Use of prior knowledge in biological systems modelling
Reshetova, P.

Citation for published version (APA):
Reshetova, P. V. (2017). Use of prior knowledge in biological systems modelling

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 5
The evolution of B-cell lineage trees during affinity maturation

B-cell affinity maturation in germinal centres (GCs) during an immune response is a crucial mechanism of our immune system, which is, however, hard to assess experimentally. Lineage trees reconstructed from B-cells subjected to affinity maturation have been reconstructed from experimental data. The resulting shape parameters of the tree have been shown to reflect the properties of affinity maturation. In this study we used a computational model to explore the evolution of B-cells lineage trees during affinity maturation in a single GC. We analyzed lineage tree parameters such as total number of nodes, node outgoing degrees and tree length and followed changes in these parameters as the immune response progressed. Since our model also provides information about subclonal abundance (cell counts) and affinity, we integrated those quantities in the lineage tree. This provides additional information about B-cell affinity maturation which is virtually impossible to obtain using experimental data.

5.1. Background
B-cell affinity maturation is a key defence mechanism of the adaptive immune system that improves the response against pathogens. Affinity maturation takes place in germinal centres (GCs), which are specialised structures in lymphoid organs where B cells proliferate with a high rate and undergo somatic hypermutation (SHM). SHM affects the genes coding for the B-cell receptor (BCR) which is responsible for antigen (Ag) binding. The evolution of B cells as result of affinity maturation can be visualized and analysed by constructing B-cell lineage trees from sequenced BCRs. In such tree every node represents a unique subclone (variant of a clone within VJ family produced by SHM [140]) and, consequently, SHM events can be followed. Tree edges represent the number of mutations between two connected subclones. The root of the tree generally represents the un-mutated germline sequence, while the leaves represent the subclones (at the end of affinity maturation) that were not further mutated. It has been suggested that the shape of the lineage tree reflects differences in the affinity maturation process in health and disease. Lineage tree shapes can be characterized by graph parameters such as number of nodes and average outgoing degree [141, 104]. Several tools have been developed to support the reconstruction and annotation of lineage trees from experimental data [142, 132, 143, 144].

Lineage trees have been used to confirm the role of GC as the location of SHM [145, 146, 139], to identify lineage relationships between cells from independent GCs

\footnote{P. Reshetova, B.D.C. van Schaik, A.H.C. van Kampen. Manuscript in preparation.}
Chapter 5. The evolution of B-cell lineage trees during affinity maturation

92

or from different tissues [148], and to analyse broadly neutralizing HIV antibodies [149]. A general discussion about the molecular evolution of B-cell receptors may be found in [150]. The evolution of a lineage tree during the germinal centre reaction (GCR) from a limited number of experimentally derived sequences was shown by Jacob (1993) and further analysed by Dunn-Walters and co-authors [141]. They showed that during the GCR, the lineage trees gradually change to longer, more pruned, shapes due to positive and negative selection of B cells. In addition, they compared lineage trees constructed from immunoglobulin (Ig) sequences obtained from human spleen and Peyer's patches and concluded that B cells in spleen were subjected to stronger selective forces. Finally, they were able to show that a selection of tree graph parameters significantly correlated to parameters (mutation rate, selection threshold) of a simple computational model of affinity maturation. Similar analyses were conducted by Shahaf et al. who used lineage trees to compare primary and secondary immune responses in silico [104]. Based on their analysis, tree graph parameters such as the outgoing degree of the tree root could be related to the selection threshold and initial affinity. However, later it was argued that these correlation may not be significant since experimental factors were not accounted for and because some of the graph parameters measures may reflect differences in the overall population size between tested conditions [151]. Instead, Uduman et al proposed to incorporate tree shape measures into statistical tests to detect selection of BCR sequences because shape parameters alone may not be very reliable.

