



Article

# Computational Model Reveals a Stochastic Mechanism behind Germinal Center Clonal Bursts

Aurélien Péliissier <sup>1,2,\*</sup> , Youcef Akrouit <sup>3</sup>, Katharina Jahn <sup>2</sup>, Jack Kuipers <sup>2</sup>, Ulf Klein <sup>4</sup>, Niko Beerenwinkel <sup>2</sup> and María Rodríguez Martínez <sup>1,\*</sup> 

<sup>1</sup> IBM Research Zurich, 8803 Rüschlikon, Switzerland

<sup>2</sup> Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland; katharina.jahn@bsse.ethz.ch (K.J.); jack.kuipers@bsse.ethz.ch (J.K.); niko.beerenwinkel@bsse.ethz.ch (N.B.)

<sup>3</sup> École Normale Supérieure, 75005 Paris, France; youcef.akrouit@ens.fr

<sup>4</sup> Leeds Institute of Medical Research at St. James's, University of Leeds, Leeds LS9 7TF, UK; u.p.klein@leeds.ac.uk

\* Correspondence: pel@zurich.ibm.com (A.P.); mrm@zurich.ibm.com (M.R.M.)

Received: 15 April 2020; Accepted: 8 June 2020; Published: 10 June 2020



**Abstract:** Germinal centers (GCs) are specialized compartments within the secondary lymphoid organs where B cells proliferate, differentiate, and mutate their antibody genes in response to the presence of foreign antigens. Through the GC lifespan, interclonal competition between B cells leads to increased affinity of the B cell receptors for antigens accompanied by a loss of clonal diversity, although the mechanisms underlying clonal dynamics are not completely understood. We present here a multi-scale quantitative model of the GC reaction that integrates an intracellular component, accounting for the genetic events that shape B cell differentiation, and an extracellular stochastic component, which accounts for the random cellular interactions within the GC. In addition, B cell receptors are represented as sequences of nucleotides that mature and diversify through somatic hypermutations. We exploit extensive experimental characterizations of the GC dynamics to parameterize our model, and visualize affinity maturation by means of evolutionary phylogenetic trees. Our explicit modeling of B cell maturation enables us to characterise the evolutionary processes and competition at the heart of the GC dynamics, and explains the emergence of clonal dominance as a result of initially small stochastic advantages in the affinity to antigen. Interestingly, a subset of the GC undergoes massive expansion of higher-affinity B cell variants (clonal bursts), leading to a loss of clonal diversity at a significantly faster rate than in GCs that do not exhibit clonal dominance. Our work contributes towards an *in silico* vaccine design, and has implications for the better understanding of the mechanisms underlying autoimmune disease and GC-derived lymphomas.

**Keywords:** GC; B cell receptor; somatic hypermutation; nucleotide sequence; antibody affinity; affinity maturation; immunoglobulin; clonal competition; clonal burst; clonal dominance; stochastic models; affinity models

## 1. Introduction

Germinal centers (GCs) [1–3] are specialized microanatomical structures that emerge within the secondary lymphoid organs upon infection or immunization. A GC typically contains up to a few thousand B cells that rapidly proliferate and mutate the immunoglobulin genes of their B cell receptors (BCRs). GCs play a central role in mounting an effective immune response against infectious pathogens, and failures within their tightly regulated environment can lead to the development of autoimmune diseases [4] and cancer [5]. Hence, a detailed investigation of the processes that regulate the dynamics of the GC reaction is of crucial importance to understand both healthy and pathological immune responses.

GCs are histologically divided into two different compartments, the dark zone (DZ) and light zone (LZ), which play distinct and important roles in the maturation and selection of B cells [6]. The DZ consists primarily of a tight cluster of highly proliferative B cells, known as centroblasts (CBs), while the LZ is less compact and more diverse, including GC B cells, known as centrocytes (CCs), antigen-presenting follicular dendritic cells (FDCs) and a smaller population of T follicular helper cells ( $T_{FH}$ ). A GC reaction is initiated when naive B cells are activated by  $T_{FH}$ s. Upon activation, naive B cells initiate a phase of monoclonal expansion characterized by very rapid proliferation rates and somatic hypermutation (SHM) of their immunoglobulin genes [7,8]. After a few days, GCs become polarized into the DZ and LZ, and CBs migrate to the LZ, where they undergo a phase of selection. CCs in the light zone are committed to apoptosis and compete for survival signals delivered by  $T_{FH}$ s. To receive such signals, CCs must first acquire antigen from FDCs. Higher affinity CCs can capture larger amounts of antigen, which is then engulfed, digested into peptides and returned to the B-cell surface bound to the major histocompatibility MHC class II.  $T_{FH}$ s that bind with enough strength to the peptide–MHC complex (pMHC) deliver signals that stop apoptosis, upon which a CC can leave the GC and terminally differentiate into a plasma cell (PC), responsible for secreting antibodies, or into a long-lived memory B cell (MBC) that keeps memory of past infections and can rapidly respond to repeated antigen exposure. Low affinity cells that do not receive enough  $T_{FH}$  signaling are eliminated by apoptosis in a process that replicates Darwinian evolution at the cellular level.

In addition, a fraction of CCs return to the DZ for additional rounds of cell division and BCR maturation [9]. The speed of the cell cycle in the DZ is regulated by the amount of signalling received from the  $T_{FH}$ s [10], likely by upregulating cell-cycle regulators such as MYC [11–13]. As a result, high-affinity cells that receive strong  $T_{FH}$ s signals undergo accelerated cell cycles and can replicate up to 6 times, while lower affinity cells that capture less antigen divide fewer times [14]. The regulation of the cell cycle critically contributes to the selection and clonal expansion of high-affinity cells as well as to the observed progressive decline of clonal diversity in at least a subset of GCs [15], although detailed quantitative models are still needed to understand mechanisms behind clonal evolution, competition and clonal burst induction.

**Quantitative modelling of GCs:** At the molecular level, the intracellular mechanisms associated with regulation of the B cells,  $T_{FH}$ s and FDCs interactions implicates more than 100 transcription factors [16], most of which interact in highly regulated non-linear networks [17], making the precise quantitative modeling of GC reaction tremendously complex. As GCs are stochastic systems that display a high level of variability even within the same lymph node of the same individual [18], mathematical models have been widely used to deepen our understanding of the cellular and molecular processes characterising these complex dynamic systems [19]. In particular, multi-scale stochastic [20] and spatial agent-based models have been proposed [21–23]. The advantage of such models is their faithful replication of the probabilistic interactions between the different cellular populations in the GC. Spatial models can capture the spatial dynamics and cellular flow between the two GC compartments, although they are encumbered with several methodological challenges and computational complexity. In comparison to spatial models, stochastic models offer fast and efficient computation of the main statistical properties of the GC with the theoretical guarantees of convergence to the exact probabilistic cellular distributions. Alternatively, computational models based on ordinary differential equations (ODEs) tracking the evolution of individual cells have also been proposed, and concluded that there is limited correlation between subclone abundance and affinity [24]. Other ODE models [25] were used to look at clonal diversity with a simple birth, death and mutation model.

While these models have successfully reproduced the GC dynamics and B cell maturation process, the accurate investigation of clonal diversity and burst emergence requires detailed modelling of the affinity maturation process based on more realistic representations of the BCRs, where the impact of newly acquired somatic mutations on the antigen binding capacity can be assessed. Advances in next-generation sequencing of immunoglobulin genes (Ig-seq) have revealed the dynamics of BCR sequence diversification across different B cell types in healthy and antigen-stimulated B cell donors [26]; however, structural

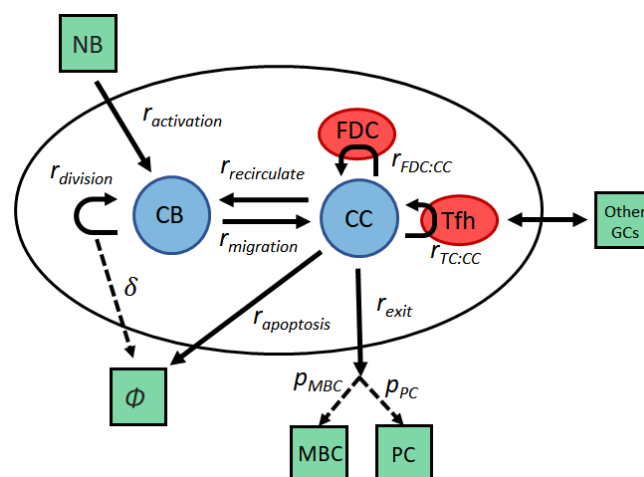
information about the BCR, which is crucial to accurately model antibody binding capability [27], is not available in a sequencing experiment. Preliminary work on BCR structural representations that model the BCR as a short amino acid chain of ( $\sim 10$  AAs) on a rectangular lattice has recently been developed [28]. Such detailed modelling comes however at a high computational cost—a single GC simulation takes a few hours, which currently limits its application to GC simulations.

In this paper we build on a previous multi-scale GC model, where the cellular interactions of B cells,  $T_{FH}$  and FDCs were modelled stochastically using the Gillespie algorithm [20]. We improve the model by adding an abstract molecular representation of the BCR based on its nucleotide sequence that allows us to track the effect of SHM on the BCR affinity through time. Our quantitative GC model is driven by 10 stochastic cellular interactions, which are parameterized using recent experimental characterizations of the GC dynamics. We expand the previous model by adding new regulatory mechanisms, such as centroblast division regulation by MYC, or centroblast apoptosis due to nonproductive BCRs, and use the extended model to explore different differentiation scenarios of B cells. Our sequence-based BCR representation enables a detailed investigation of clonal evolution using phylogenetic trees [29], and an analysis of the factors that determine the emergence of clonal dominance. Our paper summarizes the current quantitative knowledge about the GC responses, and gives additional information about clonal diversity and B-cell affinity maturation, which is particularly challenging to obtain from experimental data.

## 2. Methods: A Probabilistic Model of GC Dynamics

### 2.1. Stochastic Model of the GC

Our GC model includes reactions that occur in the low frequency regime where stochastic effects cannot be neglected, and hence cannot be simulated using ordinary differential equations. Instead, we use the formalism of chemical kinetics, where each reaction represents a stochastic process between the following reactants: naive B cells (NB), centroblasts (CB), centrocytes (CC), selected centrocytes ( $CC_{sel}$ ), memory B cells (MBC), plasma cells (PC), T follicular helper cells ( $T_{FH}$ ), and follicular dendritic cells (FDC). The set of reactions are given in Table 1, and graphically depicted in Figure 1.



**Figure 1.** Probabilistic model for the germinal center (GC) dynamics. Activated naive B cells (NB) enter the GC as centroblasts (CBs), which rapidly divide and mutate their immunoglobulin variable regions. If mutation renders a BCR nonproductive, the CB commits apoptosis ( $\emptyset$ ). After a few hours, CBs migrate to the light zone and become centrocytes (CCs). High-affinity CCs can acquire large amounts of antigen from follicular dendritic cells (FDCs), which makes them more likely to receive survival signals from helper T cells ( $T_{FH}$ ). Low affinity CCs, which acquire little or no antigen, cannot compete for T cell help and die through apoptosis ( $\emptyset$ ). CCs that have received sufficient  $T_{FH}$  signals may either recirculate into the dark zone for another round of division and hypermutation, or leave the GC as a memory B cell (MBC) or a plasma cell (PC).

**Table 1.** Set of reactions used in our germinal center model, along with the reactions rate values obtained after optimization, as detailed in Section 2.10.

Reaction	Description	Parameter Value
NB $\xrightarrow{r_{\text{activation}}}$ CB	B cell activation	$r_{\text{activation}} = 3.94$ /h
CB $\xrightarrow{r_{\text{division}}}$ 2CB where each daughter CB undergoes apoptosis (CB $\rightarrow \emptyset$ ) with probability $\delta$ per acquired mutation.	Centroblast division	$r_{\text{division}} = 0.134$ /h $\delta = 0.52$
CB $\xrightarrow{r_{\text{migration}}}$ CC	Centroblast migration to LZ	$r_{\text{migration}} = 3.75$ /h
CC $\xrightarrow{r_{\text{apoptosis}}}$ $\emptyset$	Centrocyte apoptosis	$r_{\text{apoptosis}} = 0.084$ /h
CC + FDC $\xrightarrow{r_{\text{FDC:CC}}}$ CC(pMHC) + FDC	Centrocyte antigen uptake	$r_{\text{FDC:CC}} = \frac{40}{N_{\text{CC}}}$ /h
CC + T <sub>FH</sub> $\xrightarrow{r_{\text{TC:CC}}}$ [CCTC]	Centrocyte binding to T <sub>FH</sub>	$r_{\text{TC:CC}} = \frac{145}{N_{\text{CC}}}$ /h
[CC <sub>1</sub> TC] + CC <sub>2</sub> $\xrightarrow{r_{\text{TC:CC}}}$ CC <sub>1</sub> + [CC <sub>2</sub> TC]	Centrocyte T <sub>FH</sub> switch	
[CCTC] $\xrightarrow{r_{\text{unbinding}}}$ CC <sub>sel</sub> + T <sub>FH</sub>	Centrocyte spontaneous unbinding	$r_{\text{division}} = 2$ /h
CC <sub>sel</sub> $\xrightarrow{r_{\text{recirculation}}}$ CB	Centrocyte recirculation to DZ	$r_{\text{recirculate}} = 3.75$ /h
CC <sub>sel</sub> $\xrightarrow{r_{\text{exit}}}$ MBC or PC with probability $p_{\text{MBC}}$ : CC $\rightarrow$ MBC, with probability $p_{\text{PC}}$ : CC $\rightarrow$ PC.	Centrocyte exit	$r_{\text{exit}} = 1.6$ /h see Section 2.7 for descriptions of $p_{\text{MBC}}$ and $p_{\text{PC}}$

To model the set of equations, we use the Gillespie algorithm [30], an algorithm that generates statistically correct trajectories of a stochastic system of equations according to simple mass-action kinetics. Briefly, at each time step, a propensity  $p_i$ , or instantaneous probability that a reaction takes place, is computed as  $p_i = \prod_{j=1}^s N_{ij} * r_i$ , where  $N_{ij}$  are the number of reactants involved in reaction  $i$  and  $r_i$  is the reaction rate of the channel. The algorithm iteratively draws two random numbers: The first one is used to define a time interval according to an exponential distribution, and the second one is used to choose the reaction that takes place in that time interval. It can be proven that the Gillespie algorithm converges to the exact probabilistic cellular distribution of all reactants. Furthermore, it is computationally cheap compared to alternative simulation methods, and has been extensively applied to model biomolecular interactions [31–33].

