche etablierten Impfungen zugrunde liegen, nicht vollständig verstanden. Im ersten Teil dieser Arbeit entwickle ich daher ein dateninspiriertes mathematisches Modell der B-Zell-Gedächtnisantwort, welches auf einem Ensemble aus computersimulierten Keimzentren basiert. Durch verschiedene pathogen- und patientenspezifische Parameter erlaubt dieses Modell die Beschreibung von Impfszenarien unterschiedlicher Schwierigkeit. Anhand des Modells zeige ich, dass die Verfügbarkeit hochwertiger Vorläuferzellen als Basis für klonale Selektion und die Effizienz der Affinitätsreifung in Abhängigkeit der Bindungskomplexität separate Engpässe für eine erfolgreiche Impfung darstellen. Gemeinsam mit experimentellen Kollaborationspartnern nutzen wir diese Resultate zur Erklärung von Einzelzell-Immunoglobulin-Sequenzdaten aus einer Impfstudie gegen den Malaria-Parasiten *Plasmodium falciparum* (*Pf*). Wie für komplexe Antigene vorhergesagt, offenbart sich die klonale Selektion wirksamer Vorläuferzellen mit Keimbahn-Sequenzen im Vergleich zur Affinitätsreifung nach wiederholter Vakzination mit *Pf* Sporozoiten als der effizientere Mechanismus, da die Mehrheit der Immunoglobulin-Genmutationen affinitätsneutral ausfällt. Diese Ergebnisse sind für die Entwicklung von potentiell personalisierten Impfprotokollen gegen strukturell komplexe Antigene von grundsätzlicher Bedeutung.

Ein quantitatives Verständnis der funktionellen Zellheterogenität von wachsenden Tumoren verspricht Einsichten in die Grundlagen der Krebsbiologie. Im zweiten Teil dieser Arbeit entwickle ich entsprechend mathematische Modelle des Glioblastom-Wachstums. Unter Verwendung einer Bayes'schen Herangehensweise an die Parameterschätzung und einer großen Menge experimenteller Daten aus Mausmodellen zeige ich, dass das exponentielle Gehirntumorstammwachstum von Gehirntumor-Stammzellen getrieben wird, während weiter ausdifferenzierte Nachkommen dieser Zellen trotz höherer Teilungsraten isoliert nicht zum Wachstum beitragen. Aus dem Vergleich einer drei-dimensionalen Tumorstammwachstumssimulation mit experimentellen Wachstumskurven leite ich ein hoch migratives Verhalten von Tumorstammzellen her. Auf der Basis von Einzelzell-Tracingdaten und einer Kombination aus deterministischen und

stochastischen Modellanteilen bestimme ich die entsprechende Migrationsrate und interpretiere experimentell beobachtete Klongrößen-Verteilungen. Schlussendlich nutze ich das nunmehr voll quantifizierte Tumorwachstumsmodell, um das Ansprechen auf therapeutische Interventionen vorherzusagen. Diese Vorhersagen wurden durch unsere Kollaborationspartner experimentell bestätigt, was impliziert, dass ein quantitatives Verständnis der hierarchischen Populationsstruktur eines Tumors zu informierten Behandlungsentscheidungen beitragen kann.

Summary

Heterogeneity is a hallmark of biological systems at every conceivable scale. In this work, I develop computational methods for describing various interacting types of biological heterogeneity. I apply them to explore two scenarios of biomedical interest: the evocation of protective B cell responses by vaccination and the growth dynamics of an aggressive brain tumour.

In the vast majority of currently licensed vaccines, antibody titres are strong correlates of vaccine-induced immunity. However, diseases like influenza, tuberculosis and malaria continue to escape efficient vaccination, and the mechanisms behind many established vaccines remain incompletely understood. In the first part of this work, I therefore develop a data-driven computational model of the B cell memory response to vaccination based on an ensemble of simulated germinal centres. This model can address immunisation problems of different difficulty levels by allowing both pathogen- and host-specific parameters to vary. Using this framework, I show that two distinct bottlenecks for successful vaccination exist: the availability of high-quality precursors for clonal selection and the efficiency of affinity maturation dependent on binding complexity. Together with experimental collaborators, we have used these results to interpret single-cell immunoglobulin sequencing data from a vaccination trial targeting the malaria parasite *Plasmodium falciparum* (*Pf*). As predicted for a complex antigen, after repeated immunisation with *Pf* sporozoites, the clonal selection of potent germline and memory B cell precursors against a major surface protein outpaces affinity maturation because the majority of immunoglobulin gene mutations are affinity-neutral. These findings have implications for the design of potentially personalised vaccination strategies to induce potent B cell responses against structurally complex antigens.

A quantitative understanding of functional cell heterogeneity in tumour growth promises insights into the fundamentals of cancer biology. In the second part of this work, I correspondingly develop mathematical models of glioblastoma growth. Employing a Bayesian approach to parameter estimation and incorporating a large body of experimental data from mouse models, I show that brain tumour stem cells drive exponential tumour growth while more differentiated tumour progenitor cells, although fast cycling, are unable to sustain expansion by themselves. Comparing a three-dimensional simulation of tumour growth to experimental growth curves, I derive that glioblastoma stem cells are highly migratory. Based on single-cell clonal tracing data and a combination of deterministic and stochastic modelling approaches, I identify their migration rate and explain experimentally observed clone size distributions. Finally, I employ the resulting fully quantified model of tumour growth to predict the response to two therapeutic interventions. These predictions were verified experimentally by our collaborators, suggesting that quantitative knowledge on the hierarchical subpopulation structure of a tumour may provide valuable guidance for treatment.

Introduction

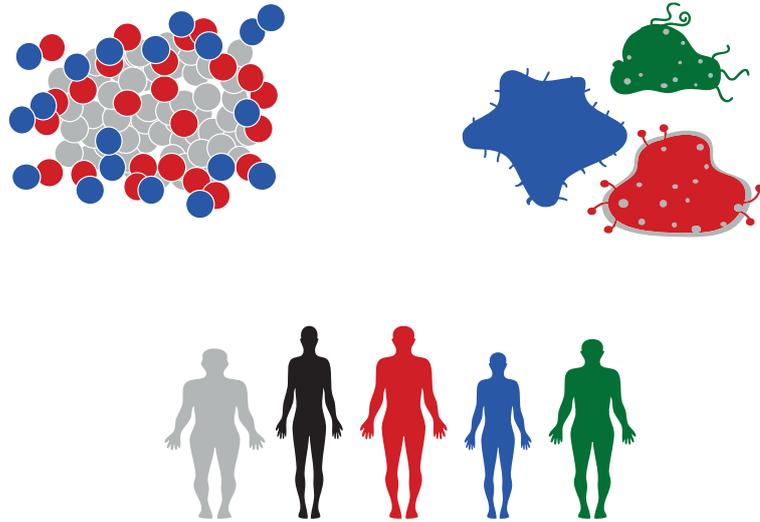
Die Wahrheit ist eben kein Kristall, den man in die Tasche stecken kann, sondern eine unendliche Flüssigkeit, in die man hineinfällt.

– Robert Musil, *Mann ohne Eigenschaften*

Heterogeneity at every conceivable scale is a hallmark of biological systems. While this is formidable when contemplating natural diversity at the genetic, cellular or species level, it leads to significant medical challenges in a variety of fields. In this work, computational methods for describing biological heterogeneity at different scales are explored and developed in two scenarios of medical interest: the aggressive brain tumour glioblastoma and the evocation of protective B cell responses by vaccination.

Cancer in general is a highly heterogeneous disease [Meacham and Morrison, 2013, Burrell et al., 2013, Melo et al., 2013, Fisher et al., 2014]. This heterogeneity begins with differences between individual cancer patients and continues in the spatio-temporal domain: one metastasis is unlike the next even within a single patient and through mutation-driven evolution two samples taken as little as a few weeks apart from each other may also be vastly different. At the cellular level, this diversity is due to both genetic and epigenetic heterogeneity as cancer cells with identical genomes may be present in different phenotypes reminiscent of specialised cell types within normal tissue. All these factors make cancer treatment a difficult problem - because of differences in early mutational events which are thus present throughout the cancer cell population, what works in one patient has no effect in the next; because of pharmacokinetic differences in drug delivery and absorption a metastasis at one site is harder to tackle than at another; and because of different cellular genomes or phenotypes some cells may be by and large protected against a treatment at hand.

Similar problems exist in the realm of vaccine development. While a number of established vaccines reliably lead to long-term protection in the vast majority of patients, this is not true in all cases [Plotkin et al., 2008, Plotkin, 2014]. Several widely spread diseases escape efficient vaccination as yet. The malaria parasite *Plasmodium falciparum* is one example in which the



Heterogeneous populations of tumour cells, pathogens and people.

most advanced vaccine candidate RTS,S AS01 leads to transient protection in only 25 to 55% of cases [Neafsey et al., 2015, Olotu et al., 2016]. Several types of heterogeneity prevent the development of a “one-protocol-fits-all” optimal vaccination strategy to be used across pathogens and patients. Pathogens are equipped with diverse strategies and tricks in order to hinder the efficient development of immune memory. These comprise particularly challenging antibody binding sites, a within-host life cycle set up to confuse the immune system, the rapid accumulation of mutations in order to continuously present different surface antigens and the similarity of crucial epitopes to self-peptides. Vaccine recipients on the other hand differ from each other due to systemic reasons such as immune responses of varying strengths as well as because of genetic factors such as the availability of correct antibody templates in their genetic code.

Aspects of both of these medical scenarios are investigated using mathematical methods in this thesis. In the first part, an agent-based simulation of the B cell immune system is developed. Agent-based computational models do not summarise individual players, in this case B cells, into populations, but allow for tracking them individually. Because individuality at the cellular level is an important aspect of the B cell response, where somatic mutations can alter the binding strengths between an individual cell’s antigen receptor and its target, this computational framework is specifically suitable here. Every computational B cell agent is equipped with a list of personal properties, one of them being its affinity for the vaccine antigen. Based on these properties individual decisions can be taken. In addition to allowing for this crucial cell-to-cell variation, the new modelling framework incorporates other types of heterogeneity, notably the diversity of pathogens. Our model is thus not tied to the description of one specific vaccine antigen, but allows to vary some of the factors causing pathogenic diversity as introduced above. These comprise epitope complexity, repertoire availability and aspects of potential pathogenic life cycles. Patient diversity can be modelled by altering the immunological repertoire with which each individual is initially equipped to fight the invader.

Agent-based models are also employed for two different purposes in the second part of the thesis in which heterogeneous cell populations within mouse model glioblastomas are investigated. While individual agents are beneficial in the B cell case in order to carry cell-specific affinities, here we make use of them because (1) they are easily extendible into the spatial world and (2) they are a suitable tool for analysing stochastic clonal fates. In the first case, tumour cell agents are tracked on a three-dimensional grid revealing a fundamental connection between tumour growth pattern and tumour architecture. In the second case, the agent-based approach is used to supplement a simple compartmental model based on ordinary differential equations. While this type of model is able to describe the diverse dynamics of the experimentally identified cell types glioblastoma stem cells, proliferating progenitors and differentiated tumour cells, it allows quantification only on the basis of population data. Using a combination of deterministic population description and stochastic clonal description, we are able to introduce single-cell clonal tracing data into the parameter estimation process.

Overall, this thesis explores several dimensions of biological heterogeneity in the context of medical decision making. Using a combination of analytical and computational methods, models incorporating such diversity are developed and employed to further mechanistic and quantitative understanding of the consequences of population heterogeneity. It emerges that extracting general principles in biology may not mean the same thing as extracting general principles in physics: while in the latter case, universal laws and universal parameters should ideally govern systems at all scales, in the former, finding general principles inevitably entails defining the right parameters for introducing heterogeneity. Mathematical approaches are necessary to identify these parameters since they largely defy exploration in the natural world.

Part I.

**B Cell Immune Response to Complex
Antigen.**

[...] der Rechenschieber, das ist ein kleines Symbol, das man in der Brusttasche trägt und als einen harten weißen Strich über dem Herzen fühlt: wenn man einen Rechenschieber besitzt, und jemand kommt mit großen Behauptungen oder großen Gefühlen, so sagt man: Bitte einen Augenblick, wir wollen vorerst die Fehlergrenzen und den wahrscheinlichsten Wert von alledem berechnen!

– Robert Musil, *Mann ohne Eigenschaften*

In the first part of this thesis, an agent-based model of the B cell immune response is developed based on a large body of experimental observations and parameters from the literature. This model is then applied to explore the effect of vaccination for antigens and patients with different characteristics; finally, properties emerging from the model are employed to explain experimental data derived from a malaria vaccination study.

Chapter 1 provides a rough introduction to the vertebrate immune system and its components with a focus on B cell immunity and the germinal centre reaction as well as to previous works modelling these processes. Following this, chapter 2 introduces the computer simulation of the B cell memory response to antigen exposure developed in this thesis. The chapter starts with a motivational section introducing the need for a new type of binding model. In chapter 3 the model is used to explore a range of varying boundary conditions concerning B cell-biologic processes, patient- and pathogen specific factors and also effects of choices in vaccination protocol design. Developing further on these results, chapter 4 shows how the model can be used to make more sense of experimental data, in this specific case using single-cell B cell information from a vaccination trial aimed against the malaria parasite *Plasmodium falciparum*. Finally, an integrative discussion of all results is provided in chapter 5.

All experimental work described in chapter 4 was performed by Rajagopal Murugan in the Division of B Cell Immunology (DKFZ Heidelberg) headed by Hedda Wardemann. Parts of sections 2.2, 2.3 and 4.1 were previously published in [Murugan et al., 2018] and have been written by the candidate herself.

1. Background - B Cell Memory Responses

In order to provide background knowledge for the development and analysis of a comprehensive simulation of the B cell memory response in chapters 2, 3 and 4, this section explores both the biological and theoretical preliminaries. In section 1.1, an overview of antibody-mediated immunity, its cellular actors and generating processes is provided. Section 1.2 discusses previous attempts at translating this complex biological system into equations and simulations.

1.1. The Humoral Immune Response

The antibody-mediated part of vertebrate immunity is an element of humoral immunity which builds on processes requiring the presence of macromolecules such as antibodies in extracellular fluids (for example blood, lymph) called *humors* in ancient greek medicine. Antibodies are produced by B cells, a constituent of the adaptive immune system. In the following sections, B cells and associated processes will be introduced shortly. Following a placement of the B cell immune response within the entire vertebrate immune system in section 1.1.1, development and functions of B cells will be discussed in section 1.1.2. An important process within the B cell cosmos, the germinal centre reaction, will be introduced in section 1.1.3.

1.1.1. Role within the Vertebrate Immune System

In general, the vertebrate immune system is composed of several layers of defence systems against pathogens (viruses, bacteria, fungi and parasites in order of increasing size). These comprise innate and adaptive immune systems both of which rely on the actions of leukocytes (white blood cells). All leukocytes are differentiated progenitors of hematopoietic stem cells and thus parts of the blood-forming system.

Innate Immunity Innate immune cells are characterised by the absence of antigen-specific receptors on their surfaces and by their ability to respond quickly (within minutes) to threats entering the body. The innate immune cell types such as macrophages, neutrophils and dendritic cells carry pattern recognition receptors on their surfaces which allow them to recognize and bind to pathogen-associated molecular patterns such as lipopolysaccharides of the bacterial cell wall which are not found in the vertebrate host. While macrophages are mostly tissue-resident and responsible for taking up and killing invading microorganisms as well as infected host cells, the granulocyte subtypes eosinophils and basophils release toxic chemicals into the periphery which help combat invaders too large for uptake. The remaining and most abundant type of granulocytes, the neutrophils, as well as the dendritic cells, also take up invaders by phagocytosis. Additionally, dendritic cells enzymatically digest the pathogen they

have taken up and present it to cells of the adaptive immune system which are thus activated and undergo proliferation [Murphy and Weaver, 2016].

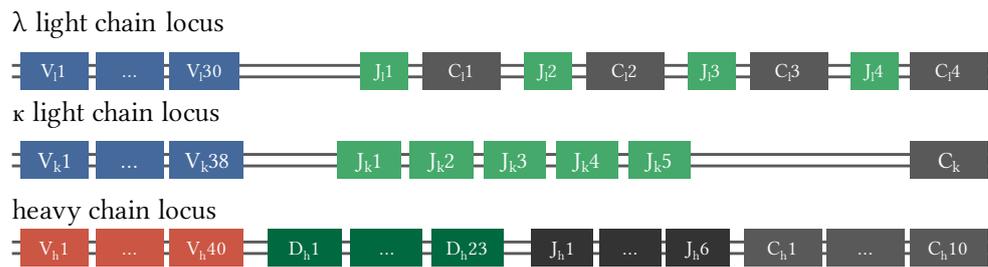
Adaptive Immunity The adaptive part of the immune system in turn consists of so-called lymphocytes which display antigen-specific receptors on their surface. T lymphocytes are the mediators of cellular immunity as they interact directly with other host cells in order to initialise immune processes. The T cell receptor is a protein complex composed of two parts, predominantly the α - and β -chains. These are assembled from a variety of possible gene segments during the maturation of T cell in thymus in a process referred to as VDJ-recombination (outlined for B cells in section 1.1.2). As a result, a theoretical diversity of up to 10^{20} different receptors is generated [Miles et al., 2011], thus increasing the probability of recognizing any given pathogen. T cells are subdivided into three major forms: cytotoxic T cells are responsible for sending apoptosis signals to other host cells which are infected with intracellular pathogens; helper T cells interact with other host cells such as B cells and prime them to take action; and regulatory T cells have a role in limiting the extent of immune responses and thus avoiding potential damage through overreaction. Through contact with antigen presented by dendritic cells, T cells become activated, expand clonally and partly differentiate into memory cells which help retain immunological memory of a past infection for many months or years [Murphy and Weaver, 2016].

In the following sections and chapters, we will focus on the second pillar of adaptive immunity, the B cell immune system. Both branches of the adaptive immune response have in common that compared to the innate responses it takes longer to activate their cellular players (several hours) and even longer (several days) to fully develop and mature them by means of clonal selection and expansion, differentiation into effector cells and, in the case of B cells, germinal centre reactions (see section 1.1.3). In turn, these responses last longer, are more specific to the threat at hand and have the potential of leading to immunological memory thus diminishing the effects of reinfection [Murphy and Weaver, 2016].

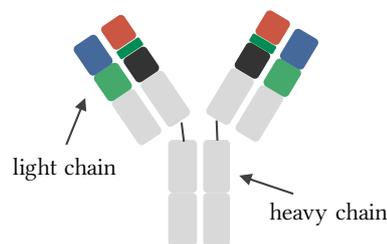
1.1.2. Development, Characteristics and Function of B Cells

Before leaving the bone as mature naïve lymphocytes, B cells move through several developmental stages within the bone marrow where, most importantly, their B cell receptor gene locus is rearranged. The mature naïve phase of their life ends once they encounter foreign antigen which leads to their activation and a choice of fate which may include entering a germinal centre or direct differentiation into an effector cell type. In the following, the developmental steps of B cells and antibodies as well as their functions are outlined.

V(D)J-recombination and B cell development B cells are ultimately derived from hematopoietic stem cells which reside in the bone marrow. These cells first differentiate into multipotent progenitors which in turn give rise to common lymphoid progenitors. The latter are able to produce lymphocytes of the B and T lineages as well as to natural killer cells [Höfer et al., 2016]. Once a common lymphoid progenitor is committed to the B cell lineage, a num-



- (a) Composition of the genetic loci from which the antibody constituents, the κ or λ light chain and the heavy chain, are composed. The combinatorial diversity from random segment usage alone is on the order of 300 for the light chains and 11,000 for the heavy chain [Murphy and Weaver, 2016].



- (b) Schematic structure of an antibody. Antibodies are complexes of four proteins, two copies each of the so called light and heavy chains.

Figure 1.1.: Schemes of immunoglobulin gene loci and an antibody. Colours of the genetic loci in part a match the colours of the corresponding regions in part b.

ber of further differentiation steps ensues [Welinder et al., 2011]. The developmental stages of B cells until maturity are termed pro-B and pre-B cells in order of occurrence [Hardy et al., 1991, Allman et al., 1999].

As introduced above, highly specific antigen-receptors are a hallmark of adaptive immunity. As sketched in figure 1.1b, the B cell receptor which is similar in structure to its secreted counterpart, the antibody, is a complex consisting of four proteins: two identical copies of a protein termed heavy chain and two identical copies of a corresponding light chain. For the light chain, two different gene loci, the κ and the λ locus, are present in the genome. It is of major importance for the functioning of humoral immunity that these B cell receptors are not alike among naïve B cells but instead may assume one of around 10^{18} unique forms, as much as 10^{14} of which can be generated during development [Xu and Davis, 2000, Yaari and Kleinstein, 2015, Elhanati et al., 2015] making use of a process called VDJ-recombination [Hozumi and Tonegawa, 1976]. The germline loci coding for the heavy and light chains contain several different variants of each of three segments necessary for constituting antibody variable regions (see figure 1.1a). These three types of segments are called V, D and J standing for *variable*, *diversity* and *joining* segments respectively. Recombination begins at the pro-B differentiation stage.

While the immunoglobulin heavy chain is composed of V, D and J segments, the light chain does not possess a diversity region. The rearrangement of the heavy chain ensues first; only

cells with a viable fully recombined heavy chain pass from the pro-B to the pre-B stage. Here, the light chain segments are recombined and checks for compatibility with the already existing heavy chain are performed. Successful recombination of both chains prompts allelic exclusion so that the recombination process on the other chromosome is abolished. This is necessary to ensure that every B cell expresses receptors from a single rearrangement of heavy and light chain genes only. An important factor contributing to the overall resulting diversity of the B cell receptor repertoire is that the V, D and J segments are not merged faithfully but in an error-prone way allowing for the introduction or loss of nucleotides between them [Roth, 2014]. As the quasi-random recombination process has the potential to lead to auto-reactive B cells and indeed this occurs at a high rate, several checkpoints eliminating such cells are part of the maturation process and the B cell life cycle [Hartley et al., 1993, Wardemann et al., 2003].

Structure and Function of Antibodies B cell receptors are the membrane-bound form of antibodies which themselves are secreted into the extracellular space [Reth, 1992]. Antibodies lack the transmembrane domain of B cell receptors - this is achieved through alternative splicing. As shown in figure 1.1b they are composed of two heavy chains and two light chains which are connected to each other covalently by disulfide bridges to form a Y-shape [Amzel and Poljak, 1979]. Each chain has a variable domain which is composed via V(D)J-recombination as described above and a constant domain which is always identical in the case of the light chain but may assume one of five forms for the heavy chain. These different heavy chain constant regions lead to different antibody classes (Immunoglobulin M (IgM), IgD, IgE, IgG and IgA) which differ in their precise functions [Schroeder and Cavacini, 2010]. The variable regions of both the heavy and the light chain are comprised of three complementarity determining regions (CDRs) alternating with framework regions (FWRs). Most of the diversity of the antibodies is contained within the CDR regions with CDR3 being the most diverse, as it contains all of the D segment and the surrounding VD- and DJ-junctions in the heavy chain and the VJ-junction in the light chain [Xu and Davis, 2000]. In turn, framework regions are mostly, though not exclusively, responsible for the structural stability of the proteins while showing less variability. The region with which antibodies bind antigenic epitopes is also called paratope and is mostly located at the end of the arms of the Y-shape where heavy and light chain CDRs come together [Davies and Cohen, 1996].

Antibodies act through a variety of mechanisms. Their most straight forward function is the direct neutralisation of a pathogen by binding to its surface. This antibody coating inhibits further interaction with the host such as bacterial or viral entry into host cells. Further, during opsonisation, antibodies mark invading agents as pathogenic for phagocytic immune cells which are thereby primed to engulf the antibody-coated organism. Antibodies with the constant regions IgM and IgG can also activate the complement system ultimately leading to phagocytic clearance or cell destruction. Through antibody-dependent cell-mediated cytotoxicity, natural killer cells are primed to induce cell death of infected host cells. Additional functions of antibodies include the inhibition of biofilm formation and cooperation in the presentation of antigen by follicular dendritic cells to B cells [Lu et al., 2018].

Types and Functions of Effector B cells Following activation by cognate antigen, naïve B cells undergo differentiation into one of two main effector cell types, antibody-secreting cells, also called plasma cells, and memory B cells, which do not continuously perform an effector function but can be restimulated easily in a secondary infection [Yoshida et al., 2010]. It is generally accepted that most differentiated B cells derive from germinal centre reactions (see also section 1.1.3 and figure 1.2), a relationship first suggested in the sixties [Thorbecke et al., 1962], but there is evidence that some types of memory B cells and plasma cells can be generated in the absence of germinal centres as well [Matsumoto et al., 1996, Kato et al., 1998, Karrer et al., 2000]. On the way to becoming a plasma cell starting from a naïve or memory B cell, cells pass through a precursor stage termed plasmablast. These plasmablasts have short life times of about one week unless they successfully migrate to a survival niche such as the bone marrow or inflamed tissue where they may persist [Radbruch et al., 2006].

Memory B cells can rapidly differentiate into plasma cells upon secondary infection, notably faster than naïve cells [DiLillo et al., 2008]. In this context, it has been suggested that long-lived humoral immunity could be due to the constant proliferation and differentiation of surviving memory B cells into antibody-secreting cells primed by persistent antigen and polyclonal stimuli, thereby offering a way of maintaining life-long immunity [Ochsenbein et al., 2000, Bernasconi et al., 2002]. However, there is no consensus across antigens or studies as to whether circulating memory B cell numbers and serum titres are correlated [Yoshida et al., 2010]. In addition to providing a quick pathway to new plasma cells upon reinfection, memory B cells are importantly also able to join secondary germinal centre reactions. Their memory function here is a different one: by supplying the germinal centre with cellular material which has already been selected for and potentially been mutationally improved during previous infections, germinal centres can produce high-quality effector cells more efficiently [Dogan et al., 2009, Pape et al., 2011, Shlomchik, 2018].

Plasma B cells were traditionally believed to be mostly short-lived because of their poor survival in culture and the observation that only low levels persist in secondary lymphoid organs over time [Nossal, 1962, Miller, 1964, Cassese et al., 2003]. However, it was later discovered that the bone marrow acts as the major niche for the maintenance of plasma cells [Benner et al., 1974] with additional evidence showing that individual plasma cells can live as long as memory B cells [Manz et al., 1997]. Here, it was also shown that these long-lived plasma cells are responsible for the maintenance of serological memory. The continued presence of protective antibodies in the serum is of major importance for preventing reinfection as it blocks pathogens even before cells of the innate or adaptive immune system need to be activated. Serum antibody titres against a number of diseases remain stable for decades even in the total absence of restimulating antigen [Hammarlund et al., 2003, Amanna et al., 2007].

1.1.3. Germinal Centre Reactions

Germinal centres are transient agglomerations of immune cells which form in secondary lymphoid organs such as lymph nodes and the spleen following exposure to T cell-dependent antigen.

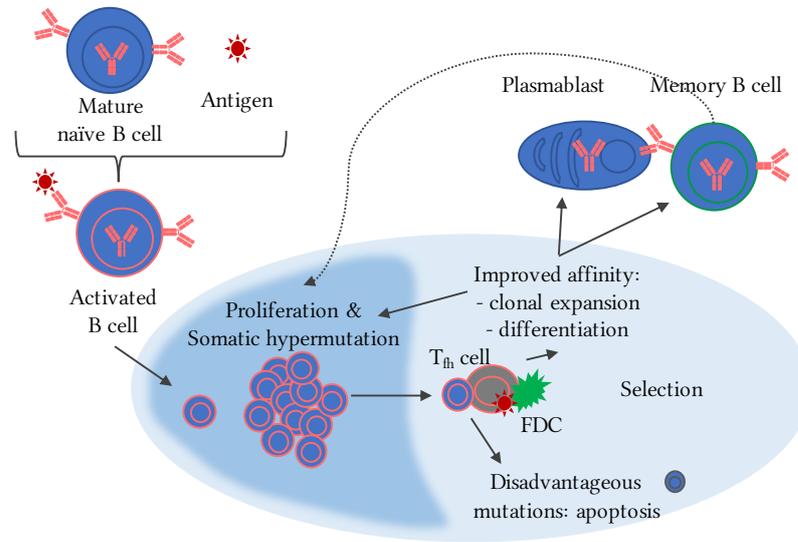


Figure 1.2.: Scheme of the germinal centre reaction. Mature naïve and memory B cells get activated to join the reaction by contact with antigen. Inside the germinal centre, they divide and potentially undergo somatic hypermutation. Based on their affinity to the antigen presented on follicular dendritic cells (FDCs) and their ability to compete for T follicular helper (T_{fh}) cell help, they are selected for further expansion or differentiation into effector cell types or removed from the process via apoptosis. See also section 1.1.3.

They consist of germinal centre B cells, follicular helper T cells (T_{fh} cells) and follicular dendritic cells. Figure 1.2 shows an overview of germinal centre architecture. Germinal centre reactions are the main sites of affinity maturation driven by somatic hypermutation and play an important role in efficient clonal selection and expansion [Shlomchik and Weisel, 2012]. In the following, the initiation of the reaction as well as the proceedings in the fully developed germinal centre are outlined.

Initiation of the Germinal Centre Reaction The construction of germinal centres follows the architecture of the underlying lymphatic tissue which consists of B cell-rich follicles and T cell zones. Germinal centres form within the follicles which are build up around a network of follicular dendritic cells [De Silva and Klein, 2015]. Following antigen exposure and the transport of this antigen to the follicular site, cognate B cells and T cells each become activated within the lymph node [Batista and Harwood, 2009], followed by B cell migration towards the border area between B and T cell zones. Here, they begin to form contacts with each other leading to full activation [Okada et al., 2005]. Not all thus activated B cells move on to become germinal centre participants as some also directly differentiate into plasmablasts or unswitched memory B cells [Paus et al., 2006, O'Connor et al., 2006, Taylor et al., 2012]. Intravital microscopy studies have shown that the migration and activation processes described here happen as early as one day after vaccination [Kerfoot et al., 2011, Kitano et al., 2011]. In addition, T cells can be primed to become T_{fh} cells by contact with follicular dendritic cells at this early

stage of germinal centre development [Choi et al., 2011]. These new T_{fh} cells then migrate to the centre of the follicle around day 3 after initial exposure and are followed by the new germinal centre B cells on day 4 [De Silva and Klein, 2015]. This completes the formation of early germinal centres which can now be identified histologically in lymph node sections.

Following the initial formation phase, germinal centre B cells begin to rapidly divide and fill up the entire follicular space defined by the network of follicular dendritic cells [MacLennan, 1994]. The size of the cell agglomeration continuously increases because of this and reaches a maximum size around day 7 after exposure [Wittenbrink et al., 2011]. By this time, two different zones of the germinal centre can be determined histologically which are referred to as light and dark zones respectively because of their appearance under light microscopy [Victora and Nussenzweig, 2012]. The dark zone contains a higher density of B cells which actively proliferate while the light zone is a more heterogeneous cell pool including B cells, T_{fh} cells and follicular dendritic cells. It is in this zone that competition for survival signals and subsequent B cell selection occur (see figure 1.2).

Structure and Dynamics of the Germinal Centre Reaction Within the dark zone, B cells divide at high rates and undergo diversification of their immunoglobulin gene regions through a process called somatic hypermutation (SHM) [Di Noia and Neuberger, 2007]. The process is governed by the enzyme activation-induced cytidine deaminase (AID), which creates errors during DNA transcription and thus induces base pair exchanges [Muramatsu et al., 2000]. The rate with which these mutations are induced is estimated at around 10^{-3} per base pair and cell division in the variable regions of the immunoglobulin genes and is thus one million times higher than the natural error rate during cell division elsewhere in the body [McKean et al., 1984, Berek and Milstein, 1987]. Through these point mutations, variants of B cells from the same clone with altered affinities may arise. AID is also responsible for the exchange of the antibody heavy chain constant domains and thus for class-switching between the different antibody isotypes IgM, IgG, IgA and IgE [Victora and Nussenzweig, 2012].

During the germinal centre reaction, B cells continuously move between the dark and light zones. Following division and potential diversification in the dark zone, they transit to the light zone in order to compete for antigen on follicular dendritic cells and for survival and division signals from T_{fh} cells. These cycles of diversification and selection enable affinity improvement of the germinal centre B cell population and drive affinity maturation, a hallmark of adaptive B cell immunity [De Silva and Klein, 2015]. While it was originally believed that antigen depots on follicular dendritic cells represent the most limiting factor for B cell selection [MacLennan, 1994], it has recently emerged that T cell help may in fact pose a stronger constraint [Victora and Nussenzweig, 2012].

The mechanism underlying this selection works via B cells taking up and digesting antigen in order to present it to T_{fh} cells on major histocompatibility complexes (MHCs). If the B cell receptor's affinity for the antigen is higher, more of it is taken up and consequently displayed, thereby directly leading to a higher stimulation from the T_{fh} cell [Victora et al., 2010]. Because

the contacts between B and T_{fh} cells are transient, a single T_{fh} cell can survey and compare many B cells in the light zone at the same time thus identifying the best candidates in the entire population [Shulman et al., 2014]. As a result of T_{fh} stimulation, selected B cells either reenter the dark zone for further clonal expansion or differentiate into an effector phenotype [Liu et al., 2015]. Unsuccessful B cells are stringently eliminated from the reaction via apoptosis. Because of the random nature of SHM, B cells may acquire self-reactivity as a part of the germinal centre reaction. Some checkpoints to eliminate these cells are in place, though they are not able to detect all types of self-reactivity [Chan et al., 2012].

1.2. Computational Models of the B Cell Immune Response

The main goal of the first part of this thesis is the development of a computational model of the B cell immune response building on an array of germinal centres as discussed in chapters 2, 3 and 4. A diverse set of theoretical germinal centre models exists in the literature and is briefly reviewed here starting with an overview of commonly used antibody-antigen binding models in section 1.2.1. This is followed by a summary of agent-based and analytical germinal centre models in sections 1.2.2 and 1.2.3. Importantly, in all models reviewed here, “agent-based” also means “stochastic”, while “analytic” translates to “deterministic”.

1.2.1. Antibody-Antigen Binding Models

A quantitative view of the binding strengths between an antibody or B cell receptor and its cognate antigen is at the base of each attempt to model a B cell response. A vast number of different abstractions of this binding process exist, mostly specialised for a chosen modelling purpose, and some of them are presented below.

Immunological Shape Space Arguably the historically most influential descriptions of antibody-antigen binding strength are the shape space-type models first introduced by [Perelson and Oster, 1979]. Here, the antibody as well as the antigenic epitope are represented by points in an n -dimensional space. The antibody is additionally equipped with a binding radius enabling it to bind to every antigen within, but not outside of, the resulting ball of stimulation. Two major classes of shape space binding models exist: Euclidean shape space, in which antibody and antigen identities are represented as n -dimensional vectors of real numbers, and Hamming shape space, in which the binding partners are characterised by sequences of length n in which each position is filled by one of k symbols. To complete the definition of a shape space model, for both cases a specific radius r needs to be chosen for the ball of stimulation. In the case of the Hamming shape space, this translates to the maximum allowed Hamming distance (number of positions in which corresponding string positions are different) between the two sequences. Much of the popularity of these shape space models is due to the fact that [Smith et al., 1997] have estimated sets of $\{n, (k), r\}$ from experimental data for both Euclidean and Hamming shape spaces. For the former, this results in either $\{n = 20, k = 4, r = 5\}$ or $\{n = 25, k = 3, r = 6\}$; for the latter, $n \in \{5, 6, 7, 8\}$ and $r \in [0.15, 0.22]$ are recommended.

As a result of these estimations, a range of later works build on them or on minor variations thereof. For example, [Meyer-Hermann et al., 2006] as well as [Chaudhury et al., 2014] use Hamming shape space models with $n = 20$ and $k = 4$ allowing for approximately 10^{12} unique sequences. Using a fixed epitope sequence, the quality of these antibody sequences is then scaled to the physiological range based on the antigen-antibody Hamming distance. [Meyer-Hermann et al., 2012] also use a four-letter shape space but enhance it by convolving the originally uniform ball of stimulation with a Gaussian distribution such that antibody sequences located further from the epitope have a smaller probability of becoming activated by it.

A variation of the Hamming-distance approach is used by [Luo and Perelson, 2015], who model an antibody as a single string and HIV as a combination of a conserved and a variable sequence in an attempt to analyse the co-evolution between immune system and virus. Each sequence is of length eight and each position is occupied by one of four letters. In order to determine binding strength, they identify the longest common substring between antibody and epitope string representations. Importantly, the matches need to be consecutive here. An even simpler variant of this method building on longest complementary matches and using two characters only has been used in the context of natural killer cell receptors and corresponding MHC molecules [Carrillo-Bustamante et al., 2015]. Variants of Euclidean shape spaces have likewise been used in the recent modelling literature [Shaffer et al., 2016, Amitai et al., 2017].

Random Energy Landscapes Shape-space models have been criticised for representing “a rosy view of evolution” [Childs et al., 2015] since mutational walks monotonically increasing affinity towards an optimum are possible without surprising drawbacks. Random energy landscapes as originally proposed for the maturation of the immune response by [Kauffman and Weinberger, 1989] represent an alternative. Here, N represents the length of the sequence while K is a parameter allowing to tune the roughness of the energy landscape between a smooth setting with a single peak and a multi-peaked noisy scenario. A given value of K implies that the functional contribution of each amino acid in the sequence is affected by K others. As a result, walks to local optima in the rugged fitness landscape decrease in length with increasing K . This model captures several experimental observations concerning affinity maturation, for example the fact that some mutations have much stronger effects on affinity than others and the observation that several matured endpoints can be reached from the same starting sequence [Kauffman and Weinberger, 1989]. [Deem and Lee, 2003] and [Sun et al., 2005] have extended the method in the immune context. Recently, [Childs et al., 2015] have used an NK-fitness model with a five-letter alphabet, a sequence length of 100 letters and an interacting neighbour number $K = 5$ in a germinal centre simulation. Concerning genetic evolution in a broader context, real fitness landscapes mapping genotypes to phenotypes have started to become available thus priming new developments in the modelling of multidimensional epistasis [de Visser and Krug, 2014]. An example for a comprehensive mutational study yielding an antibody fitness landscape is presented in [Adams et al., 2016].

Affinity Classes and Semi-Empiric Mutation Effects Specifically in the realm of analytical descriptions of the germinal centre reaction as introduced in section 1.2.3, as individual cells to which affinities need to be assigned are absent, affinity classes are broadly used. B cells

are categorised into a number of populations in order to allow a description with coupled ordinary differential equations instead of having to introduce a continuous affinity dimension and thus the need for a partial differential description. A classic treatment of somatic hypermutation processes using affinity classes of germinal centre B cells is given by [Kepler and Perelson, 1993]. [Oprea and Perelson, 1997] do not fix a number of affinity classes from the beginning of their calculations, but let subsequent affinity classes emerge as soon as the current best class is sufficiently populated. This leads to the appearance of on the order of three improved classes over the course of a germinal centre reaction. [Iber and Maini, 2002] restrict themselves to two affinity classes only, describing “low” and “high” affinity cells. That affinity classes can also be of use in agent-based approaches is shown by the implementation of [Keşmir and De Boer, 2003], where B cell agents can move up or down one class following stochastically occurring mutations. A typical observation in all of these models is that each better affinity class comes to dominate the germinal centre shortly after emergence before being superseded by its successor; a major drawback is that germline cells are typically placed in a low affinity class by definition, thus not allowing for the possibility of varying precursor qualities.

Not all analytic descriptions employ affinity classes though, as shown by [Zhang and Shakhnovich, 2010]. Here, a distribution of affinity changes caused by single mutations taken from a protein interaction database is used in order to govern the movement of B cell mass through affinity space based on a given mutation rate. An empirical component is likewise part of the sequence-based binding model developed by [Wang et al., 2015b] for antibodies binding to a HIV sequence composed of a variable and a conserved part. In this case, each sequence position may be occupied by the wild type amino acid or by a mutant, the binding energy change between the two being drawn from an asymmetric distribution with a long tail of deleterious events. Yet a different approach is taken by [Kleinstein and Singh, 2003], who are interested in the emergence of key mutant clones in the anti-hapten responses against phOx and NP. Because in these simple binding scenarios, the identification of a small number of key mutations is crucial for outcompeting clonal peers their binding model takes the shape of a decision tree including mutations in framework and complementarity determining regions, replacement and silent mutations as well as death, key and blocking events. In order to explain the low experimentally observed number of key mutants as compared to previous models, they suggest a high number of blocking mutations which hinder further affinity maturation. This concept is also used in the binding model developed in this thesis and described in section 2.2.

1.2.2. Agent-Based Germinal Centre Models

Agent-based simulations of the germinal centre reaction in which each cell is represented by a computational object which is able to store properties like mutational status and affinity are highly suitable for modelling affinity maturation because they allow individuality at a cellular level by definition. In the last decades, such models have been used both for exploring general processes within germinal centres and more specific questions building on models tailored towards given antigens or situations. Some examples for both applications are given in the following.

Early and General Germinal Centre Simulations Agent-based germinal centre models started to gain popularity around the turn of the millennium with computational power increasing and programming languages becoming more accessible to a wider user base. Early germinal centre simulations typically focus on integrating available experimental observations in order to facilitate exploration of physiological parameter spaces. An example is provided by [Shlomchik et al., 1998] who describe their simulation as “a data interpretation tool”. They combine dynamic and mutational information about germinal centres and affinity maturation such as mutation pattern, replacement-to-silent mutation ratio and the number of cells participating in a reaction and identify physiological parameters satisfying the resulting constraints. One of the outcomes of this work is the suggestion that a germinal centre may be seeded by many more cells than previously claimed based on experimental observations [Jacob et al., 1993, Kroese et al., 1987]. This is of special interest as most recent germinal centre models still insist on a low number of seeder cells (for example [Childs et al., 2015, Chaudhury et al., 2014, Wang et al., 2015b]). A work having picked up on the recent strong evidence that much higher numbers of seeders cells may in fact be involved [Tas et al., 2016] is presented by [Amitai et al., 2017]. Based on this scenario, they compare death- and birth-limited selection mechanisms and conclude that the former are more likely to give rise to the observed patterns of clonal diversity in germinal centres. In a spirit similar to [Shlomchik et al., 1998], [Kleinstein and Singh, 2003] take a striking experimental observation - the frequency with which antibody key mutations in anti-hapten responses arise - and explore which features of a model are necessary in order to reproduce the finding. This way, they identify the need for blocking mutations as part of their affinity maturation model as discussed in section 1.2.1.

In addition to general dynamic germinal centre models, several attempts at depicting the spatial structure of the reaction have been published. An early example is given by [Keşmir and De Boer, 2003] who focus on the affinity-dependent cellular adhesion of B cells on the surface of follicular dendritic cells. In an attempt to understand which selection mechanism drives germinal centre B cell competition, [Meyer-Hermann et al., 2006] build a comprehensive spatial agent-based model using more than 30 experimentally reported parameters. They conclude that in the physiological parameter regime, T_{fh} cell help alone, and not competition for antigen, can drive affinity maturation. This has since been experimentally confirmed [Victoria and Nussenzweig, 2012].

Targeting Specific Questions with Agent-Based Models In contrast to the examples given in the previous paragraph, most recent studies including germinal centre simulations focus on specific questions related to, for example, a special pathogen or vaccination schedule. Below, five examples are discussed in order of appearance. [Chaudhury et al., 2014] focus on emerging antibody cross-reactivity following administration of the polyvalent malaria vaccine AMA1. They ask how polyvalency alters specificity and cross-reactivity of the antibody response and suggest that polyvalent vaccines containing only a small number of strains may nevertheless be able to elicit broadly neutralising antibodies by altering selection pressure during affinity maturation such that cross-reactive antibodies are favoured. [Wang et al., 2015b] are likewise interested in the efficient induction of broadly neutralising antibodies albeit in the

case of the quickly mutating pathogen HIV. They find that if a cocktail of several strains is administered, affinity maturation may be driven into a frustrated state characterised by conflicting selection forces. Instead, they propose to give different strains sequentially in order to increase the probability of a cross-reactive antibody arising.

Somewhat more generally, [Childs et al., 2015] ask about the trade-offs that come with affinity maturation towards complex antigens which they define as displaying more than one immunogenic epitope. They find that in the presence of several antigens or epitopes, both the affinity and the relative breadth of the antibody repertoire are reduced. Additionally, not all epitopes are covered equally well by their simulated B cell immunity as patterns of immunodominance emerge. They conclude that the joint display of several epitopes may convey an evolutionary advantage to pathogens. In contrast to most other studies which focus on repeated realisations of individual germinal centres, [Childs et al., 2015] also marginally touch upon effects emerging in systems of several parallel reactions which their model can accommodate. Both [Luo and Perelson, 2015] and [Shaffer et al., 2016] are also concerned with pathogenic strategies as they focus on quickly mutating viral invaders. Asking whether there is way in which broadly neutralising antibodies against HIV could be driven to emerge earlier than they do during natural infection, [Luo and Perelson, 2015] find that although these antibodies have a chance of arising early, they are mostly outcompeted by more specific high-affinity strains during the initial stage. They show that computational removal of this competition leads to faster emergence of broad antibodies and suggest that vaccines containing more than one founder strain may experimentally convey the same effect. [Shaffer et al., 2016] study immunisation with diverse mixtures of theoretical strains for the same reason. They identify a “principle of optimal frustration” asserting that a well-composed antigen cocktail should contain strains that are separated by relatively large mutational distances in order to optimise the probability of eliciting broadly neutralising antibodies.

1.2.3. Analytical Germinal Centre Models

Before the emergence of agent-based simulations, analytical descriptions of germinal centre reactions were used to analyse the reaction’s properties and even today they remain useful for some types of questions. An early example of an analytic approach to hypermutation processes is given in [Kepler and Perelson, 1993] where the authors draw up germinal centre mutations as an optimal control problem. The total affinity appears as a functional of the mutation rate which in itself is a function of time; this functional is being optimised. They find a regime in which expansive mutation-free bursts and phases of fast mutation alternate and interpret this in a spatial context: cells supposedly enter and leave a part of the germinal centre in which mutations occur in cyclical fashion. In a much simpler model, [Oprea and Perelson, 1997] cast B cells into affinity classes and draw up sets of coupled ordinary differential equations of germinal centre B cells in the light and dark zones respectively and memory B cells for each class. Using this model, they show that recycling between the two zones of the germinal centre provides an efficient way of generating high affinity memory cells. The general dynamics of the reaction are of interest to [Keşmir and De Boer, 2003]. Without the need for an affinity model, they describe the emergence, duration and termination of a typical germinal centre

reaction. They find that both the total size of the germinal centre cell population and the duration of the reaction are only mildly dependent on the initial amount of antigen as the limiting effect of T_{FH} cells dominates. [Iber and Maini, 2002] introduce antibody masking into their compartmental model of germinal centre populations with low and high affinity. They suggest that this type of antibody feedback can drive affinity maturation as well as introduce outcome stability *vis-à-vis* variations in the initial concentration of antigen. The optimality of mutation and selection processes in the germinal centre as discussed by [Kepler and Perelson, 1993] has more recently been revisited by [Zhang and Shakhnovich, 2010]. They present an analytic solution to their elaborate deterministic description of a single germinal centre reaction and find that the optimal mutation rate results from a trade-off between accumulating beneficial and avoiding deleterious mutations. As far as selection is concerned, an optimal stringency is dictated by a trade-off between selecting high-affinity cells and avoiding clonal collapse.

2. Development of a Flexible Computational Model of B Cell Memory Formation

In this chapter, the agent-based computer model of the B cell memory response to vaccinations developed in this thesis is introduced. This model offers a large amount of flexibility regarding various parameters which characterise individual immunisation processes. These may have to do with specifics of the pathogen against which a B cell response should be elicited as well as with factors concerning the vaccinee. The need for a model providing this flexibility is discussed in section 2.1. Much of the flexibility concerning characteristics of the pathogen, or more specifically its epitopes, is encoded in the antibody-antigen binding model developed here and presented in section 2.2. It allows for varying complexities of binding interfaces in an abstracted way while maintaining key features of mutational processes such as the conservatism of the genetic code. Building on this binding model, section 2.3 introduces the computational description of the germinal centre reaction as developed in this work. Here, care has been taken to incorporate recent experimental discoveries while describing key processes in a condensed way. The chapter's appendix, section 2.A, contains additional information relevant for the binding model, a pseudocode version of the simulation algorithm and all simulation parameters with experimental references where applicable. Parts of sections 2.2 and 2.3 were previously published in [Murugan et al., 2018] and have been written by the candidate herself.

2.1. Motivation

As outlined in section 1.2, a variety of computational models of the B cell immune response based on germinal centre models already exist. While older models often try to deduce universal truths applicable to every vaccination process, newer models tend to focus on more specific vaccination cases by incorporating details regarding their binding interfaces or specific dynamics. This section motivates the need for a more flexible modelling approach which allows variation and comparison of pathogen-, patient- and protocol-specific parameters within a single framework.

Which Mechanisms Drive Successful Vaccination? Historically, most vaccines were developed empirically, employing principles of trial and error [Plotkin, 2014]. A common feature of these early vaccinations and newer developments relying on more rational design principles is that antibody titres are correlates of vaccine-induced immunity in the vast majority of cases [Plotkin et al., 2008]. The uppermost panel of figure 2.1 shows a sketch of the ideal titre dynamics: following antigen exposure, be it through natural infection or a vaccination, the titre

2.1. MOTIVATION

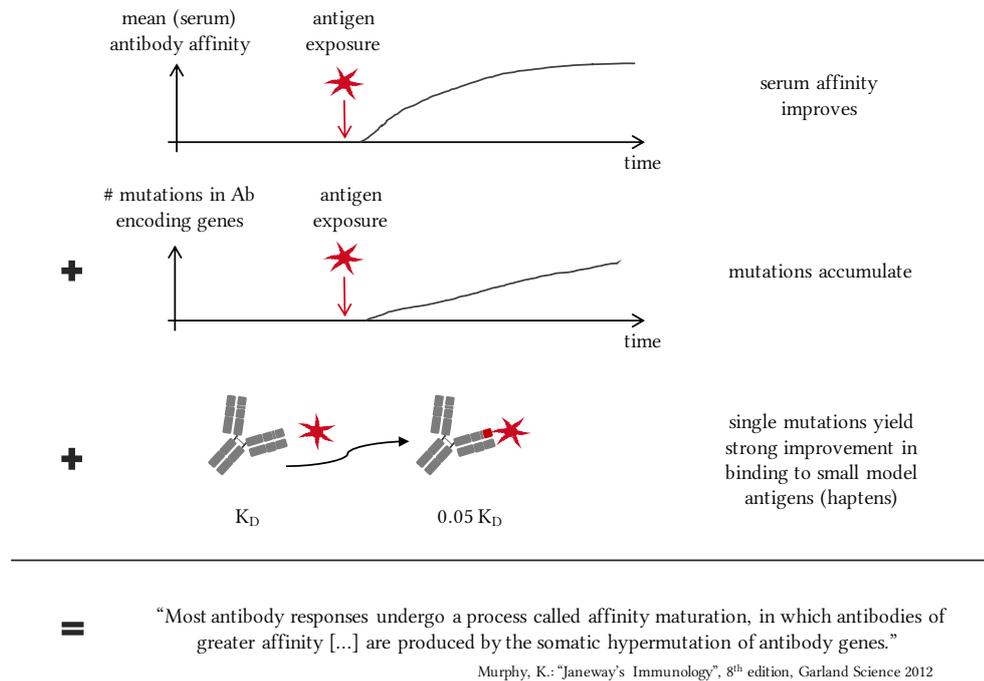


Figure 2.1.: Pieces of evidence having led to the common belief that intraclonal affinity maturation is chiefly responsible for the rise of serum antibody titres observed after vaccination.

begins to rise, reaches a level sufficient for sterile protection and ideally does not decay over many months or years.

With the discovery of somatic hypermutation as an important process in the B cell immune response and the increasing availability of sequencing methods for nucleic acids, evidence emerged that effector B cells such as memory B cells or long-lived plasmablasts may carry high numbers of mutations. The mean number of mutations of cells within germinal centres and memory cells in the periphery rises as a function of time following vaccination [Allen et al., 1987, Berek, 1992]. Additionally, closer analysis of stereotypical immune responses to haptens such as 4-hydroxy-3-nitrophenylacetic acid (NP) revealed that, at the level of individual B cell receptors, the affinity towards a given antigen may rise by more than one order of magnitude through a single nucleotide exchange alone [Allen et al., 1988, Berek et al., 1991]. Both observations are likewise sketched in figure 2.1.

In combination, these findings have led to the common interpretation that affinity maturation driven by somatic hypermutation is mainly responsible for the improvement of mean binding following vaccination. While this interpretation is likely true for some cases, it should be pointed out that it is not a logically necessary result of the pieces of evidence provided above. This is because antibody titres are commonly measured from blood serum and thus represent bulk measurements. Even though mutations within a B cell’s immunoglobulin genes are measured at a single cell level, the resulting sequences have commonly not been linked with

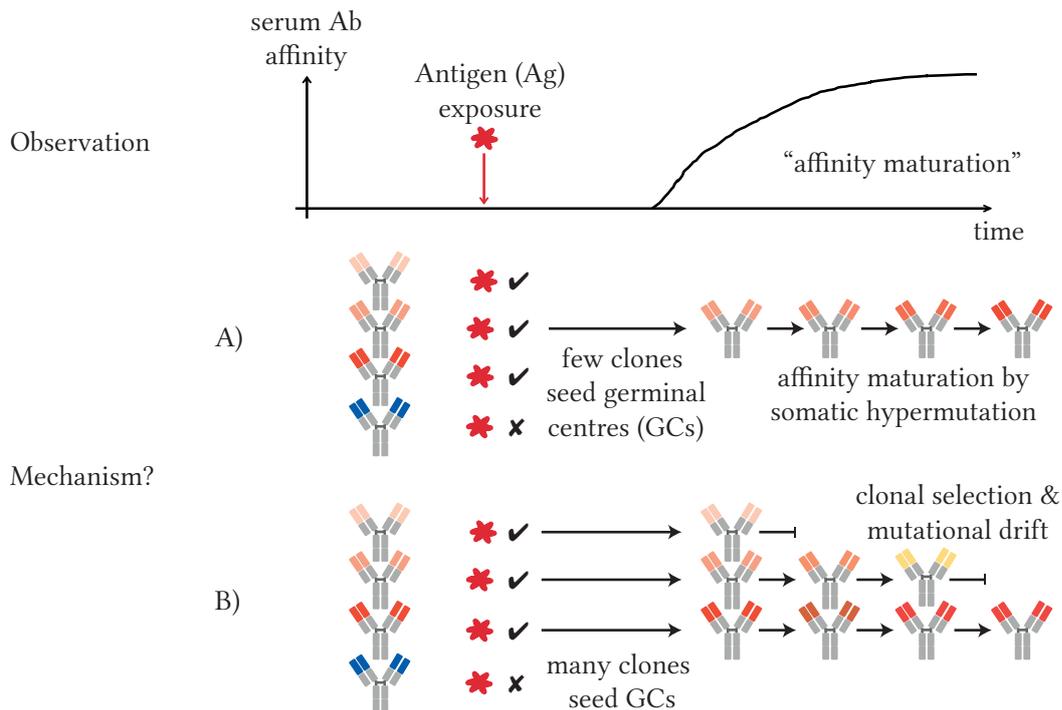


Figure 2.2.: Two basic mechanisms can underlie an observed rise in antibody serum titre; they are not mutually exclusive. Mechanism A) shows a scenario in which few clones seed each germinal centre and mean affinity improvement is then driven by somatic hypermutation and following selective expansion. This mechanism historically dominates explanations of observations. In mechanism B), a higher number of B cells seed each germinal centre, thus providing a larger base for clonal selection from the beginning. In this scenario, mutations are not necessarily required to yield improvement, they may rather lead to clonal drift and diversification while the selection and expansion of *ab initio* high-quality binders leads to the observed mean improvement.

single-cell antibody-antigen affinity measurements and much less with monoclonal protection assays. Few studies comprising this entire set of information at the single-cell level exist and in their absence, existing knowledge on haptens, where indeed individual mutations are linked to strong functional improvement, has been vastly extrapolated.

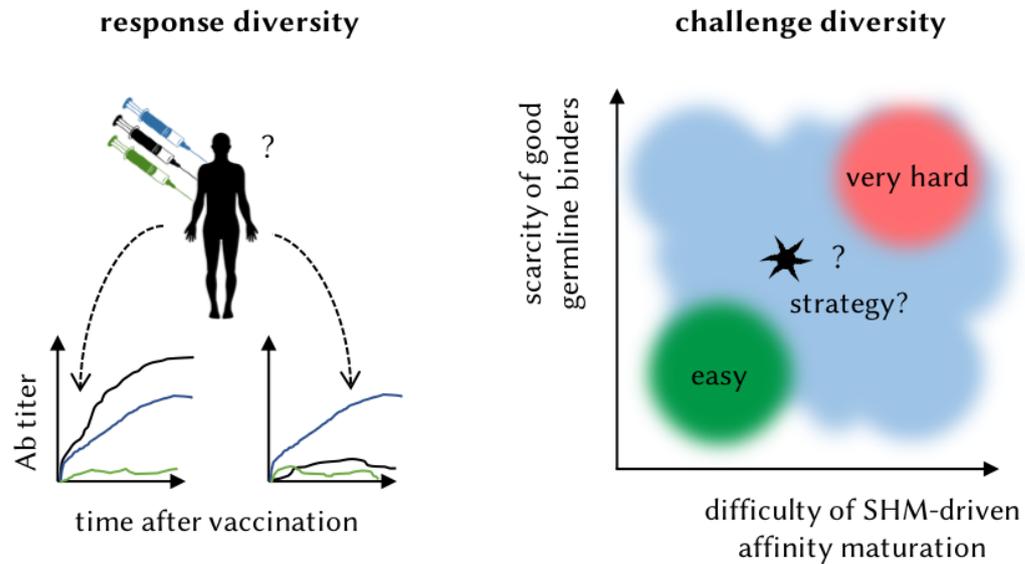
However, as figure 2.2 illustrates, two basic mechanisms can explain the observed rise in serum titre after making contact with an antigen. The interpretation discussed above stresses the importance of intraclonal affinity improvement by somatic hypermutation. On the level of a germinal centre reaction, this interpretation requires only a few weak precursor B cells to seed the reaction, affinity improvement of the population is then brought about by reliably happening beneficial mutations and concomitant expansion of the improved clones. In a second interpretation, a vast variety of binding qualities may already exist in the precursor repertoire even though low-quality binders dominate. This leads to an initially low serum titre. Upon antigen exposure, many precursor B cells are primed to enter the lymph node and participate

in germinal centre reactions. The high number of participants increases the chance of including a rare high-quality precursor. This cell then only needs to be identified and expanded relatively to its competitors; even in the hypothetical absence of somatic hypermutation, in such a system the average affinity of the population could be greatly improved over time. In this scenario, mutations are not required to be beneficial to yield the observed improvement. Instead, they may be affinity-neutral (“mutational drift”), potentially serving to keep up the diversity of the repertoire while large numbers of clonal members are being produced. Both processes are of course not mutually exclusive; in any given vaccination scenario, they are likely happening in parallel while contributing to the overall success with varying proportions.

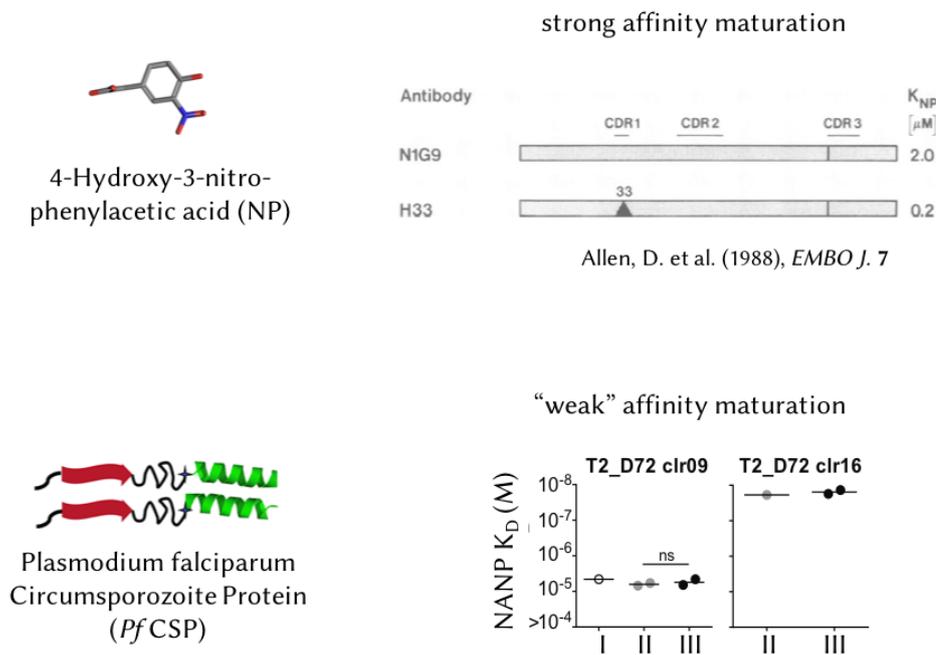
Patient- and Pathogen-Specific Factors The discussion of the previous paragraph invites a closer look at factors which may differ between individual vaccinees’ responses to antigenic contact. While many standard vaccines have a high response rate among human subjects [Peltola et al., 1994], the success rate for others remains poor [RTSS Clinical Trials Partnership, 2015, Belongia et al., 2016] for reasons that are by and large unclear. Factors that complicate the development of a stable immunological memory may be specific to a pathogen but can also depend on characteristics of the patient, thus leading to variable responses in a population. The problem and a summary of underlying factors are sketched in figure 2.3a.