Stern and co-workers analysed B-cell lineage trees to address the mechanism of B-cell maturation and trafficking between the central nervous system (CNS) and secondary lymphoid organs in multiple sclerosis (MS) [152]. It was suggested that MS CNS B cells encounter antigen and improve their affinity in the secondary lymphoid tissue. Part of these B cells then populate the CNS but continue trafficking between the CNS and periphery. Similarly, lineage tree analysis was used to study diversification of B cells found in inflamed intestinal tissue of two ulcerative colitis patients as well as B cells from mucosa-associated lymph nodes (LN) [153]. Tree shapes revealed active clonal diversification in ulcerative colitis patients. Moreover, B cells from intestinal tissues and the associated lymph nodes were shown to be clonally related, thus supplying evidence for B-cell trafficking between gut and associated lymph nodes. More recently, lineage trees were used to analyse two time-scales in affinity maturation of naive and reactivated B cells [96]. Naive B cells start with a germline state without mutations. In contrast B memory cells that are reactivated begin with a mutated, affinity-matured receptor, which is then further diversified during a GCR. B-cell lineage trees also revealed changes in affinity maturation with age supporting the observation that elderly produce increased levels of antibodies to autologous antigens and are less able to make high-affinity antibodies to foreign antigens [154]. This study showed also demonstrated tissue-specific differences between Peyer's patch GC and splenic GC with age. Lineage tree analysis applied to study autoimmune diseases showed that tree sizes in these diseases are larger compared to normal controls indicating that more mutations accumulated [155, 156]. This was expected from the chronic nature of autoimmune disorders. However, based on the analysis of the outgoing degree (see below) these studies also showed that autoimmune diseases and normal controls experienced similar selection pressure.

In this study we explore the evolution of B cells during affinity maturation in a sin-
gle GC by using a computational model which we developed in Chapter 4. In contrast to previous studies that investigated tree evolution from a limited set of experimentally-derived sequences [139, 141], our computational model enables the construction of trees including every subclone produced by SHM during affinity maturation. We analysed the change in graph parameters during the GCR and show that the total number of nodes and tree length behave in a similar fashion with experimentally derived trees. One advantage of our computational model is that it also provides information about subclonal abundance and affinity. We show how this can be integrated in the lineage tree to provide information that currently is virtually impossible to derive from experimental data. By repeating our simulations we also obtain insight in the variability of the tree shapes as result of SHM.

5.2. Methods

Software

We developed a mathematical model using ordinary differential equations (ODEs) to describe the dynamics of individual subclones during the GCR. This model is implemented in the R statistical environment version 3.2.2 [110] using R packages deSolve (version 1.12) [111], R6 (version 2.1.2), ggplot 2.0, igraph 1.0.1 and beeswarm 0.2.1. The software is freely available as open source (GPLv3) on request from the author.

Computational model

Here we describe the main aspects of our computational model. Further details can be found in Chapter 4. The GC and affinity maturation are reviewed in (Victora, 2012; Silva and Klein; 2015). Our model starts with a monoclonal expansion of B cells (centroblasts; CB) at day 0 to over 10,000 cells at day 4. During this expansion phase and the remainder of the GCR, the CBs differentiate to centrocytes (CC). SHM and the production of plasma B cells and memory B cells is initiated at day 4. At day 21 the GCR is terminated. The Ag and T follicular help (Tfh) survival signals are modelled with sigmoidal functions $S_d$ and $S_a$ that affect the rate of CB to CC differentiation, and the rate of CC apoptosis respectively. Effectively, these functions induce competition between the subclones through positive and negative selection. SHM with a rate of $10^{-3}$ per bp per division is modelled with a Poisson distribution $m = \text{Poisson}(\lambda = 0.6)$. The fate of each mutation is determined a decision tree involving silent (synonymous mutations), lethal FWR, and affinity changing CDR mutations [103, 120] (Figure 5.1). The lineage trees that we construct during the simulation are based on unique subclones that are defined as having a unique CDR protein sequence. Each new subclone created by SHM starts as a CB and, subsequently proliferates and differentiates to and co-exist as CB, CC, memory cell and plasma cell at succeeding time points. Subclones that with CB cell counts $\leq 1$ are kept in our simulation be are not further be affected by SHM to avoid the generation of new clones from these cells. Each subclone in our model has a unique BCR with an absolute affinity $\sigma$ for the Ag. The affinities of the three single cell founder CBs are set to arbitrary but different low affinity values (0.1, 0.3, and 0.5 $\mu M$). For each affinity changing mutation (Figure 5.1) the affinity of the affected subclone is updated according to $\sigma_{\text{new subclone}} = \sigma_{\text{parent}} + \Delta \sigma$ where $\Delta \sigma$ is drawn from a distribution $f(s, r, \sigma)$ with
probability density function:

\[ f(\sigma) = g(s = 3.0, r = 0.3) - \mu - (\sigma_{\text{parent}} * 0.1), \quad (5.1) \]

where \( g(s, r) \) is the inverse gamma distribution. \( \mu \) is the expected value of \( g \) and used to center the distribution around zero resulting in about equal chances for decreasing and increasing the affinity of mutated subclones. To decrease the chance for high affinity subclones to further improve their affinity we shift distribution \( f \) to the left as a function of the parent cell affinity.

**Ordinary Differential Equations**

Each subclone \( i \) can assume 4 phenotypes: centrocytes \((CC)_i\), centroblasts \((CB)_i\), memory cells \((M)_i\), and plasma cells \((P)_i\) (Figure 5.2). The temporal dynamics of each individual subclone is described by a set of ordinary differential equations (ODEs) representing these four phenotypes (Equations 5.2a to 5.2d; Table 5.1).

\[
\frac{d(CB)_i}{dt} = \rho_{CB} \cdot \left( \frac{A^h}{CP_{\text{total}}^h + A^h} \right) \cdot (CB)_i + \eta_{CC\rightarrow CB} \cdot (CC)_i \\
- \left( 1 - S_{\sigma}(\sigma_{\text{rel}, i}) \right) \cdot \eta_{CB\rightarrow CC} \cdot (CB)_i
\quad (5.2a)
\]
Figure 5.2: Representation of the GC model. Phenotypes of a single subclone are shown: centrocytes $CC_i$, centroblasts $CB_i$, plasma $P_i$ and memory $M_i$ cells, where $i$ is a subclone index. Arrows represent cell proliferation rate ($\rho_{cb}$), differentiation rates ($\eta_{CB\rightarrow CC}$, $\eta_{CC\rightarrow CB}$, $\eta_{CC\rightarrow P}$, $\eta_{CC\rightarrow M}$) and death rates ($\mu_{cc}$, $\mu_{p}$, $\mu_{m}$). $S_d$ and $S_a$ represent the survival signals affecting differentiation and apoptosis. $\sigma_i$ and $\sigma_{i,rel}$ represent the subclone absolute and relative affinities.

\[
\begin{align*}
\frac{dCC_i}{dt} &= \left(1 - S_d(\sigma_{rel,i})\right) \cdot \eta_{CB\rightarrow CC} \cdot CB_i - \eta_{CC\rightarrow CB} \cdot CC_i \\
&\quad - \left(1 - S_a(\sigma_{rel,i})\right) \cdot \mu_{CC} \cdot CC_i - \eta_{CC\rightarrow M} \cdot CC_i \\
&\quad - \eta_{CC\rightarrow P} \cdot \sigma_i \cdot CC_i \\
\frac{dM_i}{dt} &= \eta_{CC\rightarrow M} \cdot CC_i - \mu_{M} \cdot M_i \\
\frac{dP_i}{dt} &= \eta_{CC\rightarrow P} \cdot \sigma_i \cdot CC_i - \mu_{P} \cdot P_i
\end{align*}
\] (5.2b, 5.2c, 5.2d)

To facilitate monoclonal expansion to about 10,000 cells the signal $S$ is set to 0.9 for the first 4 days of the simulation to minimize apoptosis of CCs and differentiation to of CBs to CCs. The CB equation includes a density dependent expansion term defining nonspecific resource competition between the B cells, reducing their proliferation rate if the number of cells approaches $A$. The CC apoptosis rate and the CB to CC differentiation rate are multiplied by $\left(1 - S_a(\sigma_{rel,i})\right)$ to facilitate B-cell selection. Plasma B-cell differentiation depends on the absolute affinity $\sigma_i$ to reduce their production at earlier stages of the GCR. During the simulation we calculate the differential equations for periods of six hours (the duration of one CB division). After each period we impose SHM and update the population of subclones. For each non-lethal SHM a new subclone and an additional set of four ODEs is created. The CB cell count for new subclones is set to one, while the corresponding cell counts for the CCs, memory cells and plasma B cells are set to zero. The CB cell count of the parent subclone is reduced by one. If the sum
Table 5.1: Model parameters.