The standard Gillespie algorithm assumes that all particles are identical, and hence, it cannot be applied to model cellular systems where individual cellular properties need to be distinguished, such as the amount of antigen acquired by a CC, or its BCR sequence which determines its affinity and probability of capturing antigen. To overcome this limitation, a modified Gillespie [20] algorithm was developed that enables the efficient simulation of particles with individual properties by exploiting Poisson thinning, a process whereby propensities that include undesired reactions are computed and subsequently rejected. The modified algorithm was applied to simulate the GC reaction where it demonstrated speed and accuracy. Here we use the modified Gillespie algorithm to account for the following reactions: (i) CB division, (ii) CB migration, (iii) CC-T<sub>FH</sub> binding, (iv) CC competition for T<sub>FH</sub> help, and (v) CC spontaneous unbinding. A brief description of each one of the cellular interactions follows:

**GC initiation and founder cells:** NB  $\xrightarrow{r_{\text{activation}}}$  CB

A GC starts roughly 4 days after initial exposure to an antigen, after naive B cells are activated by exposure to exogenous antigen within the lymph nodes. Early interclonal competition for T cell signals selects a subset of B cells with the highest relative affinity to enter the GC reaction, contributing to the early establishment of the oligoclonality in the GC [34]. Our GC model starts empty at day 4 after immunization, after which time B cells start to enter the GC reaction at a rate of  $r_{\text{activation}}$  per hour

during the next 5 days (implying  $5 \cdot 24 \cdot r_{\text{activation}}$  seeder cells) [22,35]. The propensity of this reaction can be computed as:

$$\text{propensity} = \begin{cases} r_{\text{activation}}, & \text{between day 4 and day 9,} \\ 0, & \text{otherwise.} \end{cases}$$

In our simulation, we use a division counter [14] that takes values between 0 and 6, and is reduced by one each time a B cell divides. Once the division counter falls to zero, the cell can no longer divide and migrates to the LZ. Founding cells are initialized with the maximum division counter, i.e., 6. We assume an initial affinity for BCRs  $\simeq 0.4$  (see Supplementary Information for discussion about the choice of parameters).

**Centroblast division:**  $\text{CB} \xrightarrow{r_{\text{division}}} 2\text{CB}$

CBs proliferate up to 6 times in the DZ [14], and randomly acquire mutations in their BCR genes through SHM after each division. Similarly to seeder cells, we endow CBs with a division counter, which allows them to divide up to 6 times depending on the amount of  $T_{FH}$  help received. Cells are only allowed to migrate to the LZ once the counter has fallen to 0. The propensity of this reaction is:

$$\text{propensity} = \begin{cases} N_{\text{CB}} \cdot r_{\text{division}}, & \text{if division counter is } > 0, \\ 0, & \text{if division counter is } 0. \end{cases}$$

Each daughter CB undergoes apoptosis ( $\text{CB} \rightarrow \emptyset$ ) with probability  $\delta$  per mutation (see Section 2.2).

**Centroblast migration DZ  $\rightarrow$  LZ:**  $\text{CB} \xrightarrow{r_{\text{migration}}} \text{CC}$

Once a CB division counter drops to 0, the cell can migrate to the LZ, where its regulatory program significantly changes. The propensity of this reaction is:

$$\text{propensity} = \begin{cases} N_{\text{CB}} \cdot r_{\text{migration}}, & \text{if division counter is } 0, \\ 0, & \text{if division counter is } > 0. \end{cases}$$

Experiments suggest that class II MHC molecules are rapidly turned over on DZ B cells by ubiquitin-mediated degradation, ensuring that the peptide-MHC II density is reset after every LZ-DZ cycle [36]. Therefore, we assume that the number of antigen-MHC complexes is reset to 0 at the time of migration, which ensures fair selection against other newly activated CCs in the LZ.

**Centrocyte antigen uptake from FDC:**  $\text{CC} + \text{FDC} \xrightarrow{r_{\text{FDC:CC}}} \text{CC(pMHC)} + \text{FDC}$

Centrocytes diffuse in the LZ until they encounter an FDC. Since the encounter rate is inversely proportional to the LZ volume,  $r_{\text{FDC-CC}} \propto 1/\text{vol}_{\text{LZ}}$ , i.e., it is more likely for a CC to encounter a given FDC when the GC volume is small. As  $\text{vol}_{\text{LZ}}$  is proportional to the number of CCs,  $N_{\text{CC}}$ , the propensity can be written as:

$$\text{propensity} = r_{\text{FDC:CC}} \cdot N_{\text{FDC}} \cdot N_{\text{CC}} \propto N_{\text{FDC}}.$$

We assume that antigen uptake occurs when a CC encounters an FDC, upon which, the CC acquires antigens in an affinity-dependent manner and re-exposes on its surface as a pMHC complex. Mature B cells have been reported to carry up to  $10^5$  BCRs on their cell surface [37], and FDCs present a very large amount of antigen [38], hence we assume that there is no competition for antigen and a B cell can acquire as much antigen as the affinity of its BCR allows. Therefore, upon contact with an FDC, the pMHC complex of the centrocyte is updated with  $\text{pMHC} = \text{affinity}_{\text{CC}}$ . In the event of the CC having previously acquired antigens from another FDC, the pMHC complex is replaced. The interaction of GC B cells with FDCs is of the order of a second [6,39], and hence, we do not explicitly model the binding time between B cells and FDCs. Alternative models that take into account spatial motion have included a refractory time during which a B cell cannot attempt antigen binding

again [22]. This interval is necessary to give time for the CC to diffuse away from the FDC and prevent repetitive binding. However this is not necessary in our model, as the Gillespie algorithm models cellular distributions, and the probability of the same CC–FDC pair binding more than once is very small,  $\sim (r_{\text{FDC:CC}} / \sum_i \text{propensity}(\text{reaction}_i))^2$ .

**Centrocyte– $T_{FH}$  binding:**  $\text{CC} + T_{FH} \xrightarrow{r_{\text{TC:CC}}} [\text{CCTC}]$

After acquiring antigen, CCs diffuse into the LZ until they encounter a free  $T_{FH}$ . Upon encounter, CCs attempt to bind their pMHCs ligands to the T cell receptors expressed on the surface of the  $T_{FH}$ s. The propensity is derived in the same manner as for the encounters with FDCs:

$$\text{propensity} = \begin{cases} r_{\text{TC:CC}} \cdot (N_{T_{FH}})_{\text{unbound}} \cdot N_{\text{CC}}, & \text{if pMHC} > 0, \\ 0, & \text{if pMHC} = 0. \end{cases}$$

Regarding the number of unbound  $T_{FH}$ s,  $T_{FH}$ s enter and leave the GC during the whole reaction [40]. Therefore, our model assumes that a homeostatic flow of  $T_{FH}$ s has been established that leads to an approximately constant density of  $T_{FH}$ s during the whole reaction [18] (see Section 2.4 for a more detailed discussion). Since the volume of the LZ is proportional to the number of CCs, we can write:

$$(N_{T_{FH}})_{\text{unbound}} + N_{[\text{CCTC}]} = (N_{T_{FH}})_{\text{total}} = \alpha_{\text{TC}} \cdot N_{\text{CC}}$$

To implement this constraint, we dynamically adjust the number of  $T_{FH}$ s at every iteration of the Gillespie algorithm such that the above equation is always verified.

**Centrocyte competition for  $T_{FH}$  help:**  $[\text{CC}_1\text{TC}] + \text{CC}_2 \xrightarrow{r_{\text{TC:CC}}} \text{CC}_1 + [\text{CC}_2\text{TC}]$

Two-photon microscopy imaging reveals that the majority of  $T_{FH}$ s in GCs are not engaged in stable interactions with B cells, but instead, most contacts between GC B cells and  $T_{FH}$ s are of short duration (<5 min) [6]. In our model, we replicate this phenomenon by introducing a competition between the CCs to bind to  $T_{FH}$ s. We assume that, if during the time a  $\text{CC}_1$  is bound to a  $T_{FH}$ s a second  $\text{CC}_2$  attempts to bind to the same  $T_{FH}$ , the CC with the larger amount of pMHC will remain bound and the other CC will be rejected [20]. The propensity can be written as:

$$\text{propensity} = \begin{cases} r_{\text{TC:CC}} \cdot N_{[\text{CCTC}]} \cdot N_{\text{CC}}, & \text{if pMHC}(\text{CC}_2) > \text{pMHC}(\text{CC}_1), \\ 0, & \text{otherwise.} \end{cases}$$

As a result of this competition, a CC can interact with a  $T_{FH}$  multiple times for shorter amounts of time in our model. We assume that all these signals are cumulative and contribute towards fate decision [23,41].

Furthermore, a CC substitution reaction happens in the timescale of a diffusion process, which is typically much faster than the timescale of the GC reaction. Hence, we neglect the time associated with the unbinding of  $\text{CC}_1$  and binding of  $\text{CC}_2$ , and assume that this reaction has the same rate as a standard CC– $T_{FH}$  binding.

**Centrocyte– $T_{FH}$  spontaneous unbinding:**  $[\text{CCTC}] \xrightarrow{r_{\text{unbinding}}} \text{CC}_{\text{sel}} + T_{FH}$

A CC bound to a  $T_{FH}$  can unbind spontaneously after having received a sufficient amount of  $T_{FH}$  help:

$$\text{propensity} = \begin{cases} N_{[\text{CCTC}]} \cdot r_{\text{unbinding}}, & \text{if it was in contact with } T_{FH} \text{ for more than 30 min,} \\ 0, & \text{otherwise.} \end{cases}$$

Once a CC and  $T_{FH}$  unbind after the CC has received enough  $T_{FH}$  signal, the  $T_{FH}$  cell is released and the CC switches to the state selected.

**Centrocyte apoptosis:**  $\text{CC} \xrightarrow{r_{\text{apoptosis}}} \emptyset$

All CCs are subjected to a timed program regardless of their BCR affinity, and undergo apoptosis with rate  $r_{\text{apoptosis}}$  if not rescued by  $T_{FH}$  signals:

$$\text{propensity} = r_{\text{apoptosis}} \cdot N_{\text{CC}}.$$

**Centrocyte recirculation:**  $\text{CC}_{\text{sel}} \xrightarrow{r_{\text{recirculate}}} \text{CB}$

Selected CCs can recirculate into the DZ for additional rounds of division and affinity maturation. Their division counter is reset to 1–6 depending on the amount of  $T_{FH}$  signals received (Section 2.6).

$$\text{propensity} = N_{\text{CC}_{\text{sel}}} \cdot r_{\text{recirculate}}.$$

**Centrocyte exit:**  $\text{CC}_{\text{sel}} \xrightarrow{r_{\text{exit}}} \text{MBC or PC}$

Similarly,  $\text{CC}_{\text{sel}}$  can terminally differentiate as a PC or MBC with the following propensity:

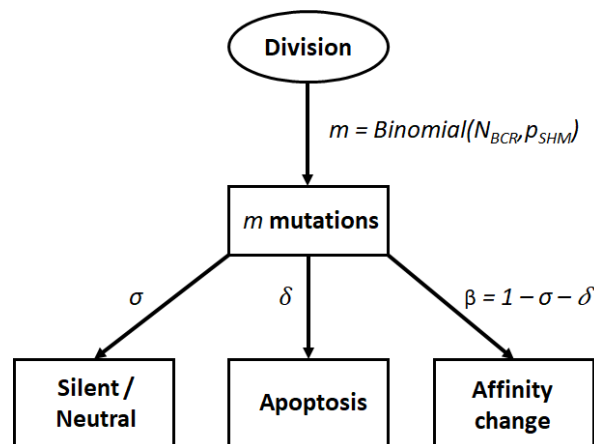
$$\text{propensity} = N_{\text{CC}_{\text{sel}}} \cdot r_{\text{exit}}.$$

We assume that once a CC has been selected to differentiate, it becomes a PM or a MBC in a pMHC-dependent manner. In Section 2.7, we computationally investigate three alternative scenarios to determine fate choice.

### 2.2. Cell Fate after Somatic Hypermutation

Most of the variability of an antibody is concentrated into three hypervariable loops called complementarity determining regions, CDR1, CDR2, and CDR3, which are responsible for the extensive range of antigens with which antibodies can interact [42]. A mutation in the CDR typically modifies the antigen binding properties of the BCR [43], while a mutation in the surrounding framework regions (FWR) might compromise its structural support and potentially lead to apoptosis. Not surprisingly, CDR chains show higher variability than FWR chains [44], although this might be due to posterior selection [6].