On the patient-specific side, varying immunoglobulin genes available in the germline may decide whether good naïve precursor B cells are available to a person. Large-scale studies on structural variation of human genomes have repeatedly found immunoglobulin genes enriched in the most variable portions of the genome and differences between ethnic groups have likewise been suggested [Roederer et al., 2015, Sudmant et al., 2015, Auton et al., 2015, Galson et al., 2015]. Additionally, as not only naïve B cells, but also previously formed memory B cells can participate in novel immune responses, the exposure history of each individual may co-determine vaccination success. Previous exposure to structurally similar antigens may have led to an enrichment of memory B cells that can be used to fight a different (related as well potentially unrelated) pathogen. Factors unrelated to the available repertoire such as strength of the immune response in general and resulting factors like abundance and size of germinal centres may additionally account for inter-patient differences. Finally, because of stochastic effects in B cell activation, recruitment and selection, differences in post-vaccination repertoires can be observed even in cases with identical starting genomes (twin studies) [Wang et al., 2015a].

As far as pathogen-specific factors are concerned, bacteria, viruses, parasites and other threats to the human host have evolved to use a vast range of tactics against the innate as well as the adaptive immune system [Finlay and McFadden, 2006]. As described for *Plasmodium falciparum* sporozoites in section 4.1, pathogens may evade immune recognition by quickly hiding from sight in tissue cells; like the blood stage parasite of *Plasmodium falciparum* they may continuously exchange the surface antigens displayed to the immune system; like HIV, they may mutate these antigens with a high rate thus never allowing the immune system to reach a fully neutralising solution. Similarity of pathogenic antigens to self-antigens may allow longer immune evasion as the precursor repertoire exhibits holes in these areas [Mayer et al., 2015]. In summary, the challenge of developing broadly successful vaccine antigens and ac-



(a) Left: Exposed to the same vaccination protocol, individual patients may exhibit vastly different antibody responses. Right: Immunisation problems vary in difficulty depending on the availability of good B cells and the efficiency of SHM-driven improvement.



(b) While specific mutations yielding affinity improvements of one order of magnitude are routinely identified in anti-hapten-responses, for example against NP, similarly strong improvements may not always be possible in responses to more complex targets. This example shows affinity of memory B cell receptors sampled at intervals of four weeks following each of three vaccination shots (I, II and III) with *Plasmodium falciparum* parasites and assigned to B cell clusters. See also section 4. Lower panel data: Rajagopal Murugan.

Figure 2.3.: Illustrations of patient- and pathogen-specific diversity and potential implications for immunisation.

companying schedules comes in diverse levels of difficulty. Two main axes of the problem are shown in figure 2.3a: availability of good germline binders (for example driven by a person's specific repertoire, prior exposures and similarity of the antigen to self-peptides) and difficulty of reaching a good binder via somatic hypermutation (for example determined by pathogen escape strategies or patient-specific functionality of germinal centres).

An additional factor that likely contributes to the axis of difficulty of successful affinity maturation is the structural complexity of the antigen - not all antigens represent similarly simple binding targets for antibodies. This assumption is driven by an observation illustrated in figure 2.3b: while haptens such as NP, which are often used as model antigens in immunological studies, are small and fit into pockets in antibody binding sites, protein antigens like the *Plasmodium falciparum* sporozoite surface antigen Circumsporozoite Protein (CSP) are larger and thus require a more complex binding interface. NP is a classic example used to demonstrate the efficacy of affinity maturation in improving antibody-antigen binding [Allen et al., 1988]. In contrast, as described in detail in chapter 4, in response to *Pf* CSP, despite strong clonal diversification, significant affinity jumps are scarcely observed. Figure 2.3b shows the affinity of members of two exemplary B cell clones sampled after each of three vaccination boosters. Thus, structural complexity of the antigen is likely to play a major role for the efficiency of affinity maturation in germinal centres.

Benefit of a Computational Model In this work, some of the factors governing vaccine response diversity as outlined above are being explored and quantified with the help of a computational model of the B cell immune response. The basic ideas of this approach are introduced in the following using a toy model; a detailed introduction to the more sophisticated main model follows in sections 2.2 and 2.3. For the motivational result shown in figure 2.4, the two axes of vaccination challenge difficulty, availability of germline binders and efficiency of affinity maturation, were collapsed into one simple parameter each.

In line with the discussion concerning haptens and protein antigens above, the difficulty of affinity maturation is represented by an abstraction of the structural complexity of antibody-antigen binding. For an immunological challenge to be considered solved in this picture, a given number of key positions in an antibody must be represented by exactly the right amino acid. Mutations in simulated germinal centres can lead to conversion between correct and false amino acids at each of these key positions. Increasing complexity of the binding interface is encoded in an increasing number of such key positions. Figure 2.4 shows results for 1, 2 and 3 key positions. Concerning the second axis of challenge difficulty, the availability of good germline binders, this simple model incorporates a binary representation: either the correct amino acid at the correct key position can be supplied in naïve cells from germline or it cannot. Results for both cases are shown for each level of binding complexity in figure 2.4. Underlying germinal centre dynamics and selection mechanisms for this toy model were simulated as described in section 2.3.

In order to quantify the result of these simulated immune challenges, we run many such problems within individual germinal centres and measure how long it takes until the cor-

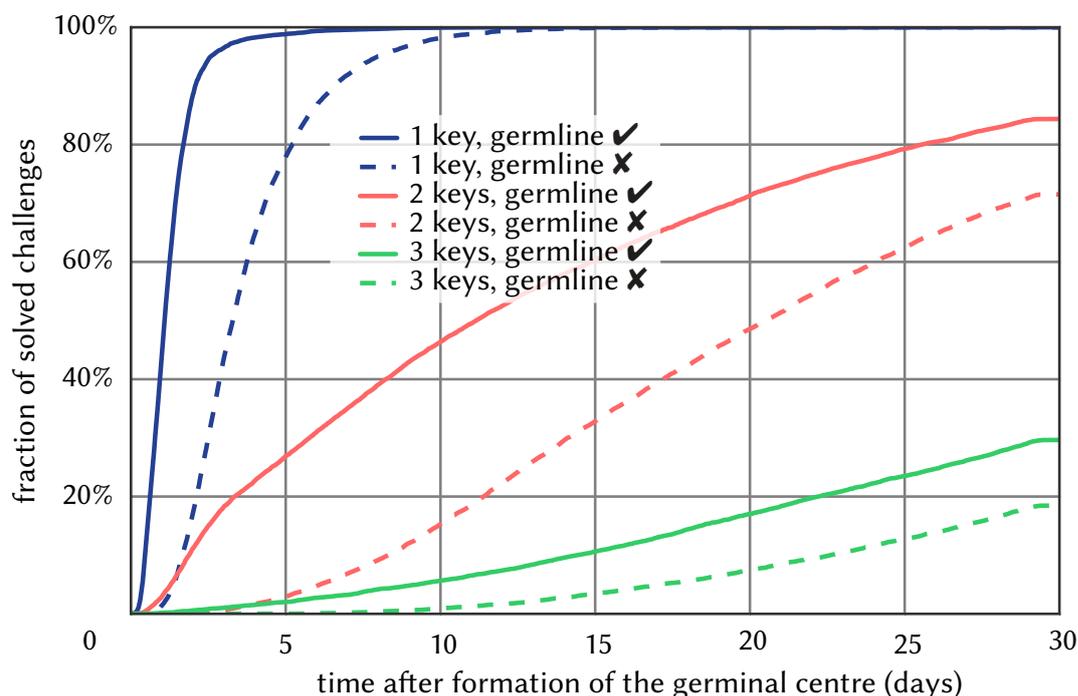


Figure 2.4.: Fraction of germinal centres that were able to solve a specific antigenic challenge assuming a simple computational model as described in section 2.1 until a given time. Three different levels of difficulty (1, 2 and 3 necessary key positions) are compared; correct keys may be available from germline (solid lines) or only be accessible via somatic hypermutation (dashed lines).

rect solution (all key positions correctly occupied) is found in each case. Figure 2.4 illustrates the main results: more complex challenges take longer to solve and patients missing useful germline input are at a disadvantage compared to their peers who can supply high-quality B cells from the bone marrow. This simple result illustrates how computational modelling allows to explore parameter effects on outcomes of interest. The main model as discussed in the following sections provides flexibility for theoretical exploration of a wide range of biologically interpretable parameters and ultimately assists in developing promising vaccination strategies.

2.2. Binding Model Allowing for Different Antigenic Complexities

During B cell production in the bone marrow a large variety of different antibody-encoding sequences is generated. Heavy and light chain genes are recombined from different available V, (D) and J segments. Variety is further increased by insertion and deletion of random nucleotides at the junctions between these segments. Diversity and variability are not distributed evenly over the variable region of heavy and light chains (both about 110 amino acids in length), but peak in three discrete regions of hypervariability in each chain. Of these six complementarity determining regions (CDRs), CDR3 on the heavy chain is by far the most diverse as it includes

both the V-D and the D-J junction and hence a high number of random amino acids [Murphy and Weaver, 2016]. In the folded antibody, the CDRs appear as protruding loops at the arms of the Y-shaped antibody. They are separated by so-called framework regions (FWRs) responsible for the stability and shape of the folded antibody. Residues within the CDRs and even more so in CDRH3 are responsible for determining binding strength to a given antigenic epitope [Burkovitz et al., 2014].

In general protein-protein binding scenarios, only a fraction of the many residues constituting either protein directly participate in binding. Their importance can be quantified using alanine scanning mutagenesis which reveals the existence of a small number of hotspots significantly contributing to the binding free energy of the complex [Moreira et al., 2007]. Shape and electrostatic complementarity as well as the ability to form hydrogen bonds between binding partners are essential for stable complex formation.

A significant part of current textbook knowledge on the beneficial effects of somatic hypermutations in antibody affinity maturation is derived from studies of immune responses against haptens. Here, recurring patterns in VDJ-segment usage as well as in mutations in specific positions along the immunoglobulin gene are observed. Together with the compact form of haptens which enter into pockets on the antibody surface during binding this has given rise to a “lock-and-key” picture in which perfect shape complementarity between antibody and antigen is decided by few key amino acids. Recently, convergent repertoire properties have been identified also in the context of medically relevant licensed vaccines or vaccination studies [Galson et al., 2014], including tetanus [Trück et al., 2015], influenza [Jackson et al., 2014], dengue [Parameswaran et al., 2013] and malaria [Murugan et al., 2018]. However, signatures in these cases are much more ambiguous than in the hapten responses and are mostly characterised by biased segment usage and/or a specific length of the CDR3 region. While relevant properties of the CDR3 region were identified in the case of dengue fever using principal component analysis in amino acid physicochemical space [Parameswaran et al., 2013], specifically beneficial single mutations have not been described.

Binding Model Concept In this work, an antibody-antigen binding model which picks up on the idea of key mutations, but expands it to accommodate more complex epitopes, is developed. Inspired by the experimental observation that immunoglobulin genes composed of certain VDJ segments preferentially appear in the pool of antibodies binding a given antigen, this model incorporates two levels of complementarity between the binding partners. Most amino acids in the antibody variable regions are involved in determining the general shape of the binding site and therefore whether binding is at all possible or not. These amino acids are referred to as “scaffold” amino acids in the context of the binding model. Biologically, this scaffold may roughly, but not exclusively, correspond to the V and J segment-encoded framework regions. In the simulation, only one epitope is being treated at a time, thus agreement of this kind is initially assumed for all B cells participating in the simulation. Scaffold-determined shape agreement can however be lost by detrimental mutations as will be introduced below. This also means that the corresponding approximately 200 amino acids do not have to be mod-

elled explicitly, a major computational advantage.

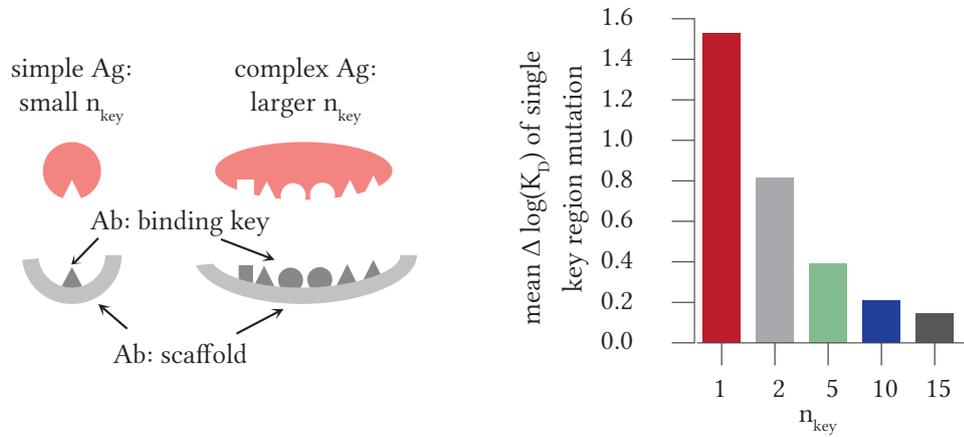
The binding quality of these shape-matched antibodies differs due to the second level of binding which is determined by an explicitly modelled sequence of amino acids. The two levels of binding are illustrated in figure 2.5a. As shown there, it is the second level of binding which allows for antigenic determinants of various complexities: for an epitope characterised by n_{key} amino acid residues on its surface, an equal number of key amino acid positions in the antibody are considered strongly relevant for binding. This number may be very low for haptens, where one key mutation can strongly increase affinity, but is higher for larger and more complex epitopes. To compare B cell immune responses to antigens of different complexities, this model therefore allows varying numbers of key residues to determine binding strength. The full physiological amino acid alphabet is being used to fill the n_{key} explicitly modelled key region positions in the antibody. Here, as a simplification, each codon (and not each amino acid) is given equal generation probability. An antigenic epitope of complexity n_{key} is similarly modelled by n_{key} random codons.

In order to derive binding energy-like values for each combination of epitope and antibody key sequences, these are aligned and for every resulting pair of amino acids a free-energy like value is read off a contact potential matrix derived statistically from a large set of folded protein structures [Thomas and Dill, 1996]. The contact potential matrix is given in table 2.5. As the lowest and highest possible binding energies are known in this model, these values are used to linearly rescale all binding energies into normalized energy values E_{bind} between 0 (worst binder) and 1 (best binder). As a result, exchanges of individual amino acids lead to smaller improvements or deteriorations of antibody affinity in cases with longer key sequences as compared to simpler epitopes with few key residues. This is illustrated in figures 2.5b and 2.5c: while a single beneficial amino acid exchange goes a long way on the normalised binding scale in the case of a single key amino acid, the process is less straight forward if several residues jointly determine binding. Here, while improvements are possible in principle, this random walk in higher-dimensional space is much less likely to reach high affinity.

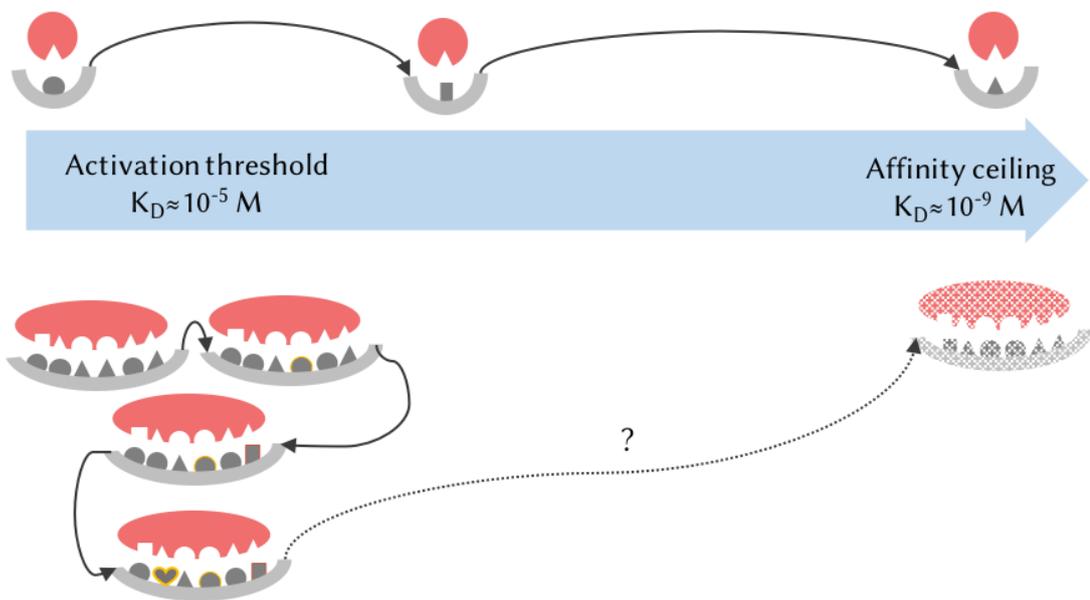
As a generic consequence of the law of large numbers, during random generation of antibody sequences, good binders become increasingly rare with rising numbers of key residues while the peak at intermediate binding energies becomes sharper, see figure 2.6a. This needs to be taken into account when comparing antigenic challenges of different n_{key} (see section 2.3). Using the experimental observation that antibody affinities usually lie in the range of $K_D = \{10^{-5}\text{M}, 10^{-9}\text{M}\}$ an energy scale in units of $k_B T$, where k_B is Boltzmann's constant and T is temperature, can be deduced. This is calculated using the relation between the standard Gibbs free energy change and standard equilibrium constant of a binding process,

$$\Delta G^\ominus = -k_B T \ln K^\ominus, \quad (2.1)$$

which evaluates to $\Delta G(K_D = 10^{-5}\text{M}) = -10.5 k_B T$ and $\Delta G(K_D = 10^{-9}\text{M}) = -21.1 k_B T$. We fix the lower end of this energy scale to the normalised energy value $E_{\text{bind}} = 0.6$ and the upper end to the ideal binder at $E_{\text{bind}} = 1$. This setting introduces degeneracy at the lower end of



- (a) Antigens are modelled as sequences of amino acids; antibodies consist of a scaffold region determining general shape agreement and a key region whose length matches that of the antigen sequence.
- (b) Depending on the length of the antibody key region, each individual mutation to one of the key amino acids contributes a larger or smaller effect to the overall affinity.



- (c) While the identification of correct key amino acids is straight forward in a mutation process acting on a simple binding interface (top), finding a perfectly matching sequence is much more difficult in a complex interface (bottom).

Figure 2.5.: A simple antibody-antigen binding model allowing for different complexities in binding interfaces.

the binding scale and moves the sequences commonly competing and dividing in the germinal centre into an energy region where beneficial mutations are less likely than deleterious ones as is experimentally observed [Adams et al., 2016], see also figure 2.6b. Weaker binders may still participate in the reaction but will mostly be outcompeted quickly.

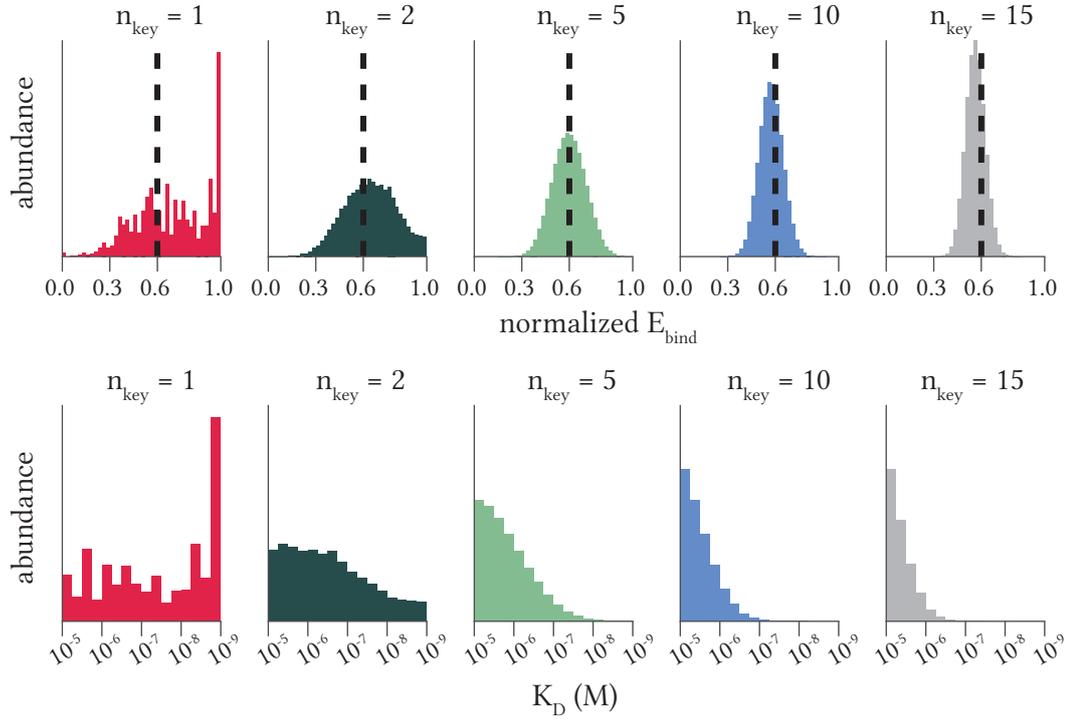
Mutations When a germinal centre B cell divides, mutations are introduced into the sequence at a rate of $p_{bp} = 0.001$ per base pair and division (see appendix 2.A). Here, the mutation rate is assumed to be uniform over all 220 amino acids in the variable regions. Using the assumption that every point mutation is independent of all others, the probabilities for 0 to 3 base pair mutations within one codon are

$$\begin{aligned}
 p_0 &= (1 - p_{bp})^3 = 0.997 \\
 p_1 &= 3 \cdot (1 - p_{bp})^2 \cdot p_{bp} = 2.994 \cdot 10^{-3} \\
 p_2 &= 3 \cdot (1 - p_{bp}) \cdot p_{bp}^2 = 2.997 \cdot 10^{-6} \\
 p_3 &= p_{bp}^3 = 10^{-9} \quad .
 \end{aligned} \tag{2.2}$$

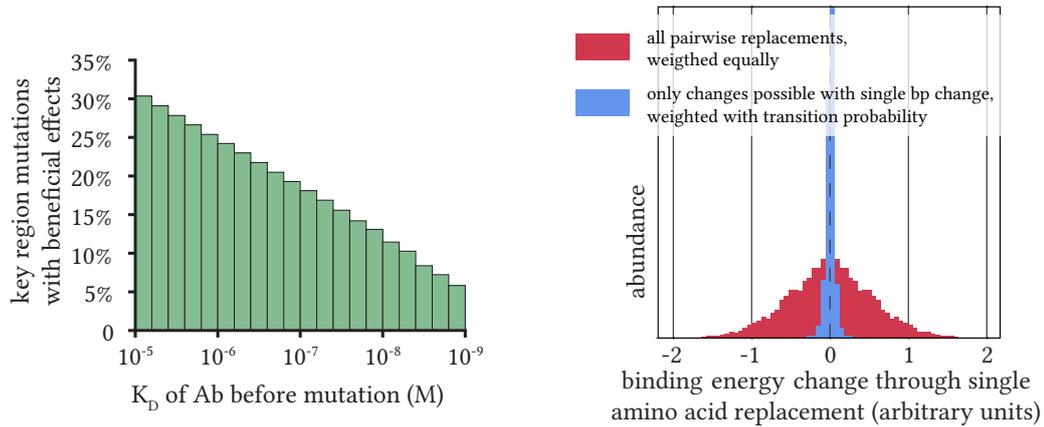
This shows that it is very unlikely to acquire more than one point mutation within a single codon and division and this possibility is thus ignored in the following. The error probability per codon is taken to be $p_{codon} \equiv p_1 \approx 3 \cdot p_{bp} = 0.003$. Using this value, the number of mutations happening within scaffold (key) regions is drawn from a binomial distribution with $n = n_{scaffold} = 220 - n_{key}$ ($n = n_{key}$) and $p = p_{codon}$ at every division. For $n_{key} = 10$, this results in mean values of $p_{codon} \cdot n_{scaffold} = 0.63$ mutations in the scaffold and $p_{codon} \cdot n_{key} = 0.03$ mutations in the key region. This simple calculation shows that even though error rates within the GC are 10^6 times higher than during normal replication, mutations in the antibody encoding genes, even before considering only relevant or even beneficial changes, are still rare events.

Mutations in the key amino acids may change the identity of the specifically modelled residue in question. Transition probabilities between amino acids are calculated taking into account all codons coding for both the start and the end amino acid and using only transitions accessible by single base pair exchanges. All base pair exchanges are assumed to be equally likely. The transition probabilities calculated this way are shown in table 2.6. Using a full amino acid alphabet with codon-determined transition probabilities and an interaction energy matrix based on real protein-protein interactions naturally introduces the conservatism of the genetic code as silent mutations as well as replacement mutations by a functionally similar amino acid are much more likely than in a random replacement model. This is shown in figure 2.6c, which compares the energy distribution resulting from random pairwise replacements with that derived using the transition probability matrix in 2.6. Figure 2.6b shows another important aspect of this binding model: the higher the affinity of a given key sequence already is, the less likely it becomes to identify further beneficial mutations as their proportion within the space of all key region mutations declines as a function of current affinity. Importantly, mutations within the key region can lead to better, worse or equal affinity according to the stochastically chosen amino acid transition. In the model they cannot, however, lead to unfoldable antibodies and

2.2. BINDING MODEL ALLOWING FOR DIFFERENT ANTIGENIC COMPLEXITIES



(a) Binding energy distributions resulting from 1000 randomly drawn antigen sequences of lengths n_{key} and 1000 randomly drawn antibody binding partners each after normalisation as described in section 2.2. Upper row: entire distribution, lower row: after truncation and mapping onto affinity scale between 10^{-5}M and 10^{-9}M .



(b) Fraction of key region mutations yielding beneficial effects as a function of the starting affinity of the antibody. The higher the affinity already is, the lower the probability to discover additional improvements.

(c) Conservatism of the genetic code: Histograms of binding potential changes after a single amino acid replacement according to table 2.5 if all pairwise exchanges are considered (red) and if only those possible via a single base pair (bp) replacement according to table 2.6 are considered (blue).

Figure 2.6.: Properties of the binding model introduced in figure 2.5.

thus to death of the corresponding B cell unless a STOP codon is generated during replacement. Conceptually, this is justified because key amino acids are not part of the structural scaffold but imagined as protruding residues at the interaction surface.

As discussed, the scaffold region consisting of the $(220 - n_{\text{key}})$ non-key amino acids is responsible for general shape correspondence with a given epitope and stability of the antibody. Mutations to this area can lead to one of three different outcomes. On average, 25% of mutations are silent and are not considered further. A fraction p_{death} of replacement mutations results in unfoldable antibodies or total loss of antigen-binding and thus in cell death because the corresponding B cell can no longer receive survival signals via its B cell receptors. For non-lethal amino acid exchanges, p_{block} governs the probability of blocking mutations in antibodies that cannot bind the target with higher affinity than their germline ancestor even if beneficial key region mutations are present. This concept was first introduced by [Kleinstein and Singh, 2003] and represents a simplified way of introducing epistasis into the model.

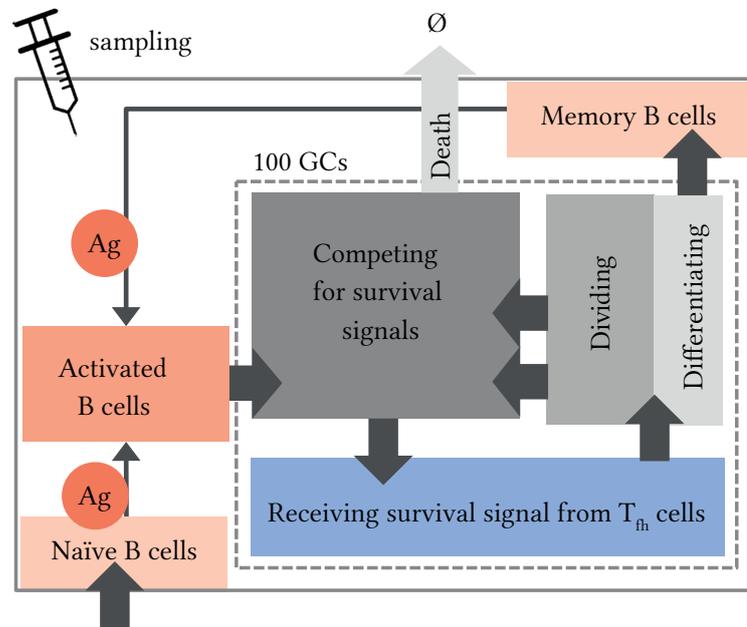
2.3. Germinal Centre-Based Model of B Cell Memory Formation

An overview of the simulation scheme described in the following is given in figure 2.7a. The simulation proceeds in discrete time steps of adjustable length t_{step} and, unless specified otherwise, most processes like the migration of B cells to the follicular site or the division of B cells within the germinal centre proceed with deterministic duration. For an overview of all simulation parameters and their experimental sources where applicable, see the appendix of this chapter in section 2.A.

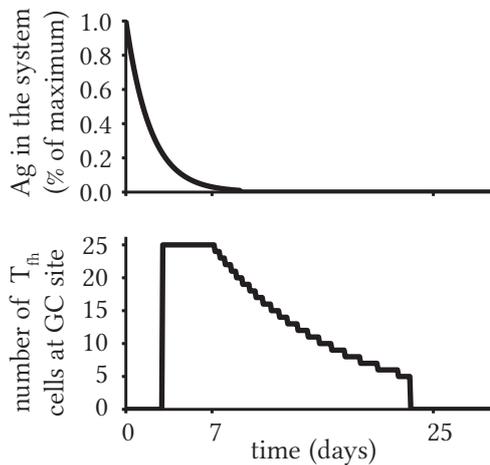
Simulation Rules As a first simulation step, an epitope sequence of length n_{key} is generated. Next, a reservoir list of antibody key sequences is engineered such that the distribution of binding energies between sequences in the reservoir and the epitope sequence conform to a discretised Gaussian distribution with mean $E_{\text{bind}} = 0.5$ and standard deviation 0.1, truncated to values above the activation threshold $E_{\text{bind}} = 0.6$. This mimics the distribution naturally arising for randomly drawn sequences for $n_{\text{key}} = 10$. The possibility of enforcing a given binding energy distribution has been introduced to allow comparison of germinal centre efficiencies across challenges of different complexities without artefacts introduced by unequal quality initial repertoires (as illustrated by the randomly drawn repertoires for different key region lengths in figure 2.6a). New naïve and unspecific memory B cells are then equipped with randomly picked sequences from the pre-generated reservoir throughout the simulation. Memory cells are immortal on the simulated time scale of 100 days whereas naïve cells die after a given lifetime $t_{\text{n life}}$ unless activated. To maintain a constant pool size, naïve cells are added to the system with rate

$$r_{\text{naïve}} = \frac{n_{\text{naïve}} \cdot n_{\text{GC}}}{t_{\text{n life}}}, \quad (2.3)$$

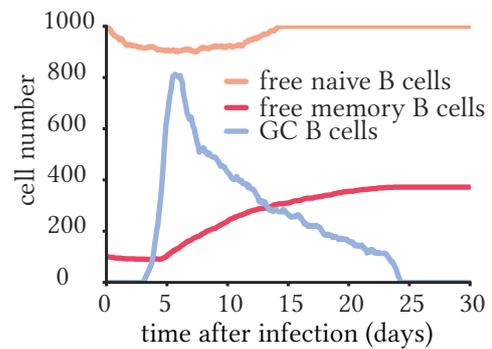
where $n_{\text{naïve}}$ describes the size of the reactive naïve cell pool per germinal centre.



(a) Scheme of the germinal centre-based simulation of the B cell immune response to vaccination, for details see section 2.3.



(b) Dynamics of antigen presence in the system (top) and presence of T_{fh} cells at the follicular site, determining the size of the germinal centre forming there (bottom).



(c) Population dynamics of peripheral naïve B and memory B cells as well as germinal centre (GC) B cells in a simulated system with a single germinal centre.

Figure 2.7.: Coarse overview of basic implementation details.

When antigen is added to the system at the i th vaccination time point t_{vacc}^i , its concentration decays exponentially from a starting value determined by the assumed vaccination dose with decay rate τ_{decay} giving the antigen time curve

$$a(t - t_{\text{vacc}}^i) = e^{-(t - t_{\text{vacc}}^i)/\tau_{\text{decay}}} \quad (2.4)$$

for $t > t_{\text{vacc}}^i$ per antigen exposure. Similarly the number of T_{fh} cells at a follicular site, which determines the size of the germinal centre forming there by governing how many B cells can receive survival signals, is at its maximum value n_{LF} between the time needed for both B and T_{fh} cells to migrate there t_{init} , and $t_{\text{GC, max}}$ whereafter their number decreases exponentially with decay constant $\tau_{\text{GC decay}}$. This roughly mimics experimentally observed GC growth and decline dynamics [Wittenbrink et al., 2011]. While antigen is present in the system, every free B cell (naïve and memory) may get activated and primed to join a germinal centre with a probability proportional to the antigen concentration at each time step. This activation step is independent of affinity and competition for antigen because it is assumed to occur in the dilute settings of the periphery or the lymphoid system.

Once arrived at the follicular site, activated cells enter the germinal centre where they compete for T_{fh} cells' signals to either divide and recycle or exit as differentiated effector cells. Competition is based on each cell's antigen affinity and the number of currently available T_{fh} cells. Dynamics of antigen and T_{fh} cells at the follicular sites following infection are shown in figure 2.7b. Winners are drawn based on a Boltzmann distribution at every time step: the probability p_i for a cell with binding energy E_i to be selected in a pool of J cells is

$$p_i = \frac{e^{-k_B T / \tilde{E}_i}}{\sum_j^J e^{-k_B T / \tilde{E}_j}}, \quad (2.5)$$

where the energy values \tilde{E} are the result of mapping the normalised energies E_{bind} between 0.6. and 1 to the energy range determined by experimentally observed affinities as described. After receiving signals for a set amount of time t_{help} , a fraction p_{recycle} of cells is randomly chosen to divide and recycle in the germinal centre while the remaining cells differentiate into effector cells and exit the reaction. Of note, because it has been observed that germinal centre B cells divide twice on average after having received T_{fh} signals once [Gitlin et al., 2014], this is implemented into the model by letting each cell divide once before deciding whether it will divide again or differentiate into an effector cell in a second round. Cells that do not successfully compete for survival and division signals within a fixed time $t_{\text{GC life}}$ are removed from the model simulating their death. Thus, selection and expansion is death-driven rather than division-driven in this model as has been argued previously based on both experimental [Anderson et al., 2009] and theoretical [Amitai et al., 2017] evidence. In agreement with this hypothesis, division rates in the model presented here do not depend on the B cells affinity for antigen but are set to a fixed time t_{div} .

Plasma cells are not modelled explicitly here and therefore cells choosing this fate disappear from the simulation. Memory B cells on the other hand rejoin the peripheral B cell pool for

activation and priming to reenter germinal centres during future immune responses. The dynamics of peripheral naïve B cells, peripheral memory B cells and germinal centre B cells are shown in figure 2.7c: the number of compatible naïve cells initially drops because they are recruited into the germinal centre with rate which is higher than their production rate. Because of this recruitment and the fast proliferation of germinal centre B cells, the population within the germinal centre quickly rises before declining again as the number of T_{fh} cells drops (figure 2.7b). The number of free memory cells begins to rise after the germinal centre has formed because they are being produced by the germinal centre reaction and enter the periphery.

Experimental observations show that mutations in germinal centres start only after a period of mutation-free expansion [Kleinstein and Singh, 2003]. Therefore, divisions happening in the time frame t_{AID} after a B cell clone first entered the germinal centre are exempt from the mutational procedure described above.

Simulation Output With the simulation implemented as outlined above, both the population dynamics of naïve, memory and germinal centre B cells (as sketched for a small system with a single germinal centre in figure 2.7c) and the clonal dynamics within individual germinal centres can be examined in detail. In order to record clonal dynamics, each simulated naïve B cell which is assumed to be uniquely produced in the bone marrow is outfitted not only with a cellular identification number, but also with a family ID. Upon division within the germinal centre, both daughters get distinct cellular IDs but inherit their mother’s family ID. Like this, the number of cells derived from the same original ancestor and present within a germinal centre (or likewise the entire memory cell pool) can be queried.

In order to make the selection and diversification process intuitively accessible, the simulation script produces visualisations of the type shown in the left column of figure 2.8. Here, every family ID as described above is associated with a certain colour. As initially many different naïve cells and thus different families seed a germinal centre reaction, many colours are observed in the first two weeks of the process. As affinity-based selection then identifies the B cell families with the highest binding potential, most clones disappear and the germinal centre becomes dominated by one or a few families. This closely matches recent experimental observations [Tas et al., 2016]. In addition to the clonal dynamics, the underlying affinities are also accessible. In figure 2.8 the dissociation constants of individual memory cells produced by the depicted GC over time are shown in the bottom row. Each dot represents a single memory cell here and can be attributed to the clones shown in the top row via the matching colours. The selection dynamics shown there are mirrored by the memory cell output: initially, many different families are present and thus produce memory cells. The initial seeder families have different affinities to the antigen sequence (governed by the distribution shown for $n_{key} = 10$ in figure 2.6a). Until approximately two weeks after antigen exposure, the mean affinity of the germinal centre’s memory output increases, almost solely driven by selection. Eventually, the repertoire within the centre is reduced to the strongest clone, depicted in dark blue in this example. Shifts in the y-direction are now due to mutations; it is directly apparent that in this specific example, no strong improvements are identified. This conclusion is underlined by the

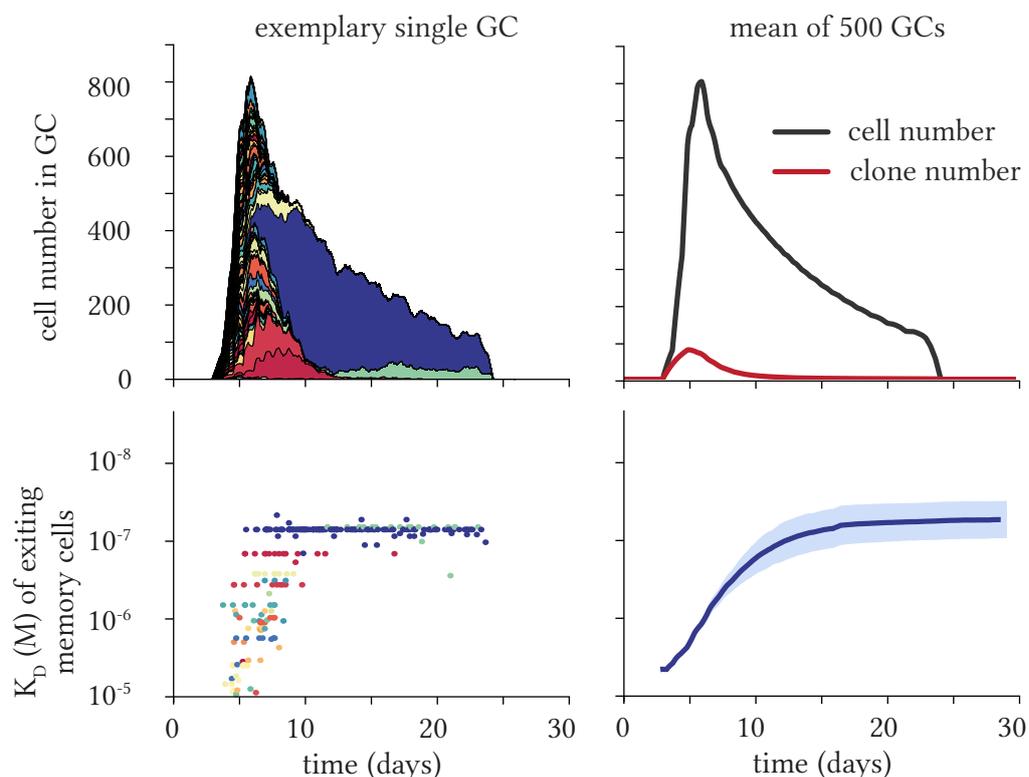


Figure 2.8.: Upper left: insight into a single germinal centre reaction seeded by about 100 different clones. Each clone is represented by a different colour so that selective expansion of the highest affinity clone (dark blue) can be visualised. The number of cells within the germinal centre reaches a peak at around 800 and then declines so that the germinal centre disappears by four weeks after the vaccination. Lower left: Memory cell output of the germinal centre shown above. Each dot represents one memory cell that has left the germinal centre and gives the production time (x-axis) and affinity (y-axis) of this cell. Initially, memory cell output is polyclonal and mean quality rises quickly as selection acts on the seeder pool. Later, in the pauciclonal phase, mutational drift of the strongest clone (dark blue) occurs. Upper right: Mean cell and clone number derived from 500 individual simulated germinal centres. Lower right: Mean quality of the memory output of these 500 germinal centres. All simulations run with $n_{\text{key}} = 10$.

bottom right panel which shows the average quality of the memory output of 500 single GC reactions. The basic pattern shown by the single example is preserved: while there is a fast selection-driven increase initially, later mutation-driven improvement continues at a much lower speed.

The parameters used for this example are the default settings listed in tables 2.1-2.4. However, many parameters, such as the antigen complexity n_{key} , the number of T_{fh} cells at the base of germinal centre size, the shape of the affinity distribution in the naïve repertoire and the mutation frequency among others are interesting subjects of investigation. With this simulation set-up at hand, these parameter spaces as well as multi-shot vaccination schedules will be explored in the chapter 3.

2.A. Supplement: Binding Model Matrices, Pseudocode, Parameters

This section contains additional information on the B cell memory simulation in described sections 2.1, 2.2 and 2.3. It comprises a pseudocode version of the implementation (algorithm 1), the default set of parameters used for running simulations together with experimental sources where applicable in tables 2.1 to 2.4 and matrices associated with the binding model in tables 2.5 and 2.6.

The full simulation code, which was written in python 2.7, is publicly available at https://github.com/LiBuchauer/gc_memo under the permanent digital object identifier **10.5281/zenodo.1048052** (<http://doi.org/10.5281/zenodo.1048052>).

Algorithm 1 Pseudocode of the germinal centre-based simulation introduced in section 2.3.

```

1: function GERMINAL CENTRE MODEL

2:   Set  $t_{\text{now}} = 0$ .

3:   Create an antigenic determinant sequence of length  $n_{\text{key}}$ .
4:   Create a large (see text) list  $L_{\text{seq}}$  of antibody sequences of length  $n_{\text{key}}$  so that the distribution of binding
   energies between these sequences and the antigen sequence is as requested (see text).
5:   Create a list of  $n_{\text{naive}} \cdot n_{\text{GC}}$  naïve B cells with sequences drawn from  $L_{\text{seq}}$  and mutation count  $m_V = 0$ .
6:   Create a list of  $n_{\text{freemem}} \cdot n_{\text{GC}}$  unpecific B cells with sequences drawn from  $L_{\text{seq}}$  and
    $m_V = \text{UniformDistribution}[0, 40]$ .
7:   Create a list of  $n_{\text{GC}}$  empty waiting lists.

8:   Calculate the time curve  $Ag(t)$  of antigenic presence in the system (see text).
9:   Calculate the time curve  $LF(t)$  of limiting factor presence in the follicular sites (see text).
10:  Open an empty event list for each GC to store events that are executed with a time delay.

11:  while  $t_{\text{now}} < t_{\text{max}}$  do

12:    Remove inactive naïve cells that are older than  $t_{\text{life, naive}}$ .
13:    if number of free naïve cells  $< n_{\text{naive}} \cdot n_{\text{GC}}$  then
14:      Create naïve B cells with rate  $n_{\text{naive}} \cdot n_{\text{GC}} / t_{\text{life, naive}}$ .

15:    In each GC, remove waiting B cells that have been there for longer than  $t_{\text{life, GC}}$ .

    % EVENTS consist of (event type, execution time, GC ID, list of cells concerned by the event).
16:    if event list contains events with  $t_{\text{execution}} = t_{\text{now}}$  then
17:      for every one of these events do
18:        if event is of type 'ENTER' then
19:          Distribute the cells randomly to the GC waiting lists.
20:        else if event is of type 'DIVIDE' then
21:          Make two possibly mutated daughter cells from every mother (see text).
22:          Append the viable daughter cells to the GC's waiting list.
23:        else if event is of type 'DIFFERENTIATE' then
24:          Append the cells to the free memory list.
25:        Discard event.

26:    if antigen is present in the system at  $t_{\text{now}}$  then
27:      Create empty list  $L_{\text{act}}$  for newly activated cells.
28:      for every cell in the free naïve and memory pools do
29:        Activate with probability  $Ag(t_{\text{now}}) \cdot p_{\text{base}}$ .
30:        if activation is successful then append cell to  $L_{\text{act}}$ 
31:      Create event of type 'ENTER' with  $t_{\text{execution}} = t_{\text{now}} + t_{\text{init}}$  and  $L_{\text{act}}$ .
32:      Append event to event list.

33:    if limiting factors are present in the follicles at  $t_{\text{now}}$  then
34:      for every GC do
35:        if there are B cells waiting for survival signals then
36:          Choose  $LF(t_{\text{now}})$  waiting cells for survival according to Boltzmann-distributed selection
          probabilities (see text).
37:          In order to incorporate double division after selection, directly make two possibly mutated
          daughter sequences from every mother (see text).
38:          Create event of type 'DIVIDE' with  $t_{\text{execution}} = t_{\text{now}} + 2t_{\text{div}}$  and a randomly selected fraction
           $p_{\text{recycle}}$  of the viable daughters from the first division round.
39:          Create event of type 'DIFFERENTIATE' with  $t_{\text{execution}} = t_{\text{now}} + t_{\text{div}} + t_{\text{diff}}$  and the remaining
          chosen cells.
40:          Append events to the event list.

41:    Set  $t_{\text{now}} = t_{\text{now}} + t_{\text{step}}$ .

```

Table 2.1.: Simulation parameters: Vaccination schedule, antigen an limiting factor dynamics.

symbol	default value	meaning	source / reason
t_{vacc}^i	[0, 28, 56] days	Vaccination time points	CHMI trial discussed in chapter 4, [Mordmüller et al., 2017]
τ_{decay}	2 days	Time constant of exponential Ag decay/removal in the system	[Mandel et al., 1980]
p_{base}	0.005	Base activation probability per free B cell and timestep	choice leading to 10% activation
$t_{\text{GC,max}}$	day 7	Day until which the GC stays at maximum size	onset of <i>Plasmodium falciparum</i> blood stage [Keitany et al., 2016]
$\tau_{\text{GC decay}}$	10 days	Time constant of exponential limiting factor decay after $t_{\text{GC,max}}$	set to reduce GC size faster than normal [Wittenbrink et al., 2011, Weisel et al., 2016] due to <i>Plasmodium falciparum</i> blood stage antigens [Keitany et al., 2016]

Table 2.2.: Simulation parameters: Simulation size.

symbol	default value	meaning	source / reason
n_{GC}	50	Number of individual GCs in the system	see chapter 3
n_{GC}	25	Number of limiting factors (T_{fh} cells) per GC	leads to GCs of 500 cells in steady state, in range of [Wittenbrink et al., 2010]
$n_{\text{naïve}}$	1000	Steady state number of potentially binding free naïve cells per GC	estimation: (B cells in mouse [Matsumoto and Futamura, 1995]) · (precursor frequency [Perelson and Oster, 1979])/ $n_{\text{GC}} = 10^9 \cdot 10^{-4} \cdot 10^{-2} = \mathcal{O}(10^3)$
$n_{\text{free mem}}$	100	Number of binding free memory cells from former infections per GC at $t = 0$	[Perez-Andres et al., 2010] and text
t_{step}	2 hours	Simulation timestep	chosen to match smallest simulated time unit (t_{help})

Table 2.3.: Simulation parameters: GC dynamics.

symbol	default value	meaning	source / reason
t_{init}	3 days	Time needed by B cells to become fully activated and migrate to the follicle to form a GC after initial antigenic contact	[De Silva and Klein, 2015]
t_{AID}	3 days	Time needed for cells having newly joined the GC before they start acquiring mutations during divisions	[Berek, 1992, Jacob et al., 1993, Kleinstein and Singh, 2003]
t_{help}	2 hours	Time that selected B cells need to receive survival signals before deciding to divide or differentiate	order of [Allen et al., 2007] (several contacts of 10-60 minutes)
n_{div}	2	number of divisions B cells undergo following selection	mean in [Gitlin et al., 2014]
t_{div}	8 hours	Time needed for cell division in the GC after having received survival signal	citeAllen2007b
t_{diff}	8 hours	Time needed for differentiation into a memory cell after having received survival signal	chosen to match t_{div}
$t_{\text{GC life}}$	8 hours	Maximum survival time of GC cells while unsuccessfully competing for survival signals	order of [Liu et al., 1989]
$t_{\text{n life}}$	14 days	Lifetime of naïve cells that have not been activated; new naïve cells are continuously introduced into the simulation to keep up the poolsize $n_{\text{naïve}} \cdot n_{\text{GC}}$	[Macallan, 2005]
r	0.9	Fraction of cells choosing to divide and recycle instead of leaving as a differentiated cell after receiving survival signal	[Meyer-Hermann et al., 2012, Kleinstein and Singh, 2003]

Table 2.4.: Simulation parameters: Binding model and mutations.

symbol	default value	meaning	source / reason
n_{key}	[1 ...15]	Number of residues jointly determining affinity towards a given epitope sequence	see section 2.2
n_V	220	Combined length of the V_H and V_L segments	[Murphy and Weaver, 2016]
b^{thr}	0.6	Binding threshold in normalized energy units	see section 2.2
K_D^{thr}	10^{-5} M	Corresponding binding threshold as dissociation constant	[Batista and Neuberger, 1998, Childs et al., 2015]
b^{top}	1	Maximum binding strength in normalized energy units	see section 2.2
K_D^{top}	10^{-9} M	Corresponding maximum binding strength as dissociation constant	[Batista and Neuberger, 1998, Childs et al., 2015]
p_{err}	0.003	Error probability per codon and division	[McKean et al., 1984]
p_{death}	0.5	Probability that a replacement mutation in a residue relevant for stability leads to a non-functional Ab (and hence cell death)	[Kleinstein and Singh, 2003]
p_{block}	0.55	Probability that a mutation in a non-key residue results in blocking of affinity maturation	mean of values given in [Kleinstein and Singh, 2003]

	C	M	F	I	L	V	W	Y	A	G	T	S	Q	N	E	D	H	R	K	P
C	-1.79	-1.23	-0.98	-0.48	-0.69	-0.94	-0.3	-0.96	-0.3	-0.42	-0.38	-0.2	-0.49	-0.32	0.04	0.55	-0.82	-0.4	0.	0.07
M	-1.23	0.36	-1.03	-0.41	-0.31	-0.94	-0.07	-1.1	0.05	0.	0.06	-0.47	-0.54	0.31	0.02	1.07	-0.35	-0.43	0.55	-0.25
F	-0.98	-1.03	-0.61	-0.66	-1.02	-0.78	-0.89	-0.82	-0.05	0.21	-0.19	0.14	0.1	-0.02	0.19	0.2	-0.75	-0.22	-0.17	-0.43
I	-0.48	-0.41	-0.66	-0.71	-1.04	-0.98	-0.89	-0.87	-0.64	0.4	-0.29	-0.13	-0.39	0.39	-0.2	0.04	-0.52	-0.08	-0.26	0.25
L	-0.69	-0.31	-1.02	-1.04	-1.14	-1.03	-0.97	-0.6	-0.57	-0.08	-0.39	-0.07	-0.13	-0.1	-0.05	0.5	-0.36	-0.1	0.1	0.09
V	-0.94	-0.94	-0.78	-0.98	-1.03	-1.15	-0.6	-0.7	-0.6	-0.2	0.06	-0.31	-0.09	-0.24	-0.02	0.25	-0.35	-0.48	-0.08	-0.08
W	-0.3	-0.07	-0.89	-0.89	-0.97	-0.6	0.02	-0.99	-0.08	-0.14	0.07	-0.2	0.4	-0.68	0.32	0.24	-0.41	-0.78	-0.3	-0.44
Y	-0.96	-1.1	-0.82	-0.87	-0.6	-0.7	-0.99	0.35	-0.37	-0.32	-0.23	0.25	-0.39	-0.74	0.22	0.11	-0.67	0.21	-0.2	-0.45
A	-0.3	0.05	-0.05	-0.64	-0.57	-0.6	-0.08	-0.37	-0.08	-0.09	-0.22	-0.01	-0.11	-0.14	0.03	0.1	-0.15	0.07	0.	0.41
G	-0.42	0.	0.21	0.4	-0.08	-0.2	-0.14	-0.32	-0.09	0.04	0.13	-0.04	0.12	-0.18	0.4	-0.06	0.	-0.15	0.1	0.4
T	-0.38	0.06	-0.19	-0.29	-0.39	0.06	0.07	-0.23	-0.22	0.13	0.26	0.05	-0.17	-0.27	0.15	-0.03	-0.27	-0.17	0.09	0.36
S	-0.2	-0.47	0.14	-0.13	-0.07	-0.31	-0.2	0.25	-0.01	-0.04	0.05	-0.13	0.4	0.37	0.3	-0.09	-0.59	0.61	0.18	0.44
Q	-0.49	-0.54	0.1	-0.39	-0.13	-0.09	0.4	-0.39	-0.11	0.12	-0.17	0.4	-0.08	-0.05	0.62	0.46	0.05	0.62	0.04	-0.21
N	-0.32	0.31	-0.02	0.39	-0.1	-0.24	-0.68	-0.74	-0.14	-0.18	-0.27	0.37	-0.05	-0.86	-0.25	-0.12	0.06	0.04	0.18	0.11
E	0.04	0.02	0.19	-0.2	-0.05	-0.02	0.32	0.22	0.03	0.4	0.15	0.3	0.62	-0.25	0.21	0.68	-0.53	-0.26	-0.09	0.84
D	0.55	1.07	0.2	0.04	0.5	0.25	0.24	0.11	0.1	-0.06	-0.03	-0.09	0.46	-0.12	0.68	0.6	-0.06	-0.15	-0.09	0.84
H	-0.82	-0.35	-0.75	-0.52	-0.36	-0.35	-0.41	-0.67	-0.15	0.	-0.27	-0.59	0.05	0.06	-0.53	-0.06	0.14	-0.01	0.14	-0.22
R	-0.4	-0.43	-0.22	-0.08	-0.1	-0.48	-0.78	0.21	0.07	-0.15	-0.17	0.61	0.62	0.04	-0.26	-0.15	-0.01	0.23	0.3	-0.02
K	0.	0.55	-0.17	-0.26	0.1	-0.08	-0.3	-0.2	0.	0.1	0.09	0.18	0.04	0.18	-0.09	-0.09	0.14	0.3	1.45	0.51
P	0.07	-0.25	-0.43	0.25	0.09	-0.08	-0.44	-0.45	0.41	0.4	0.36	0.44	-0.21	0.11	0.84	0.84	-0.22	-0.02	0.51	0.28

Table 2.5.: Pairwise contact potentials between all 20 amino acids of the standard genetic code [Thomas and Dill, 1996]. Values represent statistical potentials which are conceptually most related to free energies of interaction. They are invariant to scale and thus given in arbitrary units. The matrix is symmetric.

	C	M	F	I	L	V	W	Y	A	G	T	S	Q	N	E	D	H	R	K	P	STOP
C	0.11	0	0.11	0	0	0	0.11	0.11	0	0.11	0	0.22	0	0	0	0	0	0.11	0	0	0.11
M	0	0	0	0.33	0.22	0.11	0	0	0	0	0.11	0	0	0	0	0	0	0.11	0.11	0	0
F	0.11	0	0.11	0.11	0.33	0.11	0	0.11	0	0	0	0.11	0	0	0	0	0	0	0	0	0
I	0	0.11	0.07	0.22	0.15	0.11	0	0	0	0	0.11	0.07	0	0.07	0	0	0	0.04	0.04	0	0
L	0	0.04	0.11	0.07	0.33	0.11	0.02	0	0	0	0	0.04	0.04	0	0	0	0.04	0.07	0	0.07	0.06
V	0	0.03	0.06	0.08	0.17	0.33	0	0	0.11	0.11	0	0	0	0	0.06	0.06	0	0	0	0	0
W	0.22	0	0	0	0.11	0	0	0	0	0.11	0	0.11	0	0	0	0	0	0.22	0	0	0.22
Y	0.11	0	0.11	0	0	0	0	0.11	0	0	0	0.11	0	0.11	0	0.11	0.11	0	0	0	0.22
A	0	0	0	0	0	0.11	0	0	0.33	0.11	0.11	0.11	0	0	0.06	0.06	0	0	0	0	0.11
G	0.06	0	0	0	0	0.11	0.03	0	0.11	0.33	0.33	0	0.06	0	0.06	0.06	0	0.17	0	0	0.03
T	0	0.03	0	0.08	0	0	0	0	0.11	0.33	0.33	0.17	0	0.06	0	0	0	0.06	0.06	0.11	0
S	0.07	0	0.04	0.04	0.04	0	0.02	0.04	0.07	0.04	0.11	0.26	0	0.04	0	0	0	0.11	0	0.07	0.06
Q	0	0	0	0	0.11	0	0	0	0	0	0	0	0.11	0	0.11	0	0.22	0.11	0.11	0.11	0.11
N	0	0	0	0.11	0	0	0	0.11	0	0.11	0.11	0	0	0.11	0	0.11	0.11	0	0.22	0	0
E	0	0	0	0	0.11	0	0	0	0.11	0.11	0	0	0.11	0	0.11	0.22	0	0	0.11	0	0.11
D	0	0	0	0	0.11	0	0	0.11	0.11	0.11	0	0	0	0.11	0.22	0.11	0.11	0	0	0	0
H	0	0	0	0	0.11	0	0	0.11	0	0	0	0	0.22	0.11	0	0.11	0.11	0	0	0	0.11
R	0.04	0.02	0	0.02	0.07	0	0.04	0	0.11	0.04	0.11	0.11	0.04	0	0	0.04	0.33	0.04	0.07	0.04	0.04
K	0	0.06	0	0.06	0	0	0	0	0	0.11	0	0.11	0	0.11	0	0.11	0.11	0	0.11	0	0.11
P	0	0	0	0	0.11	0	0	0	0.11	0	0.11	0.11	0.06	0	0	0	0.06	0.11	0	0.33	0

Table 2.6.: Algebraic transition probabilities between amino acids based on the underlying codons. Each replacement of a nucleotide by another one during somatic hypermutation is assumed to be equally likely. As each codon consists of three base pairs, transition into each of its $(3+3+3)$ neighbours at Hamming distance 1 has equal probability of $1/9$. The transition matrix includes values which are not multiples of $1/9$ because most amino acids are initially encoded by more than one codon. This transition probability matrix is not symmetric (and also not quadratic), but to be read in the following way: the transition probability upon a single base pair exchange of the amino acid in line i into the candidate in column j is given by the entry $\{i, j\}$ (and not vice versa). The last column gives transition probabilities into the STOP codon. Grey background indicates transition probability into a different codon coding for the same amino acid.

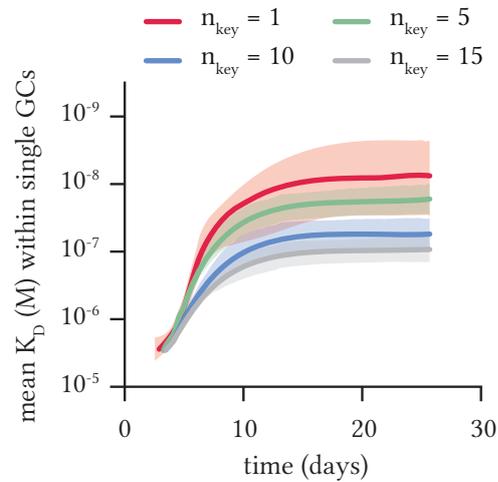
3. Application of the Model to Vaccination Scenarios with Various Boundary Conditions

In this chapter, simulation results for varying parameter regimes and boundary conditions are explored. These comprise the effects of varying antigen complexity on the affinity maturation within individual germinal centres in section 3.1 and the influence of germinal centre- and vaccination protocol-specific parameters in systems of many interacting germinal centres in section 3.2. In section 3.3, a mechanism underlying successful expansion is introduced and the influence of patient-specific repertoire differences is examined.