<table>
<thead>
<tr>
<th>B-cell type</th>
<th>Proliferation rate $(day^{-1})$</th>
<th>Differentiation rate $(day^{-1})$</th>
<th>Apoptosis rate $(day^{-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centroblast (CB)</td>
<td>$\rho_{CB} = 4$ [129, 124, 130]</td>
<td>$\eta_{CB \rightarrow CC} = 6$ [91]</td>
<td></td>
</tr>
<tr>
<td>Centrocyte (CC)</td>
<td></td>
<td>$\eta_{CC \rightarrow M} = 1$ [131]</td>
<td>$\eta_{CC \rightarrow P} = 0.1$ [131]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\eta_{CC \rightarrow CB} = 1$ [91]</td>
</tr>
<tr>
<td>Plasma cell (P)</td>
<td></td>
<td></td>
<td>$\mu_{P} = 0.25$ [131]</td>
</tr>
<tr>
<td>Memory cell (M)</td>
<td></td>
<td></td>
<td>$\mu_{M} = 0.01$ [131]</td>
</tr>
</tbody>
</table>

Other parameters

| Capacity | A = 8000 |
| Number of founder cells | 3 |
| Initial affinities | 0.1, 0.3, 0.5 $mol^{-1}$ |
| s | 0.3 |
| r | 0.3 |
| h | 20 |
| affinity shift | 0.1 |

of CC and CB counts for subclone $i$ is less than 0.1 cells we remove the subclone and corresponding equations from the system.

**Lineage tree construction**

Lineage trees reconstructed from experimentally derived sequence are mainly based on CCs (since CBs do not, or at very low levels express the BCR). Consequently, in our simulation we construct lineage trees from CCs only. We only used CCs with cell counts $\geq 1$.

In our simulation we incrementally build the lineage trees for each of the three founder B cells while new subclones are being produced by non-lethal SHM. Therefore, in contrast to tree reconstruction from experimental data we did not have to use any special tree reconstruction method. Moreover, we were able to construct lineage trees where the difference between any connected two nodes is a single mutation.

**Calculation of graph parameters**

We analysed lineage trees resulting from the founder cell with the highest initial affinity. We repeated the simulation 14 times to obtain information about the change in variability during the GCR for each individual graph parameter.

For every lineage tree we calculated and discussed a subset of nine graph parameters that changed during the GCR (Table 5.2, Figure 5.3):

The graph parameters can be interpreted in terms of affinity maturation. The total number of nodes (N) represents the number of subclones produced through the GCR. The total number of leaves (L) represents the number of distinct subclones that were not further affected by SHM. Internal nodes (IN) include all nodes except the root and leaves. Part of these internal nodes represent living subclones that still reside in the population of subclones. Pass through nodes (PTNs) represent subclones with exactly one child. Although a subclone (a PTN) could have produced descendants during the GCR, these subclones did not survive the selection process or the mutations producing these subclones were lethal. In contrast, split nodes (S) have two or more descendants. The outgoing degree represents the number child subclones produced by SHM for each non-leave
Table 5.2: Graph parameters.