We model the effect of a mutation using a decision tree [24]. In this representation, a nucleotide change can either: (1) be neutral or silent, i.e., the mutation has no impact on affinity, with probability  $\sigma$ ; (2) result in a nonproductive BCR that leads to apoptosis with probability  $\delta$ ; or (3) change the affinity of the BCR with probability  $\beta = 1 - \sigma - \delta$  (Figure 2). In our simulation, we assume that a mutation can happen in both the FWR and CDR regions, although only mutations in the CDR region can lead to a change in affinity. To model the effect of a mutation, we draw two random numbers. The first number determines the number of single nucleotide changes in the whole BCR sequence according to a binomial distribution with parameters  $N_{\text{BCR}} = 660$  [45] and  $p_{\text{SHM}} = 1 \times 10^{-3}$  [46]. The second random number determines the impact of each mutation according to Figure 2.



**Figure 2.** Decision tree to model cell fate after somatic hypermutation (SHM). The number of mutations in the BCR after each centroblast division is sampled from a binomial distribution. Each mutation can produce a BCR of similar affinity, lead to cell apoptosis or produce a BCR of different affinity with probability  $\sigma$ ,  $\delta$ , and  $\beta$  respectively.

### 2.3. B Cell Receptor Affinity to Epitope

To model changes in affinity, we only consider the sequence  $\mathcal{S}_{\text{CDR}}$  of the CDR region of each BCR. Each sequence  $\mathcal{S}_{\text{CDR}}$  is composed of  $N_{\text{CDR}}$  sites that can take 4 possible values  $\in [A, T, G, C]$ . To simplify sequence comparison, we further assume that all the BCRs sequences are of the same length ( $N_{\text{CDR}} = 25$ ). We define the affinity of a sequence by using the normalized Hamming distance between the sequence  $\mathcal{S}_{\text{CDR}}$  and the optimal CDR sequence  $\mathcal{S}_0$  to bind the target antigen, as follows:

$$a_{\text{CDR}} = 1 - \frac{\|\mathcal{S}_{\text{CDR}} - \mathcal{S}_0\|_{\text{H}}}{N_{\text{CDR}}}. \quad (1)$$

While this representation does not consider the structural properties of the BCR, it has the advantage of modelling the effect of individual nucleotide changes and being computationally very cheap to compute, which enables the running of exhaustive simulations of the GC with a realistic number of cells. Furthermore, Equation (1) is bounded by  $0 \leq a_{\text{CDR}} \leq 1$ , and hence,  $a_{\text{CDR}}$  is endowed with the general properties expected of any physiologically relevant affinity function. For instance, it becomes probabilistically less likely to increase the affinity of a sequence that is close to the upper bound. Conversely, if a sequence exhibits low affinity, there is a wide range of mutations that can significantly increase the affinity. Our definition of affinity has strong similarities with previously postulated affinity measures [47], with the important difference being that we use the Hamming distance instead of the Euclidean distance, which is more adequate to work with strings of characters (nucleotides). It also agrees with other postulated properties of the affinity function such as: homogeneity, i.e., the function applies equally well to all regions of the shape space; complementarity, i.e., the optimal sequence can also be represented as a sequence, where shortest distance to the antigen sequence corresponds to greater affinity; and smoothness, i.e., small changes in the distance correspond to small changes in the affinity [48].

### 2.4. $T_{\text{FH}}$ Dynamics

Two-photon microscopy experiments have shown that  $T_{\text{FHs}}$  exist in a state of dynamic exchange between different GCs in a lymph node, and newly activated  $T_{\text{FHs}}$  can invade pre-established GCs and contribute to B cell selection and plasmablast differentiation [40]. Furthermore,  $T_{\text{FHs}}$  also proliferate [18], which results in cell count variability across time points. However, the ratio of GC B cells to  $T_{\text{FHs}}$  in the LZ appears to be maintained at a relatively constant ratio of  $\sim 7:1$  through the entire GC reaction [18]. We also note that a constant ratio of CC: $T_{\text{FHs}}$  results in an approximately constant probability of a CC encountering a  $T_{\text{FH}}$  through the whole GC response.

In a typical GC,  $T_{\text{FHs}}$  are activated by a variety of different antigens, and hence, CCs can only successfully bind to a subset of the available  $T_{\text{FHs}}$ . Since our model assumes that only one type of antigen is presented by the FDCs and acquired by CCs, we decrease the experimentally determined CC: $T_{\text{FHs}}$  ratio to a lower number, which we optimise using time resolved measurements on GC kinetics data (Section 2.10). Our estimation is that the optimal effective ratio is  $\approx 1/46$ , suggesting that only 15% of measured  $T_{\text{FHs}}$  are specialized to a specific antigen.

### 2.5. $T_{\text{FH}}$ Survival Signals

The signaling interaction between a CC and a  $T_{\text{FH}}$  is non-linear: ICOSL on B cells promotes upregulation of CD40L on T cells, which in turn further upregulates ICOSL on the B cell, creating a positive feedback loop that promotes a mode of intimate contact in which a large area of the B and T cell surfaces are juxtaposed [49]. Thus, small differences in pMHC density between B cells of different affinities can be nonlinearly amplified into large differences in signaling that allow for a more efficient selection of high affinity B cells. To reproduce this behavior, we describe the strength of the binding signal using a family of monotonically increasing functions with parameter  $n$ , such as:

$$\text{signal strength (pMHC)} = \exp(\text{pMHC}^n) - 1,$$



where the parameter  $n$  controls the strength of the nonlinear interaction: higher values of  $n$  make it harder for cells with low or moderate affinity to acquire  $T_{FH}$  signals (see the Supplementary Information for a more detailed discussion). At every unbinding event, the cumulative signal of the CC acquired from  $T_{FH}$  is incremented with:

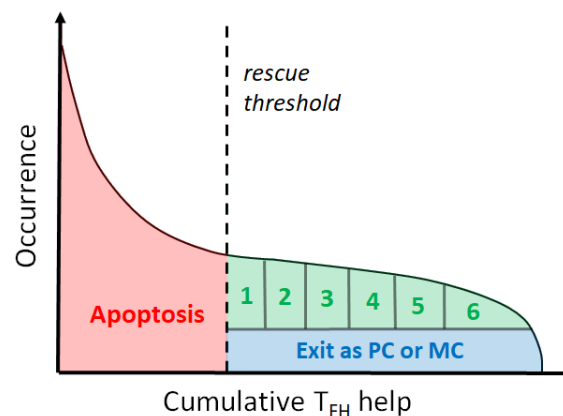
$$\Delta T_{\text{help}} = \text{time in contact with } T_{FH} \cdot \text{signal strength (pMHC)} .$$

Thus, in our model, the total amount of help received from a  $T_{FHs}$  is the cumulative signal obtained by aggregating the signals received through each encounter, however small, that a CC has with a  $T_{FH}$  in the LZ. In practice, most encounters are of short duration and typically increase the cumulative  $T_{FHs}$  help by only a small amount (70% of the bindings last for less than 5 min). This mimics the observation that a CC probes and attempts to bind to many different  $T_{FHs}$  endowed with different receptors, before receiving a substantial amount of signalling from a  $T_{FH}$  with a compatible receptor.

### 2.6. Regulation of the Cell Cycle

The process of acquiring T cell help ends when a CC unbinds spontaneously from a  $T_{FH}$  after having received signals for  $\sim 30$  min, upon which, the CC is considered selected and can either leave the GC and terminally differentiate into a PC or a MBC, or recirculate into the DZ for further rounds of division and BCR maturation. Cells that recirculate into the DZ regulate their cell cycle speed according to the amount of  $T_{FHs}$  help received, with cells that have received strong signals being able to replicate up to 6 times [10,14]. The regulation of the cell cycle is likely to be induced by MYC [11,13], a cell-cycle regulator that is upregulated by  $T_{FH}$  help in cells selected for reentry into the DZ [12]. MYC is also known to play a critical role in the regulation of division bursts in the DZ [11].

To reproduce the regulation of the cell cycle, we use a division counter [14] that takes values between 1 and 6 depending on the amount of  $T_{FH}$  help received. To implement the division counter, we first run a short simulation to precompute the distribution of cumulative help from  $T_{FH}$ . This initial distribution is kept fixed throughout the GC simulation and used to define 6 equal percentiles, such that a cell in the  $n$ th percentile undergoes  $n$  divisions (Figure 3). Such definition corresponds to an average of  $\sim 3.5$  divisions per recirculating cell in our model, very close to the experimentally determined average of 3 divisions per recirculating cell [14]. Besides increasing in number due to the higher number of cellular divisions, high-affinity cells also acquire a higher number of mutations than those that undergo fewer divisions [14], which enhances their chances of developing more effective BCRs.



**Figure 3.** Fate of B cells after interaction with  $T_{FH}$ . Cells that do not receive sufficient help from  $T_{FHs}$  undergo apoptosis. The remaining cells either leave the GC and differentiate into a PC or MBC, or cycle back to the dark zone, where they undergo up to 6 divisions depending on the amount of  $T_{FH}$  signals received.

### 2.7. Plasma Cell and Memory B Cell Differentiation

Upon exit from the GC, B cells terminally differentiate into PCs or MBCs. Temporal data has demonstrated the existence of a temporal switch, where MBCs tend to be produced during the early phases and PCs are produced at later times of the response [50]. The precise mechanism that determines whether a cell differentiates into a PC or MBC is not completely understood [1], although recent work suggests that affinity to antigen plays a central role in selecting the cells that enter the PC pool, while lower affinity selected B cells progress to MBCs [51]. In this scenario,  $T_{FH}$  help stimulates CD40 signaling and IRF4 transcription, which in turn regulates, in a dose-dependent manner, the choice between differentiating into a PC or cycling back into the DZ [52]. This model has been computationally investigated using a multi-scale probabilistic model and shown to lead to good agreement with the experimentally determined GC output [20]. Regarding the differentiation into MBCs, the determinants that predispose cells to enter the MBC pool or the markers that identify cells committed to the MBC lineage are not known. A fraction of LZ B cells with lower affinity are prone to progress to resting MBCs in a Bach2 TF-dependent manner [53,54], although high-affinity cells might also become MBCs. This suggests that stochastic effects might be at play in the selection of the MBC fate.

We use our model to investigate alternative differentiation scenarios. Similarly to previous work, we assume that IRF4 regulates cell fate in a dose-dependent manner, where high-affinity cells that express high amounts of IRF4 are directed to the PC pool, and medium affinity cells either cycle back into the DZ or differentiate as a MBC. Once a cell is positively selected, we simulate and analyse three possible differentiation scenarios:

- **Antigen-driven differentiation model (model 1):** In this first scenario, we consider that affinity determines B cell fate in a deterministic manner. Namely, there is a threshold  $pMHC_{\text{thresh}}$  above which a CC becomes a PC. Below that threshold, a cell becomes a MBC.
- **Dynamic antigen-driven differentiation model (model 2):** A more sophisticated hypothesis is that T cells progressively become less sensitive to the pMHC complex and, as the average affinity of B cells increases in the GC, a larger amount of pMHC is needed to induce the same amount of IRF4. In the second scenario, the choice between PC and MBC is also deterministic, although according to a dynamic threshold that increases linearly with time, i.e.,  $pMHC_{\text{thresh}} = \text{affinity}_{GC}(t - t_0) + a_0$ .
- **Stochastic differentiation model (model 3):** Finally, we also investigate a probabilistic scenario where, once a B cell has been selected for differentiation, it becomes a PC or MBC according to some fixed probabilities that depend on the amount of expressed pMHC as follows:  $p_{MBC} = e^{-k(pMHC - p_0)}$  and  $p_{PC} = 1 - p_{MBC}$ .

In all three scenarios, apoptosis occurs irrespective of affinity. Namely, apoptosis is modelled in the Gillespie algorithm as a reaction channel with constant kinetic rate, and hence, its probability of happening only depends on the total number of CCs. We note, however, that CCs carrying high-affinity BCRs are more likely to be positively selected and differentiate or recirculate, while lower affinity CCs are more likely to stay in the GC and undergo apoptosis, which might result in an indirect association between low affinity and apoptosis [55]. Regarding recirculation, it happens stochastically according to a Gillespie channel of constant kinetic rate once a CC has been selected, and hence, recirculation after selection is an affinity-independent event. In all scenarios, the amount of received T cell signaling is only important to stop apoptosis and to decide on the number of divisions for cells that recirculate into the DZ. In that respect, all models are based on antigen-driven selection mechanisms.

### 2.8. Phylogenetic Trees Representation

We use phylogenetic trees to visualize the evolution of the CDR sequences through SHM. In such a representation, each founder cell, associated with a unique V, D and J combination, defines the unmutated germline of a new tree, and newly acquired mutations are represented as downstream nodes. Different approaches to represent clonal trees are possible. Here, we chose to associate each

leaf with a sequenced cell and each internal node with a unique mutation of the BCR. Using our simulated data, it is trivial to reconstruct lineage trees, as we can follow each founder cell and its progeny through the whole GC reaction. Then, the topological properties of each phylogenetic tree, such as number of nodes, average depth of the leaves, trunk length, etc provide important information about the affinity maturation [29,56], and can be used to compare simulated trees to trees inferred from experimental data.