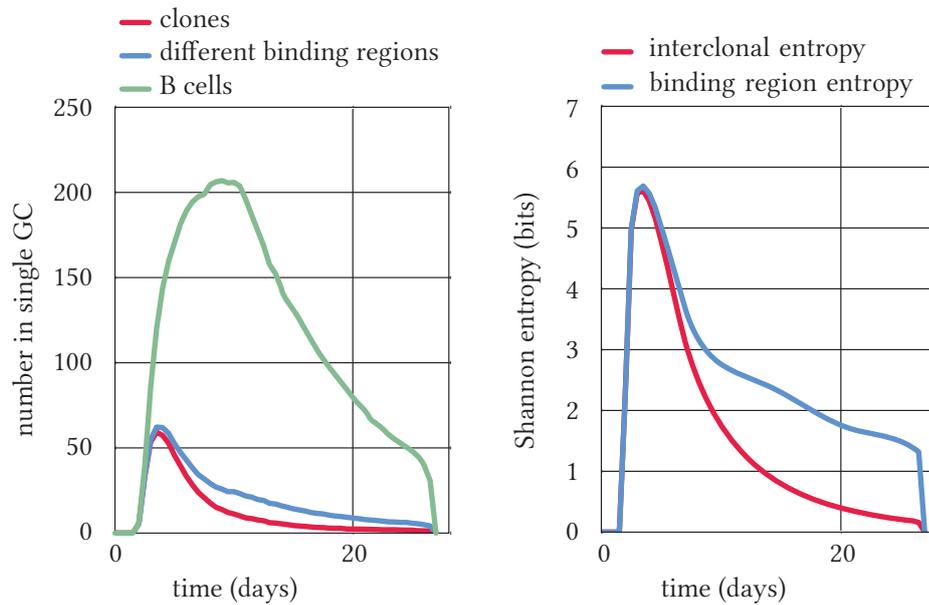
3.1. Antigen Complexity and Clonal Diversity

As outlined in section 2.1, one factor determining the difficulty of vaccination attempts is how easy it is to generate major antibody affinity improvements via somatic hypermutation in germinal centres. Because patterns of recurring mutations with strong individual effects are described for small haptens, but much less for larger protein antigens, the binding model developed in chapter 2 is set up in such a way that binding interfaces of differing size can be described by a varying number of key residues n_{key} . In the following, the effect of this parameter on the efficiency of affinity maturation within individual germinal centres is explored in detail.

Decreasing Efficiency of Affinity Maturation with Complexity In contrast to historic beliefs [Kroese et al., 1987], newer studies indicate that many seeder cells populate a germinal centre reaction initially and additional participants may also join throughout an ongoing reaction [Tas et al., 2016]. Accordingly, the default activation parameters used in this work lead to $\mathcal{O}(100)$ seeder cells as is illustrated in the introductory example in figure 2.8. In our model, these around 100 cells enter the germinal centre solely based on whether they bind the antigen presented by the follicular dendritic cells there with at least threshold affinity (without selection mechanisms acting at this stage). Therefore, the affinity distribution of seeder cells is equal to the affinity distribution of all precursor cells in the system. As described in section 2.3, our model allows to enforce specific affinity distributions in the precursor pool. According to the default distribution shown in figure 2.6a for $n_{\text{key}} = 10$, there is a high number of low affinity binders, some intermediate binders and very rare strong binders. In the following, this distribution is used for all antibody-antigen binding complexities.



(a) Mean affinity towards a given antigen within individual simulated germinal centres as a function of time after antigen exposure. Different colours show different antigen complexities; as discussed in section 2.3, affinity distributions of initial seeder cells are equal for all complexities. Thus, the final differences in mean affinity are not due to pure selection effects on the seeder repertoire, but necessarily due to mutations.



(b) Left: number of B cells, B cell clones and different binding regions (cells from the same clone may have different key regions due to hypermutations) within a single small germinal centre as a function of time after antigen exposure. The number of clones drops much faster than the number of binding regions. Right: Resulting interclonal and binding region entropies in a single germinal centre as a function of time after antigen exposure. While the reduction of clone number leads to a drastic reduction in clonal entropy, somatic hypermutations diversify the binding region pool leading to a higher diversity towards the end of the reaction.

Figure 3.1.: Affinity and diversity development within a single simulated germinal centre.

Figure 3.1a shows what happens when $\mathcal{O}(100)$ randomly sampled cells from this affinity distribution seed a germinal centre reaction which then proceeds for around 4 weeks. Each line gives the mean affinity of all germinal centre B cells as a function of time for one of the complexities $n_{\text{key}} = \{1, 5, 10, 15\}$. All four scenarios start with the same mean affinity because the seeder cells are drawn from the same distribution, but evolve differently afterwards. The lower the complexity, the steeper the rise in mean affinity and the better the final result. The overall rise in affinity is brought about by the joint action of beneficial hypermutations and selective expansion of pre-existing high-affinity clones. However, the effects based on pre-existing clones are limited to the diversity in the seeder repertoire and are therefore of equal magnitude for all four complexity cases shown here. This mechanism cannot account for the observed differences in final affinity outcome. Consequently, the diverse results for the four difficulties support the initial assumption that affinity maturation against some antigens may proceed more smoothly than against others - if more residues jointly determine affinity, identifying all changes necessary for a strong binder while at the same time avoiding detrimental or completely destructive mutations becomes an increasingly difficult challenge.

Somatic Hypermutations Maintain Diversity One may provocatively ask why mutation-driven affinity maturation as an immune process evolutionarily developed in the first place if it is so inefficient in cases with complex antigens. Firstly, not all pathogenic strategies are alike and cases in which the precursor repertoire does not only harbour very few adequate B cells, but none at all, do exist. Here, affinity maturation via hypermutations is the only path to high-affinity antibodies, however inefficient it may be. In addition, a second aspect of the role of hypermutations becomes apparent when one looks at the diversity within germinal centres (and likewise within the memory B cell pool produced by an immune response). Figure 3.1b shows the number of B cells, the number of clones these belong to and the number of unique key regions within a germinal centre over the course of the reaction. In this specific example, a little more than 50 clones seed the reaction. This number is quickly reduced over time until only one or a few remain. In turn, because of the ongoing diversification by hypermutation, one clone is likely comprised of individuals carrying different sequences and specifically also different binding regions once it has expanded. As a result, the number of diverse binding regions does not decrease as drastically as does the number of clones within the reaction. This effect becomes even clearer when looking at the normalised Shannon entropy within the germinal centre as defined by

$$H(\mathbf{X}) = - \sum_i P(x_i) \cdot \log_2 P(x_i), \quad (3.1)$$

where \mathbf{X} is a vector containing the number x_i of each species i within the ensemble and $P(x_i)$ is the resulting fraction of species i within the total mass. The right panel of 3.1b shows that while the interclonal entropy (the diversity brought about by different clonal origin) drops to 0, this is not true for the binding region entropy. Thus, while interclonal diversity decreases strongly due to the nature of the germinal centre reaction, intraclonal diversity rises and thus reduces the overall diversity loss.

The resulting maintenance of diversity in the memory pool emerges as an important feature

of an efficient immune system especially in the light of recent findings showing that cross-reactive memory cells participate in the fight of infections different from the ones that initially generated them. This is true for both related infections, i.e. different strands of influenza [McCarthy et al., 2018], and presumably completely unrelated infections, such as the reuse of existing memory cells in previously malaria-naïve Europeans in a malaria vaccination trial [Murugan et al., 2018]. Returning to the model, mutations may happen not only within the n_{key} mutations determining binding strengths within a specific case, but also within the scaffold region. It is conceivable that in the biologic scenario, not all antigens contact a given antibody at the same position such that amino acids which are part of the key region for one antigen are not relevant for another. As a result, in principle one should take into account the diversification of the entire sequence when judging the conservation of diversity by hypermutation. While this is not possible in the current model, where the scaffold region is not modelled as an explicit sequence, it is very likely that rigorous analysis of the development of this total entropy would reveal even higher stability in diversity..

Stochastic Differences Between Germinal Centre Reactions Because of the stochastic nature of somatic hypermutation, it is evident that two hypothetical germinal centres supplied with the exact same seeder cells would not produce the same outcome. However, the magnitude of the resulting variance depends both on the complexity of the antigen and the number of germinal seeder cells as illustrated in figure 3.2. This figure shows twelve examples for single germinal centre reactions, two each for six different parameter combinations of complexity and seeder cell number.

Looking at the default scenario with $n_{\text{key}} = 10$ and many seeder cells as used likewise in figure 2.8, the final maximum affinity reached by two independent reactions with different seeder cell ensembles does not differ strongly. This is because the comparatively high seeder number increases the likelihood of including high quality cells in the seeder pool. The slow speed of mutation-driven improvement then plays a minor role; effects caused by single mutations in the $n_{\text{key}} = 10$ scenario are frequently so small that they are not even fixed as illustrated by the two examples chosen for this case in figure 3.2 (upper left panel). Here, both in the orange and the blue dominating clones, clonal members with slightly higher affinity evolved transiently, but did not succeed in taking over the population. Reasons for this may include a quick succession of the beneficial mutation by a detrimental one, for a example a blocking mutation which reduced the affinity back to the value of the original clonal ancestor, differentiation of the improved cell into an effector cell and thus loss from the germinal centre and also failed selection of the improved candidate due to the Boltzmann-stochastic nature of the implemented selection process.

Moving to the neighbouring scenario, $n_{\text{key}} = 10$ and few key seeders, the benefit of recruiting many cells into the reaction is underlined. In the two examples shown here, only two or three clones seed the reaction; as these are drawn from a precursor affinity distribution with few high-quality cells, cases in which only low-quality binders are involved are the norm. Thus, reaching a high average affinity is more dependent on mutations and strong subject to stochastic effects. Similar general differences between the few- and the many-seeder scenario

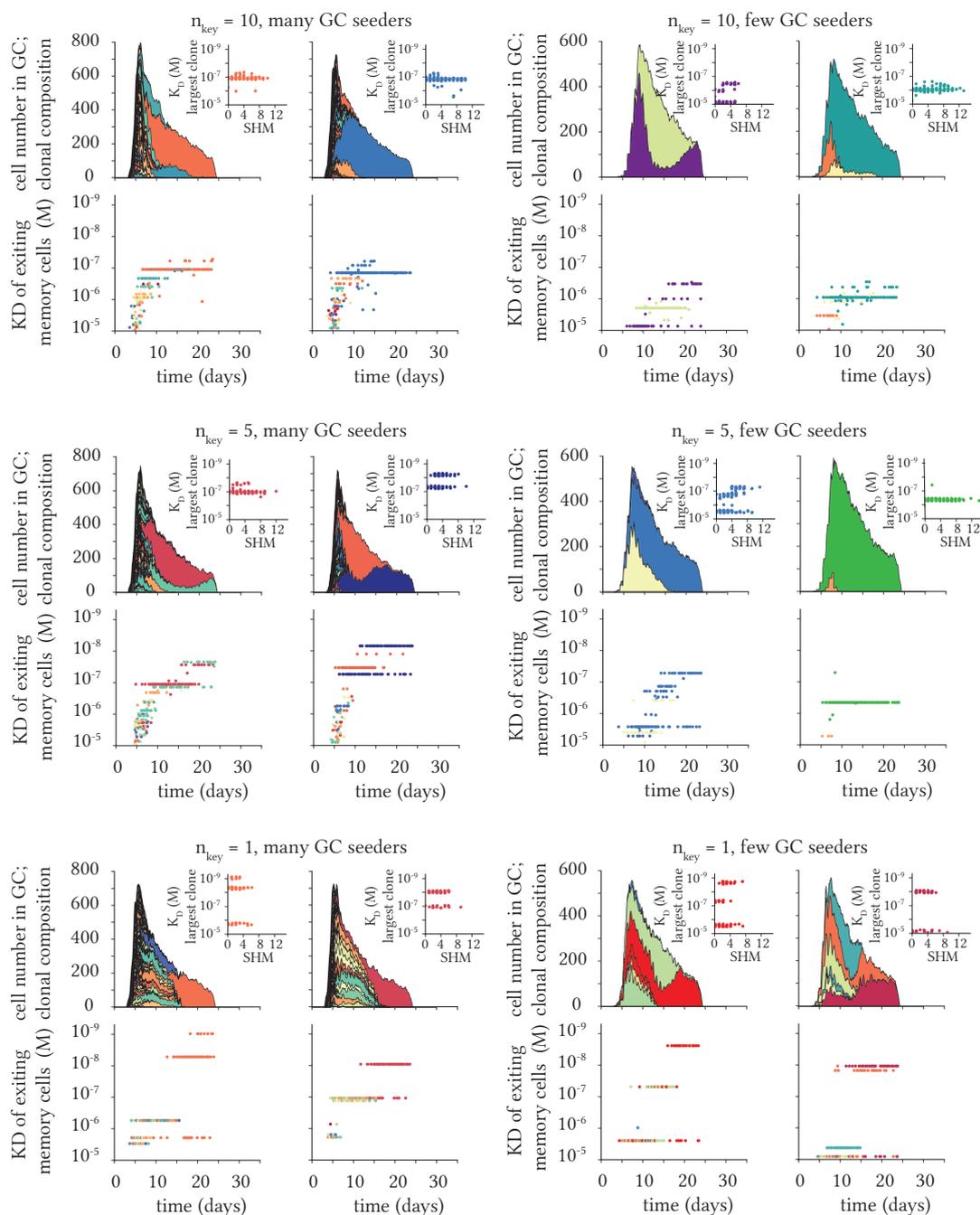


Figure 3.2.: Twelve examples of individual germinal centre reactions, visualised as introduced in figure 2.8. The inset to each upper panel additionally shows the affinity development within the dominating clone as a function of the number of somatic hypermutations it has acquired. Complexities $n_{\text{key}} = 10$, $n_{\text{key}} = 5$ and $n_{\text{key}} = 1$ are compared here; the effect of having few ($\mathcal{O} = 3$) or many ($\mathcal{O} = 100$) seeders are additionally explored.

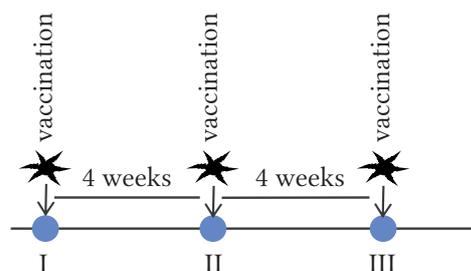
are valid for other complexities as well as illustrated for $n_{\text{key}} = 5$ and $n_{\text{key}} = 1$ here. However, as mutations lead to significant improvement more reliably if less key residues are involved in binding, the low-seeder scenarios more routinely produce high-affinity outcomes in these cases.

In summary, variability in the maximum affinity of a germinal centre can be reduced by bringing more seeder cells into each reaction, for example by increasing antigenic dose in a vaccination protocol. Keeping the seeder number constant, variation between individual reactions is also larger in cases with lower complexity because of the larger affinity jumps caused by individual mutations, but this is not a disadvantage as the mean maximum quality in cases with challenges of lower complexity is still higher than in more difficult mutation problems. Stochastic effects are further reduced by large systems in which many germinal centres interact over the course of several vaccination shots as discussed in section 3.2.

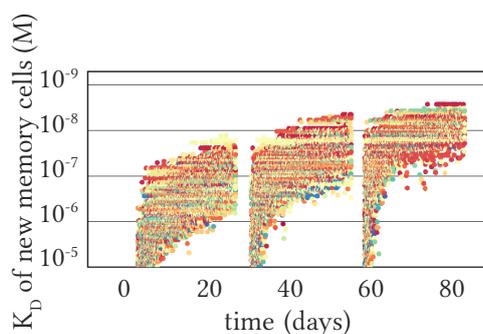
3.2. Multi-Shot Immunisation and General Parameter Space

The exploration of parameter effects in section 3.1 is limited to individual germinal centre reactions. As much as these are important building blocks of B cell immunity, they cannot individually mimic a human response to vaccination. Humans have several hundreds of lymph nodes, each of which can harbour $\mathcal{O}(10 - 100)$ germinal centres following infection or vaccination. In addition, the spleen is a major site for germinal centre reactions [Murphy and Weaver, 2016]. During a single acute response, germinal centres develop by and large independently, although some crosstalk in which the same clone is identified within two or more neighbouring reactions is possible [Tas et al., 2016]. Our model allows the analysis of larger systems with many germinal centres within the same peripheral B cell pool and the following sections will focus on such a system with 50 germinal centres. This specific choice was made on the one hand because key read-outs like mean memory pool affinity and entropy did not change by much if more centres were added after this point and on the other because this choice of system size, leading to $\mathcal{O}(10^5)$ B cell objects, is still manageable in the current implementation. The resulting runtime of one simulation with three vaccination shots and 50 germinal centres is around 30 minutes for $n_{\text{key}} = 10$.

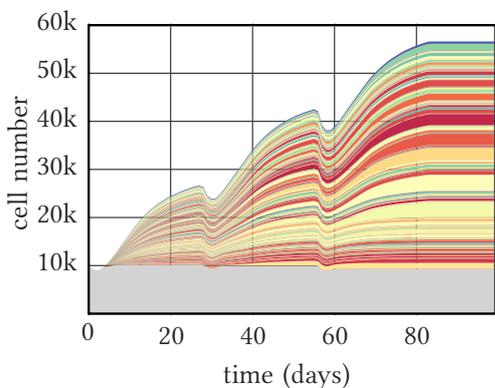
A Triple-Shot Vaccination Protocol While the model gives us the freedom to implement any vaccination schedule, we focus on a specific one for the following parameter exploration. Specifically, this is a triple-shot scenario with a primary vaccination followed by two boosters after intervals of four weeks. This specific choice is motivated firstly because of the comparison to a malaria vaccination study using this schedule in chapter 4 and secondly because many licensed vaccines likewise use a triple-shot schedule with similar time intervals, e.g. the childhood vaccines against rotavirus, polio, pneumococcus and hepatitis B [Robert-Koch-Institut, 2017]. Additionally, in the default settings, germinal centres have disappeared by four weeks after initial exposure and the three vaccinations are thus non-overlapping. Because the current implementation does not include decay processes of the generated memory (memory B cells live eternally in the model), it would currently not make sense to increase the time interval



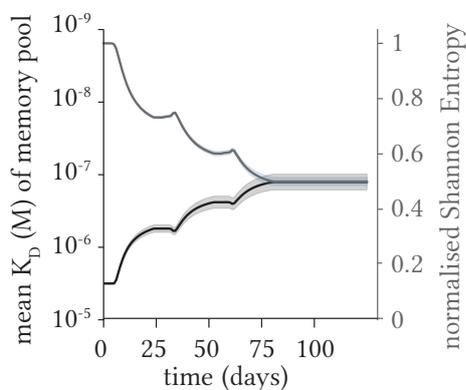
(a) Sketch of the triple-shot vaccination schedule used for the results in the other panels and most of chapter 3.



(b) Total memory cell production of a system with 50 germinal centres, each dot representing a new memory cell similar to figure 2.8.



(c) Development of the clonality of the memory pool with each colour representing a B cell clone derived from the same ancestor.



(d) Dynamics of mean affinity and normalised Shannon entropy over the time course of the triple vaccination schedule in panel a.

Figure 3.3.: Effect of a triple-shot vaccination protocol on B cell clonality and memory cell affinity in a system with 50 germinal centres governed by default parameters (tables 2.1 to 2.4).

beyond this point. The chosen default vaccination schedule is depicted in figure 3.3a.

The remaining panels of figure 3.3 show the effect of the this vaccination schedule on the memory pool for $n_{\text{key}} = 10$ and all other parameters set to their default values as given in section 2.A. Using the same picture language as introduced for individual germinal centres with colours indicating different clones, we now examine the memory population of the entire system. Looking at the affinity of each individual memory cell as a function of its production time point (figure 3.3b), a general trend towards higher affinities at later times is observable not only within the reaction to each vaccination but also globally: drawing on the advances following the first two vaccinations, the best memory cells are produced in response to the third exposure. Comparing all three vaccinations, low-quality memory cells are initially produced, but the time span during which this happens is reduced in the second and third process, as the fraction of highly competitive seeder cells from the now pre-selected memory pool is higher and thus the competitive advantage for selection to act on is larger from the start of these germinal centre reactions.

Instead of investigating the clonal distribution and the associated entropy within germinal centres like in section 3.1, we now analyse the clonal entropy within the memory B cell pool built up during the vaccination protocol. Figure 3.3c shows both the overall increase of memory cells throughout the triple-shot protocol and how it their clonal entropy is reduced as individual strong clones gain ground while weak clones are deselected. The peripheral memory cell number drops transiently directly following vaccination because some memory cells are recruited into germinal centres to fuel competition there and also to be differentiated into plasmablasts. Colours are only assigned to clones which expand to at least two cells at any point during the simulation; the grey base shown in figure 3.3c collects all cells which appear as singletons only.

The entropy of the memory pool as a function of time can be quantified using the normalised version of Shannon’s entropy which reduces effects of unequal population sizes (as they naturally appear in a growing cell population) and is defined by

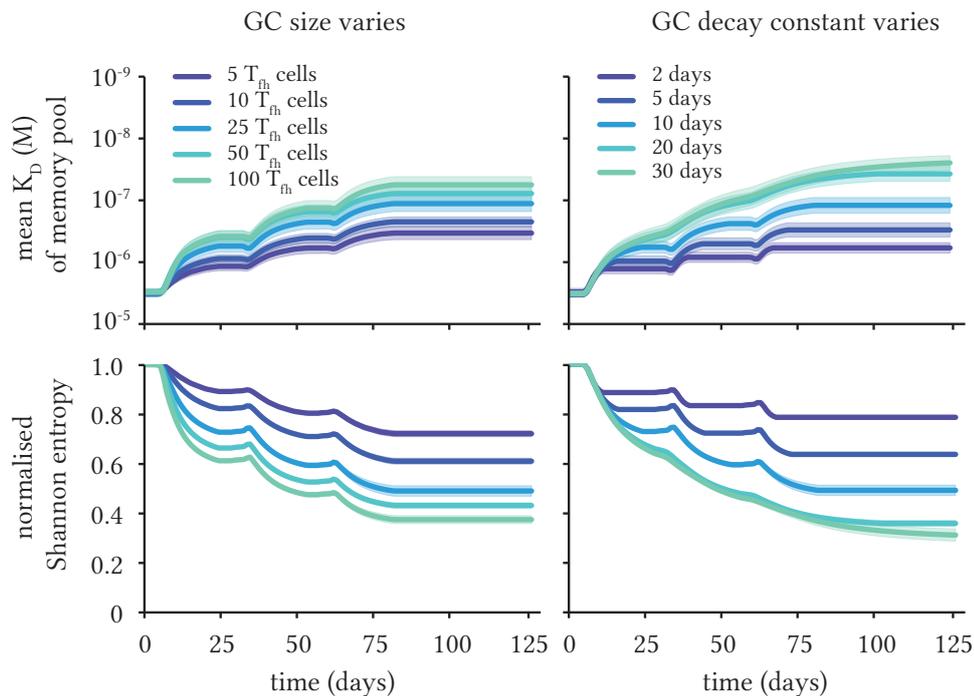
$$H(\mathbf{X}) = - \sum_i^n \frac{P(x_i) \cdot \log_2 P(x_i)}{\log_2 n}, \quad (3.2)$$

where n is the number of clonal species present. The result is shown in figure 3.3d. Even though the entropy of the pool is reduced with each vaccination, the pool remains rather diverse because of the underlying structure with many lightly interacting germinal centres. In comparison to a hypothetical system with only one or a few much larger germinal centres, this architecture is thus not only beneficial because it allows for decentralised recruitment of precursors all over the body, but also because it assists in maintaining a diverse repertoire. This in turn increases the likelihood that some of the clones fighting the current infection will also react against a different one.

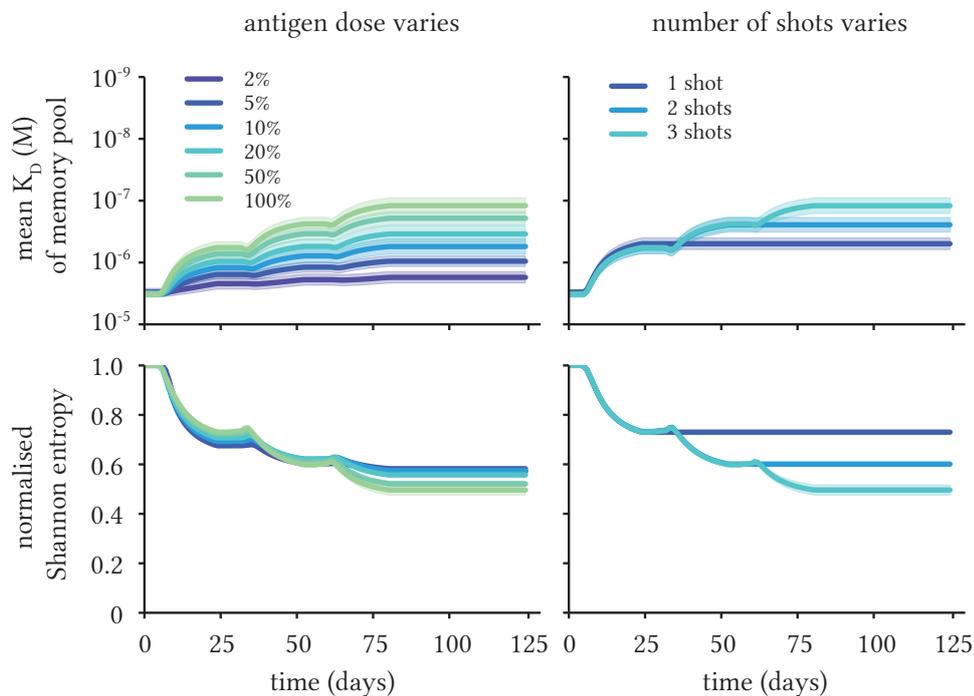
Exploring Parameter Sensitivity The sensitivity of the simulation outcome to a number of chosen model parameters is analysed in the following. For this, the chosen parameter is varied while all other parameters remain constant at their default value as given in section 2.A. The parameters discussed below include germinal centre-specific parameters as well as choices concerning the vaccination protocol.

First, we analyse the effect of germinal centre peak size on final affinity and entropy after the default triple-shot vaccination protocol. The default value of 25 T_{fh} cell was chosen because it produces germinal centres with around 500 B cells in them at steady state - this is within the physiological range reported for mice [Wittenbrink et al., 2010], albeit on the smaller end, and still allows for efficient computation. While varying the number of T_{fh} cells which act as limiting factors in germinal centre B cell survival and selection, it becomes apparent that larger germinal centres lead to slightly higher final mean affinities because the population basis of mutational processes and thus the absolute probability for beneficial events is higher (figure 3.4a). However, the effect of large germinal centres on the diversity of the pool is much stronger than their effect on affinity. Thus, keeping in mind the previous discussion on maintaining diversity as an important but often overlooked aim of a humoral immune response, intermediate germinal centre sizes are likely beneficial. Indeed, physiologically, germinal centre sizes span several orders of magnitude [Wittenbrink et al., 2010], thereby potentially combining beneficial effects of both smaller (more diversity) and larger reactions (higher affinity).

Next, we look at the effect of germinal centre lifetimes. It is reported that germinal centre reactions start a few days after initial exposure and carry on for 4-10 weeks [Wittenbrink et al., 2011, Weisel et al., 2016]. Here, we have chosen a decay constant of 10 days as a default value, which leads to disappearance of the germinal centres after 4 weeks, because of the comparison to a malaria vaccination trial as described in chapter 4. In this specific case, germinal centre reactions against the antigen of interest are likely suppressed faster because of new antigens emerging in following stages of the malaria life cycle. Generally, longer germinal centre life times lead to a higher final affinity and lower entropy of the peripheral pool (3.4a). This is expected, because longer reactions lead to a higher total production of cells from the winning clones, thus both providing a stronger basis for beneficial mutations and simply producing more effector cells for the periphery. However, in figure 3.4a, a qualitative change becomes apparent when comparing decay constants 10 and 20 days. This is because for the highest values shown here (20 and 30 days), individual germinal centre reactions caused by consecutive vaccinations are not longer cleanly separated in time, but begin to overlap. Thus, the previous reaction is still ongoing and newly activated T_{fh} cells and B cells can join it. These refuelled reactions have an advantage over newly formed ones as they already have a larger high-quality population from the previous selection round as a starting point. Again, it is a glimpse at the development of entropy that supplies a potential reason for why physiological germinal centre reactions subside after a few weeks: long, and particularly overlapping reactions, lead to strong diversity loss. However, it is as yet unclear whether the type of germinal centre overlap discussed here and occurring in simulations really plays a role in biology. Questions like how new cells join existing germinal centres, how they compete in there and whether older cells are pushed out by the newcomers remain to be investigated more closely.



(a) Effects of germinal centre (GC) size and duration of the reaction.



(b) Effects of vaccination dose and number of shots.

Figure 3.4.: Effects of germinal centre determinants and vaccination protocol on mean affinity of the peripheral memory cell pool and its diversity during a triple-shot vaccination protocol as shown in figure 3.3a.

While the parameters discussed above, germinal centre size and lifetime, are governed mainly by physiological or pathogen-related factors, other parameters such as the vaccination dose or the number of shots administered are directly within control of the experimentalist or physician. As figure 3.4b shows, both impact the final outcome strongly. A higher vaccination dose is assumed to lead to more activation in our model and thus recruits more cells into germinal centre reactions as seeders. This in turn increases the cell pool on which selection acts. In the following second and third vaccinations, high doses are additionally important to guarantee sufficient mixing between germinal centres. The resulting “avalanche effect” is further detailed in section 3.3. Interestingly, entropy is not strongly affected by higher antigenic doses, as the basic fact that one or a few clones come to dominate each reaction after a few weeks remains intact for low-dose regimes and thus for cases with weak seeders only.

Finally, analysing the effect of the number of vaccinations shows no surprises. More shots lead to a higher final affinity as both expansion of strong binders and beneficial mutation processes are possible for a longer period of time. Concomitantly, clonal diversity is reduced by each additional administration. Thinking about these results in the context of repeated natural infections, this again underlines the diversifying effects of hypermutations as a monoclonal and, more importantly, monofunctional repertoire after a couple of infections with the same pathogen would not be desirable. As we will see in chapter 3.3, repeated vaccinations are specifically important for challenges with complex antigens and a precursor distribution with existing but rare high-quality cells. Overall, as (at least in the simulated scenario) each vaccination increases the quality of the memory pool further albeit with shrinking increment, it becomes relevant to ask how good is good enough: In the physiological context, what antibody titre is required for sterile immunity?; After how many vaccinations can we stop?

3.3. Avalanche Effect and Repertoire Differences

While the previous sections have explored how pathogen-specific effects like the complexity of the antibody-antigen binding interface with a given epitope (section 3.1) or a pathogen’s influence on germinal centre life time as well as the vaccination schedule can affect outcomes (section 3.2), this section focuses on differences between individuals receiving vaccines and on what may cause differential outcomes. In order to understand these differences, underlying effects governing selection, expansion and mutation are examined first.

The Avalanche Effect In a system comprised of many weakly connected germinal centres as described above and mirroring the physiological situation, it is not directly clear how one single high-affinity B cell can be expanded into a clone of functionally significant size. Specifically in cases where intraclonal affinity improvement through hypermutations is limited because of a high number of joint changes necessary for strong improvement, individual germinal centres cannot be relied on to routinely mutate weak seeders into strong candidates. While highly beneficial changes may be found in a few reactions, scenarios are conceivable where this happens in a small fraction of germinal centres only.

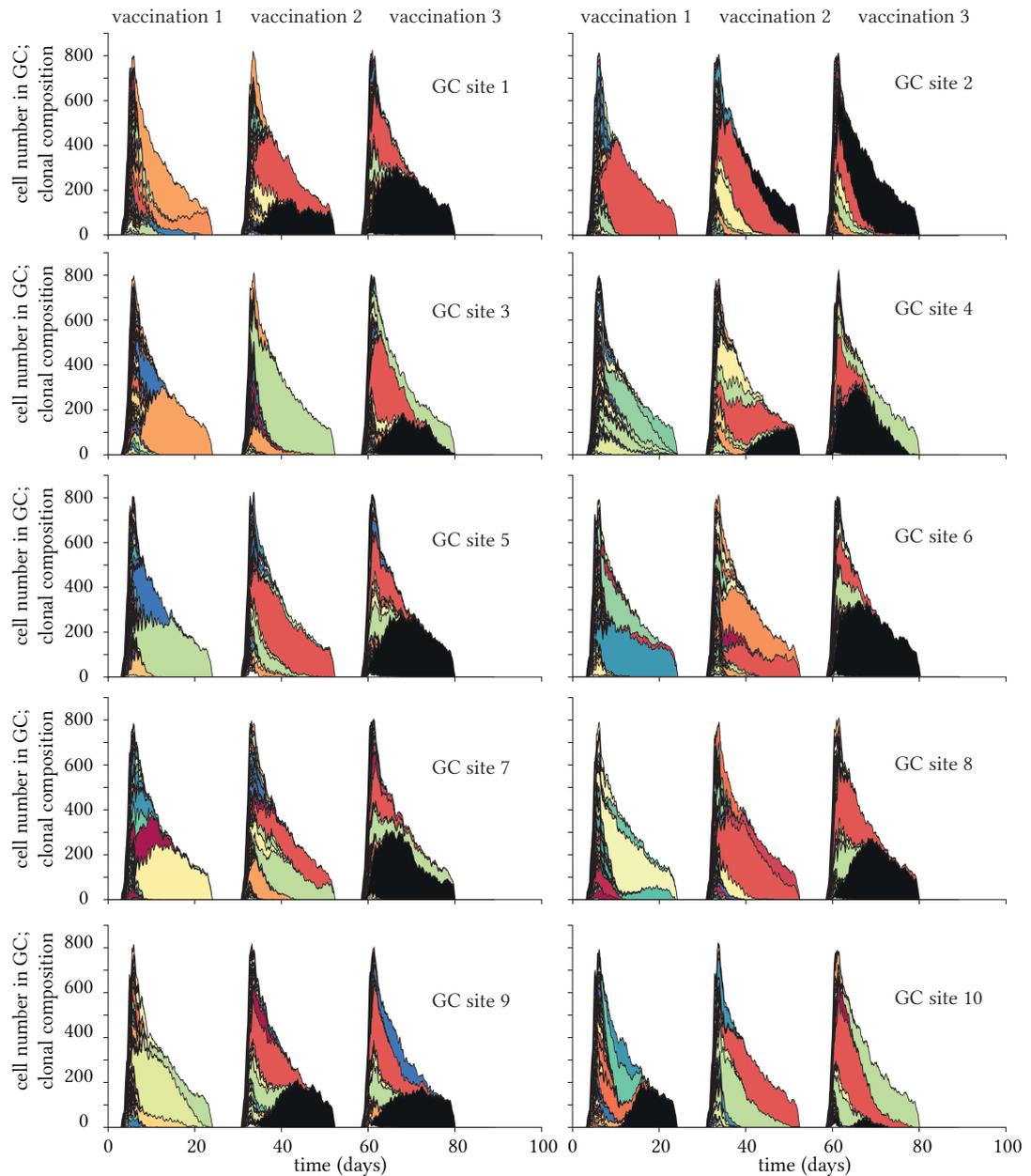


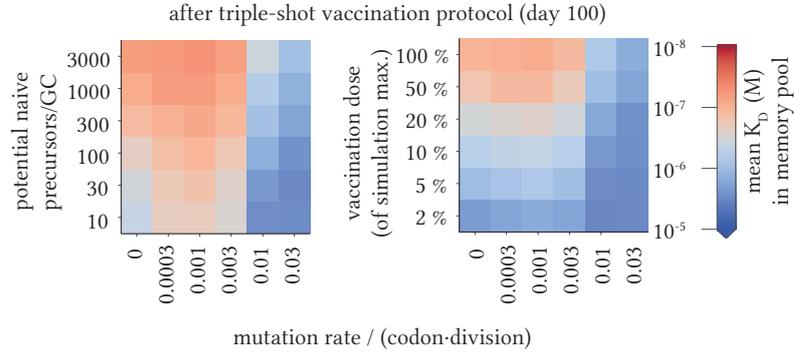
Figure 3.5.: Illustration of the avalanche effect for a toy system of ten germinal centre sites. Each panel depicts the clonal composition forming at each of the ten sites following vaccinations 1-3. During the first round, the ten germinal centres are almost without interaction, each contains its own set of clones. At site 10, the black clone eventually takes over the reaction after the first exposure. After the subsequent booster vaccinations, members of this henceforth expanded clone succeed to gain significant ground in more and more germinal centres (four after the second vaccination and all ten after the third), though they do not always become dominant.

Section 3.2 showed that high vaccination doses as well as having three rather than one or two vaccinations is strongly beneficial for the final outcome as measured by the mean affinity of the memory B cell pool in a complex challenge scenario with $n_{\text{key}} = 10$. As in this complex case not every germinal centre reaction can be trusted to produce high-affinity B cells from scratch (see also figure 3.1a), high recruitment and subsequent mixing are of pivotal importance. Here, we focus on the mixing process fuelled by several vaccinations, which we term “avalanche effect”. This effect is illustrated in figure 3.5, which shows a toy system comprised of ten germinal centre sites. One may imagine these sites as ten different lymph nodes each housing one germinal centre. Following the first infection, each of the ten centres is seeded by independent precursors, both naïve and pre-existing memory cells which have not previously been expanded in response to the antigen at hand. Each reaction then elapses as described in chapter 2 with mutation and selection processes leading to one or a few winners in each centre. Generally, these dominant clones have the most opportunity of producing effector cell offspring into the periphery and thus they come to occupy larger fractions in the free memory pool (see also figure 3.3c).

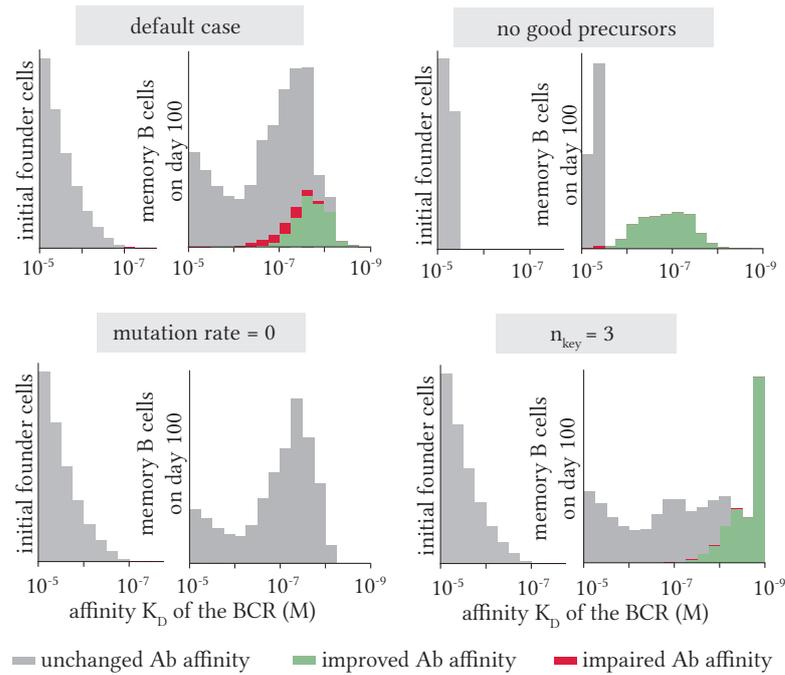
In the next infection, previously generated specific memory B cells may participate as is their role, and some of them reenter the newly formed germinal centre reactions. Thus, the best clones from each of several reactions now meet to compete much like in the next round of a tournament. In the toy example shown in figure 3.5, a black, a dark red and a mint green clone seem to appear and come to domination in several germinal centres each in the second infection. As these clones are now being expanded at multiple locations, they same goes for their effector cell progeny which thus reaches an even higher fraction in the periphery. The process is repeated in the third vaccination - in the toy example, all three clones named above have now infiltrated all ten germinal centres, mostly it is the black clone which comes to dominate. The fact that the black clone wins in most but not all reactions shows that while mutations can still improve affinity even at later stages, there is a strong memory of the affinity with which a clone entered the competition if antigen complexity is high. Because of the picture presented here, where each member of a strong clone can go on to seed and dominate a new germinal centre repeatedly, we term this amplification process the “avalanche effect”.

A Mutational Thought Experiment While the avalanche effect is an important factor in bringing about high-quality vaccination outcomes in difficult challenges, it relies partly on strong precursors and partly on beneficial mutations identified early on. With a computational experiment we can further disentangle the separate effects of mutation-driven improvement and pure selection. The underlying simulation setting here is again the triple-shot protocol with default parameters and as a read-out of success the final mean affinity of the memory pool is used. Instead of varying one parameter at a time, pairs of two parameters are being varied jointly and the final affinity is encoded in colour as shown figure 3.6a. The parameter pairs of interest here are {potential precursors per GC; mutation rate} and {vaccination dose; mutation rate}. Both the number of potential precursors per GC, i.e. the size of the free B cell pool from which seeder cells can be recruited, and the vaccination dose influence the number of germinal centre seeders and consequently the initial germinal centre diversity. While

3.3. AVALANCHE EFFECT AND REPERTOIRE DIFFERENCES



(a) Heatmaps showing the effects of hypermutation rate within germinal centres and (left) the number of potential naïve precursors per germinal centre or (right) the activation dose. Colour encodes the mean quality of the memory B cell pool after the end of a triple-shot vaccination protocol as shown in figure 3.3a. The physiological mutation rate is 0.003 base pair mutations per codon and division. Simulations were performed with the default parameters given in section 2.A with $n_{\text{key}} = 10$. Each square gives the mean of 30 runs.



(b) Final affinity distributions of memory B cells. The colour coding within each bar shows what proportion of cells with the indicated affinity has unchanged, improved or impaired affinity compared to its original ancestor; the distribution on the left side shows the available naïve repertoire. Four scenarios are compared here: in the default case, both selection and hypermutation act on the memory repertoire for a complex antigen of $n_{\text{key}} = 10$. In the next scenario, no good precursors are available. In contrast, the third scenario is mutation-free. For comparison, the last case shows the result after a vaccination campaign with $n_{\text{key}} = 3$. Here, mutational processes lead to high-end affinities in a higher fraction of cases.

Figure 3.6.: Memory pool quality and composition after a triple-shot vaccination scenario (see figure 3.3a) and how it is governed by mutation and selection processes.

these two parameters may vary because of pathogenic specifics or the vaccination protocol, it is unlikely that the hypermutation rate per codon and division varies strongly between cases (except for potential pathological cases), but varying it is nevertheless an interesting thought experiment.

The resulting scenarios without hypermutation (mutation rate = 0) are of particular interest: even in the total absence of somatic hypermutation, the final affinity can be strongly improved by recruiting as many seeder cells as possible. The physiological mutation rate has been determined to be around 0.003 mutations per codon and division; this value is close to the optimum identified in our simulation. However, there is a minor deviation from this value, as figure 3.6a suggests that a slightly smaller mutation rate is optimal. This deviation is likely due to simplifying assumptions in the computational model, specifically the assumption that mutations are equally likely in all areas of the immunoglobulin variable segments, while biologically mutational hotspots in and around the complementarity-determining regions exist [Zheng et al., 2005, Wei et al., 2015, Yeap et al., 2015]. Nevertheless, our model clearly shows that higher mutation rates would lead to worse outcomes as the probability of deathly or blocking mutations reaches levels overtaking potential positive effects of key region mutations.

Patient-Specific Repertoire Differences It has emerged as a result of previous sections that selective expansion of rare high-affinity precursors is a very important process for reaching high affinity at the population level. An alternative way of evaluating the relative contributions of mutations and selection in different vaccination scenarios is presented in figure 3.6b. Here, the distribution on the right of each panel shows how the affinities of memory B cells are spread out over the modelled affinity range after the default triple-shot vaccination schedule. On the left of each panel, the affinity distribution of the precursors and thus the germinal centre founder cells is shown. An additional colour coding in the final distributions helps identify the fraction of cells in each bin which have i) improved affinity compared to their germline ancestor, ii) impaired affinity compared to their germline ancestor or iii) unchanged affinity compared to the ancestor. Looking at the default case (all parameters as given in section 2.A with $n_{\text{key}} = 10$), the difference between the immature precursor repertoire and the memory repertoire after three vaccinations is striking, however, more than 80% of the memory cells still have the same affinity as their germline ancestor. In this case, where high-affinity precursors are available from germline, selecting and expanding them is more efficient than mutating them and hence produces the bulk of high-affinity cells at the end of the protocol.

The picture changes for the second scenario, in which no good precursors are available. As discussed in section 2.1, this may be due to genetic characteristics of the vaccinee in question leading a deficiency of high-quality germline precursors. Additionally, differences in the precursor repertoire may come about due to the infection history of the patients - even if there are no high-affinity germline B cells, previous related or unrelated infections may have shaped efficient antibodies and concomitantly immune memory which can be reactivated in the case at hand. Returning to the scenario under investigation, if no good precursors exist it is obvious that somatic hypermutation is the only way towards high-quality binders. Indeed, a small pop-

ulation of antibodies with an affinity around $K_D = 10^{-7} M$ is produced, but it is obvious that the patient simulated here would still have disadvantages compared to one with a better precursor repertoire. Additional vaccination rounds could help to further expand the now present decent precursors. In the spirit of personalized vaccines, an initial screening of precursor repertoire quality may help predict how many vaccinations at what dose are likely required until protection in each individual.

For completeness, two additional scenarios are examined in figure 3.6b. The first one is again concerned with the absence of hypermutations for a scenario with complexity $n_{\text{key}} = 10$. Compared to the truncated repertoire discussed before, the final repertoire here greatly resembles that of the default case as selection effects alone yield a competitive result. Finally, in the case of a mutationally simpler challenge with $n_{\text{key}} = 3$, a large population of highest-quality cells is produced via affinity improvement and successfully expanded.

4. Case Study: The Tübingen Controlled Human Malaria Infection Trial

In this chapter, the computational vaccination model introduced in chapters 2 and 3 is applied to increase mechanistic understanding of an experimental study on B cell memory development over the course of a malaria vaccination trial with three immunisations. Basic information on the malaria parasite *Plasmodium falciparum*, the vaccination trial and the experimental methods applied to the resulting blood samples are put together in section 4.1. Following this, several model predictions are tested against the experimental data and reasons for different observations at the population and clonal levels are discussed in the context of other recent findings in the field in section 4.2.

4.1. Introduction to Trial Protocol and Experimental Analysis

Epidemiology of Malaria and the Life Cycle of *Plasmodium falciparum* Roughly three billion people live in areas endangered by malaria, a parasitic disease caused by a parasite with a complex life cycle. Five different species of the genus *Plasmodium* are responsible for all human infections with an estimated 200 million cases and several hundred thousand deaths occurring annually. More than 90% of the latter are located in sub-Saharan Africa where the form *Plasmodium falciparum* (*Pf*) predominates and children below the age of five are most strongly affected. Symptoms are manifold and range from fatigue and muscle aches over irregular fever to coma (specifically for falciparum Malaria), liver injury and severe anaemia [White et al., 2014]. Longitudinal studies in endemic areas have shown that adults typically acquire non-sterile immunity to malaria after several parasite exposures, thus acting as asymptomatic parasitic reservoirs [Bousema et al., 2014]. While new-borns and young children are the most vulnerable to severe malaria disease and related death, asymptomatic carriage can also be observed in this age group [Males et al., 2008].

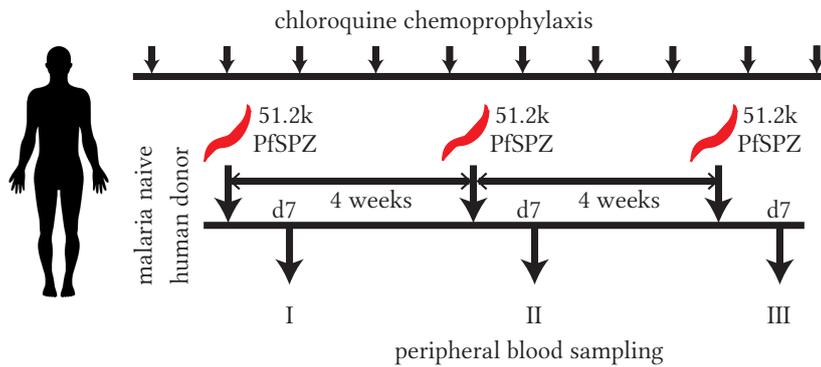
The life cycle of *Pf* takes place in the human host and the *Anopheles* mosquito. Starting with the mosquito bite, 15-40 motile *Pf* sporozoites (*Pf*SPZ) are released into the dermis [Yuda and Ishino, 2004, Amino et al., 2006]. From there, they travel to the liver via the blood stream where they invade hepatocytes in which multiply extensively. After about one week, infected hepatocytes burst open and release several thousand merozoites into the blood stream. It is these merozoites which infect erythrocytes where they continue to replicate exponentially with an infection to cell rupture period of 48 hours. This phase of malaria infection is referred to as blood stage; the symptomatic phase of the disease is associated with it. Within erythrocytes, merozoites can undergo either asexual expansion or choose to differentiate into sexual gametocytes. These are then again taken up by *Anopheles* mosquitoes during blood

meals. Sexual reproduction within the mosquito host eventually leads to the production of new sporozoites, thus completing the life cycle [Hafalla et al., 2011, White et al., 2014].

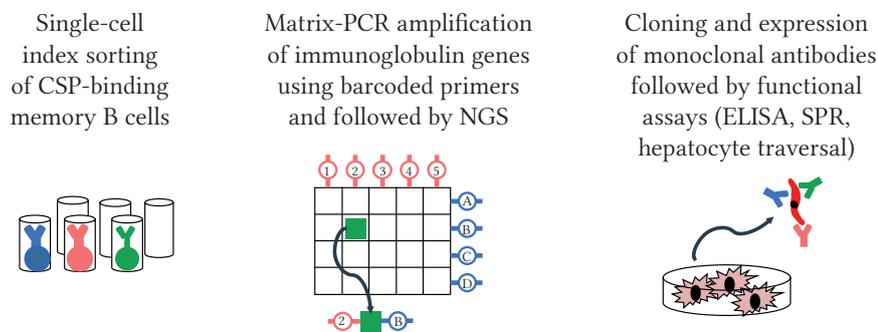
Because of their early position within the human fraction of the malaria life cycle, sporozoites are an attractive target for vaccine development attempts. As early as 1967 it was shown that sterile protection against live sporozoite challenge could be established by vaccination with radiation-attenuated *Plasmodium berghei* sporozoites in rodents [Nussenzweig et al., 1967]. Subsequently, circumsporozoite protein (CSP) - the antigen depicted in figure 2.3b - was identified as the major sporozoite surface protein and it could be shown that monoclonal antibodies against CSP blocked hepatocyte invasion in *Pf* and *P. vivax* [Hollingdale et al., 1984]. CSP consists of three structural domains: between N-terminal and C-terminal domain, a central region consisting predominantly of several repeats of the amino acid sequence NANP is placed [Dame et al., 1984]. Albeit with varied length, this repeat region is highly conserved between *Pf* strains [Zavala et al., 1983], thus increasing its attractiveness as a vaccine target. The current leading malaria vaccine candidate RTS,S AS01 (tradename Mosquirix) contains 18.5 NANP repeat units as well as the C-terminal domain of CSP with the aim of inducing protective antibodies [Neafsey et al., 2015]. Recent phase 3 clinical trials in different age groups in endemic regions have however resulted in overall low protection between 25 % and 55 % without long-term protection [Neafsey et al., 2015, Olotu et al., 2016].

Tübingen Controlled Human Malaria Infection Trial The blood samples analysed in the experimental work underlying this comparison are derived from a malaria immunisation trial, the *Tübingen Controlled Human Malaria Infection Trial*, in which sterile protection against *Pf* was induced using a chemoattenuated *Pf* sporozoite vaccine. All details are given in [Mordmüller et al., 2017]. In brief, aseptic, purified, cryopreserved, non-irradiated sporozoites were administered to malaria-naïve, healthy adult volunteers by direct intravenous inoculation. While [Mordmüller et al., 2017] present result for several doses and vaccination schedules, the data discussed in the following is derived from a protocol in which $5.12 \cdot 10^4$ sporozoites were injected three times at intervals of 28 days under chloroquine antimalarial chemoprophylaxis. Blood used in the study discussed below was collected seven days after each of these infections (see figure 4.1a). As tested by a controlled infection with a lower dose of sporozoites in the absence of chloroquine ten weeks after the last dose, the immunisation process prevented blood stage infection in all nine participants.

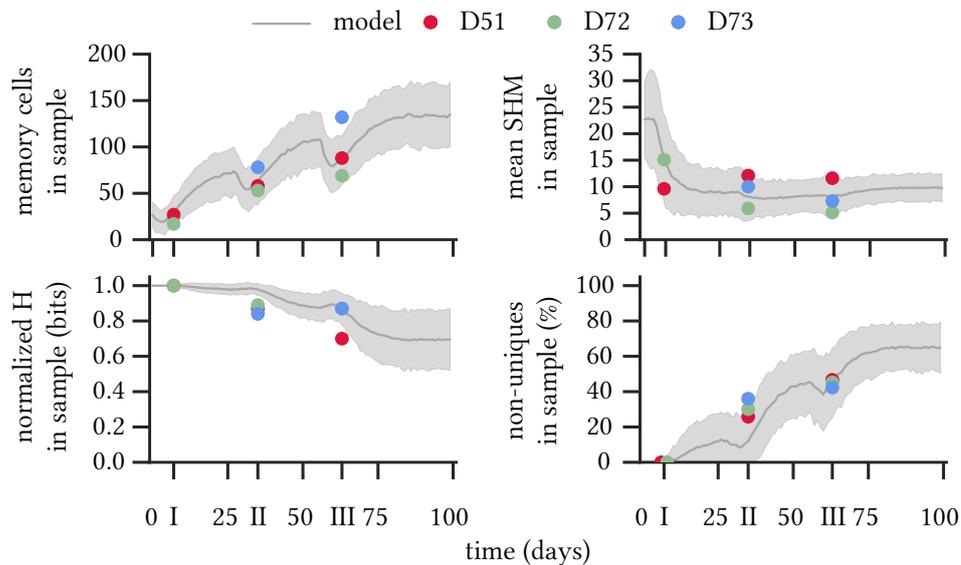
Single-cell processing of memory B cells The full experimental proceedings are discussed in [Murugan et al., 2018] and [Murugan, 2018] and summarised in figure 4.1b. In brief, single CSP-binding memory B cells ($7AAD^- CSP^+ CD19^+ \{IgG \text{ or } CD27\}^+$) and antigen-unspecific plasmablasts ($7AAD^- CD38^+ CD27^+ CD19^+$) were isolated from suspensions of peripheral blood mononuclear cells from eight out of the nine study participants using flow cytometry with index sorting. Immunoglobulin (Ig) heavy, κ and λ chain genes were amplified using a high-throughput robotic platform employing a nested matrix polymerase chain reaction with bar-coded primers [Tiller et al., 2008, Murugan et al., 2015, Busse et al., 2014] and the resulting amplicons were sequenced using 454 GS FLX+ (Roche) or MiSeq 2×300 -base pair (bp) paired-



(a) Scheme of the immunisation trial. Blood was sampled from 9 participants seven days after each immunisation.



(b) Experimental strategy applied to immune cells extracted from peripheral blood at time points I, II and III. PCR - polymerase chain reaction, NGS - next-generation sequencing.



(c) Comparison of three representatively sampled donors with model outcomes. H - Shannon's entropy, SHM - somatic hypermutation. Experimental data: Rajagopal Murugan.

Figure 4.1.: Controlled human malaria infection trial, experimental processing and comparison of data and model statistics.

end (Illumina) sequencing. Only wells with paired full-length functional Ig heavy and light chain gene sequences were analysed and matched with index-sorting information using the bioinformatics platform sciReptor [Imkeller et al., 2016]. Recombinant monoclonal antibodies were generated in human embryonic kidney cells [Tiller et al., 2008]. These antibodies were purified and their reactivity and affinity were measured using enzyme-linked immunosorbent assays (ELISA) against 10-fold repeats of NANP (NANP₁₀) and surface plasmon resonance (SPR) with NANP₅, respectively. Protective efficacy of antibodies was tested using a *Pf*SPZ hepatocyte traversal assay [Sattabongkot et al., 2006, Triller et al., 2017]. All experiments were performed by Rajagopal Murugan and colleagues from the Division of B Cell Immunology lead by Hedda Wardemann at the DKFZ in collaboration with Elena Levashina and colleagues at the Vector Biology Unit of the Max Planck Institute for Infection Biology in Berlin.

4.2. Comparison of Experimental Data and Computational Model

As described above, a comprehensive set of information at the single-cell level was collected on B cells following each of the three vaccinations during the malaria immunisation trial outlined in section 4.1. In addition to the cell phenotype, its clonal belonging derived from its antibody gene sequence as well as antibody affinity are known. In the following, this wealth of experimental information will be compared to observations and predictions generated with the model introduced and analysed in chapter 2 and 3.

Matching dynamics As an initial test, basic characteristics of the experimentally sampled memory B cells from three comprehensively sampled donors were compared to the model (figure 4.1c.) All model parameters used for the comparison are the default values shown in section 2.A with an antigenic complexity of $n_{\text{key}} = 10$ because the recently published crystal structures of two different antibody-CSP binding interfaces suggest complicated binding modes with many contacts [Triller et al., 2017]. First of all, we analyse the number of memory cells in each sample. For normalisation, the computational sample size was scaled using the mean experimental sample size at the second time point and decorated with Poisson noise. Sample sizes at time point II are on the order of 100 cells. We have previously seen in figure 3.3c that the size of the modelled memory pool increases with each infection as germinal centres help expand the population. Figure 4.1c shows that the dynamics of this increase agree well with the number of memory B cells found in the experimental samples from three representative donors with prominent anti-*Pf*CSP response. The same goes for the average number of hypermutations identified in the variable segments. It may at first seem surprising that there is no significant increase in mean mutation number over the course of the three time points neither in the model nor the experiment and reasons for this observation will be discussed below. The lower two panels of figure 4.1c are concerned with the diversity of the samples. The decreasing clonal entropy of the full simulated memory pool was previously shown in figure 3.3d and this behaviour passed down to a smaller sample. Experimentally, a similar degree of clonal expansion with concomitant diversity loss is observed.

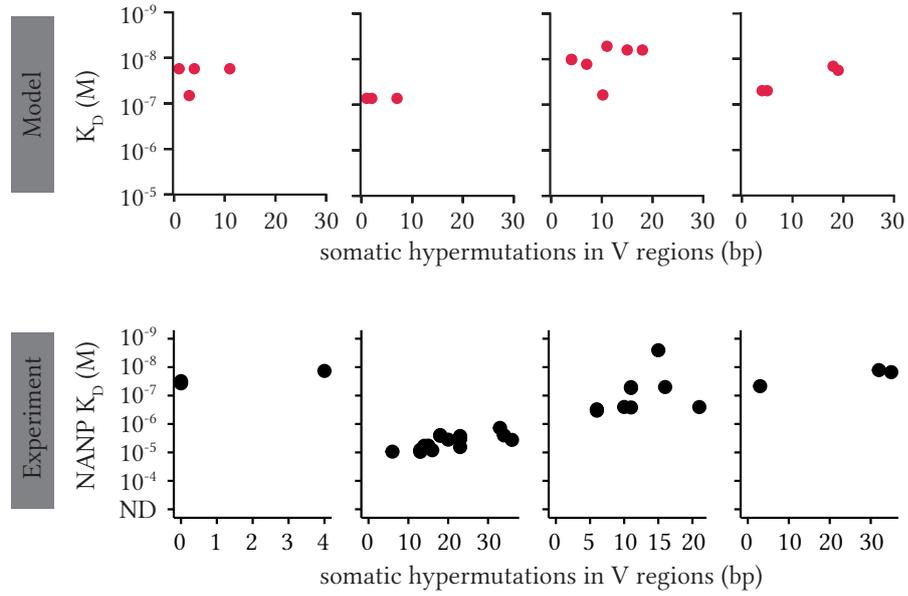
Overall, these simple comparisons underline that large agent-based models like the one at hand have the ability of integrating a vast pool of experimental knowledge in a useful way. The fact that the model builds on more than 25 parameters may seem disconcerting at first, however, almost all of these parameters are available in the experimental literature and there is thus no need to allow free variation. Such modelling approaches can therefore be viewed as a method of reassembling biological complexity which has had to be taken apart experimentally in order to determine individual quantities. A bottom-up approach of this type can help understand processes in complex systems whose individual components have already been quantified.