<table>
<thead>
<tr>
<th>Graph parameter</th>
<th>Abbreviation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of nodes</td>
<td>N</td>
<td>[47, 1235]</td>
</tr>
<tr>
<td>Total number of leaves</td>
<td>L</td>
<td>[33, 867]</td>
</tr>
<tr>
<td>Number of internal nodes</td>
<td>IN</td>
<td>[13, 483]</td>
</tr>
<tr>
<td>Number of pass through nodes</td>
<td>PTN</td>
<td>[7, 361]</td>
</tr>
<tr>
<td>Average outgoing degree (except the root)</td>
<td>avgOD</td>
<td>[4, 267]</td>
</tr>
<tr>
<td>Root outgoing degree</td>
<td>rootOD</td>
<td>[6, 711]</td>
</tr>
<tr>
<td>Average path length from the root to leave</td>
<td>avgRL</td>
<td>[2, 5]</td>
</tr>
<tr>
<td>Maximum path length from the root to the leave</td>
<td>maxRL</td>
<td>[4, 11]</td>
</tr>
<tr>
<td>Average distance between the root and any split node</td>
<td>avgRSN</td>
<td>[1, 5]</td>
</tr>
</tbody>
</table>

Figure 5.3: Graph parameters describing a B-cell lineage tree.

Subclone. In the first phase of the GCR the founder subclone monoclonally expands to thousands of cells. Consequently, once SHM is initiated the founder subclone produces a large number of descendants, which leads to an exceptionally high root outgoing degree.

To avoid skewed results we consider the average outgoing degree of all split nodes except the root (avgOD) and the root outgoing degree (rootOD) separately. PTN, L, avgOD, and rootOD provide a measure for the “bushiness” of the tree which inversely correlates with the amount of selection pressure B-cell subclones experience. Decline of rootOD during affinity maturation as a result of competition between subclones is an indicator of the progression of the GCR. Also the average path length from root to leave (avgRL) and the maximum path length from root to leave (maxRL) represents the average and maximum number of mutations in single subclone. This parameter is influenced by the
mutation rate, initial affinity (lower affinity founder cells may acquire more mutations to obtain maximum possible affinity), and length of GCR reaction. Average distance between the root and any split node (avgRSN) indicates the progress of the GCR. While the GCR progresses, early low affinity root descendants are removed from the subclonal population as a result of the selection process and avgRSN is growing.

Lineage tree graph parameters were calculated for every day starting at day 10 and ending at day 21. Due to the nature of the differential equations the cell counts may become less than 1. In our model we allow subclones with cell counts < 1 to avoid too much interference with the simulation but remove a subclone (the corresponding set of equations) if cell counts are < 0.1. However, subclones with cell counts < 1 do not have a biological interpretation therefor we construct lineage trees for subclones with cell counts ≥ 1. Trees before day 10 only contained very few nodes and, therefore, were excluded from the analysis. To gain insight in the (change in) variability of the graph parameters we repeated simulations 14 times resulting in 14 lineage trees at every time point. We report the values of the graph parameters as boxplots where the boxes indicate the 25th and 75th percentile, the horizontal line in the box represents the median, the whiskers indicate the 5th and 95th percentiles and the crosses indicate outliers (defined by extreme studentized deviate test [157]).

Expanded subclones
Subclonal expansion was defined following Chapter 4. In brief, a histogram of CC cell counts $c$ for all subclones at the end of the GCR is constructed to reflect their frequencies $F(c)$. Next we define $T$ as lowest count $c$ for which $F(c) = 0$. We assumed that $F(c ≥ T) = 0$ for the underlying but unknown null distribution of unexpanded subclones. Consequently, we define all subclones with cell counts > $T$ and $F(T) ≥ 1$ to be expanded. The expansion threshold $T$ was estimated for each individual simulation at the end of the GCR.

5.3. Results
We first visualized lineage trees development at different time points during the GCR. Secondly, we analysed the tree graph parameters during the GCR. Finally, we analysed subclonal expansion and affinity in the context of a lineage tree.

Visualization of lineage tree development during the GCR
To visualize the lineage tree evolution during the GCR we selected a representative simulation and constructed lineage trees for days 10, 13, 17, and 21 (Figure 5.4). The length of these trees (maxPL) increases from 3 at day 10 to 6 at day 17.