To construct trees from experimental data, i.e., repertoire sequencing data, we first align the variable (V), diversity (D) and joining (J) gene segments of each sequence using the IMGT [57] (for V,D,J) and VBASE2 (for V) [58] databases with BLAST [59]. In the case of discrepancy between V segment assignments from the two databases, the assignment leading to the lowest number of somatic mutations is chosen. As cells have at most 20 mutations, all assignments were unambiguous with a V(D)J alignment score over 95%. Next, functional V(D)J sequences are grouped together into clones if they shared the same V, J and D gene segments. The grouped sequences from each clone are then aligned with ClustalW [60]. This second alignment is necessary to infer a mutation matrix, as tree inference algorithms typically require sequences to be aligned and of the same length, while experimentally determined sequences have different lengths, and include insertions and deletions. The corresponding trees are computed with SCITE [61], a stochastic search algorithm based on Markov chain Monte Carlo sampling. The root node is defined by taking the unmutated V(D)J germline and filling the remaining junction region with the consensus sequence of all available sequences. Finally, the obtained trees are analyzed with the BioPhylo Python library [62] to compute the main topological tree properties, such as number of leaves, internal nodes, branches, trunk, etc.

### 2.9. Clonal Diversity

Clonal diversity has been measured using a multi-color fate mapping technology that permanently tags individual B cells and their progeny with different combinations of fluorescent proteins. The method can generate up to 10 different color combinations, thus limiting the experiment to characterising the evolution of a maximum number of 10 clones. A Normalized Dominance Score (NDS), defined as the fraction of all B cells in a GC that carry the dominant color combination, was used to characterize clonal diversity [15,63]. To simulate this experiment, we cluster founders into 10 groups, and define an NDS score as the ratio between the cell count of the dominant cluster over the total number of cells. The NDS is used only for a direct comparison with measurements, but not for other characterisations of the clonal diversity. Later in the paper, we use the clonal dominance (CD), which is simply the ratio between the cell count of a clone cluster and the total number of cells.

### 2.10. Parameter Optimization

Our models include 17 free parameters that need to be optimised. As the amount of quantitative information about the GC dynamics is still limited, we exploit the recent literature to obtain reasonable estimates about the value of some parameters as well as variability bounds (*Lower bound* and *Upper bound* columns in Table 2). Further details can be found in the Supplementary Information.

In addition, many recent works have experimentally characterized the GC, and we use these data to constrain different properties of the GC response, such as GC kinetics, output cell counts, etc. We further use data from mice experiments when human experiments are not available. We briefly discuss here the constraint bounds on each GC property.

- **Number of GC B cells:** A time-resolved study of 3457 GC sections from mice provided information about the GC area  $A_{GC}$  throughout the GC reaction [18]. Assuming that GC are spherical, then  $A_{GC} = \pi r_{GC}^2$ , from where the radius of the GC,  $r_{GC}$ , can be estimated. If we further assume that B cells are the predominant cellular type in the GC, we can estimate the number of B cells in the GC as:

$$N_{\text{B cell}} \approx \frac{V_{\text{GC}}}{V_{\text{B cell}}} = \left( \frac{r_{\text{GC}}}{r_{\text{B cell}}} \right)^3,$$

where we have considered cells to be spherical. Taking  $r_{\text{B cell}} = 6.2 \mu\text{m}$  [64] and  $r_{\text{GC}} = 80 \mu\text{m}$  (day 9) [18], we get an average of  $N_{\text{B cell}} \approx 2000$  at the GC peak. As most histological sections do not go cut GCs through the centre, estimations of individual GC volumes are only approximated. Nevertheless, it has been shown that cross-sectional area distributions of spleen sections reflect accurately the broad real-size distributions of GCs [65]. Moreover, as there is significant variability in GC B cell sizes, notably during the cell division cycle [66,67]), the previous number has to be considered only an order of magnitude approximation to the total number of cells. Interestingly, the same study found that GC undergo a fast expansion phase from day 4 to 9, and then progressively decrease in volume by a factor of 2, roughly 10 days after the initial expansion [18].

- **DZ/LZ ratio:** Despite high variability in size, a typical cellular composition with relatively stable cell ratios of resident  $T_{\text{FHS}}$ , macrophages, proliferating cells, and apoptotic nuclei seems to be maintained during the established phase of the response [18]. Time-resolved data about the relative size of DZ to LZ is also available from the same study. While the variability in the relative volumes of DZ and LZ is considerably high at all sampled time points (Figure 4B), it stabilizes after the phase of monoclonal expansion to an average of  $\sim 1$  during the entire GC reaction [18]. The high variability is consistent with some other studies that found a ratio of 2 [9]. In our simulation, the DZ/LZ ratio is defined as the number of CBs over the number of CCs; however, we only estimate this ratio after day 9 after immunization, hence avoiding the early days of the GC establishment, where the DZ and LZ are still not spatially separated.
- **Cell death:** Apoptosis is prevalent in the GC, with up to half of all GC B cells dying every 6 h [55]. We use a relatively constant cell death rate of 8%/h in both the DZ and LZ through the entire GC response.
- **SHM mutations:** Time-resolved data on sequenced BCR mutations, obtained by comparing sequences to the closest known V,D,J germline sequences, suggest that B cell clones undergo roughly 5 mutations every 10 days, and [6,44]. We use these data to further constrain our model.
- **GC cellular output:** Finally, we use time-resolve data of the GC cellular output [50] to constrain our model. Because the experiment only measured relative abundances rather than direct cell counts, we scale both the model predictions and experimental measures by the maximum value.

Next, we compute an error for each set of parameters, defined as the sum of the root mean square deviation (RMSD) between the model prediction and experimental results. For a fair measurement of the fitting performances across heterogeneous GCs, the RMSD is normalized by dividing it by the standard deviation of the experimental data averaged over the different time-points—we also average over same time-point replicates, when available, to reduce variability. As no measurement of the cell death standard deviation is available, we arbitrarily assume a standard deviation of 1% cell death/h (measured rate is 8% cell death/h). For each one of the 4 properties for which data is available, i.e., cell count, DZ/LZ ratio, cell death and number of mutations, we compute an independent NRMSD. A NRMSD lower than one indicates that the simulation error is lower than the uncertainties intrinsic to experiments.

The score function is minimized with the maxLIP0 algorithm [68], which is both parameter free and provably better than a random search. It is a good alternative to Bayesian optimization methods [69], that typically require the definition of prior assumptions about the function being optimized and thus require domain knowledge. The constraints specified for each parameter, as well as the obtained optimal value, are reported in Table 2. Due to the high computational cost required for parameter optimization with high dimensionality, a simplified ODE system was also used to speed up the maxLIP0 search (described in the Supplementary Information).

**Table 2.** Parameter bounds and optimized values used in our model. The *Lower bound* and *Upper bound* are derived from the literature; the fitted values are obtained with the maxLIP0 optimization algorithm. The *Sensitivity* column indicates how sensitive the score function is to a  $\pm 10\%$  change in the parameter value. The labels low, moderate, high correspond respectively to a  $<10\%$ ,  $<100\%$  and  $<1000\%$  change of the score function. Further details can be found in the Supplementary Information.

	Parameter	Description	Lower Bound	Upper Bound	Fitted	Sensitivity	Mainly Affects
Intercellular	$r_{\text{activation}}$ [ $\text{h}^{-1}$ ]	B cell activation rate	1	10	3.94	high	Value of GC peak
	$r_{\text{division}}$ [ $\text{h}^{-1}$ ]	CB division rate	0.08	0.16	0.134	moderate	DZ/LZ ratio
	$r_{\text{migration}}$ [ $\text{h}^{-1}$ ]	CB migration rate	0.15	3.75	3.75	low	DZ/LZ ratio
	$r_{\text{apoptosis}}$ [ $\text{h}^{-1}$ ]	CC apoptosis rate	0.06	0.16	0.084	high	GC decay
	$r_{\text{exit}}$ [ $\text{h}^{-1}$ ]	CC <sub>sel</sub> exit rate	0.41	3.75	1.6	high	GC decay
	$r_{\text{recirculate}}$ [ $\text{h}^{-1}$ ]	CC <sub>sel</sub> recirculation rate		fixed to 3.75		high	GC decay
	$r_{\text{FDC:CC}}$ [ $\text{h}^{-1}$ ]	FDC encounter rate		fixed to $40/N_{\text{CC}}$		low	Clonal competition
	$r_{\text{TC:CC}}$ [ $\text{h}^{-1}$ ]	T <sub>FH</sub> encounter rate		fixed to $145/N_{\text{CC}}$		moderate	Clonal competition
	$r_{\text{unbinding}}$ [ $\text{h}^{-1}$ ]	T <sub>FH</sub> unbinding rate		fixed to 2		high	GC decay
	$\alpha_{\text{TC}} = N_{\text{TC}}/N_{\text{CC}}$	T <sub>FH</sub> to CC ratio	1/100	1/7	1/46	high	GC decay
	$N_{\text{FDC}}$	Number of FDCs		fixed to 250		low	Clonal competition
	Intracellular	$\delta$	Lethal mutation probability	0.1	0.9	0.52	high
$\sigma$		Silent mutation probability		fixed to 0.28		low	Affinity maturation
$p_{\text{SHM}}$		Mutation rate		fixed to $1 \times 10^{-3}$		high	GC decay
$N_{\text{CDR}}$		CDR length (nucleotides)		fixed to 25		low	Affinity maturation
PC/MBC	$p_{\text{MHC}}_{\text{threshold}}$	Model 1 parameter	0	1	0.46	moderate	PC/MBC production
	$t_0$	Model 2 parameter	0	$\infty$	3 days	moderate	PC/MBC production
	$a_0$	Model 2 parameter	0	1	0.05	moderate	PC/MBC production
	$k$	Model 3 parameter	0	$\infty$	14	moderate	PC/MBC production
	$p_0$	Model 3 parameter	0	1	0.4	moderate	PC/MBC production

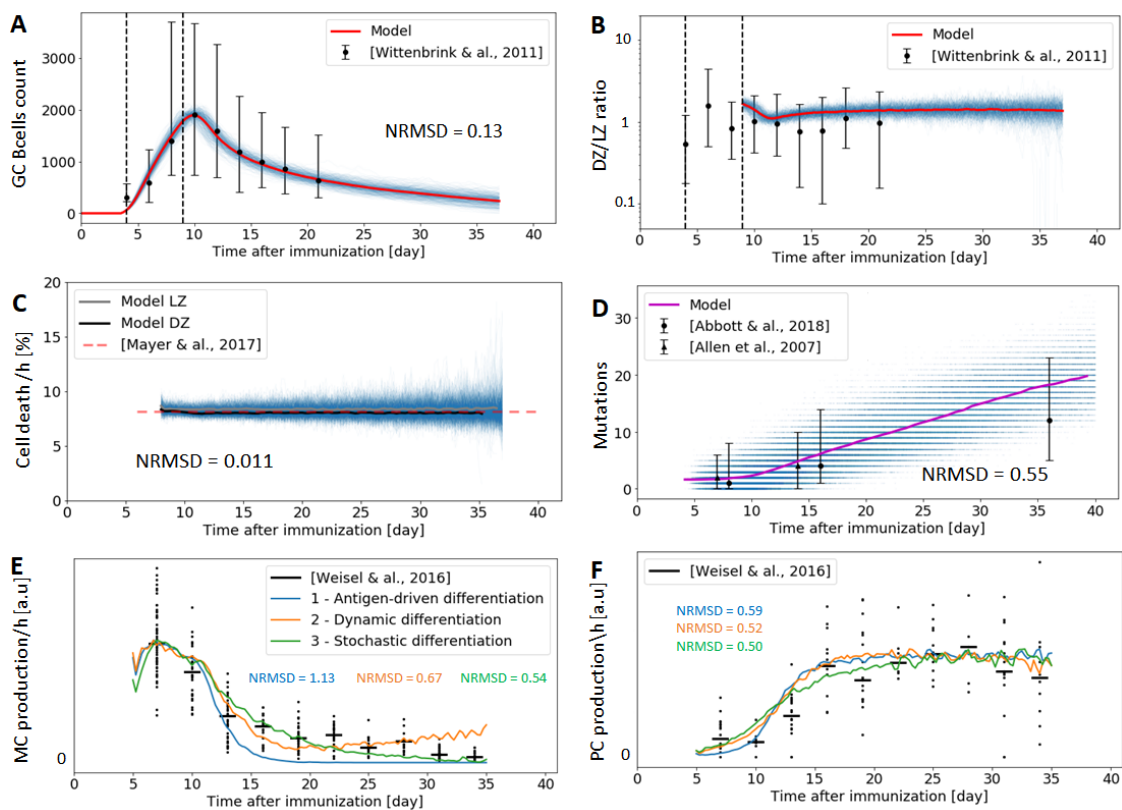
### 3. Results

#### 3.1. Germinal Center Dynamics

We simulate our GC model using a modified Gillespie algorithm [20] and the parameters optimised as discussed in Section 2.10. Namely, 1000 simulations were run and the normalized root mean square deviation (NRMSE) of each optimised GC property was computed. Results are presented in Figure 4.

Our model is in good agreement with the experimental data, having a fairly low ( $\lesssim 1$ ) NRMSD for all optimised GC properties. For instance, the NRMSD of the GC B cell count is 0.13, indicating a smaller deviation than the standard deviation of the experimentally determined number of GC output cells. A reason for this disparity is the high heterogeneity of GCs in terms of size and cell count [18,65], with measured GCs ranging from  $\sim 500$  to  $\sim 10,000$  B cells in size. As our model is trained to replicate the mean experimental values, it is expected to show lower standard deviation. Despite this difference,

our model recapitulates well the general trends of experimental measurements, e.g., the reduction of GC volume by half in  $\sim 5$  days [65], and the shutdown of the GC at around day 40.