Intra- and interclonal affinity improvement We have previously observed that the contribution of hypermutations to overall affinity improvement can be minor in comparison with clonal selection in section 3.3. In the model as well as in the experimental data, mutational effects within individual clones can be examined in addition to the bulk dynamics. A small subset of four individual clones each is shown in figure 4.2a which gives the affinity of individual B cells as a function of their mutational status. It becomes apparent that the textbook logics as shown in figure 2.1, where more hypermutations in general mean more affinity does not apply to this case. Overall, computationally as well as experimentally, strong improvements within B cell clones are scarcely observed at this level of binding complexity albeit larger jumps do occur. In the majority of cases, clones retain a strong memory of their germline ancestor's affinity towards the antigen and do not change it significantly over the course of many base pair replacements.

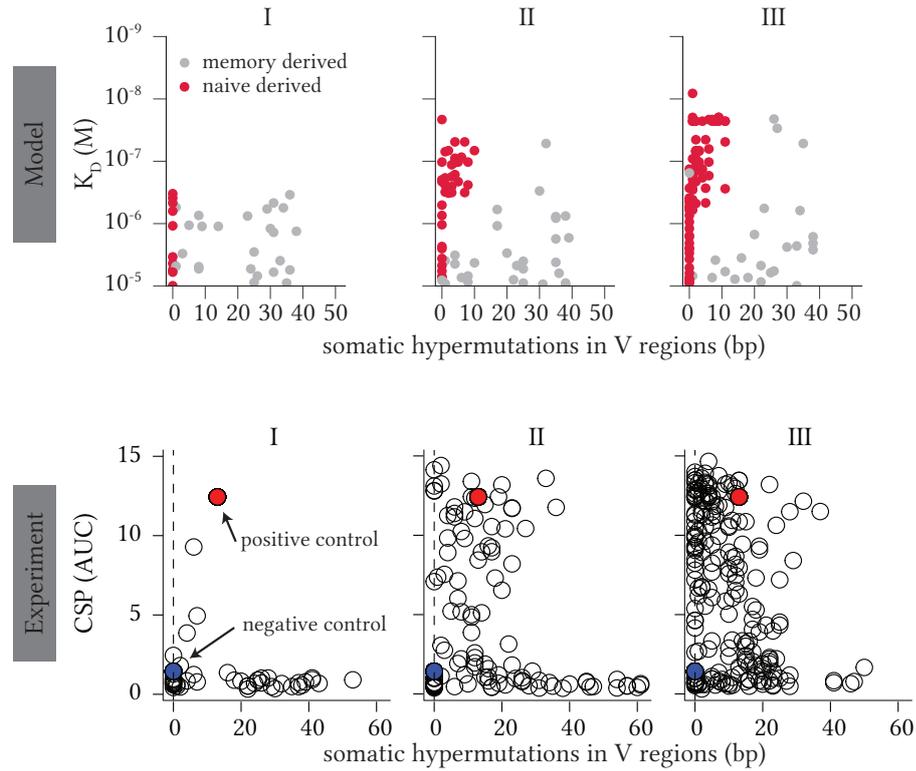
Instead of focusing on the mutational changes of affinity within individual clones, figure 4.2b shows the mutational status of the entire sampled memory pool from one representative simulation in the computational case and all eight donors pooled in the experimental case. Focusing on the model prediction first, the sample contains no high-quality binders at the first time point (seven days after infection I) although we know that they are present in the modelled repertoire by design. However, their frequency in the population is too low at this point in order to reliably appear in small samples. Moving on to the second time point, high-affinity cells suddenly appear at the lowly mutated end of the spectrum. The colour code applied to the dots representing single cells indicates whether the sampled memory cell derives from an ancestor that was recruited into the response from the naïve compartment (red) or from the population of pre-existing memory cells (grey). This shows that the newly appeared high-affinity cells with low mutation frequencies overwhelmingly stem from naïve ancestors. Following the third exposure, this population is expanded further and slowly starts to acquire low levels of hypermutations. Underlying these observations is the avalanche effect (section 3.3) - only a few good precursors need to be activated in order to multiply them into strongly contributing populations after two or three vaccination shots. Strong mutational effects are not required for improvement at the population level if such precursors are present. Moving on to the experimental data plotted in the same way in the lower panels of figure 4.2b, the model prediction is precisely confirmed. Even though no high-affinity cells are sampled at time point I, increasing numbers of lowly mutated strong binders appear after the second and third vaccination.

Another curiosity appears when examining the single-cell affinity as a function of muta-

4.2. COMPARISON OF EXPERIMENTAL DATA AND COMPUTATIONAL MODEL



(a) Affinity as a function of mutational state within individual modelled and experimental B cell clones. NANP - central repeat region of *PfCSP*.



(b) Affinity as a function of mutational state at the three sampling time points. All clones from all eight donors are pooled.

Figure 4.2.: Comparison of mutational effects on affinity between model and experiment. Each dot/circle represents one monoclonal antibody. bp - base pair. Experimental data: Rajagopal Murugan.

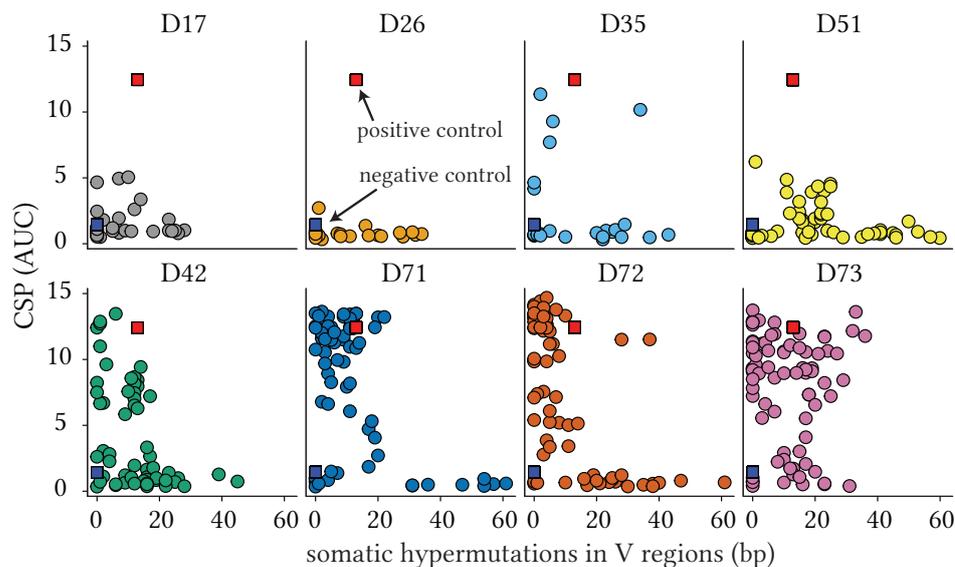


Figure 4.3.: Affinity as a function of mutational state similar to figure 4.2b. However, donors are shown individually here while all time points and clones are pooled. Each dot/circle represents one monoclonal antibody. bp - base pair. Data: Rajagopal Murugan.

tional state for each of the eight patients individually (figure 4.3). Here, all cells from all three time points were pooled. It becomes apparent that the selection-based avalanche process outlined in the previous paragraph only worked in half of the participants. While the donors {D42, D71, D72, D73} show the described expanded population with strong affinity and few mutations, this is not the case in donors {D17, D26, D35, D51}. These donors have mostly low affinity B cells distributed over the entire spectrum of hypermutation states. Looking back at the modelled repertoire distributions in figure 3.6b, specifically at the scenario where no good precursors for efficient selection are available, this is what would be expected: an overall low quality of the response with few better candidates emerging late while the majority of cells has trouble mutating away from its germline ancestor's affinity even with a high number of mutations.

It is important to point out that all eight donors developed fully protected from infection in the sporozoite challenge. First of all, analysis here focused on the central region of CSP, an important and predominant surface antigen with functional implications, but by no means the only surface antigen of *Pf* sporozoites. It is conceivable that the remaining donors developed immunity against another part of CSP or even against another surface antigen because their precursor repertoire was better equipped for these tasks. Additionally, it is a matter of current debate which branches of the immune response are mainly responsible for anti-malaria immunity and/or whether this may differ from patient to patient. The efficacy of anti-CSP antibodies for preventing infection has however been demonstrated [Triller et al., 2017, Murugan et al., 2018].

Finally, it is of interest to put these results into the context of the lower-dose variants of

the TUCHMI trial [Mordmüller et al., 2017] and also the outcomes of long-term natural exposure [Triller et al., 2017]. In the first case, the same vaccination schedule using 4 and 16 times less sporozoite induced protection in six out of nine and three out of nine donors respectively. This is in line with the finding presented here that an avalanche effect driven by high antigen doses for selection and subsequent mixing could be underlying affinity improvement processes against CSP. In the second case, the authors find that even after long-term natural *Pf* exposure of adults in Gabon only a weak anti-CSP B cell memory developed. They further examine two affinity matured NANP-reactive memory B cell antibodies and find that repeat recognition was mostly mediated by germline encoded antibody features. Hypermutations in turn lead to a stabilisation of the binding interface. A major difference between the vaccination scenario described above and the natural exposure is the number of sporozoites on display to the immune system, the number in natural infections being as low as 20 [Yuda and Ishino, 2004]. This in combination with the short time that sporozoites are available for immune recognition before homing to hepatocytes may promote their immune-escape and hinder the activation of sufficient B cells. It may also explain why, even though in general malaria has strongly shaped the human genome in regions where it is endemic [White et al., 2014] and supposedly vice versa, sporozoites remain vulnerable to antibodies encoded in the germline. Because of the short exposure at low concentration, there may not have been sufficient evolutionary pressure on the parasite's genome for removing this vulnerability. Repeated long-term exposure in endemic regions may eventually compensate for these deficits and contribute to the partial immunity often found in their inhabitants.

5. Discussion - B Cell Memory Responses

In this thesis, a flexible computational model of the humoral immune response has been developed. This model allows to vary and explore the effects of a vast range of biologically interpretable parameters connected to characteristics describing specific pathogens, patients or vaccination processes. To this end, the model integrates quantified experimental knowledge from several decades of research on B cell biology, germinal centres and antibody-antigen interaction. Using a newly developed binding model combining advantages of shape-space and NK-type binding models, the simulation introduced here allows to vary epitope complexity in an intuitively accessible way. Additionally, a novel method for visualising selection and mutation effects in simulated germinal centres is introduced and facilitates understanding and comparison to experimental observations. Using this computational model, we have been able to disentangle mutation and selection effects both at the level of individual germinal centre reactions and at the level of the resulting memory B cell population. The importance of the thus predicted avalanche effect could be demonstrated in a direct comparison to experimental data from a malaria vaccination study. The complete computational framework is available at https://github.com/LiBuchauer/gc_memo.

Modelling Diverse Challenges within a Unified Framework In chapter 2, we have introduced a new computational model of the B cell immune response in detail. The necessity for this new approach is inferred from the observation that affinity maturation, the process of B cell affinity improvement via somatic hypermutation mostly in germinal centres, does not function with equal efficiency for all antigens. While anti-hapten responses routinely lead to a small number of characteristic beneficial mutations, the same is not observed when analysing matured repertoires against more complex pathogens where larger proteins are targeted by the immune system. Because of this, we have abstracted the binding interface between antigen and antibody into a set amino acids chiefly responsible for determining relative binding strengths. This set is of size n_{key} , a crucial quantity taking low values for structurally simple binding problems with fewer contacts and higher values for larger binding interfaces typically observed in protein-protein binding.

Akin to shape-space models, our binding model essentially derives affinity from matching two sequences against each other. However, instead of taking arbitrary symbols and evaluating matches, our model uses the 20 physiological amino acids and binding strength is assessed using true energy-like quantities derived from a study on folded proteins [Thomas and Dill, 1996]. This, together with the codon-based mutation procedure implemented here, makes our model more realistic as it introduces the conservatism of the genetic code into the challenge of improving affinity mutationally. The genetic code has evolved to guarantee maximal stability

in the face of random mutations; this property is necessarily also part of affinity maturation in germinal centres. Single base pair replacements lead to replacement by the same or a functionally similar amino acid much more frequently than in a random replacement model. In the context of desired mutations in the germinal centre, this provides a dampening factor on affinity diversification.

A general flaw of sequence-based binding models that evaluate pairwise interactions only is that they have a single optimal solution in sequence space and smooth mutation paths lead towards it. This is biologically not realistic because of epistatic effects in protein-protein binding situations: amino acid replacements happening outside of the actual binding site can nevertheless affect binding quality because of resulting structural changes. Our model incorporates this threat by allowing non-lethal mutations to the scaffold region to exert blocking effects on the binding strengths of the antibody: if a blocking mutation is present in the scaffold region, the antibody is no longer able to affinity mature via key region mutations. Taken together, the codon-based mutation model, the binding energy-based binding model and the introduction of mutations blocking affinity maturation via long-range effects represent improvements to existing shape-space models because the resulting situation is closer to the reality of evolution in affinity maturation.

Our dynamic model of individual germinal centre reactions builds on a cycle of stochastic competition for survival signals based on antigen affinity, receiving these signals via T_{FH} contact, division and reentry into the competition process or differentiation into a peripheral effector cell type. Importantly, we allow the colonisation of new germinal centres not only by naïve cells but also by pre-existing memory B cells. As B cell identities are defined by their key region sequence alone in our model, there is no reason why pre-existing memory cells from unrelated infections should not be able to participate in a new primary infection. Thus, the participation of mutated memory cells in addition to fresh naïve cells is part of the model and indeed the participation of highly mutated memory cells in primary exposures of European donors to *Plasmodium falciparum* is observed experimentally (see also section 4). Concerning seeder cells, our model differs from most recent germinal centre simulations in that it picks up on the experimental finding that on the order of hundred different B cell clones likely seed a germinal centre upon formation [Tas et al., 2016]. In our model, the number of actual seeder cells depends on their concentration in the entire B cell pool as well as on the antigenic dose chosen for the simulation: more antigen increases the probability of activation for every existing B cell. Building on these straight-forward assumptions, we have shown that number of seeder cells has a strong impact on the final quality of the memory B cell pool. This indicates that individuals partly or completely lacking strong precursors have decreased likelihood of developing protective vaccine responses.

While we provide experimentally backed suggestions for the choice of default parameters, each parameter can be varied computationally and often such exploration is biologically motivated. As examples, the size and lifetime of germinal centres may depend on specifics of the antigens towards which they develop. If information about such properties is available, incorporating it into the simulation can help to find mechanistic explanations for observed

outcomes. An example is given by the vaccination study examined in chapter 4, where shorter germinal centre lifetimes are implemented because of the pathogen's life cycle - as a result, the decreased time window available for mutational changes contributes to the weak intraclonal improvement observed within B cell clusters over time.

Joint Action of Individual Players In contrast to a real life situation, a computational model allows access to all simulated processes at a high resolution. Here, we have developed an intuitive depiction of clonal selection and mutation processes during germinal centre reactions on top of the underlying simulation. These figures illustrate how germinal centres set out with a diverse repertoire of seeder cells which is consecutively narrowed down to one or a few clones because of stringent selection by T_{fh} cells and resulting expansion. The mutational effects occurring throughout the process are mirrored by the memory cell output produced by the reaction. An analysis of this output shows how, initially, the mean quality of newly produced memory cells increases quickly as selection acts upon the diverse seeder pool. However, the speed of improvement is reduced or even stalled once mono- or pauciclinality is reached and further improvements need to be brought about by mutations.

Mutational improvement is more strongly affected by stochastic effects than is selection. We have shown that individual germinal centre reactions under similar conditions (same dynamics, same epitope complexity, similar number of seeder cells) can lead to varying outcomes - while a set of beneficial mutation is discovered in one reaction and becomes fixed in its population, another germinal centre remains at low affinity over its entire time course. This outcome diversity is reduced by increasing the number of seeder cells, as this shifts the burden of population improvement from the inefficient mechanism of mutation to the more reliable workings of selection. Analysing different antigen complexities, we have shown that this is especially important for antigenic challenges with higher numbers of key residues jointly determining affinity. Because the mutational random walk proceeds in higher-dimensional space here, finding the optimum is a more difficult endeavour. The efficiency of selection on the other hand is unaffected by the increase in complexity. Thus, clonal selection from a sufficiently rich precursor repertoire outpaces affinity maturation by somatic hypermutation in cases of complex binding interfaces.

In order to understand physiological B cell reactions which rely not only on one but on several hundreds or thousands of germinal centres at different sites in the body, we have embedded on the order of 100 individual germinal centre simulations within a joint pool of free B cell objects. This allows to explore effects emerging from the presence of several weakly interacting sites of affinity maturation. Investigating a vaccination schedule comprised of three consecutive immunisation shots at intervals of four weeks, we have found that repeated exposures are important for further reduction of stochastic activation, selection and mutation effects. We have identified the underlying mechanism to work as follows: individual strong binders present at very low frequency in the naïve or pre-existing memory precursor repertoire become activated during the first exposure and succeed to join one or a few of the hundred available germinal centre reactions. Within this reaction, they become expanded to occupy $\mathcal{O}(1/100)$ of the newly generated memory pool, a much enhanced fraction compared to their previous pres-

ence. During the following immunisations, in the case that general activation is high enough, the probability that a member of this high-affinity clone joins each new reaction as a seeder increases. Consequently, its expansion is amplified. We show that even in the complete absence of somatic hypermutations, multi-exposure selection processes can produce high-quality memory pools. Notably, this situation is strongly preferred over one with working hypermutation but suffering from total absence of good precursors.

As briefly discussed in chapter 4, following the second and third immunisation, members of individual clones occupy much higher fractions of the total memory repertoire than could be explained by expansion in a single GC alone. This provides experimental confirmation of the theoretically predicted avalanche effect. Importantly, we predict that the magnitude of this multiplication effect is dependent on the strength of activation and thus on antigen dose, as this is not only crucial for initial activation of rare good precursors but also for sufficient mixing of seeder cells during booster vaccinations. During the malaria vaccination study underlying the experimental data in chapter 4, lower vaccination doses were tested likewise [Mordmüller et al., 2017]. However, they did not lead to protection in all subjects. The relevance of high doses for precursor activation and exploitation of the avalanche effect may in part explain this result.

The Importance of Maintaining Diversity Why does the size of individual germinal centres not exceed a few thousand cells physiologically? Might it not be beneficial to focus competition at one given site instead of having a distributed system of several hundreds or thousands of reactions? As part of a parameter sensitivity analysis we have found that larger germinal centres lead to higher mean affinities in the memory cell pool: At one centralised site, an individual B cell outcompeting all peers could hence be expanded much more efficiently. However, we have also looked at inter- and intraclonal diversity within both individual germinal centres and the entire population. While clonal selection quickly reduces the number of clones present within a reaction, the actions of somatic hypermutation increase the number of different binding regions present. Thus, the increase of intraclonal diversity by mutation counteracts to some extent the clonal focus introduced by selection. In our model, the amino acids with less relevance for the binding process at hand are not modelled specifically and mutations happening within this scaffold area are merely counted. It is therefore impossible here to evaluate the diversity increase introduced into the entire sequence, but it is certain that overall sequence diversity is maintained at an even higher level compared to key region diversity alone because no selection acts on it. What has been discussed for individual germinal centres is likewise true a population level, where clonal entropy decreases with each additional vaccination shot, largely due to the avalanche effect described above.

Based on theoretical as well as experimental observations, we have argued that memory cells stemming from unrelated previous infections are likely to be recruited into current B cell responses. Is it thus obvious why maintenance of repertoire diversity is beneficial - instead of filling the entire available memory B cell niche with exact duplicates of a high-quality binder, having variations of this cell with very similar affinity is a more promising strategy when looking at potential future threats. While variations within the key binding region may be

particularly relevant for related antigens such as mutational variants or regionally different strains, variability within the entire sequence could be relevant for protection towards completely unrelated threats. These may bind to the antibody in a different way at a different position on its three-dimensional structure thus defining new key amino acids. Taken together, it is likely that maintaining diversity, even though mostly treated as the poor cousin of improving affinity, is a major reason for somatic hypermutation in germinal centres. Germinal centres enable strong selection and expansion and can therefore put the diversity of the host's memory B cell repertoire as a whole at risk. Here we have shown that there are cases in which the affinity-improving effects of somatic hypermutations are minimal; at the same time, their diversification benefits remain present.

In an evolutionary light, this interpretation of the workings of affinity maturation could have led to an optimal germinal centre size. While we have shown that a set of larger germinal centres produces a memory B cell pool of higher quality, it also leads to decreased clonal diversity overall. At the same time, having a higher number of germinal centres naturally increases this diversity. As a third level, vaccination strategies comprised of several immunisations as well as repeated natural exposures likewise lead to a progressive reduction of clonal diversity with concomitant affinity increases. Based on our model we have found that even during avalanching scenarios where the same strong clone is distributed to all germinal centre sites, this clone does not necessarily take over in all of them because of stochastic selection and mutation effects. In combination, while large germinal centres and repeated exposure increase affinity, having a highly distributed system guarantees diversity.

Addressing Inter-Patient Variability In section 3.3 we have shown that differences in precursor repertoires between individuals can strongly influence vaccination success. While a triple-shot vaccination protocol is highly efficient in amplifying initially rare cell populations and can lead to decent final repertoires even without mutational help, the picture is less rosy if such precursors are absent. Here, mutation remains as the only road to better B cells, and in the case of high-complexity antigens this path requires a long time and constant simulation. The importance of efficient precursor identification and amplification could be confirmed experimentally in chapter 4 by analysing single-cell affinity as a function of mutational state over time. Additionally, looking at the same measure within individual participants of the vaccination trial revealed evidence for differing precursor repertoires similar to theoretical cases (with and without good precursors) described above: while some donors seem to have identified a high-quality low-mutated population early on in the protocol and then go onto expand it, others failed to produce good CSP-binders over the entire period of the trial even though their weak binders display up to 60 base pair mutations compared to germline by the end of the trial.

While the existence of differing vaccination outcomes has been known for a long time, an understanding at this level of detail, where single-cell sequencing data is brought together with affinity measurements of monoclonally expressed antibodies as in [Murugan et al., 2018] are a recent development. In combination with functional assays and crystal structure analyses of antibody-antigen complexes they have the potential to drive rational vaccine development.

Focusing on differences between human subjects, it may become possible to develop personalised vaccination strategies in the future. Following the identification of a protective B cell epitope on the antigen under examination, a thorough analysis of the binding interface and the effect of individual mutations can produce an estimate for the complexity of this interface and the resulting challenges for somatic hypermutation. More importantly, with increasing availability and decreasing cost of single-cell methods like antigen-specific flow cytometry or immunoglobulin gene sequencing methods it may become possible to routinely query the precursor repertoire of each individual and develop an optimal vaccination strategy based on the result. If the person has a precursor repertoire containing many useful B cells for the problem at hand, low activation doses and fewer vaccinations may be sufficient for full protection, while in cases with worse starting conditions, higher doses, longer intervals and also more immunisations may be required. Vaccine administration, specifically in the case of expensive vaccines, could thus be optimised both in economic terms and regarding comfort of the patient. Additionally, in cases where several protective vaccine targets are known, a prior screening of the available repertoire may allow an informed choice of vaccine composition thus making optimal use of the given boundary conditions. Computational models like the one introduced here and enhanced with specific information concerning antigen and patient on a case-to-case basis will help to identify optimal schedules in this vision of the future.

Part II.

**Stem Cell-Driven Glioblastoma
Growth**

Und ist schon jemals ein Ziegel so vom Dach gefallen, wie es das Gesetz vorschreibt? Niemals! Nicht einmal im Laboratorium zeigen sich die Dinge so, wie sie sein sollen. Sie weichen regellos nach allen Richtungen davon ab, und es ist einigermaßen eine Fiktion, dass wir das als Fehler der Ausführung ansehen und in ihrer Mitte einen wahren Wert vermuten.

– Robert Musil, *Mann ohne Eigenschaften*

In the second part of this thesis, a mathematical model of glioblastoma growth is developed based on qualitative experimental observations. All parameters of this model are thereafter estimated based on quantitative experimental data; finally, the quantified model is used to predict and understand further experimental observations not used for parameter estimation.

Chapter 6 provides a rough overview of the fields of both mouse models and mathematical models in glioblastoma research with a slight focus on brain cancer stem cells. Data analysis begins in chapter 7, in which experimentally observed tumour growth curves are analysed with phenomenological models and a simple three-dimensional simulation of tumour growth is developed. Chapter 8 takes mathematical modelling from the phenomenological to the mechanistic level by introducing a system of differential equations describing the different behaviours of brain tumour stem cells and their progenitors. This basic model is then modified to describe a vast set of experimental observations; ultimately, division, differentiation and death rates of the various cellular phenotypes are estimated. Based on this quantified model, chapter 9 introduces advanced experimental procedures including cancer treatments and single-cell labeling approaches and how their outcomes can be predicted as well as interpreted in the context of the simple description from chapter 8. Finally, an integrative discussion of all results is provided in chapter 10.

All experimental work described in the following chapter was performed by Muhammad Amir Khan, Yue Zhuo and Peng Zou in the Division of Molecular Neurogenetics (DKFZ Heidelberg) headed by Hai-Kun Liu.

6. Background - Cancer Stem Cells in Glioblastoma

In this chapter, the frame for the scientific work following in chapters 7, 8 and 9 is set by giving background information in three major areas. Section 6.1 introduces the malignant brain tumour glioblastoma, its properties as well as general mouse model techniques used to explore it; section 6.2 discusses mathematical modelling attempts to stem-cell driven cancers, three-dimensional solid tumours and to glioblastoma more specifically; and section 6.3 gives details of experimental techniques and mouse models used in the laboratory of Dr. Hai-Kun Liu at the DKFZ and thus directly underlying this work.

6.1. Glioblastoma Multiforme

The aggressive brain tumour *Glioblastoma multiforme*, its most prominent pathological and genomic characteristics and its response to treatment efforts are roughly introduced in section 6.1.1. As basic as well as applied research into this topic relies heavily on mouse models, the most important forms of brain tumour mouse models are discussed briefly in section 6.1.2. Finally, in section 6.1.3, the notion of cancer stem cells, specifically in the neural context, is defined and the clinical relevance of the concept is described.

6.1.1. Epidemiology and Pathology of Glioblastoma

Tumours of the central nervous system comprise heterogeneous groups of benign and malignant neoplasias, most of which are of glial origin and thus called gliomas [Ohgaki and Kleihues, 2005]. According to the specific type of glial cell these tumours originate from, a further differentiation into astrocytic, oligodendroglial and ependymal as well as hybrid forms is possible [Louis et al., 2007].

Prognosis and Pathological Features Among gliomas, the high-grade astrocytoma, also called *Glioblastoma multiforme* (GBM), is the most prevalent and aggressive form - about 45% of all gliomas fall into this category in which relative 5-year survival rates are below 5% [Ostrom et al., 2014]. Glioblastomas are characterised specifically by active angiogenesis, pseudopalisading necrosis and diffuse invasion into neighbouring brain tissue [Louis et al., 2007]. Although glioblastoma cells tend to migrate between brain hemispheres thus often causing a symmetric lesion also referred to as “butterfly glioblastoma” [Dziurzynski et al., 2012], extracranial migration is rare [Schweitzer et al., 2001].

Glioblastoma Genome In a comprehensive attempt at identifying common genomic characteristics of glioblastoma, an international consortium of scientists (*The Cancer Genome Atlas*)

analysed DNA copy number, gene expression and DNA methylation aberrations in 206 glioblastomas and additionally looked at sequence aberrations in 91 out of these [McLendon et al., 2008]. Overall, 90% of glioblastoma patients carried at least one amplification or mutation in receptor or downstream kinase genes such as EGFR, PDGFR or PIK(3)K or, alternatively, a deletion or mutation within inhibitor genes of these kinases such as PTEN. The most abundant disturbed pathway was the p53 pathway with alterations (deletion or mutation in TP53 or amplification of proteins inhibiting p53) occurring in 86% of subjects. In addition to these pathways commonly associated with well-regulated cell cycling and survival, deletion or mutation of NF1 was observed in 23 % of cases. Significant evidence for NF1 being a tumour suppressor had previously been found in mice, although its role in human remained under debate. This finding underlines the importance of large-scale collaborative efforts for the extraction of robust statistics from comprehensive data sets and the subsequent identification of potential drug targets.

Treatment The standard of care for glioblastoma comprises surgical resection, radiotherapy and administration of the chemotherapeutic temozolomide (TMZ) [Stupp et al., 2005, Malmström et al., 2012]. As a result of longitudinal studies, it has been reported that TMZ treatment leads to greatly increased numbers of mutations in the relapse tumour compared to the primary sample [Choi et al., 2018]. Interestingly, a longitudinal genomic study of 114 patients, in which only 15% showed hypermutation signatures at relapse, finds evidence for highly branched tumour evolution [Wang et al., 2016]. Notably it is estimated that most tumour clones dominating the relapse typically existed years before diagnosis, indicating that while they may not have been competitive in the original setting, a potentially higher resistance to treatment allowed them to reconstitute the regrowth.

Overall, in spite of many conceptual advances in the field of glioblastoma research during the last decades, a significant improvement of patient survival rates remains wanting. Next to genetic diversification during treatment as outlined above, a major roadblock is the high invasiveness of glioblastoma cells. They are highly motile, able to detach themselves from the tumour bulk either individually or in small groups and travel easily among various brain structures. The resulting satellites cannot be reliably distinguished from the surrounding healthy tissue during operation or using magnetic resonance imaging [Louis et al., 2007]. Thus, further investigation of the properties of migrating cells, the characteristics of potential functional subtypes within the tumour bulk as well as associated molecular targets is necessary.

6.1.2. Glioblastoma Mouse Models

Murine cancer models in general represent a useful tool for basic cancer research and pre-clinical drug study as they, in contrast to *in vitro* settings, capture complex aspects of human neoplasias including tissue-defined space constraints, vascularisation and invasion of host immune cells. Two major groups of mouse models for cancer and specifically for glioblastoma are available: xenograft models and genetically engineered models. Both types are introduced briefly in the following.

Xenograft Models Xenograft models are further subdivided into two categories, cell-line-derived (CDX) and patient-derived xenograft (PDX) models. The former rely on human or mouse cell lines which are implanted into selected organs of immunocompromised mice. CDX models are fairly easy to establish and have been key to the identification of a number of cytotoxic drugs [Day et al., 2015]. Specifically in glioblastoma, important histopathological features could be recapitulated using cell lines, among them infiltrative invasion into the brain parenchyma, palisading necrosis and angiogenesis [Camphausen et al., 2005, Wachsberger et al., 2005]. However, CDX models are frequently of little help in predicting outcomes of more advanced treatment strategies as they do not depict the entire heterogeneous complexity of a natural tumour and its immune- and microenvironments [Day et al., 2015].

Patient-derived xenograft models overcome part of these issues as intact tumour tissue preserving tumour architecture is directly implanted into immunocompromised mice [Hidalgo et al., 2014]. The additional step of *in vitro* selection and homogenisation is omitted. Examples of the use of PDX models in glioblastoma research include testing the efficacy of therapeutic antibodies [Reilly et al., 2015] and a study of emerging resistance against radiotherapy and temozolomide [Anderson et al., 2014]. In laboratory practice, the applicability of this method is limited by the availability of fresh biopsy tissue. Although experimental cohorts can be expanded by serial transplantation, the resulting tumours become successively less representative of the original setting [Day et al., 2015].

Genetically Engineered Models Another possibility of *in vivo* tumour study is offered by genetically engineered mouse models (GEMMs), in which techniques manipulating the mouse genome are exploited to introduce driver mutations into specific tissues. In general, these models provide a good approximation of cancer development as tumours can be followed from initiation through progression while being immersed in intrinsic stroma and a working immune system. If treatment is administered, evolution under potential treatment-induced selection pressure can additionally be studied. Both germline and non-germline GEMMs are available.

Germline GEMMs are mouse lines which are engineered to harbour a genetic modification such as an overexpression of oncogenes or the silencing of tumour suppressors in the germline. These modifications may be constitutive or inducible (for example via the Cre-LoxP system) and are passed on through breeding. Examples of mouse lines developing glioblastoma or other types of glioma include heterozygous mice carrying null mutations in both TP53 and NF1 [Reilly et al., 2000] and mice carrying a null mutation of TP53, a homozygous deletion of NF1 as well as a heterozygous deletion of Pten [Kwon et al., 2008]. Next to the considerable investment necessary for developing a germline mouse model, inducible systems additionally have the disadvantage of reserving the induction process for triggering driver mutations, thus blocking the possibility to investigate other gene functions this way.

Finally, in non-germline GEMMs oncogenes or silencing constructs for tumour suppressors are delivered into somatic cells by viruses. In this way, delivery can be site-specific, and multiple genes can be introduced into the host genome at once. In glioblastoma, approaches using viral transduction have been successful. A method using the RCAS/Tva-system is employed in

the experimental work underlying this study and is further described in section 6.3.1. Alternatively, DNA can be transfected non-virally, an approach which has been used increasingly since the introduction of CRISPR/Cas9-based tumour models. Here, double-strand breaks induced at specific genomic sites can be repaired using genetic material from a donor vector. Examples of mouse model tumours induced this way include myeloid malignancies [Heckl et al., 2014], liver tumours [Xue et al., 2014] and also glioblastomas [Zuckermann et al., 2015].

6.1.3. Cancer Stem Cells in Glioblastoma

Stem cells in general are defined as undifferentiated cells which are able to self-renew as well as to produce more differentiated cell types [Tannishtha et al., 2001]. In order to fulfil these tasks, stem cells can make use of symmetric as well as asymmetric division [Morrison and Kimble, 2006].

Adult Stem Cells While embryonic stem cells eventually give rise to all of the several hundreds of cell types in the human body, adult stem cells are more restricted in their possible fates. In contrast to embryonic stem cells, they are found in the body throughout the entire life of an organism and take care of replenishing tissue populations that are lost due to routine turnover. In addition, they are often capable of reconstituting damaged tissue by symmetric division and thus self-multiplication in the case of damage. A tissue may contain more than one type of stem cell as defined by the landmark capabilities of being self-renewable and giving rise to (several) differentiated cell types [Goodell et al., 2015]. Starting from the 1980s, one or several types of tissue-specific stem cells were identified, isolated and characterised in a variety of organs including hematopoietic system [Spangrude et al., 1988], intestine [Bjerknes and Cheng, 1999] and epidermis [Oshima et al., 2001, Blanpain et al., 2004]. In these examples as well as more generally, maintenance of stemness as well as stem cell number homeostasis are tightly linked to the existence of a stem cell niche, a cellular microenvironment nurturing, protecting and commanding its stem cell inhabitants [Moore and Lemischka, 2006].

Normal Neurogenesis In the mammalian brain, first evidence for continued production of neurons and astrocytes throughout adult life was uncovered in 1992, overwriting the previous belief that neurogenesis in the mammalian central nervous system came to an end in infancy [Reynolds and Weiss, 1992]. It has since been found that adult neurogenesis takes place in restricted brain regions, notably the subgranular zone (SGZ) in the dentate gyrus which produces dentate granule cells and the subventricular zone (SVZ) which generates neurons that subsequently migrate to the olfactory bulb [Doetsch et al., 1999, Ming and Song, 2011]. In the SVZ, radial glia-like cells (also called B cells in this context) lie at the apex of the stem-cell hierarchy. They give rise to oligodendrocytes as well as to transit amplifying cells (C cells) which in turn produce neuroblasts (A cells). These then migrate to their final destination and become mature neurons, also called interneurons. It is unclear whether astrocytes derive directly from B cells or their transit amplifying progeny [Ming and Song, 2011, Ponti et al., 2013]. Nestin and GFAP are common markers for the radial glia-like stem cell stage whereas proliferating progenitors (C and A cells) are characterised by the expression of Mash1 and Dlx2 [Ming and Song, 2011].

Cancer Stem Cells Following the first discovery of a subpopulation of human acute myeloid cells able to initiate cancer upon transplantation into immunodeficient mice [Lapidot et al., 1994], evidence for populations with stem cell-like capabilities in the cancer context have been identified in a vast variety of malignancies including breast, colon, lung, liver and pancreatic tumours [Pattabiraman and Weinberg, 2014]. Therefore, the general existence of tumour subpopulations which are able to faithfully reconstitute complete “tumour tissue” (for example following transplantation into a new host) while others fail to do so is no longer a matter of active debate. However, this does not rule out cases in which many or all of the tumour cells can function as tumour stem cells in this sense [Goodell et al., 2015]. What the tumour hierarchy looks like in each specific case and what fraction of cells carries malignant stemness depends on the original phenotypic hierarchy of the tissue, its dynamics and most importantly on how this hierarchy was disturbed, i.e. which normally regulated process was uncaged by the malignant transformation [Hanahan and Weinberg, 2000, Hanahan and Weinberg, 2011].

In the context of the “cancer stem cell dispute” [Pattabiraman and Weinberg, 2014] it is specifically important to point out the difference between a cancer stem cell (CSC) as introduced above and the cell-of-origin of a cancer - the latter describes the cell in which the initial transformation event occurred. This cell’s population within the normal context may, but need not be, equivalent to the corresponding population in the cancerous hierarchy [Rycaj and Tang, 2015]. The number of stem cell divisions in a given tissue over a lifetime has recently been found to be a strong correlate of the risk of developing cancer in this specific tissue [Tomasetti and Vogelstein, 2015], thus suggesting that original malignant events in the stem cell compartments are common.

An important although somewhat trivial consequence of the identification of stem-like subsets in some cancers is that genetically identical cancer cells exist in two or more alternative phenotypes: CSCs and their progeny. It follows that these phenotypes may display diverging behaviour, for example differential expression of signalling pathways. While this fact appears to be at the basis of treatment escape of CSC populations in many cases, it also opens research directions for the specific targeting of the self-renewing subpopulation [Cojoc et al., 2015].

Brain Cancer Stem Cells Initial evidence for a stem cell-like population in brain tumours based on self-renewal and sphere-forming assays appeared in 2003. The surface marker CD133 was found to discriminate between sphere-forming and non-sphere-forming cells [Singh et al., 2003]. In a follow-up study, isolated CD133⁺ glioma cells displayed the ability to initiate tumour growth when transplanted into immunodeficient mice while CD133⁻ cells were unable to do so. The tumours developed from the former exhibited histopathological similarities to human brain tumours [Singh et al., 2004]. The previously mentioned neural stem cell marker Nestin has likewise been found to differentiate between distinct glioblastoma cell populations: using a glioma-prone mouse line subjected to chemotherapeutic treatment with temozolomide (TMZ) when tumours had developed, [Chen et al., 2012] showed that the population reconstituting tumour growth following initial proliferation arrest was expressing Nestin. They hypothesize that the enhanced survival of this population is due to its relative quiescence. Next to CD133 and Nestin, the nuclear receptor Tailless (Tlx or NR2E1) which is expressed in the mouse brain

during development [Monaghan et al., 1995] is also able to discriminate between functional subsets in glioblastoma [Zhu et al., 2014].

Glioma stem cells share surface markers as well as important signalling pathways with normal neural stem cells suggesting a close relationship [Sandberg et al., 2013]. However, it has also been shown using *in vitro* reprogramming that it is possible to turn differentiated glioblastoma cells back into tumour-propagating cells by activating a set of four neurodevelopmental transcription factors [Suvà et al., 2014]. Thus, even though a cancer stem cell origin in neural stem cells may seem manifest, it can currently not be excluded that they physiologically originate from more differentiated cells that have undergone reprogramming. In any case, a variety of pathways which are differentially expressed in glioblastoma stem cells compared to neural stem cells have been identified [Purow et al., 2005, Bruggeman et al., 2007, Bleau et al., 2009, Sandberg et al., 2013]. An emerging pattern from these studies seems to be the longer term or even permanent expression of stem cell activating factors which are normally only transiently produced, thus driving stem cells into excess proliferation.

6.2. Mathematical Models of Cancer

In this section, a rough overview of mathematical modelling approaches to specific aspects of cancer is given. Section 6.2.1 introduces various models dealing specifically with glioblastoma, while sections 6.2.2 and 6.2.3 deal with general questions of tumour modelling. In section 6.2.2, spatial aspects of solid tumours and approaches to model these are introduced; section 6.2.3 briefly discusses published theoretical work on cancer stem cells.

6.2.1. Models of Glioblastoma

As will become apparent in the following, glioblastoma growth has been modelled using a wide array of mathematical methods and approaches including ordinary differential equations (ODEs), partial differential equations (PDEs), different simulation approaches, hybrid models and stochastic processes. However, a vast majority of published work falls into one of two categories: (1) purely theoretical studies in which (mostly) complicated models building on a high number of assumptions are developed and resulting hypotheses are at most qualitatively compared to data and (2) data-driven studies which aim at optimising treatment schedules or patient-specific survival prognoses using simplified descriptive models which are not always mechanistically motivated, thus leading to little biological insight.

Theoretical Exploration In the first category, recent investigations have focused on tumour glioblastoma invasiveness and angiogenesis as well as pharmacokinetic aspects of treatment delivery [Alfonso et al., 2017]. One simulation study investigated the avascular phase of tumour development including blood perfusion, vessel dilation and the distribution of chemical substances such as oxygen and growth factors thus allowing to examine impact of the location and local environment of the initial lesion [Cai et al., 2015]. In another purely theoretical study analysing the response of glioblastoma stem cells to different mathematically modelled

treatment strategies, their enhanced resistance was attributed to alleged transdifferentiation into vascular endothelial cells [Yan et al., 2017]. [Boujelben et al., 2016] build a model of the tumour environment including blood flow, vascular permeability and diffusion to examine whether antiangiogenic treatment is beneficial or detrimental for the delivery of drugs e.g. across the blood-brain barrier. Varying treatment efficacy between patients motivates [Alfonso et al., 2016] to develop a computational model built on the assumption that glioblastoma cells either proliferate or migrate. They calculate a potentially patient-specific critical proliferation/diffusion ratio which separates glioma into two differently treatable fractions. None of these examples produce testable predictions and thus illustrate a difficulty of this field.

Treatment Optimisation and Growth Predictions The fact that model results need not necessarily be correct or useful is illustrated by the first example from the second category, in which the aim is to optimise treatment strategies. Built on a model incorporating toxicity data from clinical trials, drug efficacy information from *in vitro* studies and diffusive glioblastoma growth, [Stein et al., 2018] aim to identify optimum lapatinib dosing schedules. While the theoretical result suggests that continuous dosing is superior to pulsatile administration, the authors also find that the dosage needed for tumour control is far beyond toxicity limits. In a more successful example, [Leder et al., 2014] have optimised radiation dosing schedules in experiment-theory iterations using a mouse model and a mathematical model with treatment-susceptible differentiated and more protected stem-like cancer cells. They were able to identify schedules leading to superior survival of irradiated mice. Furthermore, [Wang et al., 2009] have introduced a model allowing to deduct two key parameters characterising proliferation and invasion respectively from routinely obtained magnetic resonance images of individual patients. This characterisation allows for patient-specific growth predictions which can help schedule control examinations. There is a growing role for this type of model in the clinical setting [Jackson et al., 2015].

Investigation of Underlying Mechanisms Finally, mathematical modelling approaches are being used to examine cellular mechanisms underlying glioblastoma development, progression and growth. This includes both genomic evolution and heterogeneity and phenotypic heterogeneity. Based on glioblastoma data from *The Cancer Genome Atlas* project, [Cheng et al., 2012] have developed a mathematical method for determining the temporal order of pathway alterations during gliomagenesis. In a similar spirit, [Sottoriva et al., 2013] analyse several spatially distinct tumour samples from each of 11 glioblastoma patients and reconstruct their phylogeny. In addition to the identification of general early and late mutation events, this revealed patient-specific patterns of cancer evolution and heterogeneity. Recently, intratumoural functional heterogeneity between glioblastoma cells has been experimentally scrutinised for the first time using a lineage-tracing approach with barcoded human glioblastoma cells in immuno-deficient mice [Lan et al., 2017]. This study has shown that glioblastoma retain a proliferative hierarchy reminiscent of the normal system with comparably slow-cycling stem cells at the apex using the critical birth-death process as a model. This approach also allowed for the conclusion that faster cycling glioblastoma progenitors have extensive self-maintenance capabilities before giving rise to unproliferative terminal cells. Further studies into both gen-

omic and phenotypic heterogeneity in glioblastoma are needed to increase the knowledge base on which both theoretical explorations and treatment optimisation models as described above can draw.

6.2.2. Models of Spatial Aspects of Solid Tumour Growth

In the past 20 years, spatial models of solid tumours have received growing interest from mathematical and computational biologists with or without direct clinical application in mind. Mathematical models which explicitly take the spatial structure of tumours into account are set up to address diverse areas of interest such as early tumour development, vascularisation and angiogenesis, invasion of surrounding tissue and the response to varied treatment attempts. Recently, these models are increasingly being applied to patient-specific data and finding their way into the clinics to help determine personalised treatment strategies [Karolak et al., 2018]. In general, truly spatial models either build on agent-based stochastic simulations or sets of coupled partial differential equations (or a hybrid of the two). Here, a third related category, “pseudo-spatial models” are included, which treat spatial dimensions implicitly rather than explicitly.

Pseudo-Spatial Models The cultivation of tumour spheroids *in vitro* has long been an important part of cancer research as they reproduce features of *in vivo* tumours more faithfully than monolayer cell cultures. While spheroids start out as a small undifferentiated cell mass, characteristic features such as a necrotic core, a quiescent layer around this core and a proliferating outer shell quickly become apparent before the spheroid, in many cases, ceases to grow entirely. Thus, in a simplified approach at describing spheroid spatial structure, three coupled ODEs governing proliferating, quiescent and necrotic cells are set up. [Wallace and Guo, 2013] analyse a comprehensive set of specific modelling choices within this framework and show that a model including a growth term proportional to the surface area can explain many experimental observations. Importantly, the total tumour size as a function of time can adopt functionally different forms depending on modelling assumptions. This is of clinical relevance, as even today simple growth models are often used to predict cancer progression in the clinic. Common choices for ODE models for this task include exponential, logistic, linear, surface-proportional and Gompertzian relations which may result in differences of predicted clinical doubling times of the tumour volume as large as one order of magnitude [Murphy et al., 2016].

Agent-Based Models In agent-based modelling, each cancer cell is treated as an individual computational object acting either on a grid or a differently structured space. The approach thus allows to include both heterogeneous cell populations and cell-to-cell diversity. Because of these advantages it has been widely used in computational cancer research [Wang et al., 2015c], each model being set up with a specific research question in mind. [Enderling et al., 2009] analyse the effect of migration rates on tumour growth in two and three dimensions and find that migration is pivotal for tumour growth and progression, as migrating cells spread into the surrounding tissue to form micrometastases. [Sottoriva et al., 2010] focus on the role of cancer stem cells in a spatial setting and conclude that the joint presence of stem-like can-

cer cells as well as more differentiated cells leads to higher spatial heterogeneity and resulting therapy resistance. In an attempt to combine three-dimensional modelling with mutational progression, [Waclaw et al., 2015] develop an evolutionary simulation which predicts that cellular dispersal and turnover limit genetic heterogeneity. They, too, find that migration is crucial for fast expansion and suggest clinically targeting short-range cellular migratory activity to improve survival. Agent-based models have also been used to examine tumour metabolism and treatment schedules: [Carmona-Fontaine et al., 2017] show in an experiment-theory collaboration how altered metabolite concentrations in the tumour microenvironment affect the behaviour of invading immune cells while [Gallaher et al., 2018] use an off-lattice agent-based model to find that modulating the dose of anti-proliferative drugs leads to efficient growth control.

PDE Models and Hybrids Tumour models building on partial differential equations precede the rise of complex agent-based models in time because of the high computational power necessary for the execution of the latter. However, as the following sample of studies shows, agent-based models have by no means replaced PDE-based approaches as both strategies continue to coexist and intermix. In an early reaction-diffusion model of cancer invasion, [Gatenby and Gawlinski, 1996] let normal and malignant cells compete in a spatial setting. Tumour cells produce an acidic medium from which normal cells are not protected, thus allowing a travelling wave of tumour cells to proceed into surrounding tissue. The spheroid-modelling problem discussed in the paragraph on pseudo-spatial models was picked up by [Sherratt and Chaplain, 2001] who formulated it in terms of continuum densities of proliferating, quiescent and necrotic cells. More recently, [Hawkins-Daarud et al., 2013] extended a previously developed continuous glioma model to include edema formation during anti-angiogenic therapy. Comparing their results to magnetic-resonance images they conclude that anti-angiogenic treatment may have little effect on the growth dynamics of the overall tumour cell number. [Swan et al., 2017] have developed a patient-specific anisotropic diffusion model which allows to predict spatial tumour spread based on a diffusion tensor imaging, a novel MRI-based imaging technique. They apply their efforts to 10 tumours and find increased prediction accuracy compared to the commonly employed Proliferation-Infiltration model [Swanson et al., 2000]. [Swan et al., 2017] also outline how their results could be applied clinically in the future: using their model, regions where most invasion has occurred could be identified thus enabling informed decisions where the clinical target volume for irradiation or excision should be extended the furthest.

6.2.3. Models of Cancer Stem Cells

This section will briefly focus on compartmental mathematical models including cancer stem cells. Of course, spatial modelling approaches also allow the inclusion of several distinct cellular phenotypes and indeed some of the models introduced in section 6.2.2 have stem cell agents (e.g. [Sottoriva et al., 2010]). While cancer stem cells were initially believed to constitute only a small fraction (i.e. on the order of 1%) of the tumour bulk, later observations showed that this is not always the case [Enderling, 2015]. Compartmental models thus explore mechanisms leading to certain ratios between stem and non-stem tumour cells which include phenotype-specific kinetics and feedback loops introducing interactions between phenotypic

populations.

Compartmental Models Most compartmental models of stem-cell driven cancers incorporate three major cellular phenotypes or minor variations of these - stem cells, transit amplifying cells and differentiated cells. This is true for all models introduced in the following. [Ganguly and Puri, 2006] set up this same hierarchy for both normal and abnormal cells and allow cells within the two upstream compartments to transit from normal to malignant via mutations. They find that mutations within the stem cell compartment are more significant for the initiation of cancer than mutations within early progenitor cells. [Molina-Peña and Álvarez, 2012] place the (now overcome) assumption of small stem cell fractions at start of their study. They allow for several proliferating progenitor phenotypes with decreasing potential in their model and derive relationships between model parameters dictated by the assumption that the cancer stem cell fraction should remain below 1% of the overall cell number. Interestingly, the previously published work by [Johnston et al., 2010] already shows that this assumption is not justifiable. Based on a model comprising the same three phenotypes as above and no specific feedback loops between them they describe how cancer stem cells can make up any fraction of the tumour thus governing its aggressiveness. [Weekes et al., 2014] introduce an age structured model which allows for differentially capable progenitors. They show that the overall expansion rate of the tumour in the exponential growth phase is governed by the symmetric division rate of the stem cell compartment alone. As previously discussed, experimentally observed tumour growth can be described by a variety of different mathematical functions and Gompertzian growth is often identified. Because of this, [Liu et al., 2013] investigate how a three-compartment model can be refined such as to exhibit Gompertzian behaviour. They conclude that negative feedbacks proportional to the number of differentiated tumour cells on both the division rates of stem cells and proliferating progenitors and their self-renewal probabilities are required to control population balance. Overall, while mathematical properties of simple compartmental models have been studied excessively, research aiming at the estimation of their parameters is scarce.

Identifying Topology by Lineage Tracing In recent years, lineage-tracing and genetic labelling techniques have been increasingly used to investigate the hierarchical dynamics of stem and progenitor cells. Often, these data require advanced statistical approaches to unravel their full informative potential [Blanpain and Simons, 2013]. This combination of experimental and theoretical techniques has not only been used in developmental and homeostatic contexts, but also in cancer. [Driessens et al., 2012] have applied it to identify lineages in squamous skin tumours via an unbiased genetic labelling strategy. They do not only identify a stem-cell like population but derive that it cycles twice as fast as downstream cells. In a slightly different experimental approach again in skin tumours, [Sánchez-Danés et al., 2016] were able to initiate tumour growth by malignant transformation of stem cells but not progenitors. Transformed stem cells were characterised by increased symmetric divisions. Lately, [Lan et al., 2017] have used lentiviral barcoding to examine the lineage topology in glioblastoma as outlined above. Reasoning via stochastic processes, they recovered the three basic cellular phenotypes so prominently featuring in the compartmental models described above.

6.3. Mouse Models and Experiments Underlying this Work

This section introduces the transgenic mouse models and specific techniques underlying the experimental work on which this theoretical study is built. This includes the non-germline tumour model developed and used by our experimental collaborators, the group of Dr. Hai-Kun Liu at the DKFZ, in section 6.3.1 as well as several useful modifications of this model in sections 6.3.2 to 6.3.6.

6.3.1. Induction of Glioblastoma Growth

The glioblastoma tumours analysed here are induced in the subventricular zone of newborn mice via a procedure using targeted viral transfection of growth-related genes (described by [Zhu et al., 2014]). The avian retroviral vector RCAS (derived from avian leukosis virus (ALV), subgroup A) is engineered to express Akt and platelet-derived growth factor β (Pdgfb). This construct is then administered to a transgenic mouse line expressing the receptor for ALV-A under the promoter of Nestin, a commonly used marker of neural stem cells [Rietze et al., 2001, Kriegstein and Alvarez-Buylla, 2009, Fuentealba et al., 2012]. Akt and Pdgfb are thus exclusively implanted into neural stem cells. [Zhu et al., 2014] show that this system induces high-grade brain tumours which recapitulate important histological features of human glioblastomas. The process is summarised in figure 6.1a.

6.3.2. Co-Expression of Tlx and GFP

The nuclear receptor Tlx has been shown to be expressed specifically by neural stem cells [Liu et al., 2008]. It is important for maintaining their undifferentiated state and governs their proliferation [Shi et al., 2004]. A previously introduced transgenic mouse line features the expression of green fluorescent protein (GFP) under the Tlx-promoter [Feng et al., 2013], thus allowing the identification of stem cells in fluorescence microscopy or fluorescence-based cell sorting. [Zhu et al., 2014] have confirmed that both the usability of Tlx as a marker for stem cells and its functional implications remain valid in the context of the glioblastoma model discussed in section 6.3.1. If the gene encoding red fluorescent protein (RFP) is additionally included in the induction process outlined in section 6.3.1, this mouse model can be used to quantify the fraction of Tlx⁺ cells within the tumour bulk using flow cytometry as shown in figure 6.1b.

6.3.3. Bioluminescence Imaging of Tumour Volume

Bioluminescence imaging describes a technique which allows time series imaging of *in vivo* processes [Contag et al., 1997, Massoud et al., 2003]. The biological structure to be imaged has to express the gene for an oxidative enzyme (Luciferase), which leads to the emission of photons upon the administration of a substrate consumed in the process of bioluminescence. This technique has been used previously in the context of imaging diverse types of mouse model brain tumours, for example [Uhrbom et al., 2004, Shah et al., 2003, Hashizume et al., 2010].

For the BLI data analysed in this study, Luciferase is transfected into Nestin⁺ cells using the RCAS system described in section 6.3.1 (administered in parallel with the RCAS-driven

induction of tumour growth). The photon flux emitted by the tumour-bearing mice is thereafter monitored regularly using a CCD camera. Importantly, tumour-derived light emission has been shown to be highly correlated with tumour volumes measured using magnetic resonance imaging (MRI) [Szentirmai et al., 2006] justifying its use as a proportional proxy of cancer cell number in this study. The resulting bioluminescence time series of an exemplary animal is shown in figure 6.1c.

6.3.4. Viral dsRed-Labeling of Proliferating Progeny

Via the administration of a retrovirus carrying the gene for dsRed, a red fluorescent protein, actively dividing tumour cells and their progeny can be labelled genetically. We have found that these viruses almost exclusively enter proliferating Tx^- cells, as co-labelling of Tx^- cells (as described in section 6.3.2) with dsRed has never been observed. This is thus a useful method for labelling all Tx^- cells (and their potential progeny) which are actively proliferating at a given time point without having to sacrifice the animal (this in turn is necessary for staining for proliferation markers such as Ki67 or GFAP).

6.3.5. TAM-Induced Expression of Rosa26-Brainbow Colours

Using a transgenic mouse line combining the modifications described in sections 6.3.1 and 6.3.2 with stem-cell specific inducible genetic labelling allows tracing the fate of individual cancer stem cells and their progeny in glioblastoma. Here, inducible genetic labelling is enabled via the transgenic expression of CreERT2 under the Tx -promoter. CreERT2 is a fusion of the site-specific recombinase Cre and a mutated hormone-binding domain of the estrogen receptor. Cre itself catalyzes recombination between specific DNA recognition sites (called loxP sites) and can thus be used for excision or inversion of the loxP-flanked DNA segment [Feil et al., 2009]. In turn, CreERT2 recombinases are *per se* inactive but can be activated by administration of tamoxifen (TAM, a synthetic estrogen receptor ligand) [Feil et al., 1997]. Thus, Cre-driven recombination can be transiently activated in cells expressing the CreERT2 gene - as CreERT2 is expressed under the Tx -promoter in our mouse model, here this applies to cancer stem cells only.

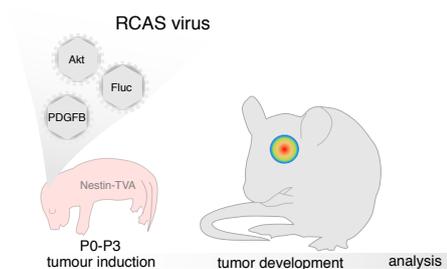
Furthermore, the DNA segment flanked by loxP sites in this labelling approach contains a Brainbow-2.1 cassette [Livet et al., 2007] under the CAGG-promoter. Upon Cre/loxP-recombination, a stop cassette is excised from the Brainbow segment and the expression of one of four fluorescent proteins (nGFP, YFP, RFP or mCFP) is started stochastically [Scheper et al., 2012]. Tx -expressing cells are thus labelled genetically and pass on the label to their progeny upon division. Importantly, the expression of the fluorescent protein is independent of the Tx -promoter, so that Tx^- progeny of the initially labelled cell can nevertheless be identified via their fluorescent colour. As the Cre recombinase is active only transiently following administration of tamoxifen, no further inversions of the colour cassettes happen after TAM has disappeared.

With the aim of allowing unambiguous identification and quantification of clones (meaning both Tx^+ and Tx^- progeny derived from one initially labelled Tx^+ cell) by fluorescence

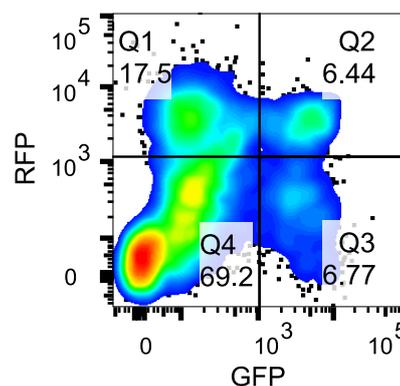
microscopy of brain tumour slices, the dose of TAM was chosen such as to ensure a sparse initial labelling thus minimising the risk of clonal overlap at later time points. Four exemplary clusters of brainbow-labelled glioblastoma cells are depicted in figure 6.1d.

6.3.6. TAM-Induced miRNA Knock-Down of Tlx and Switch of Fluorescent Protein Cassettes

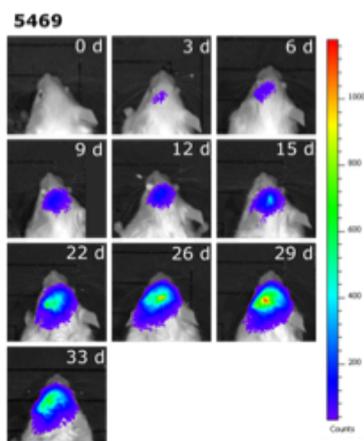
A disadvantage of the Cre-loxP system for gene knockout as described in section 6.3.5 is that for each specific gene a transgenic mouse line has to be generated, thus making the process costly. In an alternative approach, an irreversible Cre-driven flipping knockdown system (miR-Flex) allowing flexible targeting of different genes can be used [Stern et al., 2008, Koo et al., 2012]. This structure contains a forward dsRed sequence together with an inverted GFP and a likewise inverted artificial microRNA-coding sequence between opposing loxP-sites. Upon Cre-driven recombination, the entire sequence is inverted: originally red cells then turn green and start expressing the microRNA which is now in forward orientation. The entire construct is assembled using plasmid cloning and delivered to target cells using the retroviral system introduced in section 6.3.1. Using a microRNA targeting Tlx, a knockdown efficiency of 80% could be reached using this approach [Zou, 2016]. In combination with the transgenic expression of CreERT2 under the Tlx-promoter, the colour switch can be triggered specifically in Tlx⁺ cells thus marking stem cells (and their recent progeny).



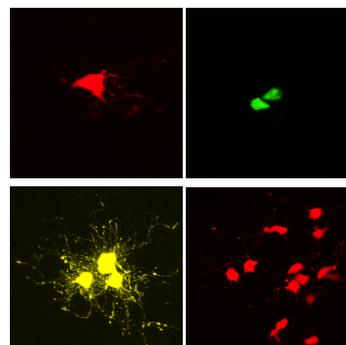
(a) Induction procedure of model glioblastoma in mice as described in section 6.3.1. Figure: Yue Zhuo.