Graph parameters N, L, PTN and IN
The median total number of subclones (N) stays relatively stable during the course of the GCR (Figure 5.5(A)) with a slight peak between days 12-14. This shows that subclonal diversity does not narrow to a single or few high affinity subclones that outcompete the other lower affinity subclones. The median total number of leaves (L) also stays relatively constant. Number of internal nodes (all nodes except the root and leaves) (IN) stays close to the number of pass through nodes (PTN) indicating that only a minor set of
Day 10

Day 13

Day 17

Day 21

Figure 5.4: Evolution of a lineage tree during the GCR. CC with cell counts ≤ 3 and their ancestors are shown. Four time points were selected from a single simulation to visualize the tree evolution. Every node represents a subclone. Each edge represents a single mutation.
nodes is able to produce more than 1 descendants (and thus are not considered PTN). The median number of PTN shows a maximum at 13 days indicating that at that stage during affinity maturation we have the largest percentage of non-diversifying subclones (Figure 5.5(B)). Graph parameter boxplots also show a variability of the trees during the GCR caused by the random process of SHM. For example, $N$ shows that the tree size at the end of the GCR varies between approximately 700 and 1000 subclones Figure 5.5. One simulation resulted in an outlier corresponding a very small tree containing only 47 subclones.

**Graph parameters rootOD and avgOD**

The outgoing degrees provide a measure for the bushiness of the lineage tree. The root outgoing degree (rootOD) and the average outgoing degree (avgOD) are decreasing while the affinity maturation is progressing (Figure 5.6).

Both outgoing degrees decrease with ongoing GCR and, therefore, are correlated to some extent. However, it remains to be established how these graph parameters pro-
Chapter 5.3. Results

<table>
<thead>
<tr>
<th>Days</th>
<th>Number of nodes</th>
<th>Root degree, founder 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 5.6: (A) The average outgoing degree (avgOD) and (B) root outgoing degree (rootOD).

Provide information about the underlying process. According to [141] each of these outgoing degrees relate to the mutation rate and selection threshold. In contrast Shahaf et al [104] related only avgOD to the mutation rate, while rootOD was related to the selection threshold although with an opposite effect than by Dunn-Walters et al. Uduman et al did not find any correlation between these ODs and model parameters [151]. However, from their research we might conclude that increasing the mutation rate would lead to trees that are less bushy.

In contrast to the increasing variance for N, L, PTN, and IN the variance for the outgoing degrees decreases with progressing GCR. Thus, although the GCR may produce trees that show large variability in size, the number of different descendant subclones produced from a single subclone decreases leading to less bushy trees. This is mainly a consequence of subclone abundance which decreases in time (see Chapter 4). Initially the abundance of the founder clones is very high as a result of monoclonal expansion. Each of these cells can be mutated and produce a new subclone represented by a single cell. These new subclones do proliferate but due to new mutation event and competition with other subclones their abundances stay relatively low, resulting in fewer possibilities to produce new subclones. If we consider rootOD then we observe that a large number of new subclones originate from the founder clone during the initial phase of the GCR (Figure 5.4). At later stages the founder clone is reduced in abundance and many of its descendants did not survive.

**Graph parameters avgRL, maxRL, avgRSN**

Figure 5.7 shows various distances in terms of acquired mutations between nodes in the lineage tree. The maximum path length from root to any leave (maxRL) shows that the length of the tree increases over time but seems to stabilize at the end of the GC. maxRL corresponds to the maximum number of mutations acquired by specific subclones. Stabilization is expected since it becomes increasingly unlikely to have mutations that further increase the affinity once it has reached the maximum binding affinity for the Ag.
Chapter 5. The evolution of B-cell lineage trees during affinity maturation

Consequently, new mutations will decrease affinity and, therefore, the survival probability of this new subclone. The maximum number of mutations (maxRL) observed from the lineage tree is 11 and is in the same order of magnitude as the previously reported maximum number of mutations occurring during the GCR [133, 135]. Shahaf et al suggested association between maxRL and the selection threshold [104]. The average distance between root and any leave (avgRL) shows that the average number of mutations acquired by the subclones is about 5 at the end of the GCR. minRL (data not shown) is either 1, or 2 at the end of the GCR and therefore is not very informative. As expected, average distance between the root and any split node (avgRSN) is growing indicating the progress of the GCR.