**Figure 4.** Experimental data vs. our model after 1000 simulations, along with the obtained normalized root mean square deviation (NRMSE) of each measured GC property. Thin blue lines represent one single Gillespie simulation, while thick lines are the average over all simulations. Experimental results are represented by a tick bar corresponding to the median and 1st-3rd quartiles, the central point being the median. Figures (A,B) depict simulations and experimental data about cell count and DZ/LZ volumetric ratio respectively, the latter derived from [18]. Figure (C) shows the death rate percentages of B cells in both the DZ and LZ (8%/h) [55]. Figure (D) shows the mutation rates at different time points, and roughly corresponds to 5 mutations every 10 days [6,44]. Figures (E,F) illustrate the temporal switch from memory B cell to plasma cell production simulated with our model using 3 different differentiation scenarios and experimentally measured [50].

Regarding PC/MBC differentiation, the time-dependent GC output of the three models investigated in Section 2.7 reproduces the temporal switch experimentally observed (Figure 4E,F), where memory cell models are produced during the first two weeks of the GC reaction, and plasma cells are produced afterwards [50]. The model based on a fixed threshold has the lowest agreement with experiments, as this model does not lead to a substantial production of MBCs after 2 weeks, while MBCs are produced during the whole GC reaction, albeit at a moderate rate at late time-points. On the other hand, both the dynamic threshold and stochastic differentiation model are compatible with experimental data (average NRMSD = 0.52 and 0.59), with the stochastic model fitting the data slightly better.

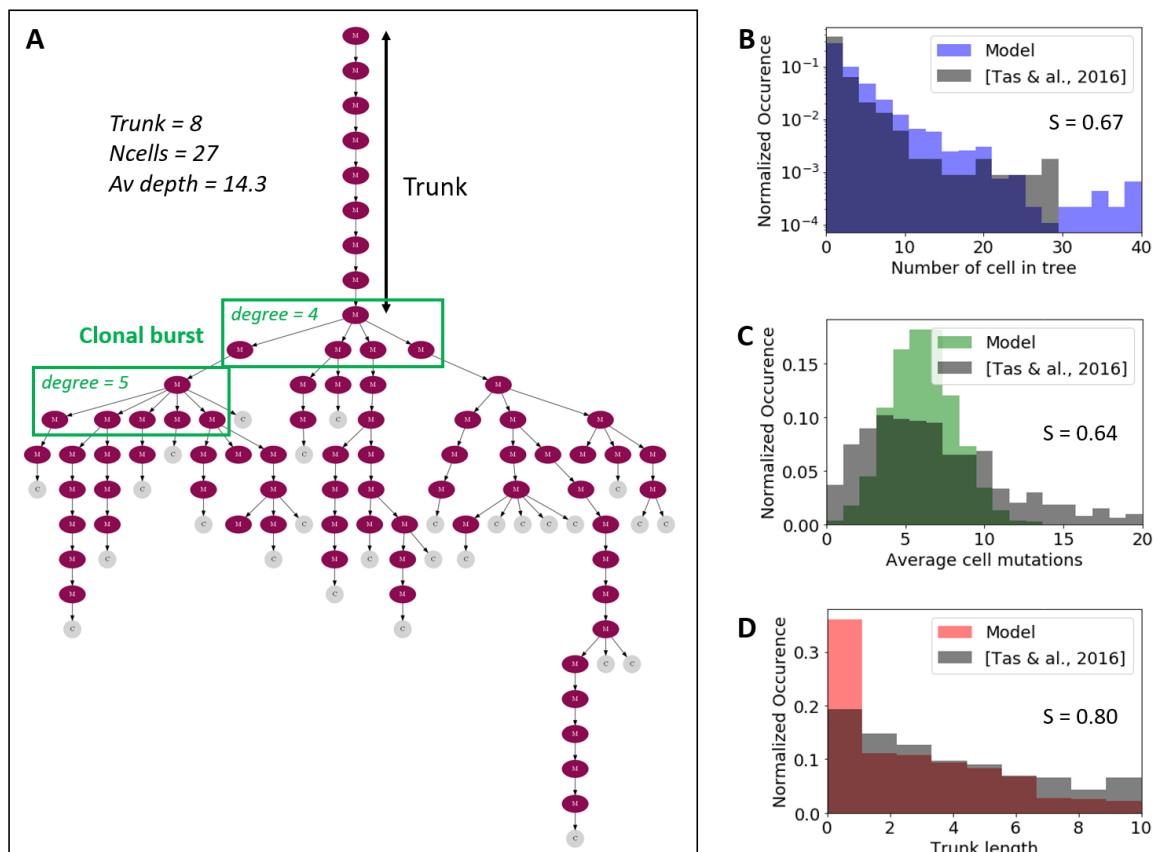
To gain insight into which parameters have a higher impact on the GC dynamics, we perform a stability analysis (Supplementary Information). The analysis reveals that the GC kinetics is most sensitive to changes in the parameters that regulate the selection process by  $T_{FHs}$ , such as the amount of  $T_{FHs}$  signals necessary for a LZ B cell to recirculate into the DZ, as well as the amount of cellular divisions it undergoes once there. Similarly, the probability  $\delta$  of dying in the DZ after a mutation leads to a nonproductive receptor critically influences the number of cells in the GC.

We also perform a sensitivity analysis to quantify the effect of changes in the parameters in the simulation output. Due to the complexity of the model, we only consider individual changes in parameters, namely, we explore changes of  $\pm 10\%$  of each parameter, while keeping the rest of the parameters unmodified [24]. The changes in the model output are quantified by measuring the change in  $\text{NRMSE}_{\text{total}}$  over all quantified GC properties (Table 2). Some parameter changes mainly affect the GC kinetics, e.g., the GC B cell counts and the DZ/LZ ratio, while others have an impact on the affinity maturation process, e.g., clonal diversity and number of acquired mutations. As an example, the lethal mutation probability  $\delta$  has a strong impact on the GC kinetics, as a 10% decrease leads to a 660% change in the total NRMSD. On the other hand, the silent mutation probability  $\sigma$  has a low impact, as a 10% decrease only leads to a change in the total NRMSD of  $<10\%$  (Table S1 in the Supplementary Information).

### 3.2. Comparison of Simulated and Experimentally Determined BCR Sequences

A direct comparison of sequences between sequenced GC B cell repertoires and sequences inferred with our model is not possible due to our abstract representation of the variable region of the BCR. As an alternative, we use phylogenetic trees to visualize the somatic evolution of B cells during the GC response. To construct experimental trees, we use single-cell BCR sequencing data obtained from individual GCs [15]. Specifically, the sequences of the Ig heavy chain of  $\sim 1500$  single B cells from 18 distinct GCs—with an average of roughly 80 cells per GC—were used to build 266 trees containing at least 2 cells. All measurements were performed on mice at day 15 after immunization with different antigens. By sampling midway through the GC reaction, intermediate BCRs that might have populated the GC at earlier time-points are lost, and the inference sometimes shows long trunks, representing the pruned evolutionary process of the sampled cells, in agreement with earlier observations [29]. Co-existing clones are only seen close to the tree leaves (Figure 5A), where selection is currently taking place. Regarding the inference from simulated data, trees can be straightforwardly reconstructed, as the whole lineage is available in a simulation. To reproduce the experimental sampling conditions with the simulated data, we uniformly sampled 80 cells at day 15 after immunization from our model.

The experimental and simulated trees were analyzed with the BioPhylo python library [62] to compute important topological properties, such as number of leaves, average leaf depth, and trunk length. The distributions of these properties in both the simulated and the experimental trees are shown in Figure 5B–D. Simulated distributions are obtained after averaging 1000 runs in our model. Overall, the distributions obtained from simulated and experimental data are in good agreement ( $S > 0.6$ , Figure 5). However, experiments suggest that the density function of cell mutations per tree is broader than what our model predicts. This could be explained by the fact that all mutations have equal contributions in our model, where one mutation can only increment the affinity by  $\frac{1}{N_{\text{CDR}}}$ , with  $N_{\text{CDR}} = 25$ . Such limitation prevents cells with a low number of mutations from having a high-affinity BCR, thus constraining the simulated density function to be narrower. Another contributing factor could be the fact that our simulation only considers the CDR region, and does not keep track of silent mutations that might occur in the FWR region.



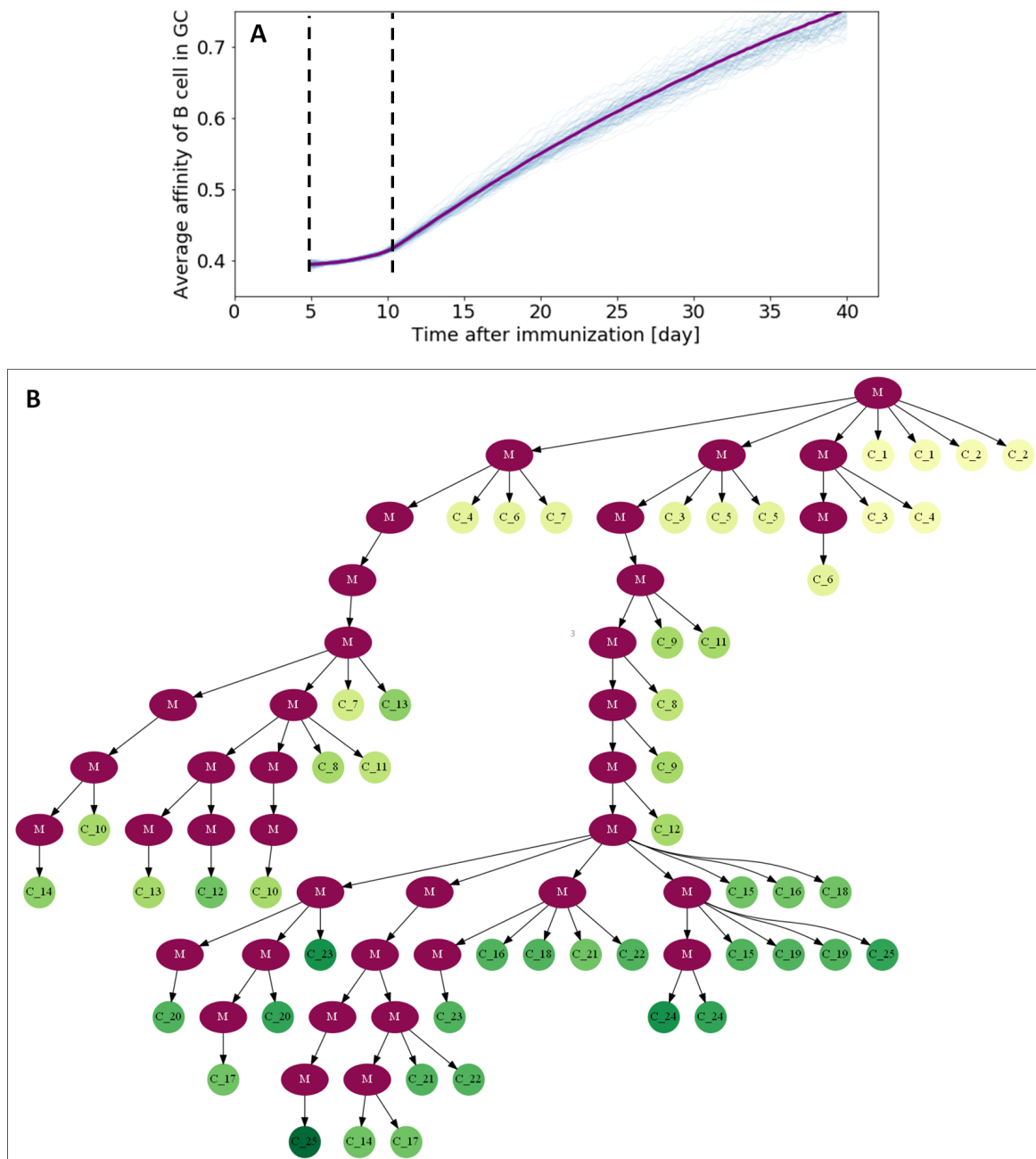
**Figure 5.** Phylogenetic trees from BCR sequences at day 15 after immunization. (A) Representative phylogenetic tree inferred from single-cell BCR sequencing extracted from individual GCs [15]. The purple nodes (M) represent mutation and the gray node (C) represent individual cells. (B–D) distribution of tree topological properties (number of leaves, average leaf depth and trunk length) in experimental and simulated trees. The score S is defined as Overlapping Area/Joint Area of the histograms.

### 3.3. Visualizing Affinity Maturation

Figure 6A depicts the evolution of the average affinity in the GC through its lifespan. Noticeably, affinity does not substantially improve during the first days of the GC reaction, corresponding to the phase of monoclonal expansion where B cells divide without selection pressure. After day 10, affinity increases linearly throughout most of the lifespan of the GC, only showing slower improvement rates near the end of the GC reaction, where the remaining B cells have presumably developed good receptors against the exposed antigens, and new mutations are likely to be detrimental to the overall affinity.

For visualization purposes, we randomly sampled two cells every day from the dominant clone, where dominance was established by analyzing clonal cell counts at the end of the simulation, and computed a tree from the obtained set of BCR sequences (Figure 6B). Since we sample uniformly throughout the lifespan of the GC, we do not observe the long trunk found in experimentally inferred trees due to pruned clones. Although the specific structure of the tree shows high variability as a consequence of the strong undersampling, the tree clearly shows how affinity matures as a result of clonal evolution. Noticeable also is the clear correlation between increased affinity and number of SHMs (the number of mutations is equal to the number of nodes connecting the root cell and a leaf). Resampling the BCR set multiple times did not significantly change the properties of the tree.