(b) Populations of stem (RFP^+GFP^+) and non-stem (RFP^+GFP^-) tumour cells after flow sorting of a dissociated tumour with characteristics as described in section 6.3.2. Data: M. Amir Khan.



(c) Exemplary time course of tumour growth as shown by bioluminescence imaging, see section 6.3.3. Figure: Peng Zou.



(d) Four examples for brainbow clusters of glioblastoma cells, presumably derived from one initially labelled cell, see section 6.3.5. Figure: M. Amir Khan.

Figure 6.1.: Illustration of selected experimental procedures.

7. Exponential Tumour Growth and Cellular Migration

Before analysing individual cellular populations within the tumour mass as well as their dynamics and roles in chapter 8, we take a coarser look at the system. As described in section 7.1, we observe exponential tumour growth in untreated mice. While this may not seem very surprising at first glance, section 7.2 shows that such a growth behaviour does not trivially come about in a spatial setting. Exponential growth has direct implications for cellular behaviour which will play important roles later on in chapter 9.

7.1. Phenomenological Modelling of Tumour Growth Curves

Figures 7.1 to 7.3 show the bioluminescence imaging (BLI) curves recorded from 23 glioblastoma-bearing animals over periods of several weeks using the mouse model introduced in section 6.3.3. The BLI values given on the y -axes therein correspond to the flux of photons emitted per second by the tumour in arbitrary units. We are interested in the dynamics of this growth phase and also want to make use of it later on for estimating parameters of glioblastoma growth in section 8.3. However, because the BLI data points from the individual animals were not taken at regular time intervals and the signal strengths vary strongly in magnitude (due to different tumour positions within the brain, varying transmission characteristics of mice of different colours and other animal-specific factors), deriving a meaningful average from the set of 23 mice is not straightforward.

In order to solve this problem, we start by fitting a number of phenomenological growth curves to each individual's tumour volume as a function of time. A wide range of mathematical models have been proposed for modelling tumour growth both *in vitro* and *in vivo* without a consensus regarding a best choice of equation having emerged across studies or diseases. Commonly employed models include exponential, Gompertzian and power-law growth [Wallace and Guo, 2013, Talkington and Durrett, 2015, Murphy et al., 2016]. Here, we have fitted these three models to the 23 experimental time series in figures 7.1 to 7.3 using the Nelder-Mead method.

In the simplest case, the exponential model, growth is governed by rate λ_0 ,

$$N_{\text{exp}}(t) = N_0 e^{\lambda_0 t}, \quad (7.1)$$

where $N_{\text{exp}}(t)$ describes the number of cells at time t assuming N_0 cells at $t = 0$. The Gom-

gompertzian growth function is given by

$$N_{\text{Gom}}(t) = K_0 e^{\log(N_0/K_0) \exp(-g_0 t)}, \quad (7.2)$$

where $N_{\text{Gom}}(t)$ is again the number of cells at time t when having started from N_0 cells at $t = 0$. The Gompertzian curve describes a sigmoidal behaviour where growth slows down at the end of the process - this behaviour is mediated by the carrying capacity K_0 . The parameter g_0 is related to the proliferation speed of cells. Lastly, the power-law model incorporates the idea that only the cells on the surface of a solid tumour are able to proliferate, for example because of nutrient constraints in the core. The tumour volume $V_{\text{pow}}(t)$ at time t thus depends on the initial radius of the tumour, R_0 , and the linear growth rate r_0 of the tumour's radius. This results in the function

$$V_{\text{pow}}(t) = \frac{4\pi}{3} (r_0 t + R_0)^3. \quad (7.3)$$

While the power-law model yields significantly worse fit results than exponential and Gompertzian growth curves by measure of comparing the resulting χ^2 values, these two models perform similarly well. However, the Gompertzian model requires an additional parameter as compared to the exponential growth model. Using Akaike's information criterion [Akaike, 1998], which includes a penalty for more complex models, we find that the exponential model is preferable in 18 out of the 23 cases analysed. In the remaining five cases, Gompertzian growth fits the data best. These results suggest that while glioblastoma growth reaches a capacity limit eventually, our study focuses on the phase of tumour expansion where growth is still exponential.

In order to describe the mean dynamics of this exponential growth phase, we take the mean of the 23 fitted values of λ_0 resulting in

$$\lambda_0 \pm \sigma_{\lambda_0} = (0.21 \pm 0.02) \text{ day}^{-1},$$

where σ_{λ_0} is the associated standard error of mean. This corresponds to a fast tumour doubling time of $t_2 = \log 2 / \lambda_0 = 3.3$ days.

When fitting each of the three models in equations (7.1), (7.2) and (7.3) to the data, the initial values N_0 (respectively R_0) were allowed to vary freely in \mathbb{R}^+ in addition to the model parameters of interest. This absorbs much of the animal-to-animal variability which may be due to different numbers of initially transformed malignant cells or the positions of the tumour in the brain with respect to the skull and measurement devices. It is likewise for this reason that their values carry no information accessible or relevant to the current discussion.

7.2. Three-dimensional Tumour Growth Simulation

In order to better understand the characteristics and constraints of biological glioblastoma progression, we implemented a computational tumour growth simulation on a three-dimensional Cartesian grid. This simulation comprises a self-renewing cell type (type *A*) and a derived cell type which is unable to divide (type *B*) as well as behavioural rules associated with cell types

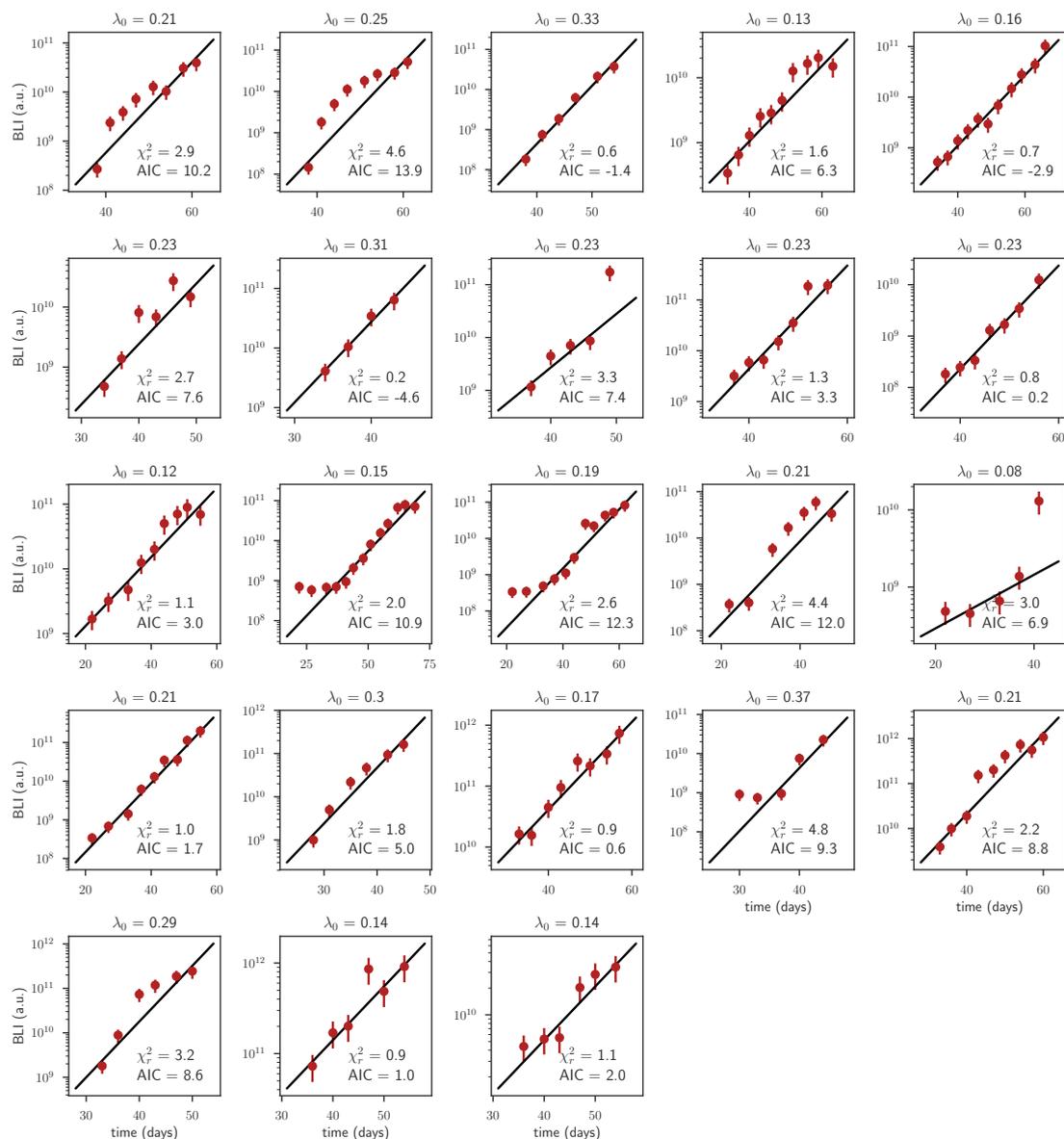


Figure 7.1.: Time course of glioblastoma growth in 23 individual animals as shown by the increase of bioluminescence over time after birth (and induction of tumour growth) in days. Red dots show measured BLI values with an assumed relative error of $1/3$, black lines show best fits of the exponential model in equation (7.1). The resulting best fit values of λ_0 are given above each individual plot in units of day^{-1} . Experimental data: Peng Zou.

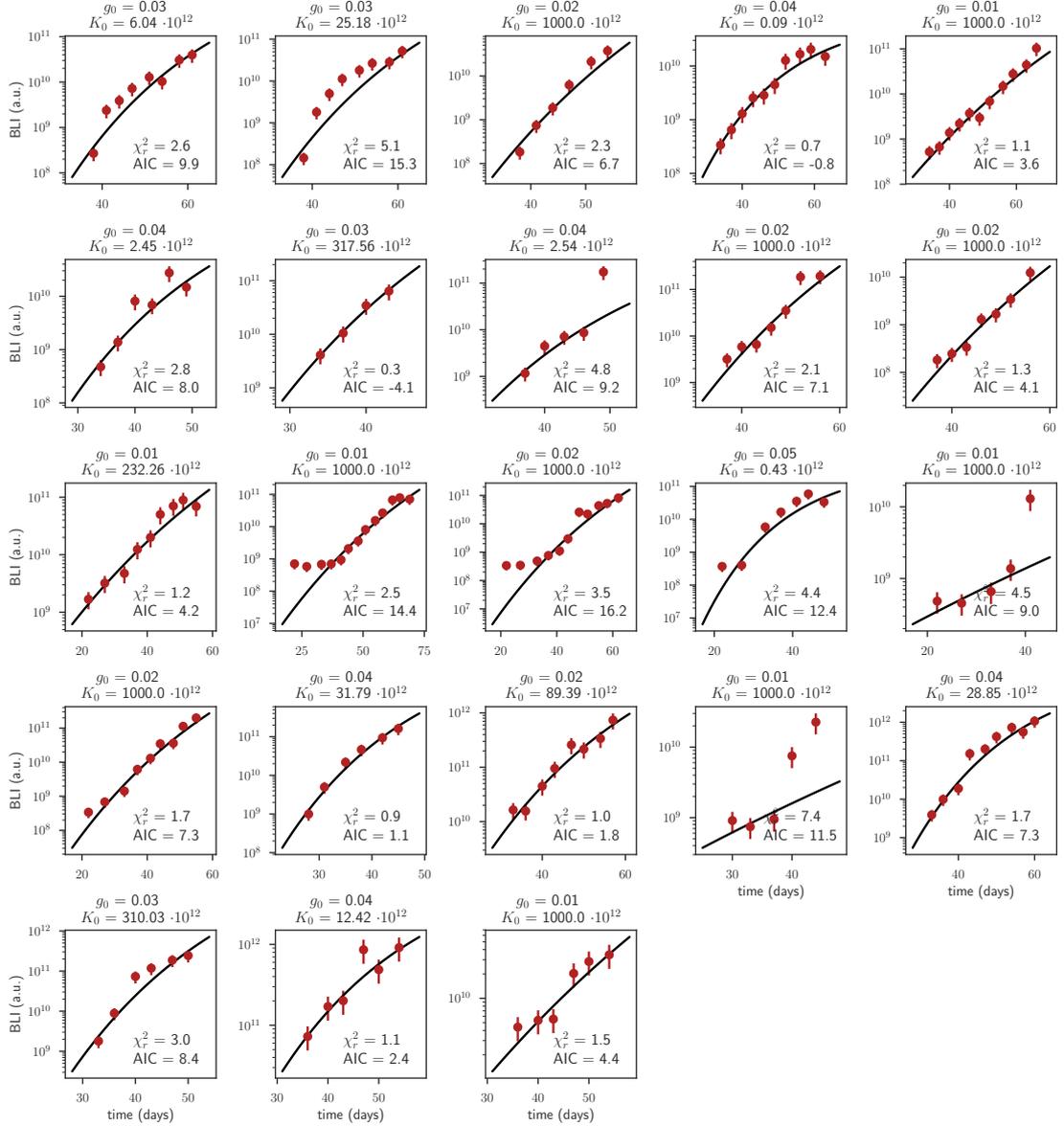


Figure 7.2.: Time course of glioblastoma growth in 23 individual animals as shown by the increase of bioluminescence over time after birth (and induction of tumour growth) in days. Red dots show measured BLI values with an assumed relative error of $1/3$, black lines show best fits of the Gompertzian model in equation (7.2). The resulting best fit values of K_0 (without unit) and g_0 (units of day^{-1}) are given above each individual plot. Experimental data: Peng Zou.

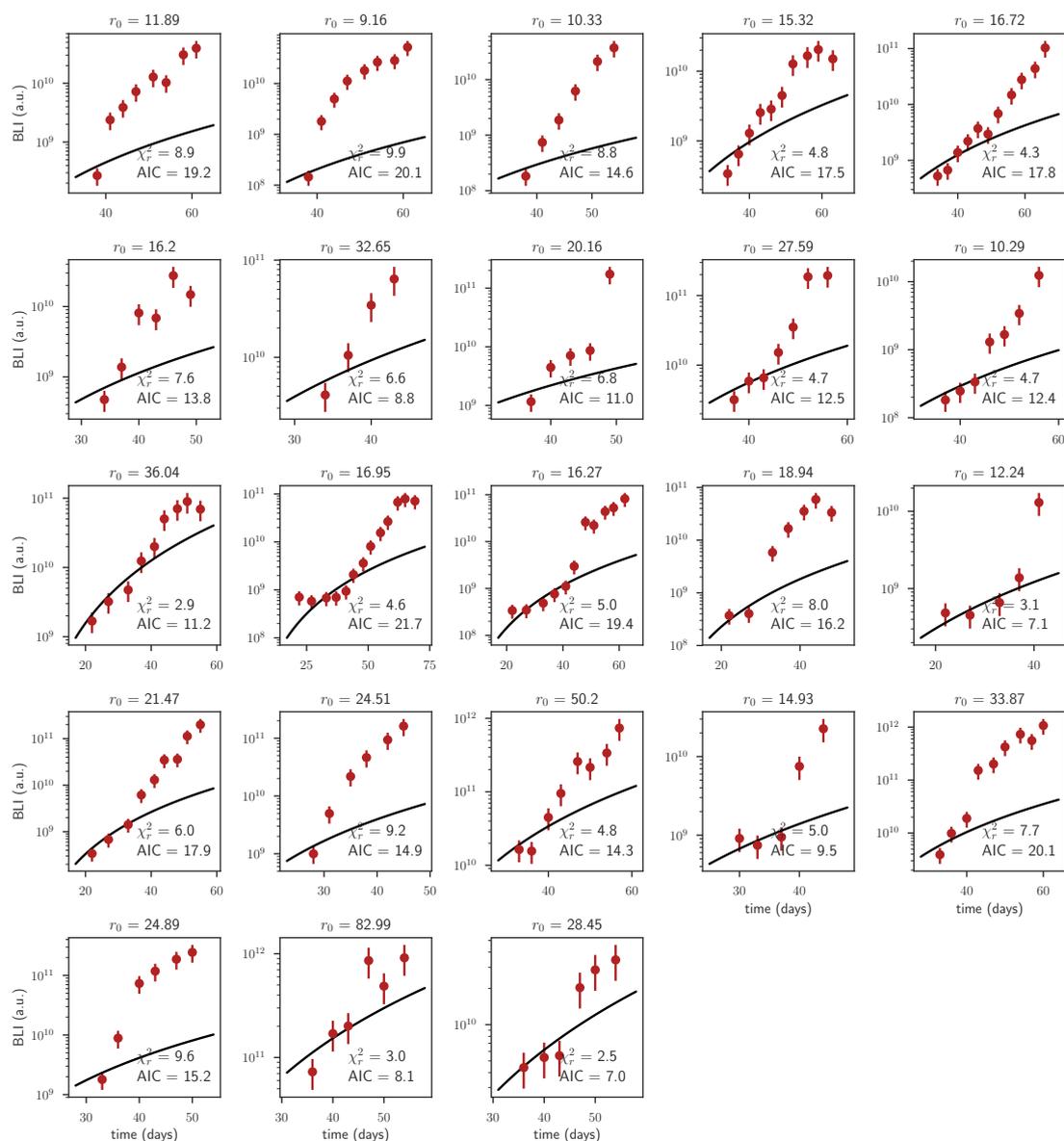


Figure 7.3.: Time course of glioblastoma growth in 23 individual animals as shown by the increase of bioluminescence over time after birth (and induction of tumour growth) in days. Red dots show measured BLI values with an assumed relative error of $1/3$, black lines show best fits of the power-law model in equation (7.3). The resulting best fit values of r_0 are given above each individual plot. Experimental data: Peng Zou.

A and B .

Self-renewing cells of type A can divide symmetrically ($A \rightarrow 2A$, rate λ_S) or asymmetrically ($A \rightarrow A + B$, rate λ_A) while cells of type B are not proliferative. Cells of type B are lost from the system via cell death while cells of type A live infinitely in this scenario. In addition, one or both of the cell types can be allowed to migrate around the grid. Different rules governing these migration processes are conceivable, and some options will be discussed below. Figure 7.4 gives an overview of the modelled processes in this simplified scenario; the algorithm in pseudo code below it summarizes the main logics of the corresponding python simulation which relies on Gillespie’s stochastic simulation algorithm [Gillespie, 1977].

Importantly, in our simulation, cells are only allowed to divide if at least one of their 26 Moore neighbours is empty. We have chosen this assumption over the competing “expanding universe”-type growth (in which cells are always able to divide and all surrounding cells get pushed outwards by the newborn daughter) because of the biological notions of nutrient scarcity at crowded interior tumour sites and cell-to-cell contact inhibition of replication [Yu et al., 2015]. Other recent grid-based simulations likewise operate on the space-limited assumption [Waclaw et al., 2015]. During migration, we do allow cells to move to grid sites which are already occupied. However, as a consequence, cells on overcrowded grid points are no longer able to divide until one of the occupants has left the site.

With this basic simulation setup at hand, we are now able to test effects of different migration scenarios on the resulting growth curves. As baseline, we start with a scenario in which neither cell type can migrate, resulting in a situation where only cells on the outer shell of the tumour can divide. As figure 7.5a shows (grey line), this scenario does not lead to sustained exponential growth. In fact, because only the outer shell can proliferate, the functional form of the resulting growth curve resembles what is described in equation (7.3), the power-law model.

As a next step, we allow cellular migration which is a known feature not only of the developing but also the adult mammalian brain. However, in the adult case, migration out of the subventricular zone (SVZ) is mostly performed by fate-restricted progenitors of neural stem cells, which themselves reside in the SVZ [Cayre et al., 2009]. Thus, in the simulated scenario, we begin by allowing differentiated cells of type B to migrate. We assume that cells move towards the outside of the tumour bulk, as this is where cell density decreases and nutrient availability improves. The direction for the simulated moves is set to the unit vector between the centre of mass of the tumour and the current position of the cell. In order to enforce faster migration and thus dispersal into the area surrounding the tumour, we set cells to take more than one step into the chosen direction. Uniform integer random noise was added to each component of the migration vector independently in order to avoid restriction of movement to the 26 Moore axes only. As shown by the red line in figure 7.5a, although this slightly increases the simulated tumour size at late time points as compared to the scenario without migration, exponential growth is not reached here. Even though the edges of the tumour are more dispersed in this scenario, the basic fact that many self-renewing cells are stuck on the dense inside of the tumour and thus cannot replicate remains intact.

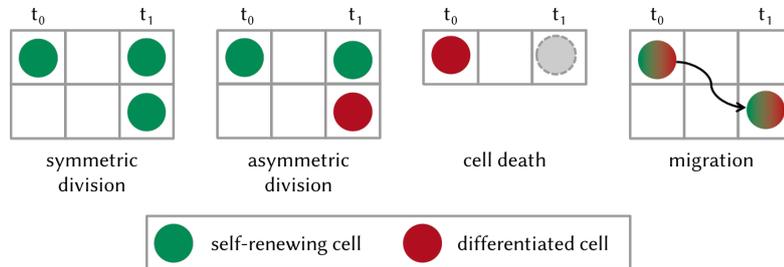


Figure 7.4.: Summary of rules of the three-dimensional tumour growth simulation on a grid as discussed in section 7.2.

Algorithm 2 Simulation algorithm of three-dimensional tumour growth with migration.

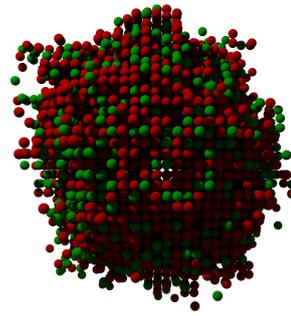
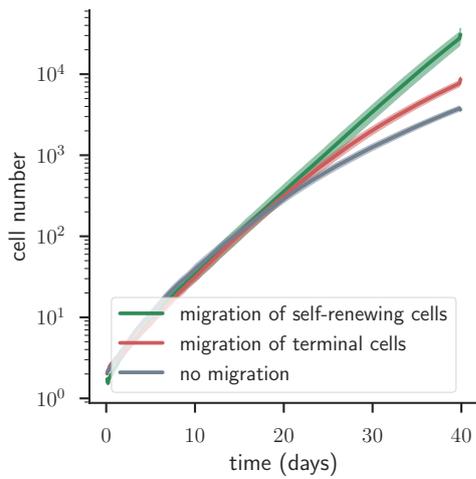
- 1: **function** TUMOUR GROWTH(symmetric division rate, asymmetric division rate, migration rate(s), desired migration behaviour (random or directed))
 - 2: Initialise an empty three-dimensional Cartesian grid.
 - 3: Put one cell of type A into the middle of the grid.
 - 4: **while** $t_{\text{now}} < t_{\text{end}}$ **do**
 - 5: Based on a random number generator, the current cell population in the grid and the rates associated with the possible events, decide which type of event will happen next and after what time increment Δt .
 - 6: **if** event is symmetric division of cell type A **then**
 - 7: Select one cell of type A at random.
 - 8: **if** this cell has one or more empty neighbour sites **then**
 - 9: Choose a random empty neighbour and add a new cell of type A there.
 - 10: **else**
 - 11: Do not divide.
 - 12: **else if** event is asymmetric division of cell type A **then**
 - 13: Select one cell of type A at random.
 - 14: **if** this cell has one or more empty neighbour sites **then**
 - 15: Choose a random empty neighbour and add a new cell of type B there.
 - 16: **else**
 - 17: Do not divide.
 - 18: **else if** event is migration of cell type A (or B) **then**
 - 19: Select one cell of type A (or B) at random.
 - 20: According to the chosen migration behaviour (random migration or migration directed outwards), move the cell to its new position.
 - 21: Update time, $t_{\text{now}} \leftarrow t_{\text{now}} + \Delta t$
 - 22: **return** the current occupancy of the grid with cells of type A and B .
-

Lastly, we allow migration of the self-renewing cells of type A instead. We find that this approach allows for sustained exponential growth, as self-renewing cells migrate away from the tumour shell and thus find a situation similar to the one encountered by the first founder cell - not surrounded by any competing malignant cells. The resulting growth curve is given by the green line in figure 7.5a. The parameter values used for all simulations underlying this figure are as follows: symmetric division of stem cells $\lambda_S = 0.25 \text{ d}^{-1}$, asymmetric division of stem cells $\lambda_A = 0.65 \text{ d}^{-1}$ and outwards migration of both stem and differentiated cells where applicable $\beta = 0.4 \text{ d}^{-1}$. The dispersal factor used is $N_{\text{disp}} = 3$ with uniform random noise $N_{\text{noise}} \in \{-2, -1, 0, 1, 2\}$ drawn separately for each spatial dimension.

This examination of different migrational behaviours shows that if persistent exponential tumour growth is observed *in vivo*, for example via bioluminescence imaging, the underlying process must permit replication of a steady fraction of the cells driving tumour growth. This is specifically not the case if cells are only allowed to proliferate on the surface of a radially growing tumour, as here the surface scales with r^2 whereas the tumour volume scales with r^3 . Two options that do allow for exponential growth are outwards migration at a rate high enough to exhibit dispersal as discussed here or “expanding-universe” growth, which we have argued against above.

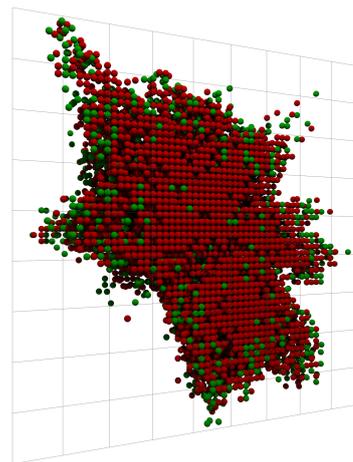
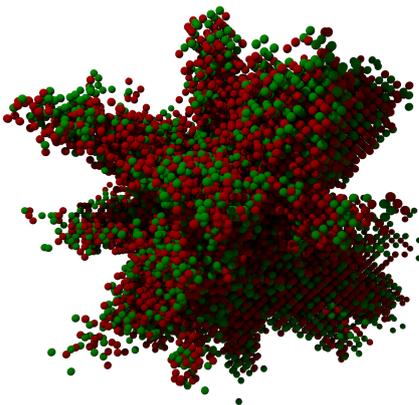
Next, we seek experimental confirmation for the theoretically derived conclusion that self-renewing cells are outwards-migrating. For this, we analyse the spatial structure of simulated tumours. Figure 7.5b shows a tumour where cells were not allowed to migrate. It has a dense, round shape; when cut open (not shown), cells of type A and B are mixed throughout the entire bulk. In comparison, a tumour produced by application of the outwards migration rule is shown in figure 7.5c. Because cells are set to migrate away from the current centre of mass, initial stochastic deviations from a spherical shape lead to the formation of main migrational axes which manifest themselves as fingers. This effect is most likely an artefact of the simulation procedure, as in the experimental setting, physiological structures in the host tissue (here the brain) would likely dominate the detailed directions taken by migrating cells. More importantly however, we observe the accumulation of stem cells on the outer shell of the tumour entity. This becomes more clear when looking at a cut through the tumour as shown in figure 7.5d. Using the mouse model introduced in section 6.3.6, experiments can directly scrutinize this prediction. Figure 7.5e shows a detail of a tumour section where the outside of the tumour is towards the lower right corner, thus confirming the proposed outwards migration of self-renewing tumour cells.

Of interest, Dr. Chunxuan Shao from the Division of Molecular Neurogenetics at DKFZ has analysed recently published RNA sequencing data from human glioblastoma samples [Darmann et al., 2017]. In this study, RNA was sequenced using the Smart-seq2 protocol and tumour samples were separated into core and surrounding tissue parts prior to the analysis. He found that the nuclear receptor Tlx is strongly enriched in the tumour shell compared to the core, indicating that these findings hold true also in the human context.



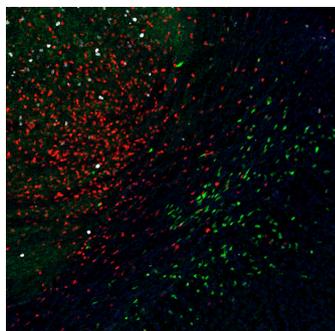
(a) Tumour growth curves for three different migration scenarios including dispersal, see section 7.2. Curves show means of 40 realisations.

(b) A tumour in which cells do not migrate after 50 days of growth.



(c) A tumour in which stem cells migrate outwards after 50 days of growth.

(d) As in part c, but cut in half to show interior structure.



(e) Cut through a mouse-model glioblastoma: stem cells and their recent progeny are labelled in green, differentiated cells in red.

Figure 7.5.: Tumour growth outcomes for different simulated migration rules.

8. Model Development and Calibration

This section introduces the mathematical model used for a joint description of several of the experimental data sets described in section 6.3 as well as the Markov Chain Monte Carlo-based framework used for parameter estimation and model-based predictions. The resulting estimates for division, differentiation and death rates of the cell populations constituting the tumour are presented with corresponding confidence regions. Because our mathematical model builds on three cellular subpopulations (stem cells S , dividing progeny P and terminally differentiated progeny D), we refer to our model as SPD-model in the following.

8.1. A Mathematical Model of Glioblastoma Growth

Based on our own earlier observations [Zhu, 2013, Zhu et al., 2014] as well as recently published findings [Lan et al., 2017], we here choose to model glioblastoma growth with a simple mathematical model comprising three cell types: cancer stem cells, their actively dividing progeny and terminally differentiated cells which are no longer able to proliferate. It is noteworthy that this hierarchy mirrors what has been found in the case of non-malignant neurogenesis in the subventricular zone which similarly displays (quiescent and active) stem cells as well as two types of further differentiated cell types with decreasing self-renewal capacity [Ponti et al., 2013].

[Zhu et al., 2014] have shown that only the Tlx^+ subset of glioblastoma cells is able to initiate tumour growth recapitulating the histological features of the primary tumour upon xenotransplantation. Furthermore, they give evidence that not all Tlx^- cells are proliferating, suggesting that these further differentiated cells eventually cease to divide and remain in a terminal state. These findings form the basis for a mathematical model assuming three distinct cell types.

[Lan et al., 2017] have named these cell types S , P and D for stem cells, proliferating cells and differentiated cells respectively and show that a simple model topology as described by figure 8.1 and equation (8.1) is able to qualitatively explain their observations of glioblastoma growth at the clonal level in the context of a critical birth-death process with immigration. Here, we set out to quantitatively determine all rates used in the model.

As illustrated by figure 8.1, we include the following cellular processes into our mathematical analysis: symmetric division of stem cells ($S \rightarrow 2S$) with rate λ_S , asymmetric division of stem cells ($S \rightarrow S + P$) with rate λ_A , symmetric division of progeny ($P \rightarrow 2P$) with rate λ_1 , differentiation of an actively dividing progeny cell into a terminally differentiated cell ($P \rightarrow D$) with rate d_1 and finally loss of terminally differentiated cells ($D \rightarrow \emptyset$) with rate μ_1 . Assuming mass-action kinetics, the dynamics of this system are formalised by the following set of

ordinary differential equations in the limit of large cell numbers,

$$\begin{aligned}
 \dot{S}(t) &= \lambda_S S(t) \\
 \dot{P}(t) &= \lambda_A S(t) + (\lambda_1 - d_1)P(t) \\
 \dot{D}(t) &= d_1 P(t) - \mu_1 D(t).
 \end{aligned} \tag{8.1}$$

Explicit solutions for this system of equations are readily available. Assuming tumour growth to start from one stem cell at time point $t = 0$ (initial conditions $S(0) = 1, P(0) = 0, D(0) = 0$), we find

$$\begin{aligned}
 S(t) &= e^{\lambda_S t} \\
 P(t) &= \frac{\lambda_A}{\lambda_S + d_1 - \lambda_1} (e^{\lambda_S t} - e^{(\lambda_1 - d_1)t}) \\
 D(t) &= \xi \left[(d_1 - \lambda_1 - \mu_1)e^{\lambda_S t} + (\lambda_S + \mu_1)e^{(\lambda_1 - d_1)t} + (\lambda_1 - d_1 - \lambda_S)e^{-\mu_1 t} \right], \\
 \xi &= \frac{\lambda_A d_1}{(\lambda_S + \mu_1)(\lambda_1 - d_1 + \mu_1)(\lambda_S + d_1 - \lambda_1)}.
 \end{aligned} \tag{8.2}$$

Because of the experimental observation by [Zhu et al., 2014] that Tx^- cells (cells of type P and D in the language of the model) cannot in general sustain tumour growth by themselves, we require that cells of type P always be lost via differentiation at least as fast as they divide,

$$\lambda_1 - d_1 \leq 0. \tag{8.3}$$

Revisiting equation (8.2) with equation (8.3) in mind, we see that the second term in $P(t)$ as well as the second and third term in $D(t)$ become irrelevant at later times because their exponents assume negative values for all times. The only remaining terms are governed by the symmetric division rate of stem cells λ_S for all three populations. Defining the overall tumour dynamics $N(t)$ via

$$N(t) = S(t) + P(t) + D(t), \tag{8.4}$$

we thus find the effective growth rate of the whole tumour to be governed only by the symmetric division rate of stem cells for $t \gg 0$,

$$N(t) \propto e^{\lambda_S t}. \tag{8.5}$$

As a further property of this model, the fractional composition of the tumour bulk reaches a steady state dependent on the five rates defining the model even though the tumour cell population grows exponentially for $\lambda_S > 0$. These stationary fractions are given by

$$\begin{aligned}
 \left\langle \frac{S}{N} \right\rangle &= \kappa (\lambda_S + d_1 - \lambda_1) (\lambda_S + \mu_1) \\
 \left\langle \frac{P}{N} \right\rangle &= \kappa \lambda_A (\lambda_S + \mu_1) \\
 \left\langle \frac{D}{N} \right\rangle &= \kappa \lambda_A d_1, \\
 \kappa &= \frac{1}{(\lambda_S + \lambda_A - \lambda_1)(\lambda_S + \mu_1) + d_1(\lambda_A + \lambda_S + \mu_1)}.
 \end{aligned} \tag{8.6}$$

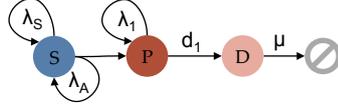


Figure 8.1.: Model of glioblastoma growth consisting of stem cells S , dividing progeny P and differentiated cells D . λ_S and λ_A describe symmetric and asymmetric division rates of stem cells respectively; λ_1 and d_1 give the symmetric division rate of progeny cells and their differentiation speed into terminally differentiated cells; μ_1 gives the death rate of these.

8.2. Simulation of Population Experiments

Here we describe how the experimentally observed quantities concerning the mean behaviour of tumour cells can be calculated using the mathematical model described in the previous section as a basis. The experiments discussed in sections 8.2.1, 8.2.2 and 8.2.3 are used for parameter estimation in this section; further scenarios simulated for purposes of prediction and qualitative comparison are presented in sections 9.1, 9.2 and 9.3.

8.2.1. BLI Growth Curves

The phenomenological growth curve description of equation (7.1) exactly matches the functional form of the mechanistically derived model growth curve of equation (8.5). Thus, in order to describe the BLI tumour growth data adequately with our model, we must seek to minimise the difference between λ_S and the experimental value of λ_0 derived above. Assuming the experimentally measured values λ_0 to stem from a Gaussian distribution, this results in the following likelihood function for the BLI data given the model parameter λ_S :

$$\mathcal{L}_{BLI}(\lambda_0, \sigma_{\lambda_0} | \lambda_S) = \frac{1}{\sqrt{2\pi\sigma_{\lambda_0}^2}} \exp\left(-\frac{(\lambda_0 - \lambda_S)^2}{2\sigma_{\lambda_0}^2}\right). \quad (8.7)$$

Thus, we find for the logarithm of the likelihood

$$\log \mathcal{L}_{BLI}(\lambda_0, \sigma_{\lambda_0} | \lambda_S) = \text{const.} + \frac{(\lambda_0 - \lambda_S)^2}{2\sigma_{\lambda_0}^2}, \quad (8.8)$$

which will form part of the composite log-likelihood function at the basis of parameter estimation in section 8.3.

8.2.2. Decline of proliferation index in the progeny compartment

In this experiment, the development of the proliferative capacity of Tlx^- (progeny) cells was measured over time. To this end, viral vectors carrying the gene for the red fluorescent protein dsRed were administered as described in section 6.3.4 approximately 30 days after birth, thus genetically labelling all currently dividing progeny cells. In the following, the time point of this labelling defines t_{dsRed} , the start of experiment and simulation. A total of 13 animals were then sacrificed at time points $t_{\text{dsRed}} + \Delta t_i$, $\Delta t_i \in [1 \text{ day} \dots 13 \text{ days}]$, and tumour sections were

stained for the proliferation marker Ki67. The fraction of dsRed⁺ cells which were stained positive for Ki67, $\mathcal{F}_{\text{exp}}^{\text{Ki67}}$, was evaluated and the result is shown in figure 8.2. Each experimental point here represents between one and three animals, the error bars show rough estimates of the standard error of the mean (SEM). For this estimation, the sample standard deviation was calculated on the time point with the largest sample size available (day 6, $n = 3$) and used as an estimator for the population standard deviation σ_{Ki67} of all time points. The SEM estimate for each time point Δt_i was then calculated using the number of animals available at this time point, $n(\Delta t_i)$, giving $\sigma_{\mathcal{F}}(\Delta t_i) = \sigma_{\text{Ki67}}/\sqrt{n(\Delta t_i)}$.

In order to simulate this experiment, we first evolve the tumour model given by equations (8.1) until day 30, the approximate starting point of the real experiment. However, it is important to note that the exact starting point of the simulated experiment does not influence the outcome, as - like in the experimental case - we are interested only in the fraction of cells still proliferating after time intervals Δt . The time course of this fraction is independent of the starting time because all rates in the model are assumed to be stationary (constant in time).

Let the viral vector introduce dsRed into all cells of type P at $t_{\text{dsRed}} = 30$ days while not labelling any stem cells or terminally differentiated cells. This gives as initial conditions for our experiment:

$$\begin{aligned} S_{\text{dsRed}}(t_{\text{dsRed}}) &= 0 \\ P_{\text{dsRed}}(t_{\text{dsRed}}) &= P(30 \text{ days}) \\ D_{\text{dsRed}}(t_{\text{dsRed}}) &= 0. \end{aligned} \quad (8.9)$$

Using these initial conditions, we solve equations (8.1) numerically resulting in solutions $P_{\text{dsRed}}(t_{\text{dsRed}} + \Delta t)$ and $D_{\text{dsRed}}(t_{\text{dsRed}} + \Delta t)$. $S_{\text{dsRed}}(t_{\text{dsRed}} + \Delta t) = \text{const.} = 0$ because no stem cells incorporate dsRed and there is no dedifferentiation in the model. The modelled proliferative fraction $\mathcal{F}_{\text{mod}}^{\text{Ki67}}$ within the dsRed population at Δt is then readily calculated as

$$\mathcal{F}_{\text{mod}}^{\text{Ki67}}(\Delta t) = \frac{P_{\text{dsRed}}(t_{\text{dsRed}} + \Delta t)}{P_{\text{dsRed}}(t_{\text{dsRed}} + \Delta t) + D_{\text{dsRed}}(t_{\text{dsRed}} + \Delta t)}. \quad (8.10)$$

In analogy to equations (8.7) and (8.8), we can now derive the parameter-dependent logarithmic likelihood of the experimental Ki67-fractions as shown in figure 8.2 for a given set of model parameters $\theta = \{\lambda_S, \lambda_A, \lambda_1, d_1, \mu_1\}$ to

$$\log \mathcal{L}_{\text{Ki67}}(\{\mathcal{F}_{\text{exp}}^{\text{Ki67}}, \sigma_{\mathcal{F}}\}|\theta) = \text{const.} + \sum_i \frac{\left(\mathcal{F}_{\text{exp}}^{\text{Ki67}}(\Delta t_i) - \mathcal{F}_{\text{mod}}^{\text{Ki67}}(\Delta t_i; \theta)\right)^2}{2(\sigma(\Delta t_i))^2}, \quad (8.11)$$

where the sum runs over all time points where experimental information is available.

8.2.3. Tumour composition information

As discussed in section 8.1, the fractional decomposition of the tumour into its three populations S , P and D is stationary for $t \gg 0$ and is given by equation (8.6). Unfortunately, the fractions of proliferating and terminal progenitors are currently not accessible experimentally.

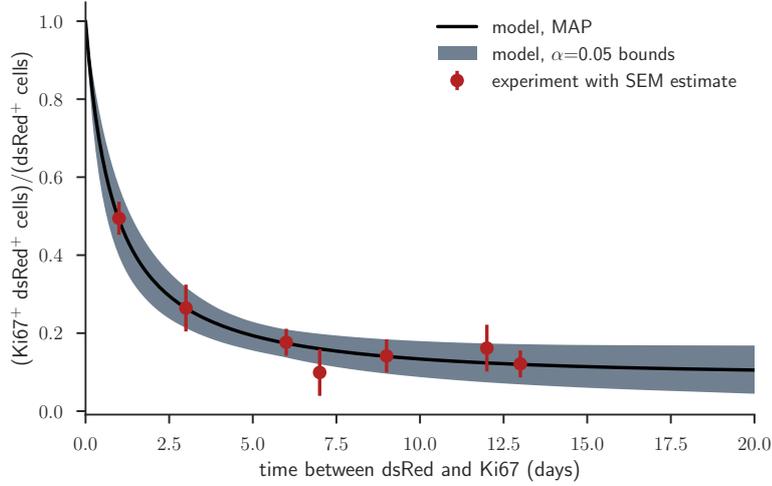


Figure 8.2.: Time course of proliferative capacity of progeny cells as a function of time. Shown is the fraction of descendants of progeny proliferating (model population P) at $t = 0$ which is still proliferating at the given later time, see also section 8.2.2. Each experimental data point (dark red) represents between one and three animals together with an estimate of the corresponding standard error of mean. The model bands derive from 20,000 simulations of equation (8.10) with parameter sets θ drawn from the posterior distribution and satisfying $p(\theta|\theta_{\text{MAP}}) > 0.05$ (see section 8.3). The simulation generated from the maximum likelihood set of parameters is shown additionally. Experimental data: Yue Zhuo.

The fraction of stem cells within the tumour bulk was assessed by flow cytometry using a mouse model which expresses RFP in all glioblastoma cells and additionally GFP in Tlx^+ cells (stem cells). Additionally, the fraction of actively proliferating cells within the tumour bulk ($\mathcal{F}_{\text{exp}}^{\text{prolif/bulk}}$) and the fraction of proliferating stem cells within the population of proliferating cells ($\mathcal{F}_{\text{exp}}^{\text{stem/prolif}}$) were measured by a combination of staining for the proliferation marker Ki67 and the mouse model co-expressing GFP and Tlx. The value of $\mathcal{F}_{\text{exp}}^{\text{stem/prolif}}$ was previously published in [Zhu, 2013] as the fraction of tumour cells stained positively for the proliferation marker PCNA; the value for $\mathcal{F}_{\text{exp}}^{\text{prolif/bulk}}$ shown here is a weighted average of values from [Zhu, 2013, Zhu et al., 2014] and new data reported here, all based on Ki67-stainings:

$$\begin{aligned}\mathcal{F}_{\text{exp}}^{\text{stem/bulk}} &= (0.23 \pm 0.04)\%, \\ \mathcal{F}_{\text{exp}}^{\text{prolif/bulk}} &= (0.22 \pm 0.04)\%, \\ \mathcal{F}_{\text{exp}}^{\text{stem/prolif}} &= (0.09 \pm 0.02)\%.\end{aligned}$$

These experimental ratios and their model counterparts are visualised in figure 8.3 together with the model estimations for the experimentally inaccessible subpopulation fractions.

In order to construct equivalent ratios from our mathematical model, we must make assumptions about the proliferative action of the three populations S , P and D . Fully differentiated cells of type D are not proliferative per model definition. Similarly, all proliferating progeny of type

P is actively dividing by construction of the model. The case of stem cells S however is less clear, as we know that neural stem cells, like stem cells in general, enter and exit the cell cycle depending on external and internal cues, thus switching between active and quiescent states [Li and Clevers, 2010, Ponti et al., 2013, Bond et al., 2015]. Here, we have chosen not to include two separate stem cell populations into the model because we have insufficient information for estimating all rates resulting from an additional cell population. Nevertheless, cell cycle analysis in our tumour model indicates that $f_{G_0} = (83 \pm 6)\%$ of Tlx^+ cells are in the G_0 -phase at any one time [Zhu, 2013]. We make use of this information directly in modelling the two experimentally observed ratios introduced above without an explicit quiescent population by assuming the remaining 17% to be actively proliferating. This value is in agreement with the fraction of Tlx^+ cells staining positively for PCNA [Zhu et al., 2014]. As a result, we define for comparison with experimental observations

$$\begin{aligned}\mathcal{F}_{\text{mod}}^{\text{stem/bulk}} &= \lim_{t \rightarrow \infty} \frac{S(t)}{N(t)}, \\ \mathcal{F}_{\text{mod}}^{\text{prolif/bulk}} &= \lim_{t \rightarrow \infty} \frac{(1 - f_{G_0})S(t) + P(t)}{N(t)}, \\ \mathcal{F}_{\text{mod}}^{\text{stem/prolif}} &= \lim_{t \rightarrow \infty} \frac{(1 - f_{G_0})S(t)}{(1 - f_{G_0})S(t) + P(t)}.\end{aligned}\tag{8.12}$$

The stationarity of the second and third model ratios follows directly from the stationarity of the subpopulation fractions given in equation (8.6). Their values were calculated numerically.

As in the previous sections, these considerations lead to a parameter-dependent logarithmic likelihood contribution from the tumour composition data of the form

$$\begin{aligned}\log \mathcal{L}_{\text{comp}}(\mathcal{F}_{\text{exp}}^{\text{s/b}}, \mathcal{F}_{\text{exp}}^{\text{p/b}}, \mathcal{F}_{\text{exp}}^{\text{s/p}}, \sigma_{\mathcal{F}_{\text{s/b}}}, \sigma_{\mathcal{F}_{\text{p/b}}}, \sigma_{\mathcal{F}_{\text{s/p}}} | \theta) &= \text{const.} + \frac{(\mathcal{F}_{\text{exp}}^{\text{s/b}} - \mathcal{F}_{\text{mod}}^{\text{s/b}}(\theta))^2}{2\sigma_{\mathcal{F}_{\text{s/b}}}^2} \\ &+ \frac{(\mathcal{F}_{\text{exp}}^{\text{p/b}} - \mathcal{F}_{\text{mod}}^{\text{p/b}}(\theta))^2}{2\sigma_{\mathcal{F}_{\text{p/b}}}^2} + \frac{(\mathcal{F}_{\text{exp}}^{\text{s/p}} - \mathcal{F}_{\text{mod}}^{\text{s/p}}(\theta))^2}{2\sigma_{\mathcal{F}_{\text{s/p}}}^2},\end{aligned}\tag{8.13}$$

which is used for joint parameter estimation in section 8.3.

8.3. Parameter Estimation, Method and Results

In order to calibrate the model introduced in section 8.1 to the data using the likelihoods derived in section 8.2, we use a Bayesian framework in which we approximate the posterior distribution of model parameters with Markov chain Monte Carlo (MCMC) sampling. This section comprises a short summary of the principles of Bayesian parameter estimation (section 8.3.1), a summary of how MCMC sampling was employed (section 8.3.2) and an overview of the resulting parameter estimates and their confidence bounds (section 8.3.3).

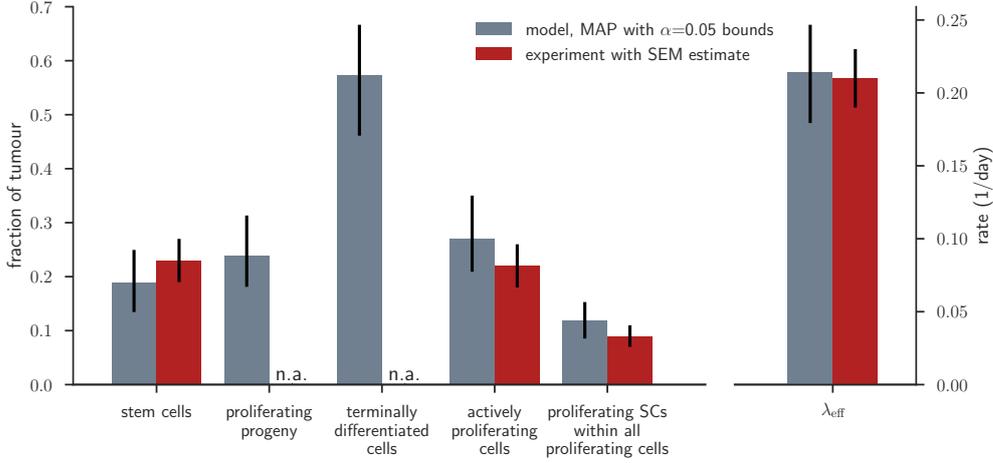


Figure 8.3.: Tumour composition (left y-axis) as discussed in section 8.2.3 and effective tumour growth rate (right y-axis) as discussed in section 8.2.1. Red bars indicate experimental values with black bars showing standard error estimates; grey bars show the result of the maximum *a posteriori* probability parameter set θ_{MAP} with black bars derived by simulating tumours from 20,000 parameter sets θ drawn from the posterior distribution and satisfying $p(\theta|\theta_{\text{MAP}}) > 0.05$ (see section 8.3). Experimental information is not available for the fractions of stem cells, proliferating progeny and differentiated cells. Experimental data: Zhe Zhu, Peng Zou, M. Amir Khan.

8.3.1. Short Introduction to Bayesian Parameter Estimation

At the heart of Bayesian parameter estimation (see for example [Sorensen and Gianola, 2002, MacKay, 2005, Box and Tiao, 2011]) lies Bayes' Theorem which provides the probability distribution of parameters θ as a function of the available experimental data D , a chosen model topology \mathcal{H} , an error model and distribution conveying prior information on the distribution of parameters,

$$P(\theta|D, \mathcal{H}) = \frac{\mathcal{L}(D|\theta, \mathcal{H}) \pi(\theta|\mathcal{H})}{P(D|\mathcal{H})}. \quad (8.14)$$

This resulting probability distribution $P(\theta|D, \mathcal{H})$ over parameters is called *posterior distribution* or simply *posterior*. Data D as well as model and error assumptions enter the calculation via the likelihood function $\mathcal{L}(D|\theta, \mathcal{H})$, potential prior information is encoded in the *prior distribution* (or *prior*) $\pi(\theta)$. The normalising constant $P(D|\mathcal{H})$ naturally gives the total probability of the data given the chosen model when taking into account all possible parameter sets θ ,

$$P(D|\mathcal{H}) = \int d\theta \mathcal{L}(D|\theta, \mathcal{H}) \pi(\theta|\mathcal{H}), \quad (8.15)$$

and is thus also called *evidence* for the chosen model \mathcal{H} . Although this constant does not affect model fitting and parameter estimation and will therefore be dropped in the following, it plays a major role in Bayesian model selection, where comparing evidence values of several models allows to rank their agreement with the data [MacKay, 2005].

We have previously introduced all components necessary for making use of Bayes' theorem as shown in equation (8.14) - a mathematical model in section 8.1, experimental data and associated error models in sections 8.2.1 to 8.2.3. Specifically we have given likelihood functions for all datasets in equations (8.8), (8.11) and (8.13). Since these experiments are independent of each other, we arrive at a combined likelihood function $\mathcal{L}_{\text{total}}$ by multiplying the individual likelihood contributions. This gives, up to a constant, a total logarithmic likelihood of

$$\begin{aligned} \log \mathcal{L}_{\text{total}}(D|\theta) = & \log \mathcal{L}_{BLI}(\lambda_0, \sigma_{\lambda_0} | \lambda_S) + \log \mathcal{L}_{Ki67}(\{\mathcal{F}_{\text{exp}}^{\text{Ki67}}, \sigma_{\mathcal{F}}\} | \theta) + \\ & \log \mathcal{L}_{\text{comp}}(\mathcal{F}_{\text{exp}}^{\text{p/b}}, \mathcal{F}_{\text{exp}}^{\text{s/p}}, \sigma_{\mathcal{F}_{\text{p/b}}}, \sigma_{\mathcal{F}_{\text{s/p}}}) | \theta), \end{aligned} \quad (8.16)$$

where $\theta = \{\lambda_S, \lambda_A, \lambda_1, d_1, \mu_1\}$ as previously. We assume uniform prior distributions within biologically reasonable bounds over these parameters. Furthermore we assume prior independence of parameters allowing us to give distributions $\pi^\dagger(\beta)$ for each parameter separately. For convenience, we choose the same bounds for all five rates (0 and 4 day⁻¹),

$$\pi^\dagger(\beta) = \begin{cases} 0 & \text{for } \alpha < 0, \\ \frac{1}{4}, & \text{for } 0 \leq \alpha \leq 4, \\ 0 & \text{for } \alpha > 4 \end{cases}, \quad \text{for } \beta \text{ in } \{\lambda_S, \lambda_A, \lambda_1, d_1, \mu_1\} \quad (8.17)$$

and construct $\pi(\theta)$ as the product of the five individual priors. As discussed above (equation (8.3)), we require $d_1 \leq \lambda_1$ because of previous observations and thus assign a prior probability of zero to any parameter set that does not fulfil this constraint. As we will see below, our data yield fully identifiable parameter estimates far away from the somewhat arbitrarily chosen upper bound of 4/day, thus confirming its minor importance for our inference. After dropping all constants from equation (8.14) in our context, we redefine

$$P(\theta|D) = \mathcal{L}(D|\theta)_{\text{total}} \pi^\dagger(\lambda_S) \pi^\dagger(\lambda_A) \pi^\dagger(\lambda_1) \pi^\dagger(d_1) \pi^\dagger(\mu_1) \quad (8.18)$$

for the following discussion. It is of note that this choice of constant prior implies that

$$P(\theta|D) \propto \mathcal{L}_{\text{total}}(D|\theta) \quad (8.19)$$

within the given bounds, illustrating that seeking the maximum of $P(\theta|D)$ is equivalent to maximising the likelihood $\mathcal{L}(D|\theta)$. Thus, in the case of uniform prior distributions, the most probable parameter set θ_{MAP} (MAP - maximum *a posteriori* probability) resulting from Bayesian analysis and derived as described in section 8.3.3 is numerically equivalent to the maximum likelihood estimate (MLE) [Sorensen and Gianola, 2002].

8.3.2. MCMC Sampling of the Posterior Distribution

Because of the complicated form of equation 8.16, the posterior parameter distribution $P(\theta|D)$ is not accessible analytically. Instead, we approximate it using an MCMC-approach. In general, MCMC-methods construct a Markov chain whose equilibrium distribution is the requested probability distribution [Sorensen and Gianola, 2002]. Samples taken from this chain after it

has reached equilibrium may thus be regarded as samples of the target distribution. The calculations underlying this work have been performed in the programming language python and rely on the MCMC-package emcee [Foreman-Mackey et al., 2013]. This package uses an affine-invariant ensemble sampling method [Goodman and Weare, 2010], which runs many Markov chains in parallel and constructs informed choices for the next step of each single chain on the basis of the current positions of all others.

Sampling the posterior parameter distribution defined by equations (8.16) and (8.17) yields a five-dimensional cloud of points encoding the probability associated with each region of parameter-space in two different ways. First, the density of points in each region is representative of its relative probability, because as discussed above it is a key feature of MCMC-sampling that it produces samples from the target distribution once equilibrated. Second, in addition to recording each sample's parameter values, we can record the value $P(\theta|D)$ associated with the parameter set, thus allowing us to locate regions of a given likelihood directly if the sample size is large enough. Because the visualisation of a 5-dimensional cloud of points is not straightforward, we resort to an analysis of one- and two-dimensional projections of the cloud. Figure 8.4 shows all five one-dimensional histograms and all ten two-dimensional histograms resulting from the five model parameters. These histograms are approximations of the corresponding marginal posterior distributions in which a subset of parameters has been integrated out, that is

$$P(\theta_1|D) = \int d\theta_2 d\theta_3 d\theta_4 d\theta_5 P(\theta|D) \quad (8.20)$$

for the marginal posterior of parameter θ_1 and

$$P(\theta_1, \theta_2|D) = \int d\theta_3 d\theta_4 d\theta_5 P(\theta|D) \quad (8.21)$$

for the joint marginal posterior of θ_1 and θ_2 . It is important to note that, in general, the components of the joint modal vector (mode of the full n -dimensional posterior, θ_{MAP}), are not equal to the modes of the n marginal distributions [Sorensen and Gianola, 2002]. This can also be observed in figure 8.4, where the red lines have been added to indicate the components of the full modal vector - deviations from the marginal modes in one dimension are evident especially for λ_A and μ_1 . Nevertheless, this visualisation of the posterior sample is useful because it conveys information about the general spread of sampled parameters and the correlations between pairs of parameters. We observe that the marginal posterior of λ_A is positively skewed and portrays a significant heavy tail, indicating that the currently available data do not strongly exclude high asymmetric division rates of stem cells. Furthermore, λ_1 and d_1 are strongly correlated indicating that even though both can assume a range of parameter values, in any given sampled set their values do not deviate much from each other. This fit result is in agreement with the criticality of the progeny compartment suggested by [Lan et al., 2017], where division and differentiation rates of proliferating progeny need be equal (or very similar).

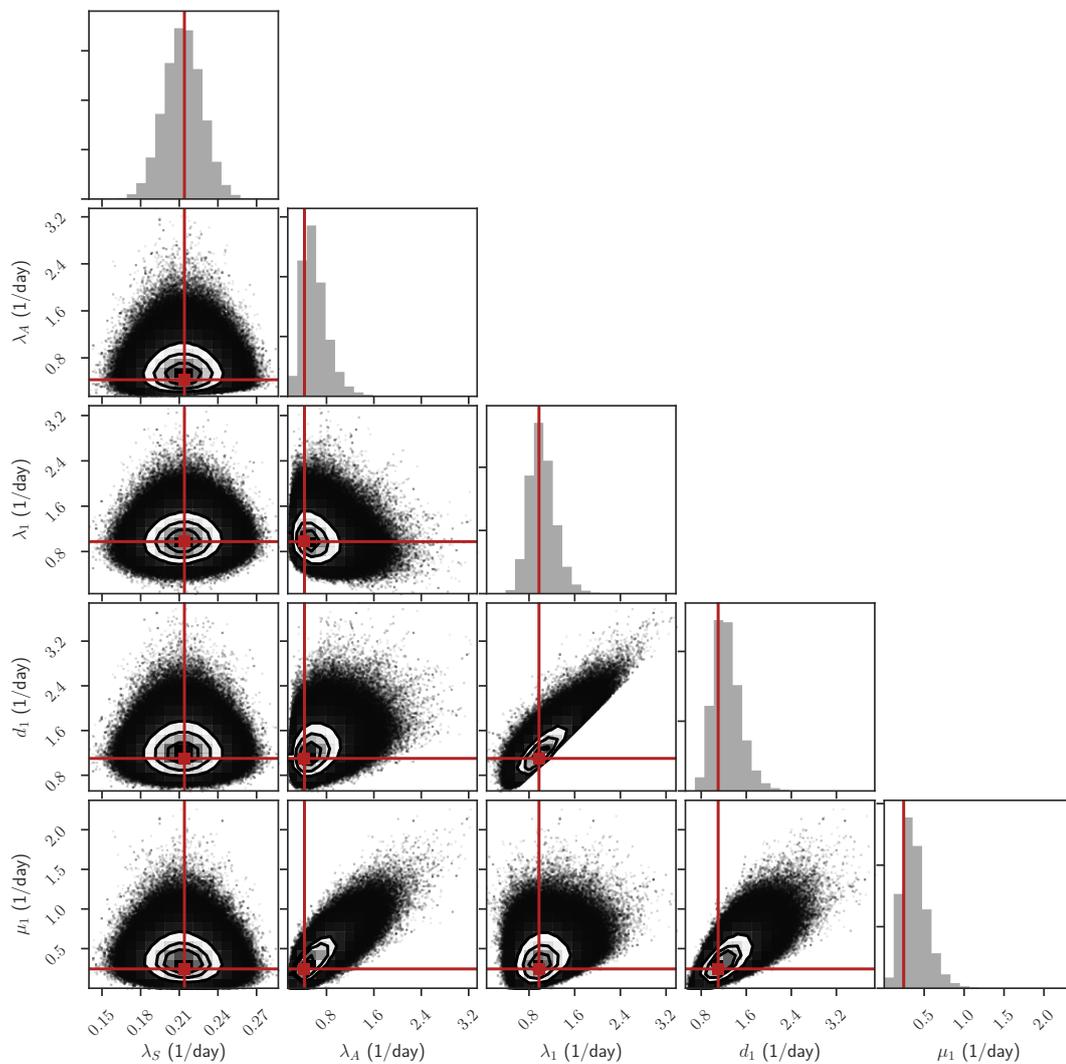


Figure 8.4.: One- and two-dimensional projections of the posterior probability distribution of the model parameters λ_S , λ_A , λ_1 , d_1 and μ_1 . The red lines indicate the position of the components of the most probable parameter set θ_{MAP} . As discussed in section 8.3, the components of the joint modal vector do not generally coincide with the marginal modes as is obvious for λ_A and μ_1 here in particular. The one and two-dimensional histograms shown here are built from close to one million samples.