Subclonal expansion and affinity in the context of lineage trees

Our simulations keeps track of the lineage tree but also of the abundance and affinity of all subclones at all time points during the GCR. This enables us to explore the dynam-
ics of subclonal expansion and affinity in the context of a lineage tree. We selected the same simulation as used for Figure 5.4. To create a representative but not overcrowded lineage tree we select all subclones with CC cell counts $\geq 10$ at the end of the GCR. Subsequently, we constructed the lineage tree from these subclones and their (unexpanded) ancestors and coloured each node according to their affinity and expansion (Figure 5.8). Effectively, this lineage tree corresponds to the tree shown in Figure 5.4 day 21 but with a sub-selection of subclones. However, because of the sub-selection criterion the resulted tree is smaller than on Figure 5.4 and is only maximum of 5 nodes long.

To identify expanded subclones we determined the threshold $T$ to be 17 cells. Consequently, all subclones with $> 17$ cells are expanded. Only ten subclones reached the expansion level at the end of the GCR. In general we observe less expansion at lower levels of the lineage tree.

Most of the branches of the tree show subclones with increasing affinity. However, few branches represent subclones with decreasing affinity as result of a SHM. As expected, subclones with low affinity extinct or have low cell counts at the end of the GCR and higher affinity subclones show increased cell counts or have been determined as expanded.

5.4. Discussion

In the current work we used a previously developed mathematical model of affinity maturation to investigate the evolution of B-cell lineage trees during affinity maturation. The model tracks individual subclones and, consequently, enables the construction of B-cell lineage trees. avgRL and maxRL demonstrate that lineage trees become longer during GCR while at the same time the outgoing degrees (rootOD, avgOD) decrease. Consequently, the lineage trees evolve from bushy trees to longer pruned trees. The total number of terminal subclones ($L$) and total subclones ($N$) stays relatively constant, around
This is in contrast to trees reconstructed and analysed by Dunn-Walters [141], based on experimental data from Jacob [139], where the number of leaves decreases. This could be a shortcoming of our computational model that does not incorporate the GC shutdown. Another possibility is that the previous study was based on a limited experimental data compared to currently available high-throughput repertoire sequencing data, which demonstrate a higher number of nodes per tree. Use of repertoire sequencing experimental data might provide a more fair comparison with our simulation results. Our simulations also demonstrate that the stochastic process of SHM is responsible for a lineage tree that largely vary in size. In contrast, the range of observed outgoing degrees becomes much smaller with proceeding affinity maturation.

We also explored subclone expansion and affinity maturation in the context of a B-cell lineage tree (Figure 5.8). The lineage tree included subclones that overcame competitors and reached relatively high cell counts at the end of the GCR. As expected, the tree demonstrates the advantage of high affinity subclones, particularly, high abundance or expansion is achieved by high affinity subclones. Noteworthy, the tree also illustrates that the affinity maturation is not necessarily a linear process and high affinity subclones may originate from not necessarily the highest affinity branch. That may be a result of SHM stochasticity, when a low affinity cell may potentially originate a high affinity subclone. Moreover, high cell counts of a subclone also do not guarantee the production of high affinity or highly expanded descendants. Further, high affinity subclones may not develop further but instead lose cell counts due to the high chance of lethal mutations.

For the best of our knowledge, such a representation of subclonal expansion and subclonal affinity in the context of a lineage tree was demonstrated for the first time. This is particularly important because information about subclonal affinity is virtually impossible to obtain from experimental data with current experimental technologies. First, the measurement of BCRs of all subclones at a specific time point during the GCR with repertoire sequencing [96, 158] would require microdissection of the GC and, consequently, would destroy the GC prohibiting further measurements. Alternatively, one could select different GCs at different time points as was done for a limited set of Ig sequences by Jacob [139] and Dunn-Walters [141] but this assumes a strong relationship and synchronicity between these GCs, which may not be true [98]. Furthermore, with current technologies it is impossible to measure the affinity of the BCR for the Ag for all subclones, which would also require that the Ag is known. To overcome the current experimental limitations in our work we successfully used a GCR model to obtain a graph representation of subclonal expansion with corresponding subclonal affinity in the context of a B-cell lineage tree.