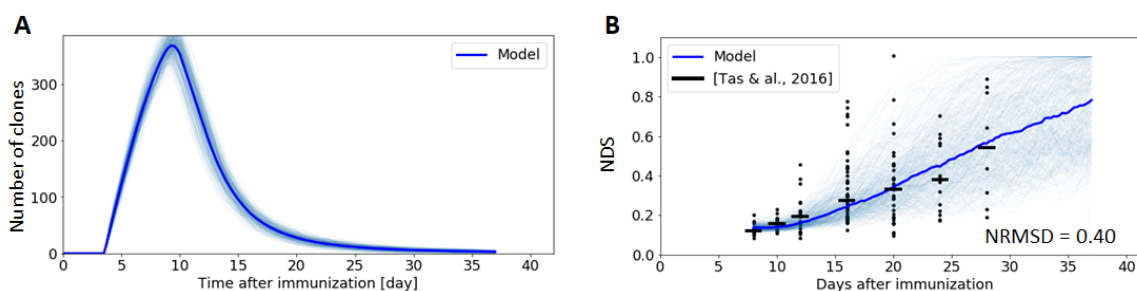
**Figure 6.** Affinity maturation in GC through its lifetime. **(A)** Mean BCR affinity as a function of time. **(B)** Tree representation of affinity maturation. Two cells were sampled every day for 25 days (50 cells in total). The purple nodes (M) represent mutation, and the color of cell nodes represents the BCR affinity (yellow = low and green = high), the cell label subscript is the day on which the cell was sampled.

### 3.4. Visualizing Clonal Competition and Clonal Burst

To better understand the complex process of clonal competition and the mechanisms behind clonal bursts observed in a subset of GCs [15], we plot the evolution of the number of clonal families in the GC as a function of time (Figure 7A), where clonal families are defined in our simulations as the cells descending from the same founder cell. Our model reveals that the number of clones increases until day 8, corresponding to the continuous entry of new founder cells in the GC. Then, shortly after the monoclonal expansion phase, the GC experiences a rapid loss of clonal diversity due to the high selection pressure exerted in the LZ on the still immature CCs, followed by a more moderate decrease that lasts until the end of the GC reaction. These results are consistent with experimental

characterizations that have revealed that clonal diversity is relatively high during the early days of the GC, followed by a phase lasting several weeks where a few clones quickly dominate the reaction [15].

While the evolution of the number of clones in the GC has not been measured quantitatively, clonal diversity has been characterized using a Normalized Dominance Score (NDS) [15] (Section 2.9). An interesting finding of this study is that, while most GCs retain substantial clonal diversity within the 3-week lifetime of the GC response, rapid and massive expansion of higher-affinity SHM variants (clonal bursts) leads to substantial loss of diversity in a subset of GCs. As illustrated in Figure 7B, our model quantitatively reproduces the measured NDS with a low NRMSD (0.40). Both the simulations and the experimental characterizations reveal a high variance in the NDS score. For instance, our model predicts that 2.5% of GCs reach clonal dominance (NDS = 1) at day 30, and 25% at day 40. This is expected if stochastic effects play a fundamental role in the emergence of dominance.

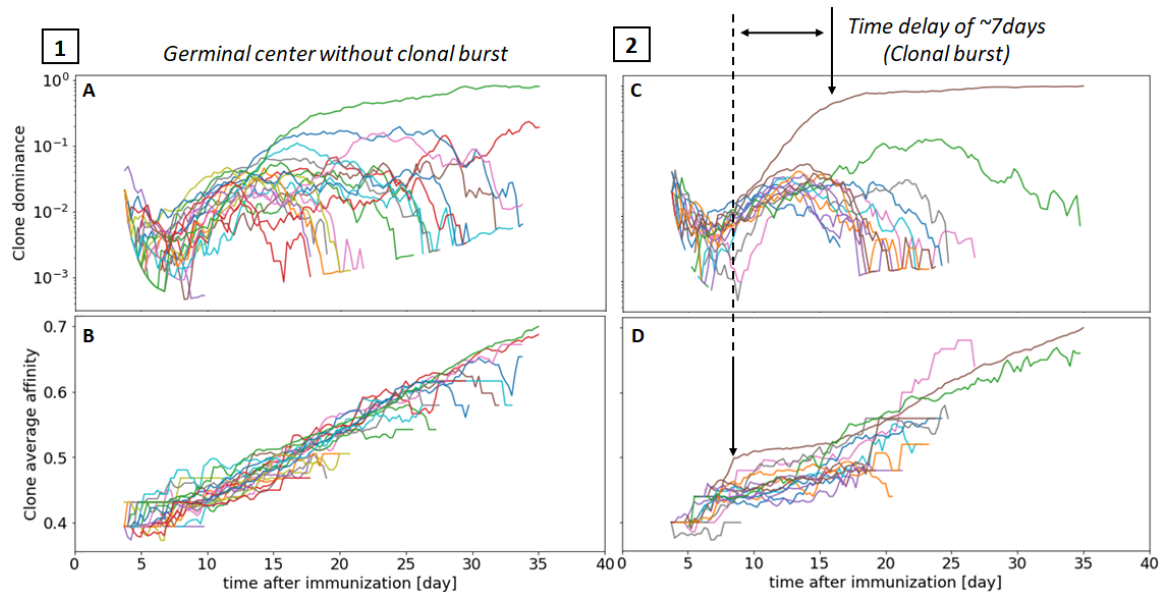


**Figure 7.** Clonal diversity evolution during the GC reaction. Thin blue lines represent one single Gillespie simulation, while thick lines are the average of 1000 simulations. Figure (A) depicts the B cell clonal evolution through the GC response, and (B) the Normalized Dominance Score (NDS), i.e., the fraction of B cells in the dominant clone, as reported by [15]. Experimental results are represented by a scatter plot, where individual points represent single measurements and the bar indicates the median.

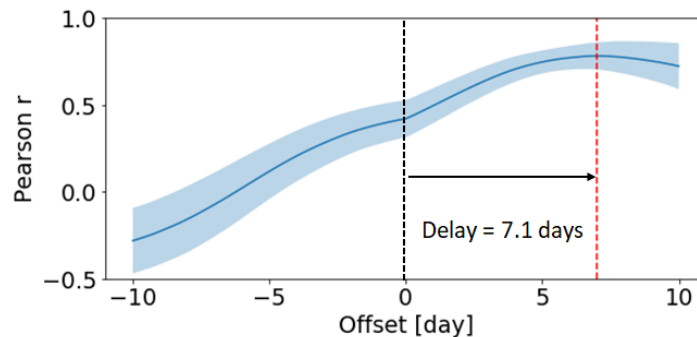
To investigate the emergence of clonal dominance in more detail, we plot the Clonal Dominance (CD), which clusters cells that descend from the same founder cell (Section 2.9), of individual families in two representative simulations in Figure 8A,C. The first simulation (1) shows a GC with no observable clonal bursts. The CD of the dominant clone (green line) only becomes dominant after day 15, but other competing clones are still significantly present until much later ( $\sim 30$  days). In the second simulation (2), the dominant clones (brown line) outcompete other clones as early as day 10, with most of the co-existing clones disappearing by day 25. It is interesting to notice that in the latter scenario, representing a GC that underwent a clonal burst, the winning clone acquired an early advantage (before day 10) in terms of affinity (Figure 8D), which enabled the clone to proliferate and gain dominance. Although other clones reached higher levels of affinity at later time points, e.g., from days 15 to 25, the early advantage that led to a higher proliferation rate was sufficient to ensure dominance throughout the whole GC reaction.

The simulation shown in Figure 8 (2) seems to suggest that there is a time delay of roughly 7 days between the initial clonal affinity advantage and the clonal burst induced by it, corresponding to multiple LZ-DZ cell cycles—in our model, a full DZ–LZ cycle lasts  $\sim 1$ – $2$  days depending on the number of divisions in the DZ. To better quantify the delay between the increase in affinity and the appearance of clonal dominance, we performed a Time Lagged Cross Correlation analysis [70] between the clonal affinity and the clonal dominance time series of each clone. Briefly, the analysis computes the Pearson correlation of one time series with the respect to the second one shifted by a time offset. By trying different values for the offset, one can find the optimal time delay that maximises the correlation. We collected time series data from 1000 clone dominance-affinity pairs by taking the 10 most abundant clones of 100 GC simulations. The results obtained, highlighted in Figure 9, confirm that the correlation is maximized for a time delay of 7.1 days, which suggests a causal relationship between a random increase in affinity and the emergent appearance of dominance one week later. A delay of 7 days corresponds to roughly 3–4 rounds of DZ–LZ B cell cycles. This delay is consistent with the

expectation that the production of high-affinity antibodies requires several rounds of cell division and affinity maturation. More specifically, with an average of three divisions per DZ–LZ cycle, our model suggests that one round of somatic hypermutation is not enough to achieve clonal dominance.



**Figure 8.** Clonal dominance and clonal affinity maturation for two representative GC simulations, without (1) and with (2) clonal burst. Figures (A,C) showcase the clonal dominance score evolution in the GC, and (B,D), the average affinity of each clonal family as a function of time. Only the most abundant 50 clones are shown to simplify the presentation.



**Figure 9.** Time Lagged Cross Correlation analysis between clonal dominance and clonal affinity time series. The Pearson correlation is averaged from the 10 most abundant clones of 100 germinal center simulations (resulting in 1000 clone dominance–affinity pairs). The shaded region corresponds to one standard deviation. The correlation peaks at 0.79 at an offset of 7.1 days.

#### 4. Discussion

We have presented a stochastic multiscale model of the GC that combines an abstract representation of individual B cell receptors (BCRs) as strings of nucleotides, a deterministic regulatory representation of the transcriptional program associated with B cell differentiation [20] and a stochastic representation of the cellular interactions that shape the GC reaction. The model parameters were optimised using time-resolved data from various experimental characterizations of the GC kinetics and GC clonal diversity [15,18,44,50,55]. Overall, our model recapitulates the GC dynamics characterized experimentally, including the GC B cell counts at different times of the GC reaction, the DZ/LZ volumetric ratio, the cell death and accumulation of mutations as a result of SHM. Compared to existing models, our model accounts for a broader range of important biological processes, e.g., centroblast apoptosis, a more realistic description of antigen capture, and faithfully replicates experimental data. A sensitivity analysis reveals

that the GC kinetics is very sensitive to small changes in some key parameters, indicating that the B cell maturation is a tightly regulated program. In particular, slight differences in the  $T_{FH}$  selection process have a high impact in the GC dynamics, and can potentially lead to a drastic increase in both the number of dividing centroblasts as well as differentiating cells. Such sensitivity could explain the high variability observed in the GC sizes [18]. Our optimisation approach enables the easy integration of new experimental data that might become available in the near future.

With the help of our model, we computationally investigated three candidate mechanisms that might potentially govern the terminal differentiation of GC B cells into PCs or MBCs. Namely, (i) a deterministic scenario, where the amount of captured antigen determines cell fate; (ii) a dynamic model, similar to the deterministic model, but where the threshold of captured antigen to become a PC increases with the lifespan of the GC; and (iii) a probabilistic scenario, where the amount of exposed antigen determines the cellular outcome in a probabilistic fashion, i.e., cells with high and low amounts of antigen have a non-zero probability of becoming either a PC or a MBC. Although the three models perform reasonably well and reproduce the observed temporal switch from memory cells to plasma cell production accurately [50], the stochastic differentiation model has the best agreement with experimental data (Figure 4E,F). One important point is that while the three models differ in the absolute counts of PCs and MBCs produced (Supplementary Figure S2), the experimental data only reports relative changes across time-points, which might hide global changes in cellular output. Additional experimental data on the GC output will help further differentiate between the 3 considered scenarios.

The amount of  $T_{FHs}$  plays a critical role in determining the GC outcome. Models where the number of  $T_{FHs}$  is fixed at a constant value tend to be stable, with the GC reaching a steady state upon which it does not decrease or increase in size, contrary to observations (Figure 4A). We therefore assume that  $T_{FHs}$  can enter and leave the GC during the whole reaction [40], which leads to having an approximately constant density of  $T_{FHs}$  [18]. With this assumption, the sizes of our simulated GCs closely resemble experimental determinations, and the GCs naturally become extinguished by day  $\sim 40$  (Figure 4A). Critical to achieving this self-shutdown is the choice of parameters. Indeed, a very rough estimation of the rate of variation of the total number of GB B cells,  $\dot{N}_B$ , shows that this number approximately evolves according to  $\dot{N}_B \simeq 0.6 \cdot r_{\text{division}} \cdot N_{CB} - r_{\text{apoptosis}} \cdot N_{CC} - N_{CC_{\text{sel}}} \cdot r_{\text{exit}}$ , where 0.6 is a correction factor that accounts for apoptotic and non-dividing centroblasts. Taking  $N_{CC}/N_{CC_{\text{sel}}} \simeq N_{CB}/N_{CC_{\text{sel}}} \simeq 100$ , as predicted by our model, our choice of parameters leads to  $\dot{N}_B \simeq -0.4 N_{CC_{\text{sel}}} < 0$ , explaining the observed reduction in  $N_B$  after day 9 (a detailed ODE analysis can be found in the Supplementary Information). Hence, after the initial phase of monoclonal expansion, our simulated GCs enter a phase of slow decrease in size until they naturally shutdown around day 40. Of course, additional mechanisms may be at play in terminating the GC reaction, such as antibody feedback [71]. Similarly, asynchronous onset and intercommunication between GCs might alter the efficiency of antibody response and lead to compromised efficiency and earlier termination of late initialized GCs [21]. Future renditions of our work will explicitly investigate the mechanisms by which affinity influences antigen uptake and antigen presentation, including its contribution to the termination of the GC reaction.