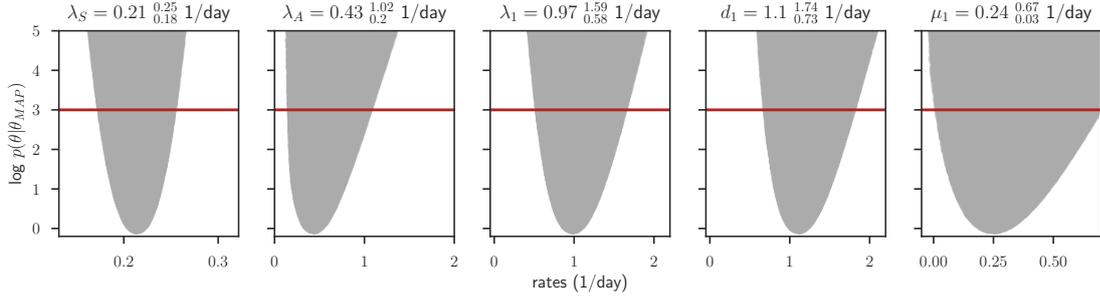


Figure 8.5.: Posterior profiles of the five model parameters λ_S , λ_A , λ_1 , d_1 and μ_1 as discussed in section 8.3. The logarithm of the relative probability as defined in equation (8.23) is shown on the y -axis, with values above the red line corresponding to less than 5% of the probability associated with the best parameter set. The shaded area shows which relative probabilities are accessible for a given parameter value (shown on the x -axis). We define the parameter values where the red line cuts the shaded area as boundaries of the credible region of 5% relative probability and give the associated values in table 8.1. The plot was constructed from close to one million MCMC-samples.

8.3.3. MAP Parameter Estimates and Confidence Intervals

In order to summarise the full parameter distribution shown in figure 8.5, we quote the maximum *a posteriori* (MAP) parameter set, i.e. the mode of the 5-dimensional posterior probability distribution. This choice of estimator is linked to the decision-theoretic loss function $G(x_0, x) = \delta(x_0 - x)$, where δ is Dirac's delta function, x_0 is the true value, but unknown, parameter value and x is our chosen estimate [Jaynes et al., 2003]. With this loss function,

$$\theta_{\text{MAP}} = \arg \max_{\theta} \int \delta(\theta - \theta') P(\theta' | D) d\theta'. \quad (8.22)$$

In practice, we make use of the probability values associated with each sampled set of the MCMC-chain for finding the most probable set θ_{MAP} and defining credible regions around it. As this method is obviously vulnerable to random effects governing the sampling procedure, we have simulated MCMC-chains long enough to ascertain two stable decimal places for all values in θ_{MAP} when repeating the entire approximation of the posterior. The results for the individual parameters are shown in figure 8.5 and can be found in table 8.1 as well.

In order to construct credible regions on these MAP estimates, we define the relative probability of every other parameter set θ as

$$p(\theta | \theta_{\text{MAP}}) = \frac{\int \delta(\theta_0 - \theta') P(\theta' | D) d\theta'}{\int \delta(\theta_{\text{MAP}} - \theta') P(\theta' | D) d\theta'}. \quad (8.23)$$

Using this definition, we can set a relative probability level α and regard as acceptable only those parameters within a credible region (CR)

$$\text{CR} = \{\theta \mid p(\theta | \theta_{\text{MAP}}) \geq \alpha\}. \quad (8.24)$$

parameter	MAP estimate	$p(\theta \theta_{\text{MAP}}) = 5\%$ credible region	probability within bounds
λ_S	0.21 day ⁻¹	[0.18, 0.25] day ⁻¹	0.99
λ_A	0.43 day ⁻¹	[0.20, 1.02] day ⁻¹	0.95
λ_1	0.97 day ⁻¹	[0.58, 1.59] day ⁻¹	0.97
d_1	1.10 day ⁻¹	[0.73, 1.74] day ⁻¹	0.96
μ_1	0.24 day ⁻¹	[0.03, 0.67] day ⁻¹	0.95
5D set θ	as above	5D hypersurface where $p(\theta \theta_{\text{MAP}}) = 0.05$	0.61
5D set θ	as above	5D hypercuboid defined by above bounds	0.88

Table 8.1.: MAP parameter estimates and information regarding their credibility, for detailed discussion see text (8.3.2, 8.3.3). For a graphical explanation of how credible regions were constructed see figure 8.5.

Like [Watson et al., 2015], we chose $\alpha = 0.05$, resulting in a cut-off of $|\log \alpha| = 3$ in logarithmic probability space. Figure 8.5 shows the application of this principle to our posterior distribution. For each of the five parameters, the figure indicates which relative probabilities can be assumed as a function of the chosen parameter if the other four parameters are allowed to vary freely. Thus, the realisations on the edge of the shaded regions in figure 8.5 represent the optimal accessible parameter set as function of each parameter in turn. The red line indicates where $\log p(\theta|\theta_{\text{MAP}}) = 0.05$ - corresponding parameter bounds are defined where this line hits the accessible region of relative probability. Here we have used our MCMC-sampled posterior to determine these profiles which are hence only close approximations of the true optimal probabilities as a function of the fixed parameter. Similar profiles can be derived by re-optimising (in a maximum likelihood sense) the problem for the fixed parameter assuming a series of values around the MLE. This is known as profile likelihood method [Venzon and Moolgavkar, 1988, Raue et al., 2009]. In analogy, we call our approach the method of *posterior profiles*. The credible regions resulting from this method are quoted in figure 8.5 and table 8.1.

Choosing a cut-off in relative posterior probability as discussed above defines a five-dimensional boundary within the posterior distribution. This boundary contains a corresponding fraction of the posterior probability. Since an alternative definition of Bayesian credible regions [Box and Tiao, 2011] uses the probability contained within specified bounds, and related frequentist methods employing constant χ^2 -boundaries [Press et al., 2007] are common, we quote the fractions of posterior probability contained within our bounds in table 8.1. For comparison, in the case of a one-dimensional Gaussian distribution over parameter β , boundaries defined by $\log p(\beta|\beta_{\text{MAP}}) = 0.05$ would contain 98 % of the total probability. The values in table 8.1 indicate that the marginal distributions of λ_A and μ_1 especially have heavier tails than a normal distribution, corresponding to what we have observed above. The last two rows of table 8.1 illustrate that, when quoting the five sets of upper and lower bounds as a summary of the posterior distribution, we are providing a conservative estimate. The five-dimensional surface defined by $p(\theta|\theta_{\text{MAP}}) = 0.05$ (meaning that all parameter sets outside of it are less than five percent as likely as the best estimate) contains 61% of the total probability, while the five-dimensional box that we describe by the one-dimensional bounds contains 88% of the total probability. This

inevitably means that the box contains a large fraction of highly unlikely parameter combinations.

Because of this, in order to derive the confidence bounds on our MAP simulations, we use only parameter sets from within the $p(\theta|\theta_{\text{MAP}}) = 0.05$ - hypersurface. Several thousands of these parameter combinations are drawn and the experiments described in sections 8.2.1 to 8.2.3 are simulated with each of them. The resulting spread in simulation outcomes defines the bounds shown in figures 8.2 and 8.3, which thus comprise all results accessible with parameter which are at least $1/20$ as likely as the maximum probability estimate.

9. Model Applications

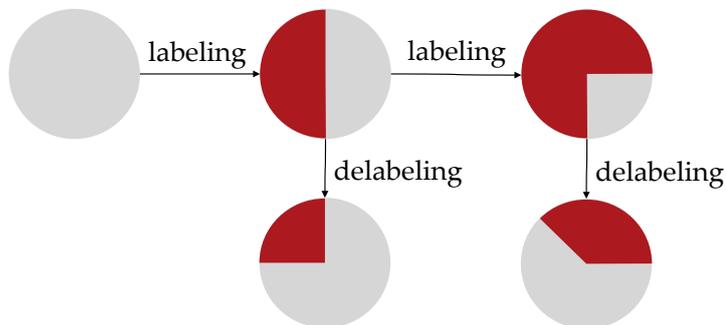
In this chapter, three advanced applications of the model introduced and calibrated in section 8 are presented. First, we predict the outcome of an experiment not used for parameter estimation - the time course of IdU-incorporation into the DNA of Tlx^+ cells - and compare the result with experimental data (section 9.1). Second, we analyse single-cell labelling data describing the fates of clones derived from individual labelled cells over time. Here, an apparent disagreement with the model introduced in section 8.1 is identified and discussed. We propose a resolution of the discord built on migration of individual cells through tumour and brain as suggested in section 7 and refit the altered model to estimate this cellular migration rate (section 9.2). Last, in section 9.3, we discuss model-based expectations of the systems response to two different treatment approaches: chemotherapy with temozolomide and a genetic knockdown approach affecting the maintenance of the stem-cell state. These are compared to experimental results.

9.1. IdU incorporation study

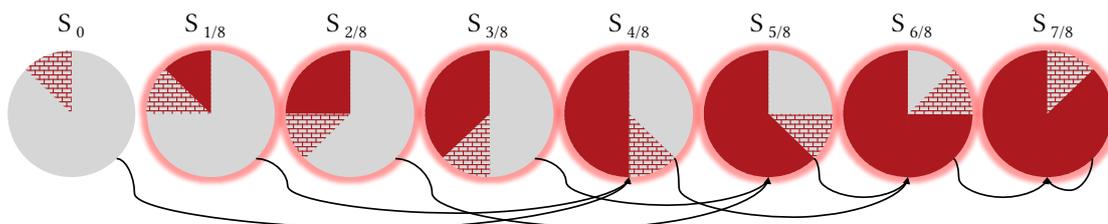
A further experiment employs the thymidine analogues IdU and CldU in order to label dividing cells while they are replicating their DNA [Cavanagh et al., 2011]. Here, IdU was administered to tumour-bearing mice carrying the Tlx -GFP construct described in section 6.3.2 once every day for a varying number of days (between once and 15 times) before mice were sacrificed and the fraction of IdU-positive cells within the GFP-positive population was analysed.

In order to model this experiment mathematically, we make use of a number of recent estimations derived by colleagues in the group of *Theoretical Systems Biology* at DKFZ. First, there is a threshold in the fraction of labelled DNA which needs to be surpassed for the cell to be detected as labelled. We take this threshold to be $1/8$ of the total DNA (Melania Barile, unpublished). Second, after injection of a thymidine analogue, its concentration drops rapidly as the component is filtered and degraded by the animal's body. For the related analogue EdU (ethynyl-deoxyuridine), uptake by dividing cells ceases by five hours after administration (Adrien Jolly, unpublished). Because of this, we assume the labelling period following IdU administration to last for 5 hours likewise. Divisions during the remaining 19 hours of the day result in synthesis of unlabelled DNA (containing regular thymidine). The basic arithmetics of accumulation and dilution of labelled DNA during labelling and delabelling phases respectively is shown in figure 9.1a.

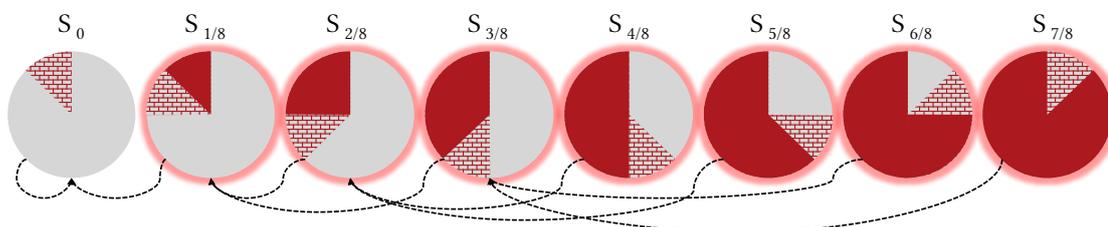
In principle, labelling and delabelling are continuous processes in the space of IdU-carrying DNA suggesting a mathematical model employing partial differential equations. However, since we are only interested in the labelled and unlabelled fraction of cells and not their precise



(a) Illustration of the basic principle of labelling with thymidine analogues. If an unlabelled cell replicates its DNA in the surplus presence of the analogue (first arrow), the newly synthesized strands are completely labelled. Assuming that the labelled and unlabelled strands are randomly distributed during mitosis, each daughter cell ends up with half its DNA labelled. After another division in presence of the label (second “labelling” arrow), three quarters of the DNA are labelled. The reverse happens if the division happens while there is no analogue available - the newly replicated DNA contains regular thymidine. Thus, the label content of the two daughter cells is halved (two downwards arrows, “delabelling”).



(b) Illustration of the labelling model shown in equation (9.1). In principle, the fraction of labelled DNA in a given stem cell can assume any value in $[0, 1)$, but we group the cells into eight bins in order to allow modeling the process with a system of eight ordinary differential equations. Each pie describes the population of cells in which a minimum of the red fraction and a maximum of the red part plus the patterned part is labelled. The grey part of each pie stands for the fraction of DNA not carrying the label. The arrows between these eight populations indicate in which population the daughter cells of a cell from a given population of origin will end up if the DNA is replicated in the presence of the thymidine analogue. Because at least $1/8$ of the DNA has to carry the label to allow detection (see text), the population S_0 is not classified as labelled (no halo). Populations $S_{1/8}$ to $S_{7/8}$ carry the label in detectable amounts (depicted by red halo).



(c) As in part b, but for the delabelling model as described in equation (9.2). Here, the arrows indicate in which population the daughter cells of a cell from a given population of origin will end up if the DNA is replicated without the thymidine analogue being present.

Figure 9.1.: Modelling of DNA-labelling with thymidine analogues. Part a shows the basic principle of the labelling and delabelling processes, parts b and c explain the rationale behind the mathematical models of labelling and delabelling phases.

IdU-content, we group them into eight bins containing between 0 and 100% of IdU instead of thymidine as shown in figures 9.1b and 9.1c. The first bin, population S_0 , contains all cells with less than 1/8 of the total DNA labelled and is the only population assumed to be undetectable in an immunostaining assay. The following population, $S_{1/8}$, contains all cells with at least 1/8 but less than 1/4 of labelled DNA *et cetera*. The choice of excluding the upper bound from the bin is in agreement with the notion that theoretically it is not possible to reach a completely labelled cell as a small part of the original unlabelled DNA always remains. The arrows in figures 9.1b and 9.1c illustrate how cells move between the bins if they divide during the labelling or the delabelling phase respectively. Following this discretisation in IdU-content, it is straight forward to formulate a set of ODEs using the fractionally labelled groups and the parameters of stem cell division introduced in section 8.1. For the labelling phase, this results in

$$\begin{aligned}
 \dot{S}_0 &= -(\lambda_S + \lambda_A)S_0 \\
 \dot{S}_{1/8} &= -(\lambda_S + \lambda_A)S_{1/8} \\
 \dot{S}_{2/8} &= -(\lambda_S + \lambda_A)S_{2/8} \\
 \dot{S}_{3/8} &= -(\lambda_S + \lambda_A)S_{3/8} \\
 \dot{S}_{4/8} &= -(\lambda_S + \lambda_A)S_{4/8} + (2\lambda_S + \lambda_A)(S_0 + S_{1/8}) \\
 \dot{S}_{5/8} &= -(\lambda_S + \lambda_A)S_{5/8} + (2\lambda_S + \lambda_A)(S_{2/8} + S_{3/8}) \\
 \dot{S}_{6/8} &= -(\lambda_S + \lambda_A)S_{6/8} + (2\lambda_S + \lambda_A)(S_{4/8} + S_{5/8}) \\
 \dot{S}_{7/8} &= \lambda_S S_{7/8} + (2\lambda_S + \lambda_A)S_{6/8},
 \end{aligned} \tag{9.1}$$

with all populations as defined in figure 9.1. Similarly we find for the delabelling phase

$$\begin{aligned}
 \dot{S}_0 &= \lambda_S S_0 + (2\lambda_S + \lambda_A)S_{1/8} \\
 \dot{S}_{1/8} &= -(\lambda_S + \lambda_A)S_{1/8} + (2\lambda_S + \lambda_A)(S_{2/8} + S_{3/8}) \\
 \dot{S}_{2/8} &= -(\lambda_S + \lambda_A)S_{2/8} + (2\lambda_S + \lambda_A)(S_{4/8} + S_{5/8}) \\
 \dot{S}_{3/8} &= -(\lambda_S + \lambda_A)S_{3/8} + (2\lambda_S + \lambda_A)(S_{6/8} + S_{7/8}) \\
 \dot{S}_{4/8} &= -(\lambda_S + \lambda_A)S_{4/8} \\
 \dot{S}_{5/8} &= -(\lambda_S + \lambda_A)S_{5/8} \\
 \dot{S}_{6/8} &= -(\lambda_S + \lambda_A)S_{6/8} \\
 \dot{S}_{7/8} &= -(\lambda_S + \lambda_A)S_{7/8}.
 \end{aligned} \tag{9.2}$$

In order to simulate the IdU-labelling experiment described above, we concatenate five hours of the labelling dynamics described in equation (9.1) with 19 hours of the delabelling dynamics in equation (9.2) per day for 20 days and solve the resulting problem numerically starting from unlabelled cells only at $t = 0$ days. From the resulting solutions for each of the labelled populations, the fraction of detectably labelled cells (cells with a red halo in figures 9.1b and 9.1c) is easily computed as

$$\mathcal{F}_{\text{mod}}^{\text{IdU}}(t) = 1 - \frac{S_0(t)}{\sum_{i=0}^7 S_{i/8}(t)}. \tag{9.3}$$

Figure 9.2 shows the resulting time curves of $\mathcal{F}_{\text{mod}}^{\text{IdU}}$ for the MAP-parameter set θ_{MAP} and bounds delimitating the region accessible with parameter sets satisfying $p(\theta|\theta_{\text{MAP}}) > 0.05$.

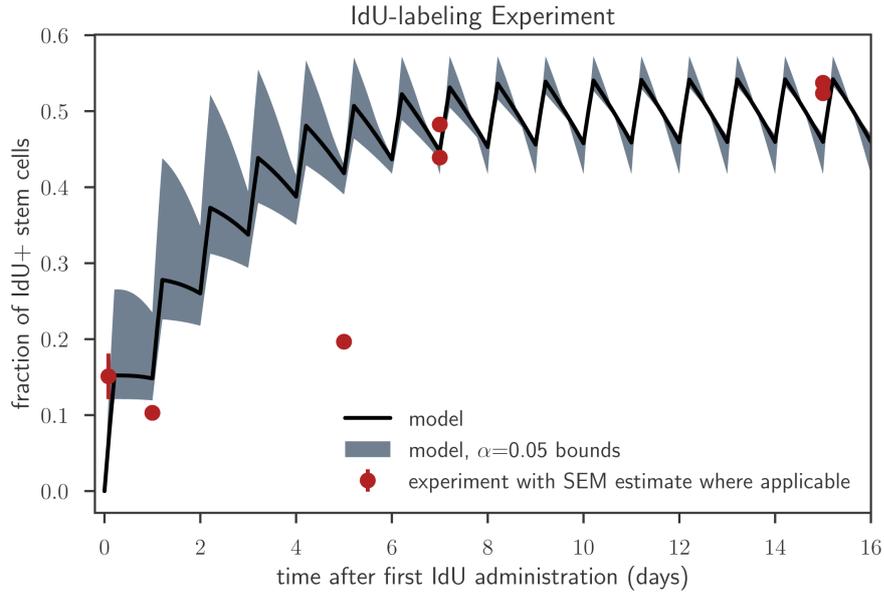


Figure 9.2.: Fraction of IdU⁺-positive Tlx⁺ cells as a function of time as discussed in section 9.1. Red dots without error bars indicate experimental results from one mouse each (six mice in total), the first red dot at 2 hours shows mean and associated standard error pooled from another six mice. MAP estimate and bounds were constructed as discussed before (see section 8.3). Experimental data: M. Amir Khan.

Additionally shown are experimental results from a total of twelve mice. Importantly, these data were not used for model calibration. Apart from one outlier on day 5, they are in good agreement with the model prediction. The “hedgehog” shape of the curve is due to the fact that after the daily injection of IdU a five-hour labelling phase (fraction of labelled cells increases) followed by a 19-hour delabelling phase (fraction of labelled cells decreases) is assumed as discussed above. The mean value of labelled cells at later times (i.e. after approximately day 7 in figure 9.2) depends solely on the ratio between the lengths of these labelling and delabelling phases and is almost completely independent of the parameters governing symmetric and asymmetric divisions of Tlx⁺ cells. These parameters are much more important for the initial transient phase where the daily mean labelled fraction is still rising (before day 7 in figure 9.2) - for this critical phase, there are currently few data points available. Still, significantly slower division rates of stem cells lead to a prolonged transient phase and are thus incompatible with the high labelled fraction found experimentally on day 7. Faster turnover rates in turn would lead to a shorter transient phase as illustrated by the upper $\alpha = 0.05$ bounds in figure 9.2, a scenario still compatible with the currently available data.

On a side note, the agreement between the four animals on days 7 and 15 and the model prediction can be interpreted as evidence in favour of the assumptions concerning the length of the labelling phase and the threshold in IdU concentration necessary for detection.

9.2. Tracing of single-cell derived clones

With the method described in section 6.3.5, individual cancer stem cells within a developed tumour can be genetically altered *in vivo* such that they start expressing one of four fluorescent proteins at random. The fluorescent protein is likewise expressed in all potential progeny of the initially labelled cell, thus enabling clonal tracing: when brain tumour slices are analysed by fluorescence microscopy a given time after the initial drug-induced labelling event, cells of the same fluorescent colour within close proximity of each other are likely derived from the same ancestor.

Here, a total of eight animals were sacrificed at $\Delta t = \{5, 5, 10, 10, 20, 26, 37\}$ days after label induction and their fluorescent clones were quantified. The resulting distributions of clone sizes are shown in the left column of figure 9.3. Each histograms is built on $\mathcal{O}(100)$ clones, a fairly low number for approximating a highly skewed distribution as will become apparent in the following. Summary statistics derived from these datasets comprise mean number of cells per clone, coefficient of variation (CV) of the clone size, skewness of the distribution as well as the relative fractions occupied by one-cell-clones, two-cell-clones and larger clones and are shown in figure 9.4. Error bars accompanying data points of individual animals in this figure show estimates of the standard error derived using bootstrapping [Efron and Tibshirani, 1994, Boos, 2003].

In a first approach at modelling this data, one would assume the observed clonal growth dynamics to be governed by the same growth laws introduced and quantified for the population data in section 8. However, a comparison of the experimentally observed mean clone sizes with the model growth curve resulting from equations (8.2) and (8.4) using the MAP-parameter set θ_{MAP} reveals a discrepancy that is both quantitative and qualitative (figure 9.5c). While the experimental mean clone size seems to reach a steady state by day 20 at the latest and never grows much beyond three cells per clone, the growth of the SPD-model starting from a single stem cell is of course exponential governed by the symmetric division rate of stem cells as discussed in section 8.2.1.

How can this contradiction be resolved? In section 7 we have shown that migration and dispersal processes, most likely of stem cells, must be a feature of the glioblastoma mouse model under investigation. Additionally, it is known that glioblastomas are highly invasive and cancer cells frequently leave the tumour bulk to spread into the surrounding brain areas where they form metastases [Cuddapah et al., 2014, Osswald et al., 2015, Hadjipanayis and Van Meir, 2009]. This behaviour is a major contributor to poor survival prognoses following glioblastoma diagnosis [Dandy, 1928, Carlsson et al., 2014, Campos et al., 2016]. Additionally, a spatially inhomogeneous distribution of actively proliferating cells as well as Tlx^+ cells has been observed in the glioblastoma mouse model discussed here [Zou, 2016] - both groups of cells are more abundant towards the outside of the solid tumour as shown in figures 9.5a and 9.5b. While a higher nutrient concentration on the outside of the tumour where normal vasculature has not been destroyed may in part account for a higher proliferative activity, it does not explain the accumulation of non-terminal cells there. Overall, the invasiveness of glioblastoma, the obser-

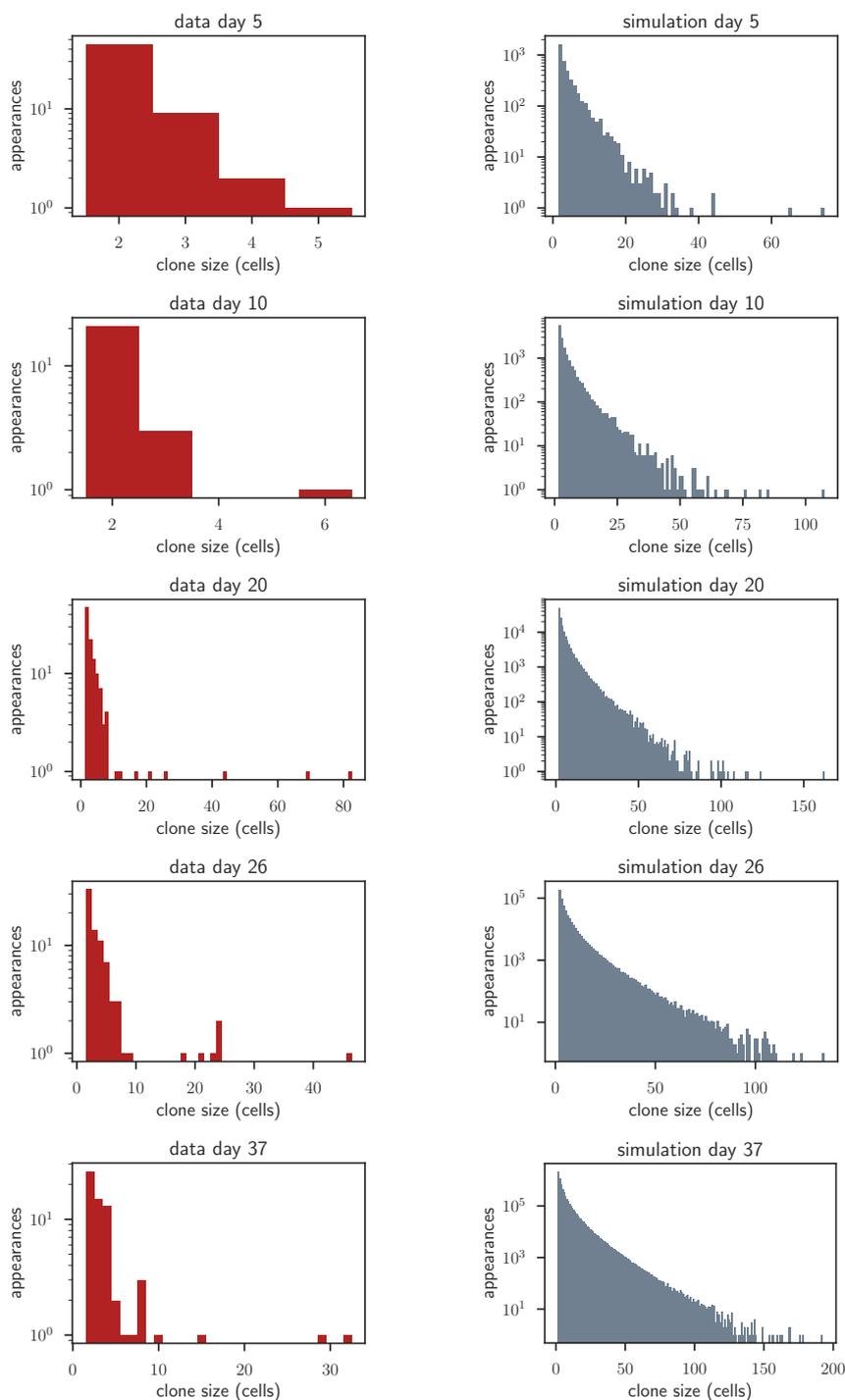


Figure 9.3.: Histograms of clone sizes found experimentally (left, red) and in a simulation including migration of cell types S and P (right, grey). The experimental figures for day 5 and 10 contain clones pooled from two animals each, the figures for days 20 to 37 contain clones from a single animal each. Simulations were performed from an initial 2000 labelled stem cells. Applying a Kolmogorov-Smirnov-test to the hypothesis that the data was drawn from the model distribution yielded p -values of 0.24, 0.86 and 0.18 for the later time points day 20, day 26 and day 37 respectively and thus no sufficient grounds for rejecting the hypothesis. For the early time points (day 5 and day 10) however, $p < 10^{-5}$ indicates that data and model are incompatible here. For discussion see the main text. Experimental data: M. Amir Khan.

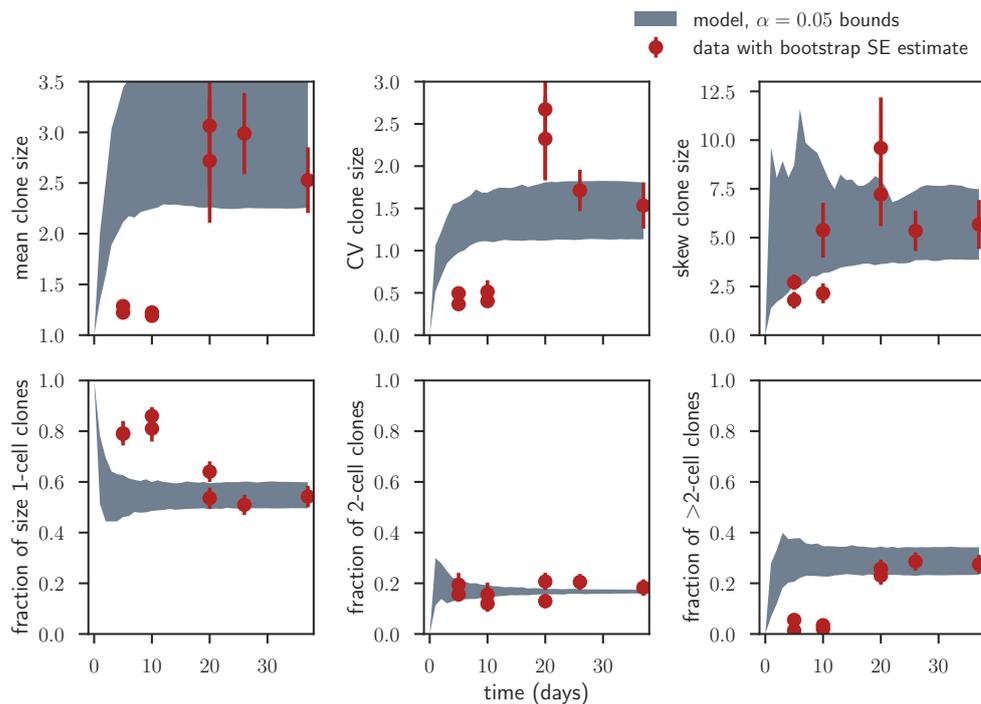


Figure 9.4.: Experimental summary statistics of the clone size histograms shown in figure 9.3 together with bootstrapped SE estimates from 10,000 resamplings (red). The grey bounds show areas accessed by 1000 stochastic model simulations using parameter sets satisfying $p(\theta|\theta_{\text{MAP}}) > 0.05$ and starting from 1,000 initially labelled stem cells. The roughness of the boundaries is due to the stochastic nature of the simulation - as the skewness is specifically sensitive to rare large clones, it remains affected by this even for the large numbers of clones simulated here. This is the reason why we do not include the experimental skewness into the likelihood used for fitting the single-cell data, see section 9.2. Experimental data: M. Amir Khan.

vation of sustained exponential tumour growth and the patterns of spatial heterogeneities in the mouse model suggest cellular migration of proliferative cells.

Because of this, we refine the SPD-model introduced in section 8 by allowing migration of cell types S (stem cells) and P (proliferating progeny). We define a migration event as a move of a cell far enough away from its clone of origin so that it is no longer identifiable as a clonal member by the observer but may instead be counted as an additional clone of its own. Figure 9.5d illustrates this principle and shows how it leads to more, but smaller, observed clones. We assign a new parameter, the migration rate β , to migration events. Stem cells and actively dividing progeny cells are assumed to leave their parent clone at equal speeds. Importantly, the original model retains validity for population-level processes like the ones discussed in sections 8 and 9.3, because cellular migration affects system properties only at the level of individual clones while leaving bulk properties like the total cell number or the cellular composition of the tumour unchanged.

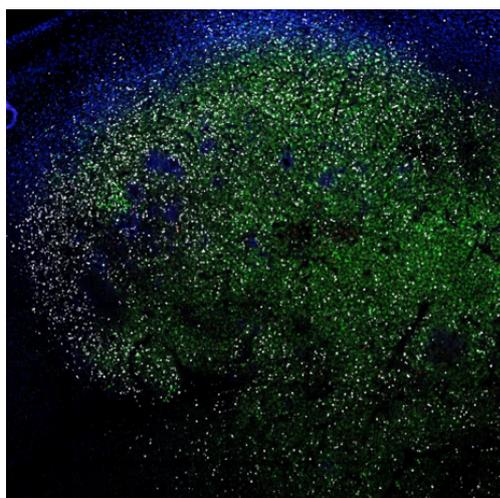
Single-cell data like the clone size distributions shown in figure 9.3 contain a wealth of information that is only fractionally conserved when using first-order summary statistics such as the mean clone size. Thus a modelling approach allowing access to higher moments of the clone size distribution is desirable. In principal, deriving a moment-generating function [Gardiner, 2009] from a master equation formulation of clonal growth could be considered as it has been used successfully used for other linear problems in biology [Floßdorf, 2013, Buchholz et al., 2013]. Here, the state space of the jump process is given by a vector containing the number of cells in each subpopulation, e.g. stem cells, progenitors and terminal cells in our example. In the case of linear systems, the method allows exact derivation of any moment of the solution probability distribution [Gillespie, 2008, Smadbeck and Kaznessis, 2013]. This probability distribution can be interpreted as the limit of a large number of experimental repetitions of independent evolutions of individual clones starting from identical initial conditions. However, because of the assumed migration events in the system discussed here, the experimental distributions found at any one time cannot be interpreted as stemming from many independent clones. Rather, some clones may have emerged at later time points than others and initial conditions for individual clonal growth may also vary because both *S*- and *P*-type cells are possible founders.

It is conceivable to instead formulate a jump process on a state space of clone sizes. A state vector could then contain the number of clones of size 1, 2, ..., n. Even though this vector is *a priori* infinitely long, the problem could in principle be addressed by truncating it to a biologically reasonable size. In this approach, new observed clones that come into existence via migration events could be added to the count for clones of size 1 with a rate proportional to the number of existing larger clones. However, because our model includes three different cell types with different properties regarding cell division and migration capacity, simply keeping track of all clone sizes in the system would not allow statements about the propensity of the formation of new clones or even the transition rates between clones. For illustration, a clone composed of five stem cells may produce new clones with propensity 5β and turn into a clone of size six with propensity $5(\lambda_S + \lambda_A)$, while a clone composed of five terminally differentiated cells can neither spawn new clones nor turn into a clone of size 6. It may, however, lose cells via cell death with propensity 5μ and thus turn into a clone of size 4.

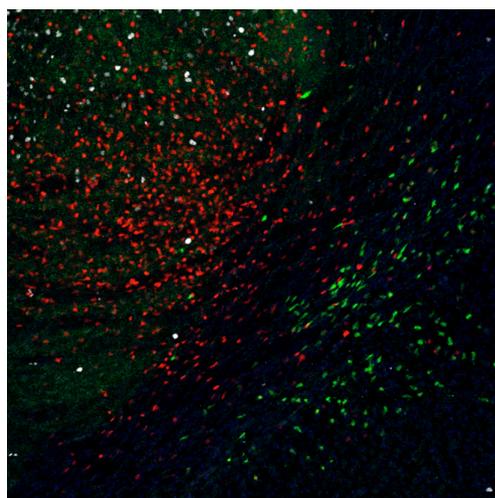
Because of the above, a comprehensive description of the stochastic evolution of a clone size distribution governed by three different cell types and allowing the emergence of new clones would require a state space that keeps track of both the number of clones at each time point and their exact composition. However, even if we cap the allowed clone size as well as the total number of clones considered, the required state space quickly becomes prohibitively large. As an example, looking at all possibilities of cellular combinations for clones between sizes 1 and 6 only, we would have to consider

$$\sum_{k=1}^6 \binom{3+k-1}{k} = 83 \text{ distinct clone types}, \quad (9.4)$$

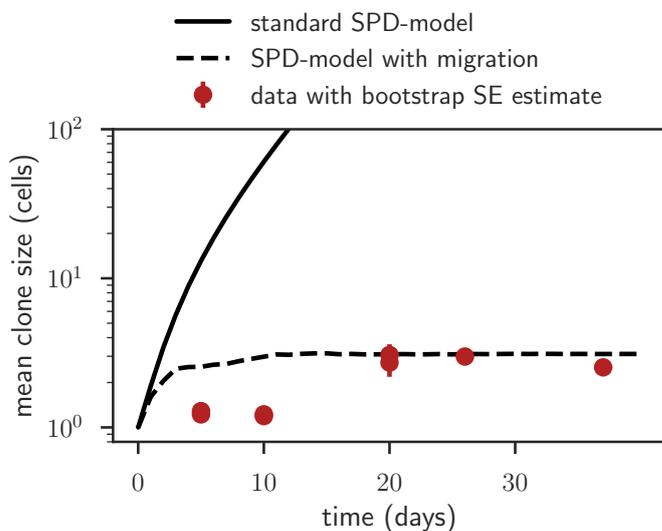
each of which has up to seven possibilities of turning into other clones and/or giving rise to



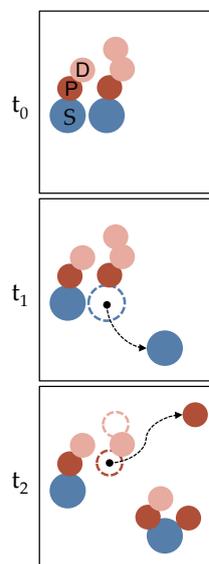
(a) Proliferating cells (stained for Ki67, white) are more abundant on the outside of the tumour (green cells). Image: Peng Zou.



(b) Tlx^+ cells and their recent progeny (green after induced switch of genetic labelling colour) are more abundant on the outside of the tumour (tumour core is towards the upper left corner in this case). Image: Peng Zou.



(c) Experimental mean clone size as a function of time (red dots) is not compatible with the bulk SPD-model from section 8, but could be explained when assuming that cellular migration leads to the observation of a higher number of smaller clones (see subfigure d). Experimental data: M. Amir Khan.



(d) Cellular migration leads to the observation of more and smaller clones.

Figure 9.5.: Spatial heterogeneity of mouse model glioblastoma (a, b) with illustrations explaining why the simple population model from section 8 is incompatible with the single-cell data (c) and how migration of individual cells can keep clone size small while increasing the number of clones identified (d).

new ones (three division events, two migration events, one differentiation and one cell death). Even disregarding the fact that a maximum clone size of six would not be of much help in attempting to describe the highly skewed data distributions in figure 9.3, operations on this state space are inconvenient and numerical cost is bound to explode. As a result, even though it may in principle be possible to define a linear master equation on a space of distributions over types of clones, there is little hope for practical benefit. Thus, we resort to stochastic simulation.

We make use of a dynamic Monte Carlo method commonly referred to as Gillespie's algorithm in Systems Biology. It produces trajectories of stochastic equations by keeping track of the exact numbers of each species present and choosing the next event as well as the waiting time until this event using a Monte Carlo step based on the species' abundance and the events' kinetic rates [Gillespie, 1977]. In order to include the emergence of new clones via migration into this framework which *a priori* deals with trajectories of individual populations, every time a cell of type S or P leaves the simulated clone, the size of this clone is reduced by one and the exact time of migration is recorded. After the original clone has been simulated until the required time, all clones spawned by migration and their offspring are likewise simulated from their respective time points of emergence. By repeating this process for a high number ($\mathcal{O}(10^3)$) of initially labelled clones and combining the information about the size of each clone, including those emerged later via migration, at chosen time points, the simulated histograms on the right side of figure 9.3 were produced.

Looking at the data summary statistics in figure 9.4, the mean clone size evolves in a peculiar fashion - while it remains below 1.5 cells per clone until day 10, the four animals sacrificed after day 20 show similar mean clone sizes between 2.5 and 3. A curve tracing all data points would thus have to be of sigmoidal shape with a sharp rise between day 10 and day 20. Our simple model does not offer this flexibility and it is also unclear which biological effects might cause such behaviour. Thus, we focus on the four animals sacrificed at later time points for quantitative comparison. As observed above, the mean clone size as well as the corresponding coefficient of variation and the fractions of size 1 and 2 clones seem to have reached saturation by day 20 as the statistics of all four late animals fall mostly within each others standard error estimates and no clear trends are discernible. Therefore, using the assumption that these four statistics have reached steady state, we calculate the mean of each of them for further comparison to the model. A conservative error estimate associated with these mean statistics is derived by Gaussian propagation of the four bootstrapped standard errors resulting in

$$\begin{aligned}\mu_{\text{exp}} &= (2.8 \pm 0.5) \\ \text{CV}_{\text{exp}} &= (2.1 \pm 0.5) \\ \mathcal{F}_{\text{exp}}^{\text{size } 1} &= (0.56 \pm 0.04) \\ \mathcal{F}_{\text{exp}}^{\text{size } 2} &= (0.18 \pm 0.03) .\end{aligned}$$

In order to use this information in parameter estimation, the stochastic simulation starting from a single labelled cell has to be repeated enough times so that the true clone size distribution at the chosen day is sufficiently approached. Here, we find that when starting from 100 labelled cells and simulating until day 20, the four chosen summary statistics fall within a

narrow interval of less than 5% deviation from the mean upon multiple repetitions of the entire process (less than 1% for the fractions of size 1 and 2 clones). Additionally, a steady behaviour of the four statistics was reached by day 20 for all tested parameter combinations θ within the bounds given in table 8.1 supplemented by different migration rates β . Thus, because the simulation uncertainty of the steady state statistics is at least one order of magnitude below the experimental error, we continue to use the parameter estimation procedure introduced in section 8.3 as an approximation for this noisy optimisation problem - specifically, we define the likelihood of the clonal tracing data set in the same way as for the bulk datasets using the experimental error to weight the squared distance between model and experimental values. Uncertainties in the model values due to the stochastic nature of their simulation are omitted. While a gradient-based optimisation algorithm would likely have problems with the remaining fluctuations in the simulation outcomes when moving through parameter space in small steps, the MCMC-approach employed here is robust to this type of noise. In the following, the set of six parameters comprising $\lambda_S, \lambda_A, \lambda_1, d_1, \mu_1$ and β is referred to as θ^\dagger . We find for the approximated logarithmic likelihood contribution of the single-cell data:

$$\begin{aligned} \log \mathcal{L}_{\text{clones}}(\mu_{\text{exp}}, \text{CV}_{\text{exp}}, \mathcal{F}_{\text{exp}}^{\text{size } 1}, \mathcal{F}_{\text{exp}}^{\text{size } 2}, \sigma_\mu, \sigma_{\text{CV}}, \sigma_{\mathcal{F}1}, \sigma_{\mathcal{F}2} | \theta^\dagger) = \text{const.} + \\ \frac{(\mu_{\text{exp}} - \mu_{\text{mod}}(\theta^\dagger))^2}{2\sigma_\mu^2} + \frac{(\text{CV}_{\text{exp}} - \text{CV}_{\text{mod}}(\theta^\dagger))^2}{2\sigma_{\text{CV}}^2} + \frac{(\mathcal{F}_{\text{exp}}^{\text{size } 1} - \mathcal{F}_{\text{mod}}^{\text{size } 1}(\theta^\dagger))^2}{2\sigma_{\mathcal{F}1}^2} + \\ \frac{(\mathcal{F}_{\text{exp}}^{\text{size } 2} - \mathcal{F}_{\text{mod}}^{\text{size } 2}(\theta^\dagger))^2}{2\sigma_{\mathcal{F}2}^2}. \end{aligned} \quad (9.5)$$

We add this logarithmic likelihood contribution to equation (8.16) to arrive at a new total likelihood including the single-cell data,

$$\begin{aligned} \log \mathcal{L}_{\text{total}}(D^* | \theta^\dagger) = \log \mathcal{L}_{\text{BLI}}(\lambda_0, \sigma_{\lambda_0} | \lambda_S) + \log \mathcal{L}_{\text{Ki67}}(\{\mathcal{F}_{\text{exp}}^{\text{Ki67}}, \sigma_{\mathcal{F}}\} | \theta) + \\ \log \mathcal{L}_{\text{comp}}(\mathcal{F}_{\text{exp}}^{\text{p/b}}, \mathcal{F}_{\text{exp}}^{\text{s/p}}, \sigma_{\mathcal{F}_{\text{p/b}}}, \sigma_{\mathcal{F}_{\text{s/p}}} | \theta) + \\ \log \mathcal{L}_{\text{clones}}(\mu_{\text{exp}}, \text{CV}_{\text{exp}}, \mathcal{F}_{\text{exp}}^{\text{size } 1}, \mathcal{F}_{\text{exp}}^{\text{size } 2}, \sigma_\mu, \sigma_{\text{CV}}, \sigma_{\mathcal{F}1}, \sigma_{\mathcal{F}2} | \theta^\dagger). \end{aligned} \quad (9.6)$$

We assume the prior distribution of the migration rate β to match those of the other five parameters, given in equation (8.17) - a uniform distribution between 0 and 4/day.

Performing the MCMC-sampling procedure in the same way as described in section 8.3 yields an approximation of the posterior distribution. By plotting each sampled parameter set's posterior probability as a function of each of the six parameters individually yields the posterior profiles shown in figure 9.6. Comparing figures 8.5 and 9.6, the fuzzy borders of the accessible posterior probability region in the latter stand out. These arise because of the stochastic nature of the simulation of clonal tracing data - two realisations of the simulation with the exact same set of parameters will not result in the exact same clone size distribution and thus not in the same summary statistics. As discussed above, we have simulated a number of initially labelled cells high enough so that the simulation variance is negligible compared to the experimental uncertainty.

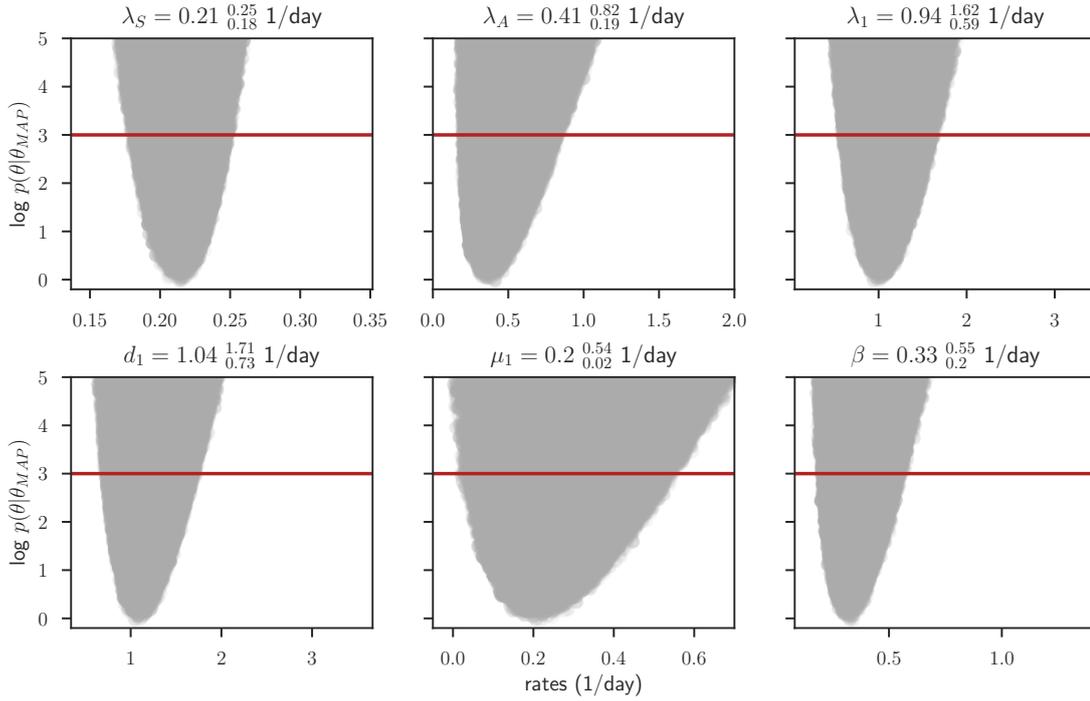


Figure 9.6.: Posterior profiles of the six model parameters λ_S , λ_A , λ_1 , d_1 , μ_1 , θ and β as discussed in section 9.2. The logarithm of the relative probability as defined in equation (8.23) is shown on the y -axis, with values above the red line corresponding to less than 5% of the probability associated with the best parameter set. As above (figure 8.5), the shaded area shows which relative probabilities are accessible for a given parameter value (shown on the x -axis). We define the parameter values where the red line cuts the shaded area as boundaries of the credible region of 5% relative probability. The fuzzy borders of the shaded area are due to the stochastic nature of the single-cell data simulation.

The parameter estimates with accompanying $\alpha = 0.05$ bounds for the more comprehensive model and data are shown above each posterior profile in figure 9.6. The previously determined values λ_S , λ_A , λ_1 , d_1 and μ_1 are in very good agreement with the earlier estimate, no significant shifts of the MAP values are observed. Through the addition of the single-cell data set, the width of the bounds on λ_A and μ_1 could be reduced; the bounds on λ_1 and d_1 are still very wide as both parameters remain strongly correlated. The newly introduced model parameter β is fully identifiable with a MAP value of 0.33 per day - this suggests that it takes cells of type S and P on average roughly 3 days to migrate far enough away that they cannot be identified as a member of the clone of origin any more by the observer. Considering that migration speeds of the order of $10 \mu\text{m}/\text{h}$ or higher are possible in the context of neural development [Hayashi et al., 2015, Cayre et al., 2009] and assuming a cellular diameter of $20 \mu\text{m}$, an order of magnitude common in mammalian cells [Milo and Phillips, 2015], 72 hours would in principle allow cells to migrate more than 30 cell diameters away from their original position. It is obvious that this may introduce problems regarding the unambiguous identification of clonal members in fluorescence microscopy even if initial labelling was sparse.

In recent years, with the emergence of single-cell labelling methods, studies exploiting the statistical information of clonal growth using stochastic modelling techniques have become increasingly popular, some employing fluorescence microscopy for *in situ* lineage tracing [Driessens et al., 2012, Blanpain and Simons, 2013, Watson et al., 2015, Davis et al., 2016] while others rely on methods such as flow cytometry or DNA sequencing outside the tissue of origin [Buchholz et al., 2013, Lan et al., 2017, Pei et al., 2017]. In the latter case, unique labels are needed for each clone because spatial information is lost during the analysis. For *in situ* lineage tracing methods however, few or even a single colour can be sufficient if initial labelling is performed sparsely. Spatial information is retained and can be exploited for identifying clonal members which are assumed to lie within close proximity of each other. However, in such methods a higher level of personal judgement of the experimentalist is required. Unambiguously deciding whether a cell a certain distance from another belongs to the same clone can be impossible in expanding tissues or systems with high cellular migration like the one at hand. Thus, when modelling such data mathematically, caution should be exercised and potential effects of migration or expansion on the model outcome should be examined.

9.3. Treatment Approaches

In this section, two approaches to treating glioblastoma are discussed and explored both within the modelling framework and experimentally. First, in section 9.3.1, the effect of chemotherapy on tumour growth is examined: despite an initial strong decrease in tumour volume, relapse is inevitable here. Second, in section 9.3.2, we look at the effects a knockdown of the stem cell marker Tlx has on the population dynamics. Here, we find that a long-term reduction of tumour volume should in principle be possible.

9.3.1. Chemotherapy with Temozolomide

Temozolomide (TMZ) is a chemotherapeutic anti-cancer drug that acts via DNA-methylation of mostly *N7* or *O6* guanine residues. During DNA replication this ultimately results in G_2/M cell cycle arrest and apoptosis [Zhang et al., 2012]. Thus, in the presence of TMZ, division attempts of cancerous cells result in cell death, while resting cells are not affected in the same way [Beier et al., 2008]. In order to model the effects of this drug, we alter the basic SPD-model given in equation (8.1) such as to incorporate them. Specifically, in the presence of the drug, we do not allow cell division to result in two daughter cells any more, but set it to lead to cell death instead. This concerns both symmetric and asymmetric division attempts of stem cells as well as divisions of progeny cells. The behaviour of fully differentiated cells in turn is not changed. This results in the following set of differential equations:

$$\begin{aligned}\dot{S}(t) &= -(\lambda_S + \lambda_A) S(t) \\ \dot{P}(t) &= -\lambda_1 P(t) - d_1 P(t) \\ \dot{D}(t) &= d_1 P(t) - \mu_1 D(t).\end{aligned}\tag{9.7}$$

Of note, we have not introduced any additional parameters into the model here.

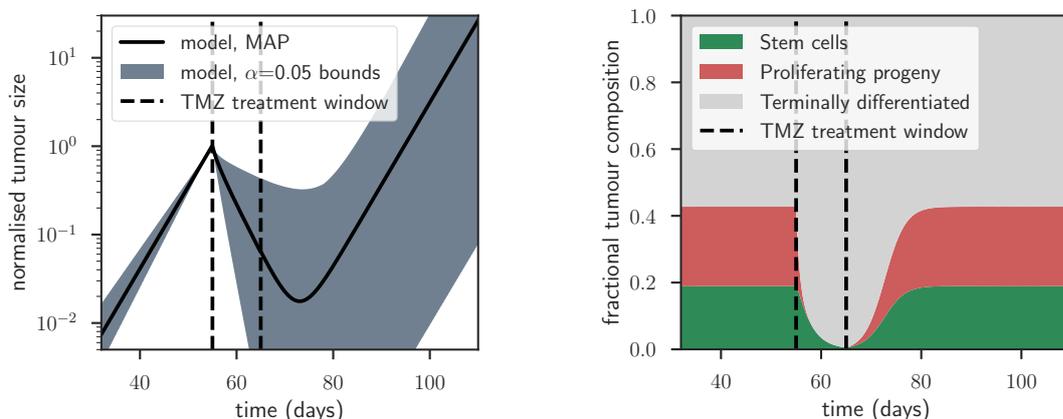
To mathematically replicate the treatment schedule, the tumour is first evolved in an undisturbed fashion until day 55, the approximate time point of experimental intervention (see figure 9.7c). For this first phase, equations (8.1) are used. As treatment is experimentally applied for 10 days, equations (9.7) are used between day 55 and day 65 before returning to the undisturbed system thereafter. The model prediction resulting from this schedule is shown in figure 9.7a for the maximum MAP parameter set $\theta_{\text{MAP}}^\dagger$ and corresponding $\alpha = 0.05$ bounds. The bounds were generated by simulating the system with a large sample of posterior parameter sets satisfying $p(\theta|\theta_{\text{MAP}}) \leq 0.05$ and then extracting the minimum and maximum resulting tumour sizes at each time point from the set of solutions.

Following the start of chemotherapy, a sharp decline in modelled tumour mass by a factor of 59 (for the MAP estimate, black curve) is observed. The reduction phase continues after the end of the treatment so that the minimum tumour volume is reached 18 days after the start of chemotherapy. Relapse then ensues following the same characteristic exponential growth displayed by the primary, and the modelled tumour regains its pre-treatment volume 40 days after treatment begins.

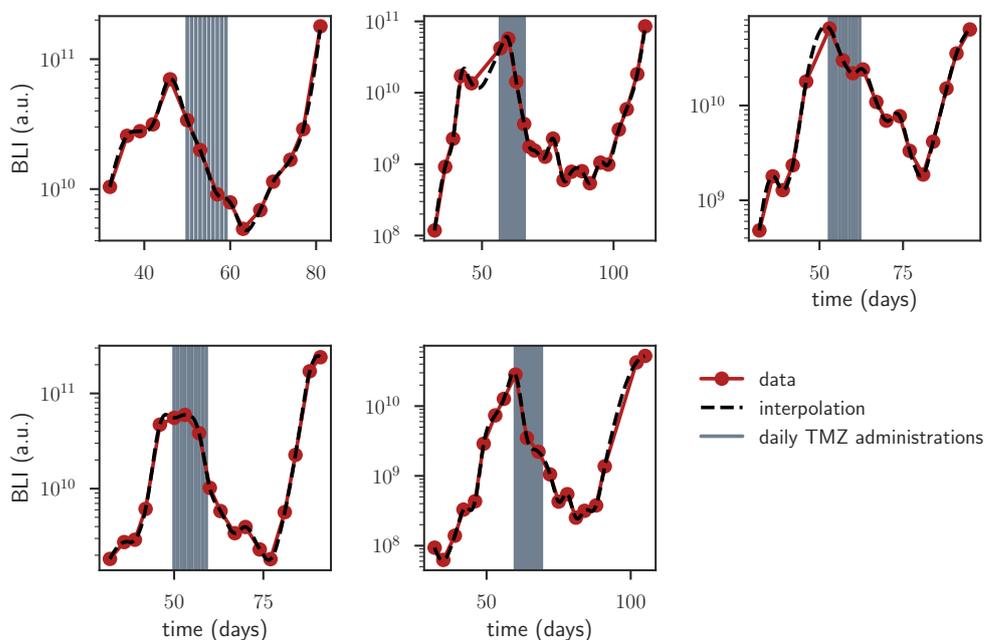
Looking at the fractional composition of the tumour following treatment illustrates how proliferating cells are depleted to very small amounts but nevertheless re-emerge after the end of the treatment phase (figure 9.7b). Following the start of chemotherapy, the fast-dividing proliferating progeny cells (red population) are quickly depleted, while the slower dividing stem cells are lost more slowly (green population). At the end of the treatment, the tumour consists almost entirely of terminally differentiated cells. However, the remaining stem cells quickly repopulate the system and the tumour returns to its original composition about a month after the start of treatment.

Experimental bioluminescence time curves of tumour-bearing mice receiving 10 days of TMZ at the age of around 55 days were recorded for five individual animals. The results are shown in figure 9.7c, with the time span of chemotherapy administration being indicated in grey. These tumours took 14 to 34 days until they reached their smallest volume and 28 to 54 days until they grew back to their original size. The smallest volume was between 7 and 115 times smaller than the volume at the beginning of the treatment. All these values are in excellent agreement with the model prediction; their rather wide range suggests significant inter-animal variability in tumour growth dynamics and treatment response.

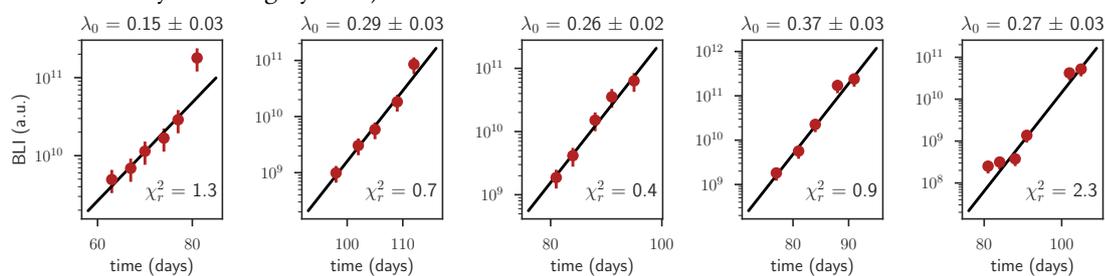
Cancer therapies such as treatment with temozolomide shape the evolution of tumour size and genetic composition via treatment-induced mutagenesis and selective pressure [Burrell et al., 2013, Venkatesan et al., 2017]. In glioblastoma, the relapsed tumour is often found to grow faster and more aggressively partly due to new TMZ-induced driver mutations [Johnson et al., 2014]. Generally, genetic heterogeneity is a major contributor to diverging and unsatisfactory therapy results [McGranahan and Swanton, 2015]. Nevertheless we have chosen not to include genetic heterogeneity into the mathematical models of the present study, as because of the



(a) Prediction using the model including chemotherapy in equations (9.7). (b) Relative model tumour composition before and after chemotherapy (shown for θ_{MAP}^+).



(c) BLI curves of tumour-bearing animals treated with temozolomide for 10 days (treatment time points indicated by vertical grey lines).



(d) Exponential fits (using equation (7.1)) to the BLI values of the relapse phase (blue to salmon circle in part d). Error model as in figure 7.1.

Figure 9.7.: Effect of 10-day chemotherapy treatment on mathematical model (parts a and b) and experiment (parts c and d). Experimental data: Peng Zou.

process of tumour induction used here in which surplus copies of important growth genes are introduced into neuronal stem cells, tumour growth is already extremely aggressive from the start. In order to further legitimise this simplification for the scope of this study, we analysed our bioluminescence data of TMZ-treated animals (figure 9.7c) for evidence that the relapse growth rate is higher than the primary growth rate. To this end, we fitted the phenomenological growth model of equation (7.1) to the relapse phase and compared the result to the previously identified primary growth rate

$$\lambda_0 \pm \sigma_{\lambda_0} = (0.21 \pm 0.02) \text{ day}^{-1} .$$

For the five TMZ-treated animals shown in figure 9.7d, we find a mean growth rate of

$$\lambda_{\text{relapse}} \pm \sigma_{\lambda_0} = (0.27 \pm 0.04) \text{ day}^{-1} ,$$

which is not significantly different from λ_0 (Student's t -test, $p = 0.18$). Thus there is no reason to alter the assumption that primary and relapse growth phase do not differ from each other in this mouse model.

9.3.2. Induced knockdown of Tlx

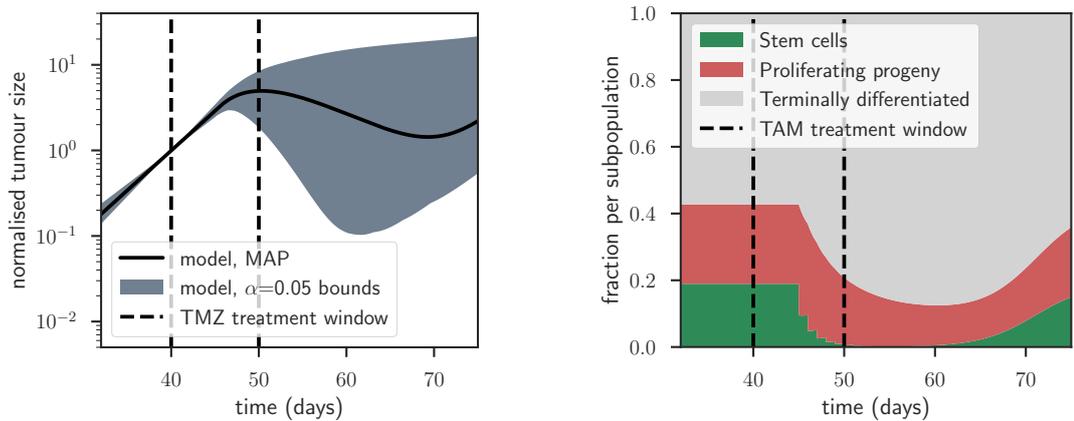
It has previously been shown that a knock-out of Tlx has strong beneficial effects on animal survival in the tumour model discussed in this study [Zhu et al., 2014]. In the present work, a miRFlex system introduced via a viral vector was used to allow inducible knock-down of Tlx. Induction was then triggered by 10 consecutive injections of tamoxifen (TAM) and a CreERT2 excision system.

In order to model this experimental approach, we assume that each administration of TAM permanently turns a fraction ϵ of the current stem cell population into non-self-renewing progenitors of type P irreversibly because of the reduction of active Tlx following the knock-down. The reason for a permanent instead of a transient switch lasting only for the time TAM is administered is that Cre drives an irreversible genetic change that remains after the end of TAM presence. However, because mRNA and protein have finite half-lives and the genetic switch is thus assumed to require some time until it takes effect, we assume that the conversion $S \rightarrow P$ in the model does not happen instantaneously but with delay τ . Otherwise, the equation system is preserved; cells behave as described in equations (8.1).

Figure 9.8a shows the model prediction resulting from this schedule for the maximum MAP parameter set $\theta_{\text{MAP}}^{\dagger}$ and corresponding $\alpha = 0.05$ bounds when using a conversion efficiency of $\epsilon = 50\%$ and a conversion delay of $\tau = 5\text{d}$. Interestingly, the tumour continues to grow for approximately 10 days following the start of the treatment. In this time, its volume is increased by a factor of around 3. Thereafter, the tumour mass starts to decline slowly. The similarity between the proliferation rate of progeny cells λ_1 and their differentiation rate d_1 plays a major role for how long stem-cell depleted tumours can retain a stable size: if $\lambda_1 = d_1$, the tumour volume can in principle remain stable infinitely. For all cases where $\lambda_1 < d_1$, a stem-cell depleted tumour will shrink with a time constant dependent on $\Delta = d_1 - \lambda_1$. For $\theta_{\text{MAP}}^{\dagger}$, λ_1 and

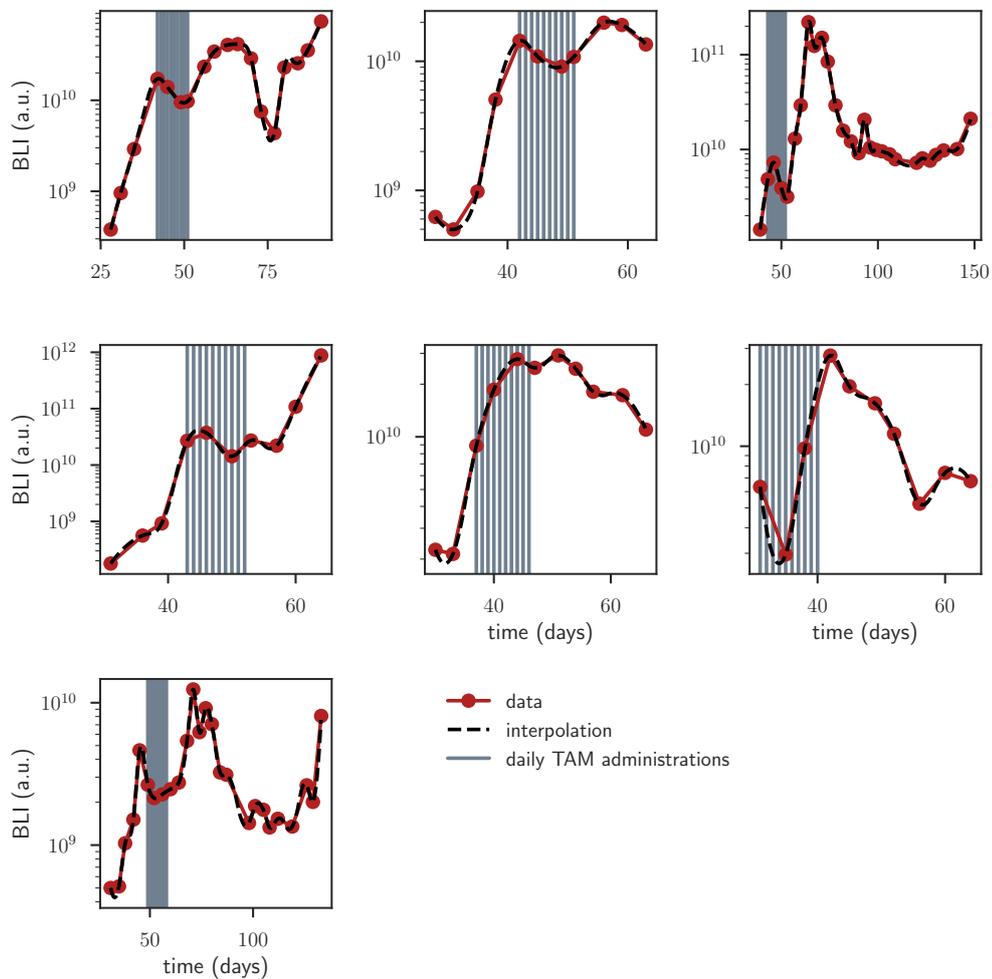
d_1 are within 15 % of each other resulting in the slow decline shown in figure 9.8a. The behaviour is illustrated in figure 9.8b - following the fast depletion of stem cells (green), the progeny compartment undergoes only a slow decay while the tumours distribution into proliferating progeny and terminally differentiated cells remains rather stable.

The experimental data in the case of Tlx-knockdown treatment is much less regular than in the case of chemotherapy with TMZ. The response of seven animals to the treatment is shown in figure 9.8c. In general, following the administration of the Cre-inducing agent TAM, the tumour volume dropped slightly before starting to grow again and reaching values generally 2 to 4 times higher than the pre-treatment volume. This peak in tumour size generally occurred 10 to 20 days after TAM administration. After this peak, tumour volume declined at varying speeds. In several cases, tumours eventually relapsed after a few weeks of remission, indicating that stem-cell depletion had worked only partially. While the model currently does not comprise a mechanism allowing the experimentally observed initial drop in tumour mass, the transient continued growth before remission is observed as predicted (expansion factor around 3 after 10 days before decline). Again, the wide range of experimentally observed behaviours in the declining phase may be indicative of heterogeneous dynamics between tumours.



(a) Prediction using a conversion efficiency of $\epsilon = 50\%$ per TAM-administration and a delay of $\tau = 5\text{d}$.

(b) Relative model tumour composition before, during and after TAM-administration.



(c) BLI curves of tumour-bearing the Tlx-MiRFlex construct and given tamoxifen at each time point marked with a grey vertical line. Behaviours are heterogeneous, but tumours continue to grow after the start of the treatment before eventually decreasing in size in most cases.

Figure 9.8.: Effect of Tlx knockdown treatment on mathematical model (parts a and b) and experiment (part c). Experimental data: Peng Zou.

10. Discussion - Cancer Stem Cells in Glioblastoma

The strength of this work lies in the direct connections it makes between simple theoretical models and experimental observations. This is true for both the three-dimensional tumour simulation and the compartmental differential equation model. While more sophisticated models have previously been developed in both categories, direct comparison to experimental data and model quantification is still largely wanting. Here, we find that cellular dispersal is a direct consequence of experimentally observed growth patterns and show that stem cells on average divide with a high rate. This in turn leads to a cancer stem cell fraction of almost 20% within tumours. In a further step, we use our quantified model to show that single-cell tracing data have to be handled with care in the absence of unambiguous clonal bar codes. Lastly, the comparison of model predictions to treatment data reveals both strengths and limitations of the model: while general features and time scales of the primary treatment response are correctly predicted, later events and details call for model extensions.

Matching Experimental and Theoretical Growth Curves In chapter 7, we have analysed tumour growth curves from 23 untreated animals recorded with the help of bioluminescence imaging. We tested which of three commonly employed tumour growth models - exponential growth, Gompertzian growth or linear radial growth - described the experimental data best and identified exponential growth as the most adequate description in the majority of cases. In order to preserve information about biological variance in tumour growth, we fitted each animal individually and thus arrived at a distribution of growth rates with a mean and standard deviation of $\lambda_0 = 0.21 \pm 0.07 \text{ day}^{-1}$. This translates to a tumour doubling time of only 3.3 days on average.

Next, we set up a three-dimensional grid-based simulation of solid tumour growth which included self-renewing and non-self-renewing species. With the help of this model we explored three different migration scenarios: (1) neither cell type is allowed to migrate, (2) non-self-renewing cells are allowed to migrate towards lower cell density and (3) self-renewing cells are allowed to migrate towards lower cell density. We find that scenarios (1) and (2) are not compatible with sustained exponential tumour growth, as self-renewing cells are unable to find enough space for continuous replication. In scenario (3), however, as self-renewing cells themselves migrate outwards, they routinely find themselves on the edge of the tumour and can thus divide without hindrance. The accumulation of Tlx⁺ self-renewing cells on the outside of mouse model glioblastoma could be verified experimentally. Likewise, analysis of single cell RNA-sequencing data from human glioblastoma samples showed Tlx to be differentially higher expressed in tumour cells extracted from the surrounding of the tumour as compared to its core.

Time-series data of mouse model brain tumours has been produced before [Szentirmai et al., 2006, Hashizume et al., 2010, Jarzabek et al., 2013] and, although this was not rigorously tested in these studies, yielded growth patterns compatible with the exponential behaviour identified here. Likewise, as introduced in section 6.2.2, spatial aspects of solid tumours including glioblastoma have received considerable attention in the literature and possible growth curves are a matter of active discussion because of their relevance for prognostics in the human setting [Murphy et al., 2016]. On the theoretical side it has been shown that migration drives tumour expansion speed [Enderling et al., 2009, Waclaw et al., 2015]. However, we are not aware of any previous study making the direct link between experimentally observed growth patterns and mechanistic assumptions. Here, we show that cellular dispersal is a necessary condition for the observed tumour expansion in our system. Furthermore, we have predicted and experimentally confirmed the resulting accumulation of self-renewing cells on the shell of the tumour. Our result suggests that such direct confrontation of simulated and observed tumour growth can help judge the likelihood of detailed underlying mechanisms as phenotype-specific migration behaviour.

A Highly Active Stem Cell Population In chapter 8 we have introduced a simple compartmental model of three coupled ordinary differential equations describing the dynamics of glioblastoma stem cells, their proliferating progeny and terminally differentiated tumour cells respectively. This model was motivated by direct experimental observations of our collaborators: in clonal tracing experiments, symmetric as well as asymmetric division of Tlx-GFP^+ cells was observed in fluorescence microscopy images; Tlx^- cells alone were unable to develop into full tumours after xenotransplantation; the proliferation index within a given set of Tlx^- cells decreases over time; and Tlx^- cells are eventually lost from the system. Mathematically, we have transcribed this into symmetric and asymmetric division of stem cells, symmetric division and differentiation of proliferating progenitors and death of terminally differentiated cells using mass-action kinetics. Importantly, this simple hierarchical model has previously been suggested in normal neurogenesis [Ponti et al., 2013] as well as in glioblastoma [Lan et al., 2017].

Basic properties of this model were then derived analytically. These include the fact that the overall growth of the tumour bulk is driven by the symmetric division rate of stem cells alone, remaining completely unaffected by the dynamics of the non-self-renewing downstream populations. Further, after a short initial development phase, the mathematical tumour displays constant fractions of the tree populations throughout the entire exponential growth phase. Making use of these distilled findings as well as of the model solutions, we set out to quantify the five event rates jointly using several experimental data sets. First of all, this included the overall tumour growth rate extracted from the bioluminescence imaging data which was directly linked to the symmetric division of stem cells in the model. Secondly, we used a data set characterising the proliferation index within an initially labelled population of proliferating progeny as a function of time. Lastly, making use of the fact that the population fractions are stationary, we included fractional measurements of cancer stem cells within the total cell number, proliferating cells within the total cell number and stem cells within all proliferating cells.

Using a Bayesian approach to parameter estimation enabled by Markov Chain Monte Carlo sampling of the parameter space we succeeded to identify all five event rates and can give upper and lower bounds for each of them. In summary, we find that stem cells divide symmetrically about once in five days while they produce a proliferating progenitor via asymmetric division about twice as often. Taken together, stem cells on average divide regularly, slightly more often than once in two days. Lower down in the hierarchy, proliferating progenitors divide symmetrically once per day and turn into terminally differentiated cells only slightly faster. While the proliferating compartment is thus not able to drive tumour expansion, it can maintain tumour volume for several days or weeks in the absence of stem cells. Our results for the progenitor compartment are in good agreement with the findings of [Ponti et al., 2013] for C and A cells (neural progenitors) in the normal system. This strengthens the hypothesis that glioblastoma progenitors by themselves resemble their normal counterparts whereas glioblastoma stem cells are continuously driven through cell cycle, having lost their ability to stay quiescent for prolonged periods of time.

Even though the historic idea of stem cells as a rarely dividing, quiescent population has recently been crumbling in its generality, specifically in the cancer context [Driessens et al., 2012], our rates do not agree with the recent results of [Lan et al., 2017] who suggest that asymmetric divisions of stem cells should happen at a much lower rate than the turnover of these progenitors themselves. However, the discrepancy is likely explained by the different models used, as [Lan et al., 2017] transplant a human glioma cell line into immuno-deficient mice - an approach resulting in much slower tumour progression. In our context, where excess growth genes have been introduced into the genome of neural stem cells, a high overall division rate and little evidence for a dominant quiescent population appear fitting.

While the existence of a quiescent stem cell population in this system was previously reported [Zhu, 2013], our current data does not contain enough information to explicitly include this into the model and subsequently identify all associated rates. We thus understand the model rates concerning the stem cell population as describing the average dynamics of this phenotype. This does not exclude the existence of a quiescent population, neither on a population level nor on the level of individual cells. In the same spirit, we have included the finding by [Zhu, 2013] that 83% of Tlx-GFP⁺ reside in the G₀ phase of the cell cycle into the calculation of actively dividing stem cells. Looking closer at the findings of [Ponti et al., 2013] in the normal system, it is striking that activated B cells (neural stem cells) display a particularly short S phase of only 4 hours. It is conceivable that this behaviour is conserved or even accentuated in the malignant case with glioblastoma stem cells rushing through short S/G₂/M phases and then remaining in a G₀-like state for slightly longer time, reminiscent of their normal behaviour. More research into existence and nature of quiescent stem cell population in glioblastoma and in the animal tumour model underlying this study in particular is required.

Cancer stem cells were historically believed to constitute only minor fractions on the order of one in a thousand or even one in a million tumour cells, a notion mostly based on assays with immuno-deficient mice [Quintana et al., 2008]. More recently, significantly higher fractions of

tumour stem cells have been reported [Quintana et al., 2008, Enderling, 2015]. Based on our quantified mathematical model, we are likewise able to quote the fraction of tumour mass occupied by each phenotype, even though they are in part not accessible experimentally. We find that stem cells make up 19% of the tumour mass while progenitors and differentiated cells contribute 24% and 57% respectively. Since the proportion of stem cells in a tumour is directly linked to its aggressiveness [Johnston et al., 2010], this high stem cell fraction may give another explanation for the poor survival of glioblastoma patients.

Migration Destroys Single-Cell Tracing Data In section 9.2 we discuss data from a clonal tracing study. Here, using inducible Cre, one of four fluorescent protein genes was randomly chosen for expression specifically within Tlx^+ cells. In the following, this genetic label was passed on to Tlx^+ as well as Tlx^- progeny of the originally labelled cell. In contrast to the data sets used for parameter estimation chapter 8, this data set is not based on an entire population of cells, but on the fates of individual stem cells. Such data is, in principle, highly informative because not only mean dynamics, but various higher order statistics describing the distribution of individual cell fates are contained. These can be exploited to deduce differentiation topology and dynamics, as has been shown in other systems [Driessens et al., 2012, Buchholz et al., 2013].

As we had described population-level growth of glioblastoma in terms of linear ordinary differential equations up to here, we naively assumed that a stochastic description based on a chemical master equation incorporating the exact same processes should be able to capture the dynamics of individual clones. The mean dynamics derived from this stochastic approach coincide with the previous deterministic description. However, looking at the experimental mean dynamics of clone sizes, we have discovered that these never go beyond a mean clone size of 3 cells, while obviously our exponentially growing population description quickly reaches orders of 10,000 or 100,000 cells. How can a population that is essentially the sum of individual clones behave this differently from its constituents? If it is not the number of cells per clone which increases exponentially, it must be the number of clones which is increasing fast. By experimental definition, where all cells derived from one initially labelled cell should be members of the same clone, this seems impossible at first glance but is quickly resolved when considering the highly migrative nature of the tumour as discussed above. With only four fluorescent colours available as clonal “bar codes”, it becomes difficult or impossible to experimentally assign cellular families. Migration gives rise to new subclones at distinct locations even if initial labelling was very sparse.

We have thus sought to integrate this difficulty into the mathematical framework. Because of the newly emerging clones with diverse cellular compositions, application of a conventional analytic approach using moment-generating functions is not practical. Instead, we have simulated the tree of subclones emerging from one initially labelled cell in a stochastic fashion. The approach requires the introduction of a new parameter, which is defined as the rate with which migrating cells move far enough away from their clone of origin to be identified as a separate clone by the experimentalist. This allows to query statistics such as mean clone size or variance based on all subclones present at a given moment, no matter whether they were spawned earlier or later. Our previously introduced framework of Markov Chain Monte Carlo

based sampling of the parameter space is able to give meaningful results even in the presence of simulation noise as we combine likelihood-based sampling with elements of approximate Bayesian computation as introduced by [Beaumont et al., 2002].

After re-sampling the parameter space comprised of the division and differentiation rates of the deterministic model as well as the new migration rate, all previously determined parameters emerge unaltered. The new migration rate is able to absorb the entire information contributed by the tracing data set while leaving other rates unchanged. This is an indicator for accuracy of the assumption that the tracing data set is a result of the same rules governing the population when additionally taking into account migration processes. The migration rate itself is estimated around 0.33/day, meaning that it takes a migrating cell on average 3 days to escape from its cellular union of origin. Following a rough estimation, in this time a migrating neural cell is able to cover as much as 30 cell diameters or more which should easily be enough to cause identification difficulties.

Two general messages can be taken from these proceedings. First, even though single-cell data is very rich in information, it is always beneficial to be in possession of population level experiments which serve as a control and prevent premature conclusions. This is especially true when the system at hand is incompletely understood. In the present case, it was the strong growth observed in the bioluminescence data which first cast doubt onto the saturating clone sizes in the single-cell tracing data set. Second, when designing single-cell experiments with the purpose of clonal lineage tracing, it is desirable to use a system allowing for high numbers of bar codes and thus unambiguous identification of clonal members. As single-cell sequencing methods become more and more readily available and affordable, such barcoding methods are likewise on the rise. Examples include lentiviral barcoding [Lan et al., 2017], TAM-inducible polylox barcoding [Pei et al., 2017] and CRISPR-Cas9 based barcoding [Alemany et al., 2018].

Is Treatment Futile? In section 9.3, we have adjusted the basic model from chapter 8 to allow the description of two alternative treatment scenarios. Adjustments to the model were kept to a minimum in order to allow quantitative predictions of treatment outcomes. In the first case, the model was adjusted to accommodate chemotherapy using the assumption that during administration of the drug, all attempted cell divisions fail leading to the immediate death of the dividing cell. All division rates in the model were turned into death rates thus avoiding the need to introduce additional parameters. In the second case, two new parameters had to be introduced in order to model the knock-down of stemness. Experimentally, the stem cell factor Tlx is knocked down using inducible micro RNAs. In the model, this is incorporated by turning a given fraction of stem cells into progenitors with a given delay upon each administration. Motivated by experimental observations, the delay was chosen to equal 5 days. For total want of information, the knock-down efficiency per single tamoxifen injection was set to 50%. Surprisingly, as outlined below, this choice leads to decent agreement with experiment.

The predictions produced by the parametrised model showed excellent agreement with experimental data in the case of chemotherapeutic treatment. The model predicted the correct magnitude of the treatment-induced drop in tumour volume as well as the correct timescales

for drop and relapse. Interestingly, this strong agreement is produced by a model which does not assume a quiescent and thus protected cell population. The fractional difference in death events during treatment is brought about solely by the difference in division rates between stem and progenitor cells and the general tendency of the latter to turn into terminal cells and, eventually, die. However, because of the simple nature of the model which does not allow for selection of cells with better survival potential or even treatment-induced mutation, the modelled tumour volume would react in exactly the same way upon a second phase of chemotherapy. Even though the growth rate of experimental tumours during the relapse phase is not significantly different from the primary growth phase, this behaviour is not observed. Here, treatment of the relapsed tumour with the same drug leads to minor volume reduction at the most. This indicates that a selection process for a resistant subpopulation is induced by the first wave of treatment; what exactly is selected for has yet to be determined.

In the case of Tlx knock-down, some but not all of the experimentally observed features in tumour volume time curves are predicted. It is particularly unclear how a regularly observed initial drop in tumour volume which appears in 6 out of 7 animals can be explained or modelled. However, the general quality and reproducibility of the knock-down data is worse than the chemotherapy data due to many uncertainties with the miRFlex strategy. Before further inquiring into features of these data, the experimental strategy should be stabilised, for example by replacing the knock-down of Tlx by inducible knock-out. With perfect or close-to-perfect efficiency however, this strategy has the potential to significantly prolong animals survival or even lead to a complete cure.

Overall, while our parametrised model succeeds in predicting elementary features of the treatment response, a more detailed knowledge of escape processes would be required and have to be implemented in order to make the model useful for treatment optimisation or true survival prediction. Such escape processes may comprise induced quiescence, mutational evolution or phenotypic plasticity. Even if the relevant mechanisms were known and incorporated into a deterministic continuous model, one question remains: how much reduction of the cancer stem cell population is enough reduction; when, if ever, will the body's defense mechanisms be able to deal with the rest?

Bibliography

- [Adams et al., 2016] Adams, R. M., Mora, T., Walczak, A. M., and Kinney, J. B. (2016). Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *eLife*, 5(DECEMBER2016):1–27.
- [Akaike, 1998] Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pages 199–213. Springer.
- [Alemany et al., 2018] Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J., and van Oudenaarden, A. (2018). Whole-organism clone tracing using single-cell sequencing. *Nature*, 556(7699):108–112.
- [Alfonso et al., 2016] Alfonso, J. C., Köhn-Luque, A., Stylianopoulos, T., Feuerhake, F., Deutsch, A., and Hatzikirou, H. (2016). Why one-size-fits-all vaso-modulatory interventions fail to control glioma invasion: In silico insights. *Scientific Reports*, 6(July):1–15.
- [Alfonso et al., 2017] Alfonso, J. C. L., Talkenberger, K., Seifert, M., Klink, B., Hawkins-Daarud, A., Swanson, K. R., Hatzikirou, H., and Deutsch, A. (2017). The biology and mathematical modelling of glioma invasion: a review. *Journal of The Royal Society Interface*, 14(136):20170490.
- [Allen et al., 2007] Allen, C. D. C., Okada, T., Tang, H. L., and Cyster, J. G. (2007). Imaging of germinal center selection events during affinity maturation. *Science (New York, N.Y.)*, 315(5811):528–531.
- [Allen et al., 1987] Allen, D., Cumano, A., Dildrop, R., Kocks., C., Rajewsky, K., Rajewsky, N., Roes, J., Sablitzky, F., and Siekevitz, M. (1987). Timing, Genetic Requirements and Functional Consequences of Somatic Hypermutation during B-Cell Development. *Immunological Reviews*, 96(1):5–22.
- [Allen et al., 1988] Allen, D., Simon, T., Sablitzky, F., Rajewsky, K., and Cumano, A. (1988). Antibody engineering for the analysis of affinity maturation of an anti-hapten response. *The EMBO Journal*, 7(7):1995–2001.
- [Allman et al., 1999] Allman, D., Li, J., and Hardy, R. R. (1999). Commitment to the B Lymphoid Lineage Occurs before D \times H \times J \times H Recombination. *The Journal of Experimental Medicine*, 189(4):735–740.
- [Amanna et al., 2007] Amanna, I. J., Carlson, N. E., and Slifka, M. K. (2007). Duration of Humoral Immunity to Common Viral and Vaccine Antigens. *New England Journal of Medicine*, 357(19):1903–1915.

- [Amino et al., 2006] Amino, R., Thiberge, S., Martin, B., Celli, S., Shorte, S., Frischknecht, F., and Ménard, R. (2006). Quantitative imaging of Plasmodium transmission from mosquito to mammal. *Nature Medicine*, 12(2):220–224.
- [Amitai et al., 2017] Amitai, A., Mesin, L., Victora, G. D., Kardar, M., and Chakraborty, A. K. (2017). A Population Dynamics Model for Clonal Diversity in a Germinal Center. *Frontiers in Microbiology*, 8(September):1–9.
- [Amzel and Poljak, 1979] Amzel, L. M. and Poljak, R. J. (1979). Three-dimensional structure of immunoglobulins. *Annual review of biochemistry*, 48(L):961–97.
- [Anderson et al., 2014] Anderson, J. C., Duarte, C. W., Welaya, K., Rohrbach, T. D., Bredel, M., Yang, E. S., Choradia, N. V., Thottassery, J. V., Yancey Gillespie, G., Bonner, J. A., and Willey, C. D. (2014). Kinomic exploration of temozolomide and radiation resistance in Glioblastoma multiforme xenolines. *Radiotherapy and Oncology*, 111(3):468–474.
- [Anderson et al., 2009] Anderson, S. M., Khalil, A., Uduman, M., Hershberg, U., Louzoun, Y., Haberman, A. M., Kleinstein, S. H., and Shlomchik, M. J. (2009). Taking Advantage: High-Affinity B Cells in the Germinal Center Have Lower Death Rates, but Similar Rates of Division, Compared to Low-Affinity Cells. *The Journal of Immunology*, 183(11):7314–7325.
- [Auton et al., 2015] Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., Mardis, E. R., Marth, G. T., McVean, G. A., Nickerson, D. A., Schmidt, J. P., Sherry, S. T., Wang, J., Wilson, R. K., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J. G., Zhu, Y., Chang, Y., Feng, Q., Fang, X., Guo, X., Jian, M., Jiang, H., Jin, X., Lan, T., Li, G., Li, J., Li, Y., Liu, S., Liu, X., Lu, Y., Ma, X., Tang, M., Wang, B., Wang, G., Wu, H., Wu, R., Xu, X., Yin, Y., Zhang, D., Zhang, W., Zhao, J., Zhao, M., Zheng, X., Gupta, N., Gharani, N., Toji, L. H., Gerry, N. P., Resch, A. M., Barker, J., Clarke, L., Gil, L., Hunt, S. E., Kelman, G., Kulesha, E., Leinonen, R., McLaren, W. M., Radhakrishnan, R., Roa, A., Smirnov, D., Smith, R. E., Streeter, I., Thormann, A., Toneva, I., Vaughan, B., Zheng-Bradley, X., Grocock, R., Humphray, S., James, T., Kingsbury, Z., Sudbrak, R., Albrecht, M. W., Amstislavskiy, V. S., Borodina, T. A., Lienhard, M., Mertes, F., Sultan, M., Timmermann, B., Yaspo, M. L., Fulton, L., Ananiev, V., Belaia, Z., Beloslyudtsev, D., Bouk, N., Chen, C., Church, D., Cohen, R., Cook, C., Garner, J., Hefferon, T., Kimelman, M., Liu, C., Lopez, J., Meric, P., O’Sullivan, C., Ostapchuk, Y., Phan, L., Ponomarov, S., Schneider, V., Shekhtman, E., Sirotkin, K., Slotta, D., Zhang, H., Balasubramaniam, S., Burton, J., Danecsek, P., Keane, T. M., Kolb-Kokocinski, A., McCarthy, S., Stalker, J., Quail, M., Davies, C. J., Gollub, J., Webster, T., Wong, B., Zhan, Y., Campbell, C. L., Kong, Y., Marcketta, A., Yu, F., Antunes, L., Bainbridge, M., Sabo, A., Huang, Z., Coin, L. J., Fang, L., Li, Q., Li, Z., Lin, H., Liu, B., Luo, R., Shao, H., Xie, Y., Ye, C., Yu, C., Zhang, F., Zheng, H., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Garrison, E. P., Kural, D., Lee, W. P., Leong, W. F., Stromberg, M., Ward, A. N., Wu, J., Zhang, M., Daly, M. J., DePristo, M. A., Handsaker, R. E., Banks, E., Bhatia, G., Del Angel, G., Genovese, G., Li, H., Kashin, S., McCarroll, S. A., Nemes, J. C., Poplin, R. E., Yoon, S. C., Lihm, J., Makarov, V., Gottipati, S., Keinan, A., Rodriguez-Flores, J. L., Rausch,

T., Fritz, M. H., Stütz, A. M., Beal, K., Datta, A., Herrero, J., Ritchie, G. R., Zerbino, D., Sabeti, P. C., Shlyakhter, I., Schaffner, S. F., Vitti, J., Cooper, D. N., Ball, E. V., Stenson, P. D., Barnes, B., Bauer, M., Cheetham, R. K., Cox, A., Eberle, M., Kahn, S., Murray, L., Peden, J., Shaw, R., Kenny, E. E., Batzer, M. A., Konkel, M. K., Walker, J. A., MacArthur, D. G., Lek, M., Herwig, R., Ding, L., Koboldt, D. C., Larson, D., Ye, K., Gravel, S., Swaroop, A., Chew, E., Lappalainen, T., Erlich, Y., Gymrek, M., Willems, T. F., Simpson, J. T., Shriver, M. D., Rosenfeld, J. A., Bustamante, C. D., Montgomery, S. B., De La Vega, F. M., Byrnes, J. K., Carroll, A. W., DeGorter, M. K., Lacroute, P., Maples, B. K., Martin, A. R., Moreno-Estrada, A., Shringarpure, S. S., Zakharia, F., Halperin, E., Baran, Y., Cerveira, E., Hwang, J., Malhotra, A., Plewczynski, D., Radew, K., Romanovitch, M., Zhang, C., Hyland, F. C., Craig, D. W., Christoforides, A., Homer, N., Izatt, T., Kurdoglu, A. A., Sinari, S. A., Squire, K., Xiao, C., Sebat, J., Antaki, D., Gujral, M., Noor, A., Ye, K., Burchard, E. G., Hernandez, R. D., Gignoux, C. R., Haussler, D., Katzman, S. J., Kent, W. J., Howie, B., Ruiz-Linares, A., Dermitzakis, E. T., Devine, S. E., Kang, H. M., Kidd, J. M., Blackwell, T., Caron, S., Chen, W., Emery, S., Fritsche, L., Fuchsberger, C., Jun, G., Li, B., Lyons, R., Scheller, C., Sidore, C., Song, S., Sliwerska, E., Taliun, D., Tan, A., Welch, R., Wing, M. K., Zhan, X., Awadalla, P., Hodgkinson, A., Li, Y., Shi, X., Quitadamo, A., Lunter, G., Marchini, J. L., Myers, S., Churchhouse, C., Delaneau, O., Gupta-Hinch, A., Kretzschmar, W., Iqbal, Z., Mathieson, I., Menelaou, A., Rimmer, A., Xifara, D. K., Oleksyk, T. K., Fu, Y., Liu, X., Xiong, M., Jorde, L., Witherspoon, D., Xing, J., Browning, B. L., Browning, S. R., Hormozdiari, F., Sudmant, P. H., Khurana, E., Tyler-Smith, C., Albers, C. A., Ayub, Q., Chen, Y., Colonna, V., Jostins, L., Walter, K., Xue, Y., Gerstein, M. B., Abyzov, A., Balasubramanian, S., Chen, J., Clarke, D., Fu, Y., Harmanci, A. O., Jin, M., Lee, D., Liu, J., Mu, X. J., Zhang, J., Zhang, Y., Hartl, C., Shakir, K., Degenhardt, J., Meiers, S., Raeder, B., Casale, F. P., Stegle, O., Lameijer, E. W., Hall, I., Bafna, V., Michaelson, J., Gardner, E. J., Mills, R. E., Dayama, G., Chen, K., Fan, X., Chong, Z., Chen, T., Chaisson, M. J., Huddleston, J., Malig, M., Nelson, B. J., Parrish, N. F., Blackburne, B., Lindsay, S. J., Ning, Z., Zhang, Y., Lam, H., Sisu, C., Challis, D., Evani, U. S., Lu, J., Nagaswamy, U., Yu, J., Li, W., Habegger, L., Yu, H., Cunningham, F., Dunham, I., Lage, K., Jespersen, J. B., Horn, H., Kim, D., Desalle, R., Narechania, A., Sayres, M. A., Mendez, F. L., Poznik, G. D., Underhill, P. A., Mittelman, D., Banerjee, R., Cerezo, M., Fitzgerald, T. W., Louzada, S., Massaia, A., Yang, F., Kalra, D., Hale, W., Dan, X., Barnes, K. C., Beiswanger, C., Cai, H., Cao, H., Henn, B., Jones, D., Kaye, J. S., Kent, A., Kerasidou, A., Mathias, R., Ossorio, P. N., Parker, M., Rotimi, C. N., Royal, C. D., Sandoval, K., Su, Y., Tian, Z., Tishkoff, S., Via, M., Wang, Y., Yang, H., Yang, L., Zhu, J., Bodmer, W., Bedoya, G., Cai, Z., Gao, Y., Chu, J., Peltonen, L., Garcia-Montero, A., Orfao, A., Dutil, J., Martinez-Cruzado, J. C., Mathias, R. A., Hennis, A., Watson, H., McKenzie, C., Qadri, F., LaRocque, R., Deng, X., Asogun, D., Folarin, O., Happi, C., Omoniwa, O., Stremlau, M., Tariyal, R., Jallow, M., Joof, F. S., Corraah, T., Rockett, K., Kwiatkowski, D., Kooner, J., Hien, T. T., Dunstan, S. J., ThuyHang, N., Fonnier, R., Garry, R., Kanneh, L., Moses, L., Schieffelin, J., Grant, D. S., Gallo, C., Poletti, G., Saleheen, D., Rasheed, A., Brooks, L. D., Felsenfeld, A. L., McEwen, J. E., Vaydylevich, Y., Duncanson, A., Dunn, M., and Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, 526(7571):68–74.

[Batista and Harwood, 2009] Batista, F. D. and Harwood, N. E. (2009). The who, how and where of antigen presentation to B cells. *Nature Reviews Immunology*, 9(1):15–27.

- [Batista and Neuberger, 1998] Batista, F. D. and Neuberger, M. S. (1998). Affinity dependence of the B cell response to antigen: A threshold, a ceiling, and the importance of off-rate. *Immunity*, 8(6):751–759.
- [Beaumont et al., 2002] Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035.
- [Beier et al., 2008] Beier, D., Rohrl, S., Pillai, D. R., Schwarz, S., Kunz-Schughart, L. A., Leukel, P., Proescholdt, M., Brawanski, A., Bogdahn, U., Trampe-Kieslich, A., Giebel, B., Wischhusen, J., Reifenberger, G., Hau, P., and Beier, C. P. (2008). Temozolomide Preferentially Depletes Cancer Stem Cells in Glioblastoma. *Cancer Research*, 68(14):5706–5715.
- [Belongia et al., 2016] Belongia, E. A., Simpson, M. D., King, J. P., Sundaram, M. E., Kelley, N. S., Osterholm, M. T., and McLean, H. Q. (2016). Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. *The Lancet Infectious Diseases*, 16(8):942 – 951.
- [Benner et al., 1974] Benner, R., Meima, F., van der Meulen, G. M., and van Muiswinkel, W. B. (1974). Antibody Formation in Mouse Bone Marrow. I. Evidence for the development of plaque-forming cells in situ. *Immunology*, 26(1971):247–255.
- [Berek, 1992] Berek, C. (1992). The development of B cells and the B cell repertoire in the microenvironment of the germinal center. *Immunol Rev*, 126(126):5–19.
- [Berek et al., 1991] Berek, C., Berger, A., and Apel, M. (1991). Maturation of the immune response in germinal centers. *Cell*, 67(6):1121–9.
- [Berek and Milstein, 1987] Berek, C. and Milstein, C. (1987). Mutation drift and repertoire shift in the maturation of the immune response. *Immunological reviews*, 96:23–41.
- [Bernasconi et al., 2002] Bernasconi, N. L., Traggiai, E., and Lanzavecchia, A. (2002). Maintenance of serological memory by polyclonal activation of human memory B cells. *Science (New York, N.Y.)*, 298(5601):2199–202.
- [Bjerknes and Cheng, 1999] Bjerknes, M. and Cheng, H. (1999). Clonal analysis of mouse intestinal epithelial progenitors. *Gastroenterology*, 116(1):7–14.
- [Blanpain et al., 2004] Blanpain, C., Lowry, W. E., Geoghegan, A., Polak, L., and Fuchs, E. (2004). Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. *Cell*, 118(5):635–648.
- [Blanpain and Simons, 2013] Blanpain, C. and Simons, B. D. (2013). Unravelling stem cell dynamics by lineage tracing. *Nature Reviews Molecular Cell Biology*, 14(8):489–502.
- [Bleau et al., 2009] Bleau, A. M., Hambardzumyan, D., Ozawa, T., Fomchenko, E. I., Huse, J. T., Brennan, C. W., and Holland, E. C. (2009). PTEN/PI3K/Akt Pathway Regulates the Side Population Phenotype and ABCG2 Activity in Glioma Tumor Stem-like Cells. *Cell Stem Cell*, 4(3):226–235.

- [Bond et al., 2015] Bond, A. M., Ming, G. L., and Song, H. (2015). Adult Mammalian Neural Stem Cells and Neurogenesis: Five Decades Later. *Cell Stem Cell*, 17(4):385–395.
- [Boos, 2003] Boos, D. D. (2003). Introduction to the Bootstrap World. *Statistical Science*, 18(2):168–174.
- [Boujelben et al., 2016] Boujelben, A., Watson, M., McDougall, S., Yen, Y.-F., Gerstner, E. R., Catana, C., Deisboeck, T., Batchelor, T. T., Boas, D., Rosen, B., Kalpathy-Cramer, J., and Chaplain, M. A. J. (2016). Multimodality imaging and mathematical modelling of drug delivery to glioblastomas. *Interface Focus*, 6(5):20160039.
- [Bousema et al., 2014] Bousema, T., Okell, L., Felger, I., and Drakeley, C. (2014). Asymptomatic malaria infections: Detectability, transmissibility and public health relevance. *Nature Reviews Microbiology*, 12(12):833–840.
- [Box and Tiao, 2011] Box, G. and Tiao, G. (2011). *Bayesian Inference in Statistical Analysis*. Wiley Classics Library. Wiley.
- [Bruggeman et al., 2007] Bruggeman, S. W. M., Hulsman, D., Tanger, E., Buckle, T., Blom, M., Zevenhoven, J., van Tellingen, O., and van Lohuizen, M. (2007). Bmi1 Controls Tumor Development in an Ink4a/Arf-Independent Manner in a Mouse Model for Glioma. *Cancer Cell*, 12(4):328–341.
- [Buchholz et al., 2013] Buchholz, V. R., Flossdorf, M., Hensel, I., Kretschmer, L., Weissbrich, B., Graf, P., Verschoor, A., Schiemann, M., Hofer, T., and Busch, D. H. (2013). Disparate Individual Fates Compose Robust CD8+ T Cell Immunity. *Science*, 340(6132):630–635.
- [Burkovitz et al., 2014] Burkovitz, A., Sela-Culang, I., and Ofran, Y. (2014). Large-scale analysis of somatic hypermutations in antibodies reveals which structural regions, positions and amino acids are modified to improve affinity. *FEBS Journal*, 281(1):306–319.
- [Burrell et al., 2013] Burrell, R. A., McGranahan, N., Bartek, J., and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, 501(7467):338–345.
- [Busse et al., 2014] Busse, C. E., Czogiel, I., Braun, P., Arndt, P. F., and Wardemann, H. (2014). Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes. *European Journal of Immunology*, 44(2):597–603.
- [Cai et al., 2015] Cai, Y., Zhou, Z. J., Wu, J., Li, Z. Y., and Long, Q. (2015). A coupled mathematical model of glioblastoma growth, pre-existing vessel co-option, angiogenesis and blood perfusion. *IFMBE Proceedings*, 52:88–91.
- [Camphausen et al., 2005] Camphausen, K., Purow, B., Sproull, M., Scott, T., Ozawa, T., Deen, D. F., and Tofilon, P. J. (2005). Orthotopic growth of human glioma cells quantitatively and qualitatively influences radiation-induced changes in gene expression. *Cancer Research*, 65(22):10389–10393.

- [Campos et al., 2016] Campos, B., Olsen, L. R., Urup, T., and Poulsen, H. (2016). A comprehensive profile of recurrent glioblastoma. *Oncogene*, 35(45):5819–5825.
- [Carlsson et al., 2014] Carlsson, S. K., Brothers, S. P., and Wahlestedt, C. (2014). Emerging treatment strategies for glioblastoma multiforme. *EMBO Molecular Medicine*, 6(11):1359–1370.
- [Carmona-Fontaine et al., 2017] Carmona-Fontaine, C., Deforet, M., Akkari, L., Thompson, C. B., Joyce, J. A., and Xavier, J. B. (2017). Metabolic origins of spatial organization in the tumor microenvironment. *Proceedings of the National Academy of Sciences*, 114(11):2934–2939.
- [Carrillo-Bustamante et al., 2015] Carrillo-Bustamante, P., Keşmir, C., and De Boer, R. J. (2015). A coevolutionary arms race between hosts and viruses drives polymorphism and polygenicity of NK cell receptors. *Molecular Biology and Evolution*, 32(8):2149–2160.
- [Cassese et al., 2003] Cassese, G., Arce, S., Hauser, A. E., Lehnert, K., Moewes, B., Mostarac, M., Muehlinghaus, G., Szyska, M., Radbruch, A., and Manz, R. A. (2003). Plasma Cell Survival Is Mediated by Synergistic Effects of Cytokines and Adhesion-Dependent Signals. *The Journal of Immunology*, 171(4):1684–1690.
- [Cavanagh et al., 2011] Cavanagh, B. L., Walker, T., Norazit, A., and Meedeniya, A. C. (2011). Thymidine analogues for tracking DNA synthesis. *Molecules*, 16(9):7980–7993.
- [Cayre et al., 2009] Cayre, M., Canoll, P., and Goldman, J. E. (2009). Cell migration in the normal and pathological postnatal mammalian brain. *Progress in Neurobiology*, 88(1):41–63.
- [Chan et al., 2012] Chan, T. D., Wood, K., Hermes, J. R., Butt, D., Jolly, C. J., Basten, A., and Brink, R. (2012). Elimination of Germinal-Center-Derived Self-Reactive B Cells Is Governed by the Location and Concentration of Self-Antigen. *Immunity*, 37(5):893–904.
- [Chaudhury et al., 2014] Chaudhury, S., Reifman, J., and Wallqvist, A. (2014). Simulation of B cell affinity maturation explains enhanced antibody cross-reactivity induced by the polyvalent malaria vaccine AMA1. *Journal of immunology (Baltimore, Md. : 1950)*, 193(5):2073–86.
- [Chen et al., 2012] Chen, J., Li, Y., Yu, T.-S., McKay, R. M., Burns, D. K., Kernie, S. G., and Parada, L. F. (2012). A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature*, 488(7412):522–526.
- [Cheng et al., 2012] Cheng, Y. K., Beroukhim, R., Levine, R. L., Mellinghoff, I. K., Holland, E. C., and Michor, F. (2012). A mathematical methodology for determining the temporal order of pathway alterations arising during gliomagenesis. *PLoS Computational Biology*, 8(1).
- [Childs et al., 2015] Childs, L. M., Baskerville, E. B., and Cobey, S. (2015). Trade-offs in antibody repertoires to complex antigens. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1676):20140245.

- [Choi et al., 2018] Choi, S., Yu, Y., Grimmer, M. R., Wahl, M., Chang, S. M., and Costello, J. F. (2018). Temozolomide-associated hypermutation in gliomas. *Neuro-Oncology*, (March):1–10.
- [Choi et al., 2011] Choi, Y. S., Kageyama, R., Eto, D., Escobar, T. C., Johnston, R. J., Monticelli, L., Lao, C., and Crotty, S. (2011). ICOS Receptor Instructs T Follicular Helper Cell versus Effector Cell Differentiation via Induction of the Transcriptional Repressor Bcl6. *Immunity*, 34(6):932–946.
- [Cojoc et al., 2015] Cojoc, M., Mäbert, K., Muders, M. H., and Dubrovskaya, A. (2015). A role for cancer stem cells in therapy resistance: Cellular and molecular mechanisms. *Seminars in Cancer Biology*, 31:16–27.
- [Contag et al., 1997] Contag, C. H., Spilman, S. D., Contag, P. R., Oshiro, M., Eames, B., Denery, P., Stevenson, D. K., and Benaron, D. a. (1997). Visualizing gene expression in living mammals using a bioluminescent reporter. *Photochemistry and photobiology*, 66(4):523–531.
- [Cuddapah et al., 2014] Cuddapah, V. A., Robel, S., Watkins, S., and Sontheimer, H. (2014). A neurocentric perspective on glioma invasion. *Nature Reviews Neuroscience*, 15(7):455–465.
- [Dame et al., 1984] Dame, J., Williams, J., McCutchan, T., Weber, J., Wirtz, R., Hockmeyer, W., Maloy, W., Haynes, J., Schneider, I., Roberts, D., and Et, A. (1984). Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. *Science*, 225(4662):593–599.
- [Dandy, 1928] Dandy, W. E. (1928). Removal of right cerebral hemisphere for certain tumors with hemiplegia: Preliminary report. *Journal of the American Medical Association*, 90(11):823–825.
- [Darmanis et al., 2017] Darmanis, S., Sloan, S. A., Croote, D., Mignardi, M., Chernikova, S., Samghabadi, P., Zhang, Y., Neff, N., Kowarsky, M., Caneda, C., Li, G., Chang, S. D., Connolly, I. D., Li, Y., Barres, B. A., Gephart, M. H., and Quake, S. R. (2017). Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Reports*, 21(5):1399–1410.
- [Davies and Cohen, 1996] Davies, D. R. and Cohen, G. H. (1996). Interactions of protein antigens with antibodies. *Proceedings of the National Academy of Sciences of the United States of America*, 93(1):7–12.
- [Davis et al., 2016] Davis, F. M., Lloyd-Lewis, B., Harris, O. B., Kozar, S., Winton, D. J., Muresan, L., and Watson, C. J. (2016). Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. *Nature Communications*, 7:13053.
- [Day et al., 2015] Day, C.-P., Merlino, G., and Van Dyke, T. (2015). Preclinical Mouse Cancer Models: A Maze of Opportunities and Challenges. *Cell*, 163(1):39–53.
- [De Silva and Klein, 2015] De Silva, N. S. and Klein, U. (2015). Dynamics of B cells in germinal centres. *Nature Reviews Immunology*, 15(3):137–148.

- [de Visser and Krug, 2014] de Visser, J. A. G. and Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*, 15(7):480–490.
- [Deem and Lee, 2003] Deem, M. W. and Lee, H. Y. (2003). Sequence Space Localization in the Immune System Response to Vaccination and Disease. *Physical Review Letters*, 91(6):8–11.
- [Di Noia and Neuberger, 2007] Di Noia, J. M. and Neuberger, M. S. (2007). Molecular Mechanisms of Antibody Somatic Hypermutation. *Annual Review of Biochemistry*, 76(1):1–22.
- [DiLillo et al., 2008] DiLillo, D. J., Hamaguchi, Y., Ueda, Y., Yang, K., Uchida, J., Haas, K. M., Kelsoe, G., and Tedder, T. F. (2008). Maintenance of Long-Lived Plasma Cells and Serological Memory Despite Mature and Memory B Cell Depletion during CD20 Immunotherapy in Mice. *The Journal of Immunology*, 180(1):361–371.
- [Doetsch et al., 1999] Doetsch, F., Caillé, I., Lim, D. A., García-Verdugo, J. M., and Alvarez-Buylla, A. (1999). Subventricular Zone Astrocytes Are Neural Stem Cells in the Adult Mammalian Brain. *Cell*, 97(6):703–716.
- [Dogan et al., 2009] Dogan, I., Bertocci, B., Vilmont, V., Delbos, F., Mégret, J., Storck, S., Reynaud, C. A., and Weill, J. C. (2009). Multiple layers of B cell memory with different effector functions. *Nature Immunology*, 10(12):1292–1299.
- [Driessens et al., 2012] Driessens, G., Beck, B., Caauwe, A., Simons, B. D., and Blanpain, C. (2012). Defining the mode of tumour growth by clonal analysis. *Nature*, 488(7412):527–530.
- [Dziurzynski et al., 2012] Dziurzynski, K., Blas-Boria, D., Suki, D., Cahill, D. P., Prabhu, S. S., Puduvalli, V., and Levine, N. (2012). Butterfly glioblastomas: A retrospective review and qualitative assessment of outcomes. *Journal of Neuro-Oncology*, 109(3):555–563.
- [Efron and Tibshirani, 1994] Efron, B. and Tibshirani, R. (1994). *An Introduction to the Bootstrap*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- [Elhanati et al., 2015] Elhanati, Y., Sethna, Z., Marcou, Q., Callan, C. G., Mora, T., and Walczak, A. M. (2015). Inferring processes underlying B-cell repertoire diversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1676):20140243–.
- [Enderling, 2015] Enderling, H. (2015). Cancer stem cells: small subpopulation or evolving fraction? *Integr. Biol.*, 7(1):14–23.
- [Enderling et al., 2009] Enderling, H., Hlatky, L., and Hahnfeldt, P. (2009). Migration rules: Tumours are conglomerates of self-metastases. *British Journal of Cancer*, 100(12):1917–1925.
- [Feil et al., 1997] Feil, R., Wagner, J., Metzger, D., and Chambon, P. (1997). Regulation of Cre Recombinase Activity by Mutated Estrogen Receptor Ligand-Binding Domains. *Biochemical and Biophysical Research Communications*, 237(3):752–757.
- [Feil et al., 2009] Feil, S., Valtcheva, N., and Feil, R. (2009). Inducible Cre Mice. In Kuehn, R. and Würst, W., editors, *Gene Knockout Protocols: Second Edition*, vol. 530, volume 530, pages 343–363. Humana Press.

- [Feng et al., 2013] Feng, W., Khan, M. A., Bellvis, P., Zhu, Z., Bernhardt, O., Herold-Mende, C., and Liu, H. K. (2013). The chromatin remodeler CHD7 regulates adult neurogenesis via activation of sox2 transcription factors. *Cell Stem Cell*, 13(1):62–72.
- [Finlay and McFadden, 2006] Finlay, B. B. and McFadden, G. (2006). Anti-immunology: Evasion of the host immune system by bacterial and viral pathogens. *Cell*, 124(4):767–782.
- [Fisher et al., 2014] Fisher, R., Pusztai, L., and Swanton, C. (2014). Cancer heterogeneity: Implications for targeted therapeutics. *Nature*, 509(7502 SUPPL.):479–485.
- [Floßdorf, 2013] Floßdorf, M. (2013). *Stochastic T cell fate decisions*. PhD thesis, Ruprecht-Karls-Universität Heidelberg.
- [Foreman-Mackey et al., 2013] Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee: The mcmc hammer. *PASP*, 125:306–312.
- [Fumentalba et al., 2012] Fumentalba, L. C., Obernier, K., and Alvarez-Buylla, A. (2012). Adult neural stem cells bridge their niche. *Cell Stem Cell*, 10(6):698–708.
- [Gallaher et al., 2018] Gallaher, J., Enriquez-Navas, P. M., Luddy, K. A., Gatenby, R. A., and Anderson, A. R. (2018). Spatial heterogeneity and evolutionary dynamics modulate time to recurrence in continuous and adaptive cancer therapies. *Cancer Research*, page canres.2649.2017.
- [Galson et al., 2014] Galson, J. D., Pollard, A. J., Trück, J., and Kelly, D. F. (2014). Studying the antibody repertoire after vaccination: Practical applications. *Trends in Immunology*, 35(7):319–331.
- [Galson et al., 2015] Galson, J. D., Trück, J., Fowler, A., Münz, M., Cerundolo, V., Pollard, A. J., Lunter, G., and Kelly, D. F. (2015). In-Depth Assessment of Within-Individual and Inter-Individual Variation in the B Cell Receptor Repertoire. *Frontiers in Immunology*, 6(OCT):1–13.
- [Ganguly and Puri, 2006] Ganguly, R. and Puri, I. K. (2006). Mathematical model for the cancer stem cell hypothesis. *Cell Prolif*, 39(1):3–14.
- [Gardiner, 2009] Gardiner, C. (2009). *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer Series in Synergetics. Springer Berlin Heidelberg.
- [Gatenby and Gawlinski, 1996] Gatenby, R. A. and Gawlinski, E. T. (1996). A Reaction-Diffusion Model of Cancer Invasion. *Cancer Research*, 56(31):5745–5753.
- [Gillespie, 2008] Gillespie, C. (2008). Moment-closure approximations for mass-action models. *IET systems biology*, 2(2):64–72.
- [Gillespie, 1977] Gillespie, D. T. (1977). Exact Stochastic Simulation of couple chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361.
- [Gitlin et al., 2014] Gitlin, A. D., Shulman, Z., and Nussenzweig, M. C. (2014). Clonal selection in the germinal centre by regulated proliferation and hypermutation. *Nature*, 509(7502):637–40.

- [Goodell et al., 2015] Goodell, M. A., Nguyen, H., and Shroyer, N. (2015). Somatic stem cell heterogeneity: Diversity in the blood, skin and intestinal stem cell compartments. *Nature Reviews Molecular Cell Biology*, 16(5):299–309.
- [Goodman and Weare, 2010] Goodman, J. and Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80.
- [Hadjipanayis and Van Meir, 2009] Hadjipanayis, C. G. and Van Meir, E. G. (2009). Tumor initiating cells in malignant gliomas: biology and implications for therapy. *Journal of Molecular Medicine*, 87(4):363–374.
- [Hafalla et al., 2011] Hafalla, J. C., Silvie, O., and Matuschewski, K. (2011). Cell biology and immunology of malaria. *Immunological reviews*, 240(1):297–316.
- [Hammarlund et al., 2003] Hammarlund, E., Lewis, M. W., Hansen, S. G., Strelow, L. I., Nelson, J. A., Sexton, G. J., Hanifin, J. M., and Slifka, M. K. (2003). Duration of antiviral immunity after smallpox vaccination. *Nature Medicine*, 9(9):1131–1137.
- [Hanahan and Weinberg, 2000] Hanahan, D. and Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, 100(1):57–70.
- [Hanahan and Weinberg, 2011] Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, 144(5):646–674.
- [Hardy et al., 1991] Hardy, R. R., Carmack, C. E., Shinton, S. A., Kemp, J. D., and Hayakawa, K. (1991). Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *The Journal of experimental medicine*, 173(5):1213–25.
- [Hartley et al., 1993] Hartley, S. B., Cooke, M. P., Fulcher, D. A., Harris, A. W., Cory, S., Basten, A., and Goodnow, C. C. (1993). Elimination of self-reactive B lymphocytes proceeds in two stages: Arrested development and cell death. *Cell*, 72(3):325–335.
- [Hashizume et al., 2010] Hashizume, R., Ozawa, T., Dinca, E. B., Banerjee, A., Prados, M. D., James, C. D., and Gupta, N. (2010). A human brainstem glioma xenograft model enabled for bioluminescence imaging. *Journal of Neuro-Oncology*, 96(2):151–159.
- [Hawkins-Daarud et al., 2013] Hawkins-Daarud, A., Rockne, R. C., Anderson, A. R. A., and Swanson, K. R. (2013). Modeling Tumor-Associated Edema in Gliomas during Anti-Angiogenic Therapy and Its Impact on Imageable Tumor. *Frontiers in Oncology*, 3(April):1–12.
- [Hayashi et al., 2015] Hayashi, K., Kubo, K. I., Kitazawa, A., and Nakajima, K. (2015). Cellular dynamics of neuronal migration in the hippocampus. *Frontiers in Neuroscience*, 9(APR):1–11.
- [Heckl et al., 2014] Heckl, D., Kowalczyk, M. S., Yudovich, D., Belizaire, R., Puram, R. V., McConkey, M. E., Thielke, A., Aster, J. C., Regev, A., and Ebert, B. L. (2014). Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nature Biotechnology*, 32(9):941–946.