We also generated BCR sequences from our simulations and compare the simulated sequences with single-cell BCR sequencing data from individual GCs by means of phylogenetic trees. Each tree represents the evolutionary process that took place in the GC and the acquisition of new mutations that conferred higher affinity to the most fitting clones. The comparison of important tree topological properties, such as the number of leaves, the average depth, or the length of the trunk, reveals good agreement between our model and experiments. We find our predicted distribution of mutations per cell to be narrower than experimentally characterized (Figure 5), which could indicate that our current model of affinity, where one mutation can only increment the affinity by  $\frac{1}{N_{\text{CDR}}}$ , constrains the density function of the mutations count to be narrower. Furthermore, because of our explicit modelling of the BCR sequences, our model enables the investigation of the clonal diversity within

the GC and the emergence of dominance. A first observation is that B cells with the same affinity can have very different outcomes, resulting in a high variability of the evolution of clonal diversity across simulations, which matches experimental observations [15] (Figure 7). Interestingly, a deeper investigation of the emergence of clonal dominance demonstrates that small advantages in the affinity to antigen acquired through random mutations are amplified and result in clonal dominance within a week (Figures 8 and 9). This delayed correlation might explain why previous computational models found limited correlation between clone abundance and affinity throughout the GC lifespan [24].

As future work, we would like to investigate approaches to overcome technical limitations of the Gillespie algorithm, such as the assumption that random interactions can be modelled as a memoryless Poissonian process. This hypothesis might not be accurate for some biological programs, such as apoptosis or cell division, which tend to be better represented using log-normal distributions [72,73]. We also plan to explore generalizations of the Gillespie algorithm that incorporate both individual properties and non-Poissonian processes, while still being statistically correct.

Our model does not include regulation by T follicular regulatory,  $T_{FR}$ , cells.  $T_{FR}$  cells have been implicated in the negative regulation of GC B cell activation, affinity maturation and the differentiation of plasma cells [74]. Furthermore,  $T_{FRs}$  may have a role in the negative selection of autoreactive B cells that make up a significant part of GC B cells (8% according to [55]). Indeed, *ex vivo* TCR repertoire analysis shows that  $T_{FHs}$  and  $T_{FRs}$  cells exhibit different TCR repertoires, with the latter more closely resembling that of regulatory T cells and being skewed toward self-antigens [75]. This finding leads to the hypothesis that  $T_{FRs}$  might negatively regulate potentially autoreactive B cells to suppress autoimmunity [76]. Further work will integrate  $T_{FRs}$  and autoreactive B cells to our model, and will provide more insight behind the mechanism of autoimmunity.

Finally, an improved model to compute BCR affinity that integrates information about BCR sequences and about the structural shape of both the antibody variable region and the antigen can significantly enhance our knowledge of the mechanisms involved in the adaptive immune response [45]. Attempts in this direction have already been made, for instance by representing BCRs as a chain of 10 amino acids on a rigid lattice [28]. Despite these promising works, a full integration of the 3D structure of both the BCR and the antigen remains challenging. An alternative approach is to focus on predicting binding affinity directly from sequence and transcriptomic data. This is already routinely performed to predict drug sensitivity, where the binding affinity between a drug compound and a protein is predicted using transcriptomic data and chemical information without explicitly accounting for the 3D structure of either molecule [77,78]. An advantage of such approaches is that they enable the *in silico* engineering of new compounds with improved biochemical properties [79]. The adaption of these models to the task of predicting the binding affinity between a receptor and an antigen or therapeutic compound, e.g., a vaccine, would open exciting opportunities for synthetic antibody engineering or vaccines *in silico* optimisation. Work in this direction could have vast repercussions in understanding the emergence of broadly neutralizing antibodies, designing serological tests or vaccines against mutating pathogens, such as SARS-CoV-2.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4409/9/6/1448/s1>. Figure S1: Germinal center B cell count with different parameters for Tcell help, Figure S2: Memory cells and plasma cells production with and without scaling, Table S1: Sensitivity analysis.

**Author Contributions:** A.P. wrote the code and analyzed the experimental data under the supervision of M.R.M. Y.A., K.J. and J.K. contributed to the phylogenetic tree analysis under the supervision of N.B. and M.R.M. U.K. contributed by giving biological insights about the model assumptions. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the COSMIC European Training Network, funded from the European Union's Horizon 2020 research and innovation program under grant agreement No 765158.

**Acknowledgments:** A.P. and M.R.M. acknowledge Marta Schips, Theinmozhi Arulraj, Philippe A. Robert, Michael Meyer-Hermann and Claus Horn for constructive discussions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mesin, L.; Ersching, J.; Victora, G.D. Germinal center B cell dynamics. *Immunity* **2016**, *45*, 471–482. [[CrossRef](#)]
2. De Silva, N.S.; Klein, U. Dynamics of B cells in germinal centres. *Nat. Rev. Immunol.* **2015**, *15*, 137–148. [[CrossRef](#)] [[PubMed](#)]
3. Victora, G.D.; Nussenzweig, M.C. Germinal centers. *Annu. Rev. Immunol.* **2012**, *30*, 429–457. [[CrossRef](#)] [[PubMed](#)]
4. Vinuesa, C.G.; Sanz, I.; Cook, M.C. Dysregulation of germinal centres in autoimmune disease. *Nat. Rev. Immunol.* **2009**, *9*, 845–857. [[CrossRef](#)] [[PubMed](#)]
5. Pasqualucci, L. Molecular pathogenesis of germinal center-derived B cell lymphomas. *Immunol. Rev.* **2019**, *288*, 240–261. [[CrossRef](#)]
6. Allen, C.D.; Okada, T.; Tang, H.L.; Cyster, J.G. Imaging of germinal center selection events during affinity maturation. *Science* **2007**, *315*, 528–531. [[CrossRef](#)]
7. Jacob, J.; Kelsoe, G.; Rajewsky, K.; Weiss, U. Intraclonal generation of antibody mutants in germinal centres. *Nature* **1991**, *354*, 389–392. [[CrossRef](#)]
8. Berek, C.; Berger, A.; Apel, M. Maturation of the immune response in germinal centers. *Cell* **1991**, *67*, 1121–1129. [[CrossRef](#)]
9. Victora, G.D.; Schwickert, T.A.; Fooksman, D.R.; Kamphorst, A.O.; Meyer-Hermann, M.; Dustin, M.L.; Nussenzweig, M.C. Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. *Cell* **2010**, *143*, 592–605. [[CrossRef](#)]
10. Gitlin, A.D.; Mayer, C.T.; Oliveira, T.Y.; Shulman, Z.; Jones, M.J.; Koren, A.; Nussenzweig, M.C. T cell help controls the speed of the cell cycle in germinal center B cells. *Science* **2015**, *349*, 643–646. [[CrossRef](#)]
11. Heinzl, S.; Giang, T.B.; Kan, A.; Marchingo, J.M.; Lye, B.K.; Corcoran, L.M.; Hodgkin, P.D. A Myc-dependent division timer complements a cell-death timer to regulate T cell and B cell responses. *Nat. Immunol.* **2017**, *18*, 96–103. [[CrossRef](#)] [[PubMed](#)]
12. Dominguez-Sola, D.; Victora, G.D.; Ying, C.Y.; Phan, R.T.; Saito, M.; Nussenzweig, M.C.; Dalla-Favera, R. The proto-oncogene MYC is required for selection in the germinal center and cyclic reentry. *Nat. Immunol.* **2012**, *13*, 1083–1091. [[CrossRef](#)] [[PubMed](#)]
13. Calado, D.P.; Sasaki, Y.; Godinho, S.A.; Pellerin, A.; Köchert, K.; Sleckman, B.P.; De Alborán, I.M.; Janz, M.; Rodig, S.; Rajewsky, K. The cell-cycle regulator c-Myc is essential for the formation and maintenance of germinal centers. *Nat. Immunol.* **2012**, *13*, 1092–1100. [[CrossRef](#)] [[PubMed](#)]
14. Gitlin, A.D.; Shulman, Z.; Nussenzweig, M.C. Clonal selection in the germinal centre by regulated proliferation and hypermutation. *Nature* **2014**, *509*, 637–640. [[CrossRef](#)] [[PubMed](#)]
15. Tas, J.M.; Mesin, L.; Pasqual, G.; Targ, S.; Jacobsen, J.T.; Mano, Y.M.; Chen, C.S.; Weill, J.C.; Reynaud, C.A.; Browne, E.P.; et al. Visualizing antibody affinity maturation in germinal centers. *Science* **2016**, *351*, 1048–1054. [[CrossRef](#)] [[PubMed](#)]
16. Song, S.; Matthias, P.D. The transcriptional regulation of germinal center formation. *Front. Immunol.* **2018**, *9*, 2026. [[CrossRef](#)]
17. Martínez, M.R.; Corradin, A.; Klein, U.; Álvarez, M.J.; Toffolo, G.M.; di Camillo, B.; Califano, A.; Stolovitzky, G.A. Quantitative modeling of the terminal differentiation of B cells and mechanisms of lymphomagenesis. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2672–2677. [[CrossRef](#)]
18. Wittenbrink, N.; Klein, A.; Weiser, A.A.; Schuchhardt, J.; Or-Guil, M. 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. *J. Immunol.* **2011**, *187*, 6185–6196. [[CrossRef](#)]
19. Buchauer, L.; Wardemann, H. Calculating Germinal Centre Reactions. *Curr. Opin. Syst. Biol.* **2019**. [[CrossRef](#)]
20. Thomas, M.J.; Klein, U.; Lygeros, J.; Rodríguez Martínez, M. A probabilistic model of the germinal center reaction. *Front. Immunol.* **2019**, *10*, 689. [[CrossRef](#)]
21. Arulraj, T.; Meyer-Hermann, M.; Binder, S.C.; Robert, P.A. Synchronous germinal centre onset impacts the efficiency of antibody responses. *Front. Immunol.* **2019**, *10*, 2116. [[CrossRef](#)]
22. Meyer-Hermann, M.; Mohr, E.; Pelletier, N.; Zhang, Y.; Victora, G.D.; Toellner, K.M. A theory of germinal center B cell selection, division, and exit. *Cell Rep.* **2012**, *2*, 162–174. [[CrossRef](#)] [[PubMed](#)]

23. Wang, P.; Shih, C.M.; Qi, H.; Lan, Y.H. A stochastic model of the germinal center integrating local antigen competition, individualistic T–B Interactions, and B cell receptor signaling. *J. Immunol.* **2016**, *197*, 1169–1182. [[CrossRef](#)]
24. Reshetova, P.; van Schaik, B.D.; Klarenbeek, P.L.; Doorenspleet, M.E.; Esveldt, R.E.; Tak, P.P.; Guikema, J.E.; de Vries, N.; van Kampen, A.H. Computational model reveals limited correlation between germinal center B-cell subclone abundancy and affinity: Implications for repertoire sequencing. *Front. Immunol.* **2017**, *8*, 221. [[CrossRef](#)] [[PubMed](#)]
25. Amitai, A.; Mesin, L.; Victora, G.D.; Kardar, M.; Chakraborty, A.K. A population dynamics model for clonal diversity in a germinal center. *Front. Microbiol.* **2017**, *8*, 1693. [[CrossRef](#)] [[PubMed](#)]
26. Ellebedy, A.H.; Jackson, K.J.L.; Kissick, H.T.; Nakaya, H.I.; Davis, C.W.; Roskin, K.M.; McElroy, A.K.; Oshansky, C.M.; Elbein, R.; Thomas, S.; et al. Defining antigen-specific plasmablast and memory B cell subsets in human blood after viral infection or vaccination. *Nat. Immunol.* **2016**, *17*, 1226–1234. [[CrossRef](#)]
27. Peng, H.P.; Lee, K.H.; Jian, J.W.; Yang, A.S. Origins of specificity and affinity in antibody–protein interactions. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E2656–E2665. [[CrossRef](#)]
28. Robert, P.A.; Meyer-Hermann, M. A 3D Structural Affinity Model for Multi-Epitope in Silico Germinal Center Simulations. *bioRxiv* **2019**. [[CrossRef](#)]
29. Dunn-Walters, D.K.; Belelovsky, A.; Edelman, H.; Banerjee, M.; Mehr, R. The dynamics of germinal centre selection as measured by graph-theoretical analysis of mutational lineage trees. *Clin. Dev. Immunol.* **2002**, *9*, 233–243. [[CrossRef](#)]
30. Gillespie, D.T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **1977**, *81*, 2340–2361. [[CrossRef](#)]
31. McAdams, H.H.; Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 814–819. [[CrossRef](#)] [[PubMed](#)]
32. Arkin, A.; Ross, J.; McAdams, H.H. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage lambda-Infected Escherichia coli Cells. *Genetics* **1998**, *149*, 1633–1648. [[PubMed](#)]
33. Rodríguez Martínez, M.; Soriano, J.; Tlustý, T.; Pilpel, Y.; Furman, I. Messenger RNA fluctuations and regulatory RNAs shape the dynamics of a negative feedback loop. *Phys. Rev. E* **2010**, *81*, 031924. [[CrossRef](#)] [[PubMed](#)]
34. Schwickert, T.A.; Victora, G.D.; Fooksman, D.R.; Kamphorst, A.O.; Mugnier, M.R.; Gitlin, A.D.; Dustin, M.L.; Nussenzweig, M.C. A dynamic T cell–limited checkpoint regulates affinity-dependent B cell entry into the germinal center. *J. Exp. Med.* **2011**, *208*, 1243–1252. [[CrossRef](#)]
35. Liu, Y.J.; Zhang, J.; Lane, P.J.; Chan, E.Y.T.; Maclennan, I.C. Sites of specific B cell activation in primary and secondary responses to T cell-dependent and T cell-independent antigens. *Eur. J. Immunol.* **1991**, *21*, 2951–2962. [[CrossRef](#)]
36. Bannard, O.; McGowan, S.J.; Ersching, J.; Ishido, S.; Victora, G.D.; Shin, J.S.; Cyster, J.G. Ubiquitin-mediated fluctuations in MHC class II facilitate efficient germinal center B cell responses. *J. Exp. Med.* **2016**, *213*, 993–1009. [[CrossRef](#)]
37. Fiala, G.J.; Kaschek, D.; Blumenthal, B.; Reth, M.; Timmer, J.; Schamel, W.W. Pre-clustering of the B cell antigen receptor demonstrated by mathematically extended electron microscopy. *Front. Immunol.* **2013**, *4*, 427. [[CrossRef](#)]
38. El Shikh, M.E.M.; El Sayed, R.M.; Sukumar, S.; Szakal, A.K.; Tew, J.G. Activation of B cells by antigens on follicular dendritic cells. *Trends Immunol.* **2010**, *31*, 205–211. [[CrossRef](#)]
39. Nowosad, C.R.; Spillane, K.M.; Tolar, P. Germinal center B cells recognize antigen through a specialized immune synapse architecture. *Nat. Immunol.* **2016**, *17*, 870–877. [[CrossRef](#)]
40. Shulman, Z.; Gitlin, A.D.; Targ, S.; Jankovic, M.; Pasqual, G.; Nussenzweig, M.C.; Victora, G.D. T follicular helper cell dynamics in germinal centers. *Science* **2013**, *341*, 673–677. [[CrossRef](#)]
41. Papa, I.; Saliba, D.; Ponzoni, M.; Bustamante, S.; Canete, P.F.; Gonzalez-Figueroa, P.; McNamara, H.A.; Valvo, S.; Grimbaldston, M.; Sweet, R.A.; et al. T FH-derived dopamine accelerates productive synapses in germinal centres. *Nature* **2017**, *547*, 318–323. [[CrossRef](#)] [[PubMed](#)]
42. Janeway, C.A., Jr.; Travers, P.; Walport, M.; Shlomchik, M.J. The structure of a typical antibody molecule. In *Immunobiology: The Immune System in Health and Disease*, 5th ed.; Garland Science: New York, NY, USA, 2001.
43. Hoehn, K.B.; Fowler, A.; Lunter, G.; Pybus, O.G. The diversity and molecular evolution of B-cell receptors during infection. *Mol. Biol. Evol.* **2016**, *33*, 1147–1157. [[CrossRef](#)]

44. Abbott, R.K.; Lee, J.H.; Menis, S.; Skog, P.; Rossi, M.; Ota, T.; Kulp, D.W.; Bhullar, D.; Kalyuzhnyi, O.; Havenar-Daughton, C.; et al. Precursor frequency and affinity determine B cell competitive fitness in germinal centers, tested with germline-targeting HIV vaccine immunogens. *Immunity* **2018**, *48*, 133–146. [[CrossRef](#)] [[PubMed](#)]
45. Kovaltsuk, A.; Krawczyk, K.; Galson, J.D.; Kelly, D.F.; Deane, C.M.; Trück, J. How B-cell receptor repertoire sequencing can be enriched with structural antibody data. *Front. Immunol.* **2017**, *8*, 1753. [[CrossRef](#)] [[PubMed](#)]
46. Kleinstein, S.H.; Louzoun, Y.; Shlomchik, M.J. Estimating hypermutation rates from clonal tree data. *J. Immunol.* **2003**, *171*, 4639–4649. [[CrossRef](#)]
47. Perelson, A.S.; Oster, G.F. Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. *J. Theor. Biol.* **1979**, *81*, 645–670. [[CrossRef](#)]
48. Meyer-Hermann, M.; Deutsch, A.; Or-Guil, M. Recycling probability and dynamical properties of germinal center reactions. *J. Theor. Biol.* **2001**, *210*, 265–285. [[CrossRef](#)]
49. Liu, D.; Xu, H.; Shih, C.; Wan, Z.; Ma, X.; Ma, W.; Luo, D.; Qi, H. T-B-cell entanglement and ICOSL-driven feed-forward regulation of germinal centre reaction. *Nature* **2015**, *517*, 214–218. [[CrossRef](#)]
50. Weisel, F.J.; Zuccarino-Catania, G.V.; Chikina, M.; Shlomchik, M.J. A temporal switch in the germinal center determines differential output of memory B and plasma cells. *Immunity* **2016**, *44*, 116–130. [[CrossRef](#)]
51. Ise, W.; Kurosaki, T. Plasma cell differentiation during the germinal center reaction. *Immunol. Rev.* **2019**, *288*, 64–74. [[CrossRef](#)]
52. Ise, W.; Fujii, K.; Shiroguchi, K.; Ito, A.; Kometani, K.; Takeda, K.; Kawakami, E.; Yamashita, K.; Suzuki, K.; Okada, T.; et al. T Follicular Helper Cell-Germinal Center B Cell Interaction Strength Regulates Entry into Plasma Cell or Recycling Germinal Center Cell Fate. *Immunity* **2018**, *48*, 702–715. [[CrossRef](#)]
53. Shinnakasu, R.; Inoue, T.; Kometani, K.; Moriyama, S.; Adachi, Y.; Nakayama, M.; Takahashi, Y.; Fukuyama, H.; Okada, T.; Kurosaki, T. Regulated selection of germinal-center cells into the memory B cell compartment. *Nat. Immunol.* **2016**, *17*, 861–869. [[CrossRef](#)]
54. Suan, D.; Kräutler, N.J.; Maag, J.L.V.; Butt, D.; Bourne, K.; Hermes, J.R.; Avery, D.T.; Young, C.; Statham, A.; Elliott, M.; et al. CCR6 Defines Memory B Cell Precursors in Mouse and Human Germinal Centers, Revealing Light-Zone Location and Predominant Low Antigen Affinity. *Immunity* **2017**, *47*, 1142–1153. [[CrossRef](#)] [[PubMed](#)]
55. Mayer, C.T.; Gazumyan, A.; Kara, E.E.; Gitlin, A.D.; Golijanin, J.; Viant, C.; Pai, J.; Oliveira, T.Y.; Wang, Q.; Escolano, A.; et al. The microanatomic segregation of selection by apoptosis in the germinal center. *Science* **2017**, *358*, eaao2602. [[CrossRef](#)] [[PubMed](#)]
56. Shahaf, G.; Barak, M.; Zuckerman, N.S.; Swerdlin, N.; Gorfine, M.; Mehr, R. Antigen-driven selection in germinal centers as reflected by the shape characteristics of immunoglobulin gene lineage trees: a large-scale simulation study. *J. Theor. Biol.* **2008**, *255*, 210–222. [[CrossRef](#)]
57. Lefranc, M.P.; Giudicelli, V.; Ginestoux, C.; Jabado-Michaloud, J.; Folch, G.; Bellahcene, F.; Wu, Y.; Gemrot, E.; Brochet, X.; Lane, J.; et al. IMGT<sup>®</sup>, the international ImMunoGeneTics information system<sup>®</sup>. *Nucleic Acids Res.* **2008**, *37*, D1006–D1012. [[CrossRef](#)] [[PubMed](#)]
58. Retter, I.; Althaus, H.H.; Münch, R.; Müller, W. VBASE2, an integrative V gene database. *Nucleic Acids Res.* **2005**, *33*, D671–D674. [[CrossRef](#)] [[PubMed](#)]
59. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
60. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [[CrossRef](#)] [[PubMed](#)]
61. Jahn, K.; Kuipers, J.; Beerenwinkel, N. Tree inference for single-cell data. *Genome Biol.* **2016**, *17*, 86. [[CrossRef](#)] [[PubMed](#)]
62. Talevich, E.; Invergo, B.M.; Cock, P.J.; Chapman, B.A. Bio. Phylo: A unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC Bioinform.* **2012**, *13*, 209. [[CrossRef](#)] [[PubMed](#)]
63. Meyer-Hermann, M.; Binder, S.C.; Mesin, L.; Vitorica, G.D. Computer simulation of multi-color brainbow staining and clonal evolution of B cells in germinal centers. *Front. Immunol.* **2018**, *9*, 2020. [[CrossRef](#)] [[PubMed](#)]



64. Schwickert, T.A.; Lindquist, R.L.; Shakhar, G.; Livshits, G.; Skokos, D.; Kosco-Vilbois, M.H.; Dustin, M.L.; Nussenzweig, M.C. In vivo imaging of germinal centres reveals a dynamic open structure. *Nature* **2007**, *446*, 83–87. [[CrossRef](#)] [[PubMed](#)]
65. Wittenbrink, N.; Weber, T.S.; Klein, A.; Weiser, A.A.; Zuschmitter, W.; Sibila, M.; Schuchhardt, J.; Or-Guil, M. Broad volume distributions indicate nonsynchronized growth and suggest sudden collapses of germinal center B cell populations. *J. Immunol.* **2010**, *184*, 1339–1347. [[CrossRef](#)] [[PubMed](#)]
66. Monroe, J.G.; Cambier, J.C. Sorting of B lymphoblasts based upon cell diameter provides cell populations enriched in different stages of cell cycle. *J. Immunol. Methods* **1983**, *63*, 45–56. [[CrossRef](#)]
67. Ersching, J.; Efeyan, A.; Mesin, L.; Jacobsen, J.T.; Pasqual, G.; Grabiner, B.C.; Dominguez-Sola, D.; Sabatini, D.M.; Victora, G.D. Germinal center selection and affinity maturation require dynamic regulation of mTORC1 kinase. *Immunity* **2017**, *46*, 1045–1058. [[CrossRef](#)]
68. Davis, K. A Global Optimization Algorithm Worth Using. *dlib C++ Library*. 2017. Available online: <http://blog.dlib.net/2017/12/a-global-optimization-algorithm-worth.html> (accessed on 1 March 2020).
69. Snoek, J.; Larochelle, H.; Adams, R.P. Practical bayesian optimization of machine learning algorithms. In Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, NV, USA, 3–6 December 2012; pp. 2951–2959.
70. Shen, C. Analysis of detrended time-lagged cross-correlation between two nonstationary time series. *Phys. Lett. A* **2015**, *379*, 680–687. [[CrossRef](#)]
71. Zhang, Y.; Meyer-Hermann, M.; George, L.A.; Figge, M.T.; Khan, M.; Goodall, M.; Young, S.P.; Reynolds, A.; Falciani, F.; Waisman, A.; et al. Germinal center B cells govern their own fate via antibody feedback. *J. Exp. Med.* **2013**, *210*, 457–464. [[CrossRef](#)]
72. Koch, A.L. The logarithm in biology 1. Mechanisms generating the log-normal distribution exactly. *J. Theor. Biol.* **1966**, *12*, 276–290. [[CrossRef](#)]
73. Dowling, M.R.; Milutinović, D.; Hodgkin, P.D. Modelling cell lifespan and proliferation: is likelihood to die or to divide independent of age? *J. R. Soc. Interface* **2005**, *2*, 517–526. [[CrossRef](#)]
74. Chung, Y.; Tanaka, S.; Chu, F.; Nurieva, R.I.; Martinez, G.J.; Rawal, S.; Wang, Y.H.; Lim, H.; Reynolds, J.M.; Zhou, X.H.; et al. Follicular regulatory T cells expressing Foxp3 and Bcl-6 suppress germinal center reactions. *Nat. Med.* **2011**, *17*, 983–988. [[CrossRef](#)] [[PubMed](#)]
75. Maceiras, A.R.; Almeida, S.C.P.; Mariotti-Ferrandiz, E.; Chacara, W.; Jebbawi, F.; Six, A.; Hori, S.; Klatzmann, D.; Faro, J.; Graca, L. T follicular helper and T follicular regulatory cells have different TCR specificity. *Nat. Commun.* **2017**, *8*, 15067. [[CrossRef](#)] [[PubMed](#)]
76. Sage, P.T.; Sharpe, A.H. T follicular regulatory cells. *Immunol. Rev.* **2016**, *271*, 246–259. [[CrossRef](#)]
77. Oskooei, A.; Manica, M.; Mathis, R.; Martínez, M.R. Network-based Biased Tree Ensembles (NetBiTE) for Drug Sensitivity Prediction and Drug Sensitivity Biomarker Identification in Cancer. *Sci. Rep.* **2019**, *9*, 1–13. [[CrossRef](#)] [[PubMed](#)]
78. Manica, M.; Oskooei, A.; Born, J.; Subramanian, V.; Sáez-Rodríguez, J.; Rodríguez Martínez, M. Toward Explainable Anticancer Compound Sensitivity Prediction via Multimodal Attention-Based Convolutional Encoders. *Mol. Pharm.* **2019**, *16*. [[CrossRef](#)] [[PubMed](#)]
79. Born, J.; Manica, M.; Oskooei, A.; Cadow, J.; Rodríguez Martínez, M. PaccMannRL: Designing Anticancer Drugs From Transcriptomic Data via Reinforcement Learning. In *Research in Computational Molecular Biology; Lecture Notes in Computer Science*; Schwartz, R., Ed.; Springer International Publishing: Cham, Switzerland, 2020; pp. 231–233. [[CrossRef](#)]

**Code Availability:** The experimental data and python code used to fit our model, are provided freely on Github under MIT license at <https://github.ibm.com/SysBio/Germinal-Center>. The GC kinetic data of [18] were derived from the online database at <http://sysimtools.eu>, and GC B cell sequences used to construct trees were made publicly available by [15]. The normalized dominance score [15] and MBC/PC output with time [50] were kindly provided by the corresponding authors of the respective studies.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).