- [Hidalgo et al., 2014] Hidalgo, M., Amant, F., Biankin, A. V., Budinská, E., Byrne, A. T., Caldas, C., Clarke, R. B., de Jong, S., Jonkers, J., Mælandsmo, G. M., Roman-Roman, S., Seoane, J., Trusolino, L., and Villanueva, A. (2014). Patient-derived Xenograft models: An emerging platform for translational cancer research. *Cancer Discovery*, 4(9):998–1013.
- [Höfer et al., 2016] Höfer, T., Busch, K., Klapproth, K., and Rodewald, H.-R. (2016). Fate Mapping and Quantitation of Hematopoiesis In Vivo. *Annual Review of Immunology*, 34(1):449–478.
- [Hollingdale et al., 1984] Hollingdale, M. R., Nardin, E. H., Tharavani, S., Schwartz, A. L., and Nussenzweig, R. S. (1984). Inhibition of entry of *Plasmodium falciparum* and *P. vivax* sporozoites into cultured cells; an in vitro assay of protective antibodies. *J Immunol*, 132(2):909–913.
- [Hozumi and Tonegawa, 1976] Hozumi, N. and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proceedings of the National Academy of Sciences*, 73(10):3628–3632.
- [Iber and Maini, 2002] Iber, D. and Maini, P. K. (2002). A mathematical model for germinal centre kinetics and affinity maturation. *Journal of Theoretical Biology*, 219(2):153–175.
- [Imkeller et al., 2016] Imkeller, K., Arndt, P. F., Wardemann, H., and Busse, C. E. (2016). sciReptor: analysis of single-cell level immunoglobulin repertoires. *BMC Bioinformatics*, 17(1):67.
- [Jackson et al., 2014] Jackson, K. J., Liu, Y., Roskin, K. M., Glanville, J., Hoh, R. A., Seo, K., Marshall, E. L., Gurley, T. C., Moody, M. A., Haynes, B. F., Walter, E. B., Liao, H. X., Albrecht, R. A., García-Sastre, A., Chaparro-Riggers, J., Rajpal, A., Pons, J., Simen, B. B., Hanczaruk, B., Dekker, C. L., Laserson, J., Koller, D., Davis, M. M., Fire, A. Z., and Boyd, S. D. (2014). Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell Host and Microbe*, 16(1):105–114.
- [Jackson et al., 2015] Jackson, P. R., Juliano, J., Hawkins-Daarud, A., Rockne, R. C., and Swanson, K. R. (2015). Patient-Specific Mathematical Neuro-Oncology: Using a Simple Proliferation and Invasion Tumor Model to Inform Clinical Practice. *Bulletin of Mathematical Biology*, 77(5):846–856.
- [Jacob et al., 1993] Jacob, J., Przylepa, J., Miller, C., and Kelsoe, G. (1993). In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. III. The kinetics of V region mutation and selection in germinal center B cells. *The Journal of experimental medicine*, 178(4):1293–307.
- [Jarzabek et al., 2013] Jarzabek, M. A., Huszthy, P. C., Skaftnesmo, K. O., McCormack, E., Dicker, P., Prehn, J. H. M., Bjerkvig, R., and Byrne, A. T. (2013). In vivo bioluminescence imaging validation of a human biopsy-derived orthotopic mouse model of glioblastoma multiforme. *Molecular Imaging*, 12(3):161–172.
- [Jaynes et al., 2003] Jaynes, E., Jaynes, E., Bretthorst, G., and Press, C. U. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

- [Johnson et al., 2014] Johnson, B. E., Mazar, T., Hong, C., Barnes, M., Aihara, K., McLean, C. Y., Fouse, S. D., Yamamoto, S., Ueda, H., Tatsuno, K., Asthana, S., Jalbert, L. E., Nelson, S. J., Bollen, A. W., Gustafson, W. C., Charron, E., Weiss, W. A., Smirnov, I. V., Song, J. S., Olshen, A. B., Cha, S., Zhao, Y., Moore, R. A., Mungall, A. J., Jones, S. J. M., Hirst, M., Marra, M. A., Saito, N., Aburatani, H., Mukasa, A., Berger, M. S., Chang, S. M., Taylor, B. S., and Costello, J. F. (2014). Mutational Analysis Reveals the Origin and Therapy-Driven Evolution of Recurrent Glioma. *Science*, 343(6167):189–193.
- [Johnston et al., 2010] Johnston, M. D., Maini, P. K., Jonathan Chapman, S., Edwards, C. M., and Bodmer, W. F. (2010). On the proportion of cancer stem cells in a tumour. *Journal of Theoretical Biology*, 266(4):708–711.
- [Karolak et al., 2018] Karolak, A., Markov, D. A., McCawley, L. J., and Rejniak, K. A. (2018). Towards personalized computational oncology: from spatial models of tumour spheroids, to organoids, to tissues. *Journal of The Royal Society Interface*, 15(138):20170703.
- [Karrer et al., 2000] Karrer, U., Lopez-Macias, C., Oxenius, A., Odermatt, B., Bachmann, M. F., Kalinke, U., Bluethmann, H., Hengartner, H., and Zinkernagel, R. M. (2000). Antiviral B Cell Memory in the Absence of Mature Follicular Dendritic Cell Networks and Classical Germinal Centers in TNFR1^{-/-} Mice. *The Journal of Immunology*, 164(2):768–778.
- [Kato et al., 1998] Kato, J., Motoyama, N., Taniuchi, I., Takeshita, H., Toyoda, M., Masuda, K., and Watanabe, T. (1998). Affinity maturation in Lyn kinase-deficient mice with defective germinal center formation. *J Immunol*, 160(10):0–4788.
- [Kauffman and Weinberger, 1989] Kauffman, S. A. and Weinberger, E. D. (1989). The NK model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of Theoretical Biology*, 141:211–245.
- [Keitany et al., 2016] Keitany, G. J., Kim, K. S., Krishnamurty, A. T., Hondowicz, B. D., Hahn, W. O., Dambrauskas, N., Sather, D. N., Vaughan, A. M., Kappe, S. H., and Pepper, M. (2016). Blood Stage Malaria Disrupts Humoral Immunity to the Pre-erythrocytic Stage Circumsporozoite Protein. *Cell Reports*, 17(12):3193–3205.
- [Kepler and Perelson, 1993] Kepler, T. B. and Perelson, A. S. (1993). Somatic hypermutation in B cells: an optimal control treatment. *Journal of theoretical biology*, 164(1):37–64.
- [Kerfoot et al., 2011] Kerfoot, S. M., Yaari, G., Patel, J. R., Johnson, K. L., Gonzalez, D. G., Kleinstein, S. H., and Haberman, A. M. (2011). Germinal Center B Cell and T Follicular Helper Cell Development Initiates in the Interfollicular Zone. *Immunity*, 34(6):947–960.
- [Keşmir and De Boer, 2003] Keşmir, C. and De Boer, R. J. (2003). 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.
- [Kitano et al., 2011] Kitano, M., Moriyama, S., Ando, Y., Hikida, M., Mori, Y., Kurosaki, T., and Okada, T. (2011). Bcl6 Protein Expression Shapes Pre-Germinal Center B Cell Dynamics and Follicular Helper T Cell Heterogeneity. *Immunity*, 34(6):961–972.

- [Kleinstein and Singh, 2003] Kleinstein, S. H. and Singh, J. P. (2003). Why are there so few key mutant clones? The influence of stochastic selection and blocking on affinity maturation in the germinal center. *International immunology*, 15(7):871–84.
- [Koo et al., 2012] Koo, B. K., Stange, D. E., Sato, T., Karthaus, W., Farin, H. F., Huch, M., Van Es, J. H., and Clevers, H. (2012). Controlled gene expression in primary Lgr5 organoid cultures. *Nature Methods*, 9(1):81–83.
- [Kriegstein and Alvarez-Buylla, 2009] Kriegstein, A. and Alvarez-Buylla, A. (2009). The Glial Nature of Embryonic and Adult Neural Stem Cells. *Annual Review of Neuroscience*, 32(1):149–184.
- [Kroese et al., 1987] Kroese, F. G., Wubbena, A. S., Seijen, H. G., and Nieuwenhuis, P. (1987). Germinal centers develop oligoclonally. *European Journal of Immunology*, 17(7):1069–1072.
- [Kwon et al., 2008] Kwon, C. H., Zhao, D., Chen, J., Alcantara, S., Li, Y., Burns, D. K., Mason, R. P., Lee, E. Y. H. P., Wu, H., and Parada, L. F. (2008). Pten haploinsufficiency accelerates formation of high-grade astrocytomas. *Cancer Research*, 68(9):3286–3294.
- [Lan et al., 2017] Lan, X., Jörg, D. J., Cavalli, F. M. G., Richards, L. M., Nguyen, L. V., Vanner, R. J., Guilhamon, P., Lee, L., Kushida, M. M., Pellacani, D., Park, N. I., Coutinho, F. J., Whetstone, H., Selvadurai, H. J., Che, C., Luu, B., Carles, A., Moksa, M., Rastegar, N., Head, R., Dolma, S., Prinos, P., Cusimano, M. D., Das, S., Bernstein, M., Arrowsmith, C. H., Mungall, A. J., Moore, R. A., Ma, Y., Gallo, M., Lupien, M., Pugh, T. J., Taylor, M. D., Hirst, M., Eaves, C. J., Simons, B. D., and Dirks, P. B. (2017). Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. *Nature*, 549(7671):227–232.
- [Lapidot et al., 1994] Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M. A., and Dick, J. E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*, 367(6464):645–648.
- [Leder et al., 2014] Leder, K., Pitter, K., Laplant, Q., Hambardzumyan, D., Ross, B. D., Chan, T. A., Holland, E. C., and Michor, F. (2014). Mathematical modeling of pdgf-driven glioblastoma reveals optimized radiation dosing schedules. *Cell*, 156(3):603–616.
- [Li and Clevers, 2010] Li, L. and Clevers, H. (2010). Coexistence of Quiescent and Active Adult Stem Cells in Mammals. *Science*, 327(5965):542–545.
- [Liu et al., 2015] Liu, D., Xu, H., Shih, C., Wan, Z., Ma, X., Ma, W., Luo, D., and Qi, H. (2015). T-B-cell entanglement and ICOSL-driven feed-forward regulation of germinal centre reaction. *Nature*, 517(7533):214–218.
- [Liu et al., 2008] Liu, H.-K., Belz, T., Bock, D., Takacs, A., Wu, H., Lichter, P., Chai, M., and Schutz, G. (2008). The nuclear receptor tailless is required for neurogenesis in the adult subventricular zone. *Genes & Development*, 22(18):2473–2478.

- [Liu et al., 2013] Liu, X., Johnson, S., Liu, S., Kanojia, D., Yue, W., Singn, U., Wang, Q., Nie, Q., and Chen, H. (2013). Nonlinear growth kinetics of breast cancer stem cells: Implications for cancer stem cell targeted therapy. *Scientific Reports*, 3:1–10.
- [Liu et al., 1989] Liu, Y. J., Joshua, D. E., Williams, G. T., Smith, C. A., Gordon, J., and MacLennan, I. C. (1989). Mechanism of antigen-driven selection in germinal centres. *Nature*, 342(6252):929–31.
- [Livet et al., 2007] Livet, J., Weissman, T. A., Kang, H., Draft, R. W., Lu, J., Bennis, R. A., Sanes, J. R., and Lichtman, J. W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, 450(7166):56–62.
- [Louis et al., 2007] Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., Scheithauer, B. W., and Kleihues, P. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathologica*, 114(2):97–109.
- [Lu et al., 2018] Lu, L. L., Suscovich, T. J., Fortune, S. M., and Alter, G. (2018). Beyond binding: Antibody effector functions in infectious diseases. *Nature Reviews Immunology*, 18(1):46–61.
- [Luo and Perelson, 2015] Luo, S. and Perelson, A. S. (2015). Competitive exclusion by autologous antibodies can prevent broad HIV-1 antibodies from arising. *Proceedings of the National Academy of Sciences*, 112(37):11654–11659.
- [Macallan, 2005] Macallan, D. C. (2005). B-cell kinetics in humans: rapid turnover of peripheral blood memory cells. *Blood*, 105(9):3633–3640.
- [MacKay, 2005] MacKay, D. J. C. (2005). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge.
- [MacLennan, 1994] MacLennan, I. C. (1994). Germinal centers. *Annual review of immunology*, 12(1):117–39.
- [Males et al., 2008] Males, S., Gaye, O., and Garcia, A. (2008). Long-Term Asymptomatic Carriage of *Plasmodium falciparum* Protects from Malaria Attacks: a Prospective Study among Senegalese Children. *Clinical Infectious Diseases*, 46(4):516–522.
- [Malmström et al., 2012] Malmström, A., Grønberg, B. H., Marosi, C., Stupp, R., Frappaz, D., Schultz, H., Abacioglu, U., Tavelin, B., Lhermitte, B., Hegi, M. E., Rosell, J., and Henriksson, R. (2012). Temozolomide versus standard 6-week radiotherapy versus hypofractionated radiotherapy in patients older than 60 years with glioblastoma: The Nordic randomised, phase 3 trial. *The Lancet Oncology*, 13(9):916–926.
- [Mandel et al., 1980] Mandel, T. E., Phipps, R. P., Abbot, a., and Tew, J. G. (1980). The follicular dendritic cell: long term antigen retention during immunity. *Immunological reviews*, 53(I):29–59.
- [Manz et al., 1997] Manz, R. A., Thiel, A., and Radbruch, A. (1997). Lifetime of plasma cells in the bone marrow. *Nature*, 388(6638):133–134.

- [Massoud et al., 2003] Massoud, T. F., Massoud, T. F., Gambhir, S. S., and Gambhir, S. S. (2003). Molecular imaging in living subjects: seeing fundamental biological processes in a new light. *Genes & Development*, 17:545–580.
- [Matsumoto and Futamura, 1995] Matsumoto, K. and Futamura, Y. (1995). Cell counts in peripheral blood and bone marrow of male CBA/N mice. *The Journal of veterinary medical science / the Japanese Society of Veterinary Science*, 57(4):755–6.
- [Matsumoto et al., 1996] Matsumoto, M., Lo, S. F., Carruthers, C. J., Min, J., Mariathasan, S., Huang, G., Plas, D. R., Martin, S. M., Geha, R. S., Nahm, M. H., and Chaplin, D. D. (1996). Affinity maturation without germinal centres in lymphotoxin-alpha-deficient mice.
- [Mayer et al., 2015] Mayer, A., Balasubramanian, V., Mora, T., and Walczak, A. M. (2015). How a well-adapted immune system is organized. *Proceedings of the National Academy of Sciences of the United States of America*, 112(19):5950–5955.
- [McCarthy et al., 2018] McCarthy, K. R., Watanabe, A., Kuraoka, M., Do, K. T., McGee, C. E., Sempowski, G. D., Kepler, T. B., Schmidt, A. G., Kelsoe, G., and Harrison, S. C. (2018). Memory B Cells that Cross-React with Group 1 and Group 2 Influenza A Viruses Are Abundant in Adult Human Repertoires. *Immunity*, 48(1):174–184.e9.
- [McGranahan and Swanton, 2015] McGranahan, N. and Swanton, C. (2015). Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell*, 27(1):15–26.
- [McKean et al., 1984] McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W., and Weigert, M. (1984). Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proceedings of the National Academy of Sciences of the United States of America*, 81(10):3180–4.
- [McLendon et al., 2008] McLendon, R., Friedman, A., Bigner, D., Van Meir, E. G., Brat, D. J., M. Mastrogiannakis, G., Olson, J. J., Mikkelsen, T., Lehman, N., Aldape, K., Alfred Yung, W. K., Bogler, O., VandenBerg, S., Berger, M., Prados, M., Muzny, D., Morgan, M., Scherer, S., Sabo, A., Nazareth, L., Lewis, L., Hall, O., Zhu, Y., Ren, Y., Alvi, O., Yao, J., Hawes, A., Jhangiani, S., Fowler, G., San Lucas, A., Kovar, C., Cree, A., Dinh, H., Santibanez, J., Joshi, V., Gonzalez-Garay, M. L., Miller, C. A., Milosavljevic, A., Donehower, L., Wheeler, D. A., Gibbs, R. A., Cibulskis, K., Sougnez, C., Fennell, T., Mahan, S., Wilkinson, J., Ziaugra, L., Onofrio, R., Bloom, T., Nicol, R., Ardlie, K., Baldwin, J., Gabriel, S., Lander, E. S., Ding, L., Fulton, R. S., McLellan, M. D., Wallis, J., Larson, D. E., Shi, X., Abbott, R., Fulton, L., Chen, K., Koboldt, D. C., Wendl, M. C., Meyer, R., Tang, Y., Lin, L., Osborne, J. R., Dunford-Shore, B. H., Miner, T. L., Delehaunty, K., Markovic, C., Swift, G., Courtney, W., Pohl, C., Abbott, S., Hawkins, A., Leong, S., Haipek, C., Schmidt, H., Wiechert, M., Vickery, T., Scott, S., Dooling, D. J., Chinwalla, A., Weinstock, G. M., Mardis, E. R., Wilson, R. K., Getz, G., Winckler, W., Verhaak, R. G. W., Lawrence, M. S., O’Kelly, M., Robinson, J., Alexe, G., Beroukhi, R., Carter, S., Chiang, D., Gould, J., Gupta, S., Korn, J., Mermel, C., Mesirov, J., Monti, S., Nguyen, H., Parkin, M., Reich, M., Stransky, N., Weir, B. A., Garraway, L., Golub, T., Meyerson, M., Chin, L., Protopopov, A., Zhang, J., Perna, I., Aronson, S., Sathiamoorthy, N., Ren, G., Yao, J.,

- Wiedemeyer, W. R., Kim, H., Won Kong, S., Xiao, Y., Kohane, I. S., Seidman, J., Park, P. J., Kucherlapati, R., Laird, P. W., Cope, L., Herman, J. G., Weisenberger, D. J., Pan, F., Van Den Berg, D., Van Neste, L., Mi Yi, J., Schuebel, K. E., Baylin, S. B., Absher, D. M., Li, J. Z., Southwick, A., Brady, S., Aggarwal, A., Chung, T., Sherlock, G., Brooks, J. D., Myers, R. M., Spellman, P. T., Purdom, E., Jakkula, L. R., Lapuk, A. V., Marr, H., Dorton, S., Gi Choi, Y., Han, J., Ray, A., Wang, V., Durinck, S., Robinson, M., Wang, N. J., Vranizan, K., Peng, V., Van Name, E., Fontenay, G. V., Ngai, J., Conboy, J. G., Parvin, B., Feiler, H. S., Speed, T. P., Gray, J. W., Brennan, C., Socci, N. D., Olshen, A., Taylor, B. S., Lash, A., Schultz, N., Reva, B., Antipin, Y., Stukalov, A., Gross, B., Cerami, E., Qing Wang, W., Qin, L.-X., Seshan, V. E., Villafania, L., Cavatore, M., Borsu, L., Viale, A., Gerald, W., Sander, C., Ladanyi, M., Perou, C. M., Neil Hayes, D., Topal, M. D., Hoadley, K. A., Qi, Y., Balu, S., Shi, Y., Wu, J., Penny, R., Bittner, M., Shelton, T., Lenkiewicz, E., Morris, S., Beasley, D., Sanders, S., Kahn, A., Sfeir, R., Chen, J., Nassau, D., Feng, L., Hickey, E., Zhang, J., Weinstein, J. N., Barker, A., Gerhard, D. S., Vockley, J., Compton, C., Vaught, J., Fielding, P., Ferguson, M. L., Schaefer, C., Madhavan, S., Buetow, K. H., Collins, F., Good, P., Guyer, M., Ozenberger, B., Peterson, J., and Thomson, E. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455(7216):1061–1068.
- [Meacham and Morrison, 2013] Meacham, C. E. and Morrison, S. J. (2013). Tumour heterogeneity and cancer cell plasticity. *Nature*, 501(7467):328–337.
- [Melo et al., 2013] Melo, F. D. S. E., Vermeulen, L., Fessler, E., and Medema, J. P. (2013). Cancer heterogeneity - A multifaceted view. *EMBO Reports*, 14(8):686–695.
- [Meyer-Hermann et al., 2012] Meyer-Hermann, M., Mohr, E., Pelletier, N., Zhang, Y., Victora, G. D., and Toellner, K.-M. (2012). A theory of germinal center B cell selection, division, and exit. *Cell reports*, 2(1):162–74.
- [Meyer-Hermann et al., 2006] Meyer-Hermann, M. E., Maini, P. K., and Iber, D. (2006). An analysis of B cell selection mechanisms in germinal centers. *Mathematical medicine and biology : a journal of the IMA*, 23(3):255–77.
- [Miles et al., 2011] Miles, J. J., Douek, D. C., and Price, D. A. (2011). Bias in the α B T-cell repertoire: Implications for disease pathogenesis and vaccination. *Immunology and Cell Biology*, 89(3):375–387.
- [Miller, 1964] Miller, J. J. (1964). An autoradiographic study of plasma cell and lymphocyte survival in rat popliteal lymph nodes. *Journal of immunology (Baltimore, Md. : 1950)*, 92:673–81.
- [Milo and Phillips, 2015] Milo, R. and Phillips, R. (2015). *Cell Biology by the Numbers*. Taylor & Francis Group.
- [Ming and Song, 2011] Ming, G. and Song, H. (2011). Adult Neurogenesis in the Mammalian Brain: Significant Answers and Significant Questions. *Neuron*, 70(4):687–702.
- [Molina-Peña and Álvarez, 2012] Molina-Peña, R. and Álvarez, M. M. (2012). A simple mathematical model based on the cancer stem cell hypothesis suggests kinetic commonalities in solid tumor growth. *PLoS ONE*, 7(2).

- [Monaghan et al., 1995] Monaghan, A. P., Grau, E., Bock, D., and Schütz, G. (1995). The mouse homolog of the orphan nuclear receptor *tailless* is expressed in the developing forebrain. *Development*, 121(3):839–53.
- [Moore and Lemischka, 2006] Moore, K. A. and Lemischka, I. R. (2006). Stem Cells and Their Niches. *Science*, 311(5769):1880–1885.
- [Mordmüller et al., 2017] Mordmüller, B., Surat, G., Lagler, H., Chakravarty, S., Ishizuka, A. S., Lalremruata, A., Gmeiner, M., Campo, J. J., Esen, M., Ruben, A. J., Held, J., Calle, C. L., Mengue, J. B., Gebru, T., Ibáñez, J., Sulyok, M., James, E. R., Billingsley, P. F., Natasha, K., Manoj, A., Murshedkar, T., Gunasekera, A., Eappen, A. G., Li, T., Stafford, R. E., Li, M., Felgner, P. L., Seder, R. A., Richie, T. L., Sim, B. K. L., Hoffman, S. L., and Kremsner, P. G. (2017). Sterile protection against human malaria by chemoattenuated PfSPZ vaccine. *Nature*, 542(7642):445–449.
- [Moreira et al., 2007] Moreira, I. S., Fernandes, P. A., and Ramos, M. J. (2007). Hot spots-A review of the protein-protein interface determinant amino-acid residues. *Proteins: Structure, Function, and Bioinformatics*, 68(4):803–812.
- [Morrison and Kimble, 2006] Morrison, S. J. and Kimble, J. (2006). Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature*, 441(7097):1068–1074.
- [Muramatsu et al., 2000] Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. *Cell*, 102(5):553–563.
- [Murphy et al., 2016] Murphy, H., Jaafari, H., and Dobrovolsky, H. M. (2016). Differences in predictions of ODE models of tumor growth: a cautionary example. *BMC Cancer*, 16(1):163.
- [Murphy and Weaver, 2016] Murphy, K. and Weaver, C. (2016). *Janeway’s Immunobiology, 9th edition*. CRC Press.
- [Murugan et al., 2018] Murugan, R., Buchauer, L., Triller, G., Kreschel, C., Costa, G., Pidelaserra Martí, G., Imkeller, K., Busse, C. E., Chakravarty, S., Sim, B. K. L., Hoffman, S. L., Levashina, E. A., Kremsner, P. G., Mordmüller, B., Höfer, T., and Wardemann, H. (2018). Clonal selection drives protective memory B cell responses in controlled human malaria infection. *Science Immunology*, 3(20):eaap8029.
- [Murugan et al., 2015] Murugan, R., Imkeller, K., Busse, C. E., and Wardemann, H. (2015). Direct high-throughput amplification and sequencing of immunoglobulin genes from single human B cells.
- [Murugan, 2018] Murugan, R. A. (2018). *Protective memory B cell response in controlled human malaria infection*. PhD thesis, Humboldt-Universität zu Berlin.
- [Neafsey et al., 2015] Neafsey, D. E., Juraska, M., Bedford, T., Benkeser, D., Valim, C., Griggs, A., Lievens, M., Abdulla, S., Adjei, S., Agbenyega, T., Agnandji, S. T., Aide, P., Anderson, S., Ansong, D., Aponte, J. J., Asante, K. P., Bejon, P., Birkett, A. J., Bruls, M., Connolly, K. M., D’Alessandro, U., Dobaño, C., Gesase, S., Greenwood, B., Grimsby, J., Tinto, H., Hamel,

- M. J., Hoffman, I., Kamthunzi, P., Kariuki, S., Kremsner, P. G., Leach, A., Lell, B., Lennon, N. J., Lusingu, J., Marsh, K., Martinson, F., Molel, J. T., Moss, E. L., Njuguna, P., Ockenhouse, C. F., Ogutu, B. R., Otieno, W., Otieno, L., Otieno, K., Owusu-Agyei, S., Park, D. J., Pellé, K., Robbins, D., Russ, C., Ryan, E. M., Sacarlal, J., Sogoloff, B., Sorgho, H., Tanner, M., Theander, T., Valea, I., Volkman, S. K., Yu, Q., Lapierre, D., Birren, B. W., Gilbert, P. B., and Wirth, D. F. (2015). Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *New England Journal of Medicine*, 373(21):2025–2037.
- [Nossal, 1962] Nossal, G. J. V. (1962). Autoradiographic studies on the immune response: I. The kinetics of plasma cell proliferation. *Journal of Experimental Medicine*, 115(1):209–230.
- [Nussenzweig et al., 1967] Nussenzweig, R. S., Vanderberg, J., Most, H., and Orton, C. (1967). Protective immunity produced by the injection of X-irradiated sporozoites of plasmodium berghei. *Nature*, 216(5111):160–162.
- [Ochsenbein et al., 2000] Ochsenbein, A. F., Pinschewer, D. D., Sierro, S., Horvath, E., Hengartner, H., and Zinkernagel, R. M. (2000). Protective long-term antibody memory by antigen-driven and T help-dependent differentiation of long-lived memory B cells to short-lived plasma cells independent of secondary lymphoid organs. *Proceedings of the National Academy of Sciences*, 97(24):13263–13268.
- [O'Connor et al., 2006] O'Connor, B. P., Vogel, L. A., Zhang, W., Loo, W., Shnider, D., Lind, E. F., Ratliff, M., Noelle, R. J., and Erickson, L. D. (2006). Imprinting the Fate of Antigen-Reactive B Cells through the Affinity of the B Cell Receptor. *The Journal of Immunology*, 177(11):7723–7732.
- [Ohgaki and Kleihues, 2005] Ohgaki, H. and Kleihues, P. (2005). Epidemiology and etiology of gliomas. *Acta Neuropathologica*, 109(1):93–108.
- [Okada et al., 2005] Okada, T., Miller, M. J., Parker, I., Krummel, M. F., Neighbors, M., Hartley, S. B., O'Garra, A., Cahalan, M. D., and Cyster, J. G. (2005). Antigen-engaged B cells undergo chemotaxis toward the T zone and form motile conjugates with helper T cells. *PLoS Biology*, 3(6):1047–1061.
- [Olotu et al., 2016] Olotu, A., Fegan, G., Wambua, J., Nyangweso, G., Leach, A., Lievens, M., Kaslow, D. C., Njuguna, P., Marsh, K., and Bejon, P. (2016). Seven-Year Efficacy of RTS,S/AS01 Malaria Vaccine among Young African Children. *New England Journal of Medicine*, 374(26):2519–2529.
- [Oprea and Perelson, 1997] Oprea, M. and Perelson, A. S. (1997). Somatic mutation leads to efficient affinity maturation when centrocytes recycle back to centroblasts. *Journal of immunology (Baltimore, Md. : 1950)*, 158(11):5155–62.
- [Oshima et al., 2001] Oshima, H., Rochat, A., Kedzia, C., Kobayashi, K., and Barrandon, Y. (2001). Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell*, 104(2):233–245.

- [Osswald et al., 2015] Osswald, M., Jung, E., Sahm, F., Solecki, G., Venkataramani, V., Blaes, J., Weil, S., Horstmann, H., Wiestler, B., Syed, M., Huang, L., Ratliff, M., Karimian Jazi, K., Kurz, F. T., Schmenger, T., Lemke, D., Gömmel, M., Pauli, M., Liao, Y., Häring, P., Pusch, S., Herl, V., Steinhäuser, C., Kronic, D., Jarahian, M., Miletic, H., Berghoff, A. S., Griesbeck, O., Kalamakis, G., Garaschuk, O., Preusser, M., Weiss, S., Liu, H., Heiland, S., Platten, M., Huber, P. E., Kuner, T., von Deimling, A., Wick, W., and Winkler, F. (2015). Brain tumour cells interconnect to a functional and resistant network. *Nature*, 528(7580):93–98.
- [Ostrom et al., 2014] Ostrom, Q. T., Bauchet, L., Davis, F. G., Deltour, I., Fisher, J. L., Langer, C. E., Pekmezci, M., Schwartzbaum, J. A., Turner, M. C., Walsh, K. M., Wrensch, M. R., and Barnholtz-Sloan, J. S. (2014). The epidemiology of glioma in adults: A state of the science review. *Neuro-Oncology*, 16(7):896–913.
- [Pape et al., 2011] Pape, K. A., Taylor, J. J., Maul, R. W., Gearhart, P. J., and Jenkins, M. K. (2011). Different B Cell Populations Mediate Early and Late Memory During an Endogenous Immune Response. *Science*, 331(6021):1203–1207.
- [Parameswaran et al., 2013] Parameswaran, P., Liu, Y., Roskin, K. M., Jackson, K. K., Dixit, V. P., Lee, J. Y., Artiles, K. L., Zompi, S., Vargas, M. J., Simen, B. B., Hanczaruk, B., McGowan, K. R., Tariq, M. A., Pourmand, N., Koller, D., Balmaseda, A., Boyd, S. D., Harris, E., and Fire, A. Z. (2013). Convergent antibody signatures in human dengue. *Cell Host and Microbe*, 13(6):691–700.
- [Pattabiraman and Weinberg, 2014] Pattabiraman, D. R. and Weinberg, R. A. (2014). Tackling the cancer stem cells-what challenges do they pose? *Nature Reviews Drug Discovery*, 13(7):497–512.
- [Paus et al., 2006] Paus, D., Phan, T. G., Chan, T. D., Gardam, S., Basten, A., and Brink, R. (2006). Antigen recognition strength regulates the choice between extrafollicular plasma cell and germinal center B cell differentiation. *The Journal of Experimental Medicine*, 203(4):1081–1091.
- [Pei et al., 2017] Pei, W., Feyerabend, T. B., Rössler, J., Wang, X., Postrach, D., Busch, K., Rode, I., Klapproth, K., Dietlein, N., Quedenau, C., Chen, W., Sauer, S., Wolf, S., Höfer, T., and Rodewald, H.-R. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature*, 548(7668):456–460.
- [Peltola et al., 1994] Peltola, H., Heinonen, O. P., Valle, M., Paunio, M., Virtanen, M., Karanko, V., and Cantell, K. (1994). The Elimination of Indigenous Measles, Mumps, and Rubella from Finland by a 12-Year, Two-Dose Vaccination Program. *New England Journal of Medicine*, 331(21):1397–1402.
- [Perelson and Oster, 1979] Perelson, A. S. and Oster, G. F. (1979). Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-non-self discrimination. *Journal of Theoretical Biology*, 81(4):645–670.
- [Perez-Andres et al., 2010] Perez-Andres, M., Paiva, B., Nieto, W. G., Caraux, A., Schmitz, A., Almeida, J., Vogt, R. F., Marti, G. E., Rawstron, A. C., Van Zelm, M. C., Van Dongen, J.

- J. M., Johnsen, H. E., Klein, B., and Orfao, A. (2010). Human peripheral blood B-Cell compartments: A crossroad in B-cell traffic. *Cytometry Part B - Clinical Cytometry*, 78(SUPPL. 1):47–60.
- [Plotkin, 2014] Plotkin, S. (2014). History of vaccination. *Proceedings of the National Academy of Sciences*, 111(34):12283–12287.
- [Plotkin et al., 2008] Plotkin, S., Orenstein, W., and Offit, P. (2008). *Vaccines*. Elsevier Health Sciences.
- [Ponti et al., 2013] Ponti, G., Obernier, K., Guinto, C., Jose, L., Bonfanti, L., and Alvarez-Buylla, A. (2013). Cell cycle and lineage progression of neural progenitors in the ventricular-subventricular zones of adult mice. *Proceedings of the National Academy of Sciences of the United States of America*, 110(11):E1045–54.
- [Press et al., 2007] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3 edition.
- [Purow et al., 2005] Purow, B. W., Haque, R. M., Noel, M. W., Su, Q., Burdick, M. J., Lee, J., Sundaresan, T., Pastorino, S., Park, J. K., Mikolaenko, I., Maric, D., Eberhart, C. G., and Fine, H. A. (2005). Expression of Notch-1 and its ligands, Delta-like-1 and Jagged=1, is critical for glioma cell survival and proliferation. *Cancer Research*, 65(6):2353–2363.
- [Quintana et al., 2008] Quintana, E., Shackleton, M., Sabel, M. S., Fullen, D. R., Johnson, T. M., and Morrison, S. J. (2008). Efficient tumour formation by single human melanoma cells. *Nature*, 456(7222):593–598.
- [Radbruch et al., 2006] Radbruch, A., Muehlinghaus, G., Luger, E. O., Inamine, A., Smith, K. G., Dörner, T., and Hiepe, F. (2006). Competence and competition: The challenge of becoming a long-lived plasma cell. *Nature Reviews Immunology*, 6(10):741–750.
- [Raue et al., 2009] Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929.
- [Reilly et al., 2015] Reilly, E. B., Phillips, A. C., Buchanan, F. G., Kingsbury, G., Zhang, Y., Meulbroek, J. A., Cole, T. B., DeVries, P. J., Falls, H. D., Beam, C., Gu, J., Digiammarino, E. L., Palma, J. P., Donawho, C. K., Goodwin, N. C., and Scott, A. M. (2015). Characterization of ABT-806, a Humanized Tumor-Specific Anti-EGFR Monoclonal Antibody. *Molecular Cancer Therapeutics*, 14(5):1141–1151.
- [Reilly et al., 2000] Reilly, K. M., Loisel, D. A., Bronson, R. T., McLaughlin, M. E., and Jacks, T. (2000). Nf1; trp53 mutant mice develop glioblastoma with evidence of strain-specific effects. *Nature genetics*, 26(1):109.
- [Reth, 1992] Reth, M. (1992). Antigen receptors on B lymphocytes. *Annual review of immunology*, 10:97–121.

- [Reynolds and Weiss, 1992] Reynolds, B. and Weiss, S. (1992). Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science*, 255(5052):1707–1710.
- [Rietze et al., 2001] Rietze, R. L., Valcanis, H., Brooker, G. F., Thomas, T., Voss, a. K., and Bartlett, P. F. (2001). Purification of a pluripotent neural stem cell from the adult mouse brain. *Nature*, 412(6848):736–739.
- [Robert-Koch-Institut, 2017] Robert-Koch-Institut (2017). Impfkalender (Standardimpfungen) für Säuglinge, Kinder, Jugendliche und Erwachsene. *Epidemiologisches Bulletin*, 34:336.
- [Roederer et al., 2015] Roederer, M., Quaye, L., Mangino, M., Beddall, M. H., Mahnke, Y., Chattopadhyay, P., Tosi, I., Napolitano, L., Terranova Barberio, M., Menni, C., Villanova, F., Di Meglio, P., Spector, T. D., and Nestle, F. O. (2015). The genetic architecture of the human immune system: A bioresource for autoimmunity and disease pathogenesis. *Cell*, 161(2):387–403.
- [Roth, 2014] Roth, D. B. (2014). V(D)J Recombination: Mechanism, Errors, and Fidelity. *Microbiology Spectrum*, 2(6):1–11.
- [RTSS Clinical Trials Partnership, 2015] RTSS Clinical Trials Partnership, T. (2015). Efficacy and safety of rts,s/as01 malaria vaccine with or without a booster dose in infants and children in africa: final results of a phase 3, individually randomised, controlled trial. *The Lancet*, 386(9988):31 – 45.
- [Rycaj and Tang, 2015] Rycaj, K. and Tang, D. G. (2015). Cell-of-origin of cancer versus cancer stem cells: Assays and interpretations. *Cancer Research*, 75(19):4003–4011.
- [Sánchez-Danés et al., 2016] Sánchez-Danés, A., Hannezo, E., Larsimont, J. C., Liagre, M., Youssef, K. K., Simons, B. D., and Blanpain, C. (2016). Defining the clonal dynamics leading to mouse skin tumour initiation. *Nature*, 536(7616):298–303.
- [Sandberg et al., 2013] Sandberg, C. J., Altschuler, G., Jeong, J., Strømme, K. K., Stangeland, B., Murrell, W., Grasmø-Wendler, U. H., Myklebost, O., Helseth, E., Vik-Mo, E. O., Hide, W., and Langmoen, I. A. (2013). Comparison of glioma stem cells to neural stem cells from the adult human brain identifies dysregulated Wnt- signaling and a fingerprint associated with clinical outcome. *Experimental Cell Research*, 319(14):2230–2243.
- [Sattabongkot et al., 2006] Sattabongkot, J., Yimamnuaychoke, N., Leelaudomlipi, S., Rasameesoraj, M., Jenwithisuk, R., Coleman, R. E., Udomsangpetch, R., Cui, L., and Brewer, T. G. (2006). Establishment of a human hepatocyte line that supports in vitro development of the exo-erythrocytic stages of the malaria parasites *Plasmodium falciparum* and *P. vivax*. *American Journal of Tropical Medicine and Hygiene*, 74(5):708–715.
- [Schepers et al., 2012] Schepers, A. G., Snippert, H. J., Stange, D. E., van den Born, M., van Es, J. H., van de Wetering, M., and Clevers, H. (2012). Lineage Tracing Reveals Lgr5+ Stem Cell Activity in Mouse Intestinal Adenomas. *Science*, 337(6095):730–735.

- [Schroeder and Cavacini, 2010] Schroeder, H. W. J. and Cavacini, L. (2010). Structure and Function of Immunoglobulins (author manuscript). *Journal of Allergy and Clinical Immunology*, 125:S41–S52.
- [Schweitzer et al., 2001] Schweitzer, G. H., Vince, C., Roosen, H. K., and Tonn, J. C. (2001). Extraneural metastases of primary brain tumors. *Journal of Neuro-Oncology*, 53(2):107–114.
- [Shaffer et al., 2016] Shaffer, J. S., Moore, P. L., Kardar, M., and Chakraborty, A. K. (2016). Optimal immunization cocktails can promote induction of broadly neutralizing Abs against highly mutable pathogens. *Proceedings of the National Academy of Sciences*, 113(45):E7039–E7048.
- [Shah et al., 2003] Shah, K., Tang, Y., Breakefield, X., and Weissleder, R. (2003). Real-time imaging of TRAIL-induced apoptosis of glioma tumors in vivo. *Oncogene*, 22(44):6865–6872.
- [Sherratt and Chaplain, 2001] Sherratt, J. and Chaplain, M. (2001). A new mathematical model for avascular tumour growth. *Journal of Mathematical Biology*, 43:291–312.
- [Shi et al., 2004] Shi, Y., Chichung Lie, D., Taupin, P., Nakashima, K., Ray, J., Yu, R. T., Gage, F. H., and Evans, R. M. (2004). Expression and function of orphan nuclear receptor TLX in adult neural stem cells. *Nature*, 427(6969):78–83.
- [Shlomchik, 2018] Shlomchik, M. J. (2018). Do Memory B Cells Form Secondary Germinal Centers? *Cold Spring Harbor Perspectives in Biology*, 10(1):a029405.
- [Shlomchik et al., 1998] Shlomchik, M. J., Watts, P., Weigert, M. G., and Litwin, S. (1998). Clone: a Monte-Carlo computer simulation of B cell clonal expansion, somatic mutation, and antigen-driven selection.
- [Shlomchik and Weisel, 2012] Shlomchik, M. J. and Weisel, F. (2012). Germinal centers. *Immunological reviews*, 247(1):5–10.
- [Shulman et al., 2014] Shulman, Z., Gitlin, A. D., Weinstein, J. S., Lainez, B., Esplugues, E., Flavell, R. A., Craft, J. E., and Nussenzweig, M. C. (2014). Dynamic signaling by T follicular helper cells during germinal center B cell selection. *Science*, 345(6200):1058–1062.
- [Singh et al., 2003] Singh, S. K., Clarke, I. D., Terasaki, M., Bonn, V. E., Hawkins, C., Squire, J., and Dirks, P. B. (2003). Identification of a cancer stem cell in human brain tumors. *Cancer Res.*, 63(18):5821–5828.
- [Singh et al., 2004] Singh, S. K., Hawkins, C., Clarke, I. D., Squire, J. A., Bayani, J., Hide, T., Henkelman, R. M., Cusimano, M. D., and Dirks, P. B. (2004). Identification of human brain tumour initiating cells. *Nature*, 432(7015):396–401.
- [Smadbeck and Kaznessis, 2013] Smadbeck, P. and Kaznessis, Y. N. (2013). A closure scheme for chemical master equations. *Proceedings of the National Academy of Sciences*, 110(35):14261–14265.

- [Smith et al., 1997] Smith, D. J., Forrest, S., Hightower, R. R., and Perelson, a. S. (1997). Deriving shape space parameters from immunological data. *Journal of theoretical biology*, 189(2):141–50.
- [Sorensen and Gianola, 2002] Sorensen, D. and Gianola, D. (2002). *Likelihood of Bayesian, and MCMC Methods in Quantitative Genetics*. Springer, New York.
- [Sottoriva et al., 2013] Sottoriva, A., Spiteri, I., Piccirillo, S. G. M., Touloumis, A., Collins, V. P., Marioni, J. C., Curtis, C., Watts, C., and Tavare, S. (2013). Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences*, 110(10):4009–4014.
- [Sottoriva et al., 2010] Sottoriva, A., Verhoeff, J. J., Borovski, T., McWeeney, S. K., Naumov, L., Medema, J. P., Sloot, P. M., and Vermeulen, L. (2010). Cancer Stem Cell Tumor Model Reveals Invasive Morphology and Increased Phenotypical Heterogeneity. *Cancer Research*, 70(1):46–56.
- [Spangrude et al., 1988] Spangrude, G., Heimfeld, S., and Weissman, I. (1988). Purification and characterization of mouse hematopoietic stem cells. *Science*, 241(4861):58–62.
- [Stein et al., 2018] Stein, S., Zhao, R., Haeno, H., Vivanco, I., and Michor, F. (2018). Mathematical modeling identifies optimum lapatinib dosing schedules for the treatment of glioblastoma patients. *PLoS Computational Biology*, 14(1):1–24.
- [Stern et al., 2008] Stern, P., Astrof, S., Erkeland, S. J., Schustak, J., Sharp, P. A., and Hynes, R. O. (2008). A system for Cre-regulated RNA interference in vivo. *Proceedings of the National Academy of Sciences*, 105(37):13895–13900.
- [Stupp et al., 2005] Stupp, R., Mason, W. P., van den Bent, M. J., Weller, M., Fisher, B., Taphoorn, M. J., Belanger, K., Brandes, A. A., Marosi, C., Bogdahn, U., Curschmann, J., Janzer, R. C., Ludwin, S. K., Gorlia, T., Allgeier, A., Lacombe, D., Cairncross, J. G., Eisenhauer, E., and Mirimanoff, R. O. (2005). Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *New England Journal of Medicine*, 352(10):987–996.
- [Sudmant et al., 2015] Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M. H. Y., Konkel, M. K., Malhotra, A., Stütz, A. M., Shi, X., Casale, F. P., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., Malig, M., Chaisson, M. J., Walter, K., Meiers, S., Kashin, S., Garrison, E., Auton, A., Lam, H. Y., Mu, X. J., Alkan, C., Antaki, D., Bae, T., Cerveira, E., Chines, P., Chong, Z., Clarke, L., Dal, E., Ding, L., Emery, S., Fan, X., Gujral, M., Kahveci, F., Kidd, J. M., Kong, Y., Lameijer, E. W., McCarthy, S., Flicek, P., Gibbs, R. A., Marth, G., Mason, C. E., Menelaou, A., Muzny, D. M., Nelson, B. J., Noor, A., Parrish, N. F., Pendleton, M., Quitadamo, A., Raeder, B., Schadt, E. E., Romanovitch, M., Schlattl, A., Sebra, R., Shabalina, A. A., Untergasser, A., Walker, J. A., Wang, M., Yu, F., Zhang, C., Zhang, J., Zheng-Bradley, X., Zhou, W., Zichner, T., Sebati, J., Batzer, M. A., McCarroll, S. A., Mills, R. E., Gerstein, M. B., Bashir, A., Stegle, O., Devine, S. E., Lee, C., Eichler, E. E., and Korb, J. O. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, 526(7571):75–81.

- [Sun et al., 2005] Sun, J., Earl, D., and Deem, M. (2005). Glassy Dynamics in the Adaptive Immune Response Prevents Autoimmune Disease. *Physical Review Letters*, 95(14):1–4.
- [Suvà et al., 2014] Suvà, M. L., Rheinbay, E., Gillespie, S. M., Patel, A. P., Wakimoto, H., Rabkin, S. D., Riggi, N., Chi, A. S., Cahill, D. P., Nahed, B. V., Curry, W. T., Martuza, R. L., Rivera, M. N., Rossetti, N., Kasif, S., Beik, S., Kadri, S., Tirosh, I., Wortman, I., Shalek, A. K., Rozenblatt-Rosen, O., Regev, A., Louis, D. N., and Bernstein, B. E. (2014). Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell*, 157(3):580–594.
- [Swan et al., 2017] Swan, A., Hillen, T., Bowman, J. C., and Murtha, A. D. (2017). A Patient-Specific Anisotropic Diffusion Model for Brain Tumour Spread. *Bulletin of Mathematical Biology*, pages 1–33.
- [Swanson et al., 2000] Swanson, K. R., Alvord, E. C., and Murray, J. D. (2000). Quantitative Model for Differential Motility of Gliomas in Grey and White Matter. *Cell Prolif.*, 33:317–329.
- [Szentirmai et al., 2006] Szentirmai, O., Baker, C. H., Lin, N., Szucs, S., Takahashi, M., Kiryu, S., Kung, A. L., Mulligan, R. C., and Carter, B. S. (2006). Noninvasive bioluminescence imaging of luciferase expressing intracranial U87 xenografts: Correlation with magnetic resonance imaging determined tumor volume and longitudinal use in assessing tumor growth and antiangiogenic treatment effect. *Neurosurgery*, 58(2):365–372.
- [Talkington and Durrett, 2015] Talkington, A. and Durrett, R. (2015). Estimating Tumor Growth Rates In Vivo. *Bulletin of Mathematical Biology*, 77(10):1934–1954.
- [Tannishtha et al., 2001] Tannishtha, R., Morrison, S. J., Clarke, M. F., and Weissman, I. L. (2001). Stem Cells, Cancer, and Cancer Stem Cells. *Nature*, 414(November):105–111.
- [Tas et al., 2016] Tas, J. M. J., Mesin, L., Pasqual, G., Targ, S., Jacobsen, J. T., Mano, Y. M., Chen, C. S., Weill, J.-C., Reynaud, C.-A., Browne, E. P., Meyer-Hermann, M., and Victorica, G. D. (2016). Visualizing antibody affinity maturation in germinal centers. *Science*, 351(6277):1048–1054.
- [Taylor et al., 2012] Taylor, J. J., Pape, K. A., and Jenkins, M. K. (2012). A germinal center-independent pathway generates unswitched memory B cells early in the primary response. *The Journal of Experimental Medicine*, 209(3):597–606.
- [Thomas and Dill, 1996] Thomas, P. D. and Dill, K. A. (1996). An iterative method for extracting energy-like quantities from protein structures. *Proceedings of the National Academy of Sciences of the United States of America*, 93(21):11628–11633.
- [Thorbecke et al., 1962] Thorbecke, G. J., Asofsky, R. M., Hochwald, G. M., and Siskind, G. W. (1962). Gamma globulin and antibody formation in vitro. III. Induction of secondary response at different intervals after the primary; the role of secondary nodules in the preparation for the secondary response. *The Journal of experimental medicine*, 116:295–310.
- [Tiller et al., 2008] Tiller, T., Meffre, E., Yurasov, S., Tsuiji, M., Nussenzweig, M. C., and Wardemann, H. (2008). Efficient generation of monoclonal antibodies from single human B cells

- by single cell RT-PCR and expression vector cloning. *Journal of Immunological Methods*, 329(1-2):112–124.
- [Tomasetti and Vogelstein, 2015] Tomasetti, C. and Vogelstein, B. (2015). Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, 347(6217):78–81.
- [Triller et al., 2017] Triller, G., Scally, S. W., Costa, G., Pissarev, M., Kreschel, C., Bosch, A., Marois, E., Sack, B. K., Murugan, R., Salman, A. M., Janse, C. J., Khan, S. M., Kappe, S. H., Adegnik, A. A., Mordmüller, B., Levashina, E. A., Julien, J.-P., and Wardemann, H. (2017). Natural Parasite Exposure Induces Protective Human Anti-Malarial Antibodies. *Immunity*, pages 1197–1209.
- [Trück et al., 2015] Trück, J., Ramasamy, M. N., Galson, J. D., Rance, R., Parkhill, J., Lunter, G., Pollard, A. J., and Kelly, D. F. (2015). Identification of Antigen-Specific B Cell Receptor Sequences Using Public Repertoire Analysis. *The Journal of Immunology*, 194(1):252–261.
- [Uhrbom et al., 2004] Uhrbom, L., Nerio, E., and Holland, E. C. (2004). Dissecting tumor maintenance requirements using bioluminescence imaging of cell proliferation in a mouse glioma model. *Nature Medicine*, 10(11):1257–1260.
- [Venkatesan et al., 2017] Venkatesan, S., Swanton, C., Taylor, B. S., and Costello, J. F. (2017). Treatment-Induced Mutagenesis and Selective Pressures Sculpt Cancer Evolution. *Cold Spring Harbor Perspectives in Medicine*, 7(8):a026617.
- [Venzon and Moolgavkar, 1988] Venzon, D. J. and Moolgavkar, S. H. (1988). A method for computing profile-likelihood-based confidence-intervals. *Appl. Stat. J. Roy. Statist. Soc. Ser. C* 37(1):87–94.
- [Victora and Nussenzweig, 2012] Victora, G. D. and Nussenzweig, M. C. (2012). Germinal Centers. *Annual Review of Immunology*, 30(1):429–457.
- [Victora et al., 2010] Victora, G. D., Schwickert, T. A., Fooksman, D. R., Kamphorst, A. O., Meyer-Hermann, M., Dustin, M. L., and Nussenzweig, M. C. (2010). Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. *Cell*, 143(4):592–605.
- [Wachsberger et al., 2005] Wachsberger, P. R., Burd, R., Marero, N., Daskalakis, C., Ryan, A., McCue, P., and Dicker, A. P. (2005). Effect of the tumor vascular-damaging agent, ZD6126, on the radioresponse of U87 glioblastoma. *Clinical Cancer Research*, 11(2 I):835–842.
- [Waclaw et al., 2015] Waclaw, B., Bozic, I., Pittman, M. E., Hruban, R. H., Vogelstein, B., and Nowak, M. A. (2015). A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*, 525(7568):261–264.
- [Wallace and Guo, 2013] Wallace, D. I. and Guo, X. (2013). Properties of tumor spheroid growth exhibited by simple mathematical models. *Frontiers in oncology*, 3(March):51.

- [Wang et al., 2015a] Wang, C., Liu, Y., Cavanagh, M. M., Le Saux, S., Qi, Q., Roskin, K. M., Looney, T. J., Lee, J.-Y., Dixit, V., Dekker, C. L., Swan, G. E., Goronzy, J. J., and Boyd, S. D. (2015a). B-cell repertoire responses to varicella-zoster vaccination in human identical twins. *Proceedings of the National Academy of Sciences*, 112(2):500–505.
- [Wang et al., 2009] Wang, C. H., Rockhill, J. K., Mrugala, M., Peacock, D. L., Lai, A., Jusenius, K., Wardlaw, J. M., Cloughesy, T., Spence, A. M., Rockne, R., Alvord, E. C., and Swanson, K. R. (2009). Prognostic significance of growth kinetics in newly diagnosed glioblastomas revealed by combining serial imaging with a novel biomathematical model. *Cancer Research*, 69(23):9133–9140.
- [Wang et al., 2016] Wang, J., Cazzato, E., Ladewig, E., Frattini, V., Rosenbloom, D. I., Zairis, S., Abate, F., Liu, Z., Elliott, O., Shin, Y. J., Lee, J. K., Lee, I. H., Park, W. Y., Eoli, M., Blumberg, A. J., Lasorella, A., Nam, D. H., Finocchiaro, G., Iavarone, A., and Rabadan, R. (2016). Clonal evolution of glioblastoma under therapy. *Nature Genetics*, 48(7):768–776.
- [Wang et al., 2015b] Wang, S., Mata-Fink, J., Kriegsman, B., Hanson, M., Irvine, D., Eisen, H., Burton, D., Wittrup, K., Kardar, M., and Chakraborty, A. (2015b). Manipulating the Selection Forces during Affinity Maturation to Generate Cross-Reactive HIV Antibodies. *Cell*, 160(4):785–797.
- [Wang et al., 2015c] Wang, Z., Butner, J. D., Kerketta, R., Cristini, V., and Deisboeck, T. S. (2015c). Simulating cancer growth with multiscale agent-based modeling. *Seminars in Cancer Biology*, 30:70–78.
- [Wardemann et al., 2003] Wardemann, H., Yurasov, S., Schaefer, a., Young, J., Meffre, E., and Nussenzweig, M. (2003). Predominant autoantibody production by early human B cell precursors. *Science*, 301(5638):1374–7.
- [Watson et al., 2015] Watson, J. K., Rulands, S., Wilkinson, A. C., Wuidart, A., Ousset, M., Van Keymeulen, A., Goettgens, B., Blanpain, C., Simons, B. D., and Rawlins, E. L. (2015). Clonal Dynamics Reveal Two Distinct Populations of Basal Cells in Slow-Turnover Airway Epithelium. *Cell Reports*, 12(1):90–101.
- [Weekes et al., 2014] Weekes, S. L., Barker, B., Bober, S., Cisneros, K., Cline, J., Thompson, A., Hlatky, L., Hahnfeldt, P., and Enderling, H. (2014). A Multicompartment Mathematical Model of Cancer Stem Cell-Driven Tumor Growth Dynamics. *Bulletin of Mathematical Biology*, 76(7):1762–1782.
- [Wei et al., 2015] Wei, L., Chahwan, R., Wang, S., Wang, X., Pham, P. T., Goodman, M. F., Bergman, A., Scharff, M. D., and MacCarthy, T. (2015). Overlapping hotspots in CDRs are critical sites for V region diversification. *Proceedings of the National Academy of Sciences*, 112(7):E728–E737.
- [Weisel et al., 2016] Weisel, F., Zuccarino-Catania, G., Chikina, M., and Shlomchik, M. (2016). A Temporal Switch in the Germinal Center Determines Differential Output of Memory B and Plasma Cells. *Immunity*, 44(1):116–130.

- [Welinder et al., 2011] Welinder, E., Åhsberg, J., and Sigvardsson, M. (2011). B-lymphocyte commitment: Identifying the point of no return. *Seminars in Immunology*, 23(5):335–340.
- [White et al., 2014] White, N. J., Pukrittayakamee, S., Hien, T. T., Faiz, M. A., Mokuolu, O. A., and Dondorp, A. M. (2014). Malaria. *Lancet (London, England)*, 383(9918):723–35.
- [Wittenbrink et al., 2011] Wittenbrink, N., Klein, A., Weiser, A. A., Schuchhardt, J., and Or-Guil, M. (2011). Is There a Typical Germinal Center? A Large-Scale Immunohistological Study on the Cellular Composition of Germinal Centers during the Hapten-Carrier-Driven Primary Immune Response in Mice. *The Journal of Immunology*, 187(12):6185–6196.
- [Wittenbrink et al., 2010] Wittenbrink, N., Weber, T. S., Klein, A., Weiser, A. A., Zuschtratter, W., Sibila, M., Schuchhardt, J., and Or-Guil, M. (2010). Broad volume distributions indicate nonsynchronized growth and suggest sudden collapses of germinal center B cell populations. *J Immunol*, 184(3):1339–47.
- [Xu and Davis, 2000] Xu, J. L. and Davis, M. M. (2000). Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity*, 13(1):37–45.
- [Xue et al., 2014] Xue, W., Chen, S., Yin, H., Tammela, T., Papagiannakopoulos, T., Joshi, N. S., Cai, W., Yang, G., Bronson, R., Crowley, D. G., Zhang, F., Anderson, D. G., Sharp, P. A., and Jacks, T. (2014). CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature*, 514(7522):380–384.
- [Yaari and Kleinstein, 2015] Yaari, G. and Kleinstein, S. H. (2015). Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Medicine*, 7(1):1–14.
- [Yan et al., 2017] Yan, H., Romero-López, M., Benitez, L. I., Di, K., Frieboes, H. B., Hughes, C. C., Bota, D. A., and Lowengrub, J. S. (2017). 3D mathematical modeling of glioblastoma suggests that transdifferentiated vascular endothelial cells mediate resistance to current standard-of-care therapy. *Cancer Research*, 77(15):4171–4184.
- [Yeap et al., 2015] Yeap, L. S., Hwang, J. K., Du, Z., Meyers, R. M., Meng, F. L., Jakubauskaite, A., Liu, M., Mani, V., Neuberg, D., Kepler, T. B., Wang, J. H., and Alt, F. W. (2015). Sequence-Intrinsic Mechanisms that Target AID Mutational Outcomes on Antibody Genes. *Cell*, 163(5):1124–1137.
- [Yoshida et al., 2010] Yoshida, T., Mei, H., Dörner, T., Hiepe, F., Radbruch, A., Fillatreau, S., and Hoyer, B. F. (2010). Memory B and memory plasma cells. *Immunological Reviews*, 237(1):117–139.
- [Yu et al., 2015] Yu, F. X., Zhao, B., and Guan, K. L. (2015). Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. *Cell*, 163(4):811–828.
- [Yuda and Ishino, 2004] Yuda, M. and Ishino, T. (2004). Liver invasion by malarial parasites - How do malarial parasites break through the host barrier? *Cellular Microbiology*, 6(12):1119–1125.

- [Zavala et al., 1983] Zavala, F., Cochrane, A. H., Nardin, E. H., Nussenzweig, R. S., and Nussenzweig, V. (1983). Circumsporozoite proteins of malaria parasites contain a single immunodominant region with two or more identical epitopes. *The Journal of experimental medicine*, 157(6):1947–57.
- [Zhang and Shakhnovich, 2010] Zhang, J. and Shakhnovich, E. I. (2010). Optimality of mutation and selection in germinal centers. *PLoS computational biology*, 6(6):e1000800.
- [Zhang et al., 2012] Zhang, J., Stevens, M. F. G., and Bradshaw, T. D. (2012). Temozolomide: mechanisms of action, repair and resistance. *Current molecular pharmacology*, 5(1):102–114.
- [Zheng et al., 2005] Zheng, N.-Y., Wilson, K., Jared, M., and Wilson, P. C. (2005). Intricate targeting of immunoglobulin somatic hypermutation maximizes the efficiency of affinity maturation. *The Journal of Experimental Medicine*, 201(9):1467–1478.
- [Zhu, 2013] Zhu, Z. (2013). *Genetic dissection of the role of self-renewal pathway in brain tumors*. PhD thesis, Ruprecht-Karls-Universität Heidelberg.
- [Zhu et al., 2014] Zhu, Z., Khan, M. A., Weiler, M., Blaes, J., Jestaedt, L., Geibert, M., Zou, P., Gronych, J., Bernhardt, O., Korshunov, A., Bugner, V., Lichter, P., Radlwimmer, B., Heiland, S., Bendszus, M., Wick, W., and Liu, H. K. (2014). Targeting self-renewal in high-grade brain tumors leads to loss of brain tumor stem cells and prolonged survival. *Cell Stem Cell*, 15(2):185–198.
- [Zou, 2016] Zou, P. (2016). *A novel mouse glioma model for robust in vivo target validation and lineage visualization*. PhD thesis, Ruprecht-Karls-Universität Heidelberg.
- [Zuckermann et al., 2015] Zuckermann, M., Hovestadt, V., Knobbe-Thomsen, C. B., Zapatka, M., Northcott, P. A., Schramm, K., Belic, J., Jones, D. T. W., Tschida, B., Moriarity, B., Largaespada, D., Roussel, M. F., Korshunov, A., Reifenberger, G., Pfister, S. M., Lichter, P., Kawauchi, D., and Gronych, J. (2015). Somatic CRISPR/Cas9-mediated tumour suppressor disruption enables versatile brain tumour modelling. *Nature Communications*, 6(1):7391.

Aber die Intellektualität ist, wie wir wissen, nur der Ausdruck oder das Werkzeug eines ausgetrockneten Lebens; [...].

– Robert Musil, *Mann ohne Eigenschaften*

I profoundly thank my supervisor Thomas Höfer for providing me with a research environment characterised by freedom and independence, for sharing funny anecdotes and deep insights, for good books and good discussions, for letting me travel half the world during these three and a half years and for giving me the feeling that there would always be support if ever the worst came to the worst.

I deeply thank Hedda Wardemann and Ilka Bischofs for taking time to listen to my occasional rants, providing thoughtful mentoring when needed and giving plenty of useful advice.

I thank my office mates Nils and Jens for many scientific, semi-scientific and not-so-scientific discussions, shared snacks and coffees and a generally very enjoyable atmosphere, even when the office was overheated (typically from March to November).

I thank all colleagues from the Höfer lab for all the good times we spent together: small moments like shared lunchtimes and after-work football as well as big events like our superb annual group retreats and beautiful Christmas parties.

I thank my previous supervisors José Crespo and Kamran Behnia for their steady interest in my progress, their kind words and their continued support with reference letters and crisis management.

I thank Hedda Wardemann and her entire group, especially Rajagopal Murugan, for the exciting and close collaboration on B cell memory development in response to Malaria, for letting me participate in their group meetings and for teaching me a great many things about immunology.

I thank Hai-Kun Liu and his group for the interesting and productive collaboration on glioblastoma growth.

I thank Nir Friedman and his entire group at the Weizmann Institute, especially Michal Polonsky, for the enjoyable collaboration on T cell diversification and the warm welcome during my research stay with them.

I thank the Bachelor's and Master's students with whom I had the pleasure to collaborate - Charlotte, Daniel, Maurice, Michael and Juan - for their good work and everything that they taught me about myself.

I thank Diana, Nick and Conny for their kind and highly professional support in almost any organisational matter.

I thank Verena and Gopal for their generous help with proofreading this thesis.

I thank Ursula Kummer and Hai-Kun Liu for agreeing to be part of my defence committee.