

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

presented by

Michael Floßdorf, M.S. (Michigan State University)

born in Bonn

Date of oral-examination: Mai 13, 2013

Stochastic T cell fate decisions

Referees: Prof. Dr. Thomas Höfer
Prof. Dr. Roland Eils

Contents

Zusammenfassung	i
Summary	iii
1. Introduction	1
2. Statistical data analysis	5
2.1. Experimental fate mapping of individual T cells	5
2.2. CD8 ⁺ T cell response is composed of highly variable single cell progenies	6
2.3. Composing the 100 cell progenies out of single cell-derived progenies	9
2.4. Experimental variability in number of transferred cells	10
3. The N→TCMp→TEMP→TEF model accounts for stochasticity of CD8⁺ T cell diversification	17
3.1. Computing moments of the cell number distributions	17
3.2. The linear N→TCMp→TEMP→TEF model reproduces the observed correlations and variability	20
3.3. Asymmetry of cell numbers in the N→TCMp→TEMP→TEF model	23
3.4. Stochastic simulations of the N→TCMp→TEMP→TEF model . . .	25
3.5. Validated predictions of the N→TCMp→TEMP→TEF model . . .	28
3.6. Identifiability of model parameters	32
4. Systematic comparison of possible differentiation models	37
4.1. Alternative unidirectional models fail to reproduce the single cell progeny data	37
4.2. Probing a general class of differentiation models	38
4.3. Identifiability of the generalized differentiation model	43
4.4. Incorporating cell death into the model	45
4.5. Incorporating asymmetric cell division into the model	49
4.6. More refined asymmetric cell division models	52
5. Implications of the N→TCMp→TEMP→TEF model	61
5.1. Timing of early division and differentiation events determines progeny population size and phenotype	61
5.2. Probability for TCMP depletion	64
5.3. Robustness of T cell response	67
5.4. Distribution of the number of cell divisions over time	69

5.5. Latecomer T cells in the N→TCMP→TEMP→TEF model	72
5.6. Simulating antigen stimulus withdrawal	78
5.7. Simulating recall responses	80
6. Discussion	85
References	96
A. Appendix	97
A.1. ODE system for first and second moments of the general model	97
A.2. ODE system for first and second moments of general asymmetric cell division model	98
A.3. ODE system for third moments of N→TCMP→TEMP→TEF model	99
A.4. Modeling the latecomer cells by assuming reduced proliferation rates	100
Abbreviations	102
Acknowledgments	103

Zusammenfassung

T-Zellen bilden ein Modellsystem für die quantitative Analyse von Zellschicksalen. Während der durch eine Infektion ausgelösten Immunantwort proliferieren die Antigen-spezifischen naiven T-Zellen massiv und differenzieren sowohl in kurzlebige Effektorzellen, die das ursächliche Pathogen bekämpfen, als auch in langlebige Gedächtniszellen, die Schutz im Falle einer Reinfektion bieten. In der Literatur finden sich widersprüchliche Modelle darüber, wie Effektor- und Gedächtniszellen aus den stimulierten naiven T-Zellen hervorgehen. Ein prinzipielles Problem bei der Untersuchung von Differenzierungsprogrammen ist die Unterscheidung zwischen Differenzierung und Selektion durch differenzielle Zellproliferation oder Zelltod. Populationsmittelwerte alleine reichen nicht aus, um diese Unterscheidung treffen zu können. Basierend auf Messungen von Einzelzellnachkommenschaften entwickeln wir in dieser Arbeit einen neuartigen theoretischen Zugang zur Aufklärung der T-Zell-Differenzierungsprogramme. Dazu kombinieren wir die stochastische Modellierung der Zellschicksale mit umfangreicher Modelldiskriminierung.

Erstens finden wir, dass ein lineares Programm mit kontinuierlich abfallendem Potenzial der Zellen zu differenzieren, umfangreiche *in vivo*-Daten beschreiben kann. Naïve Zellen (N) differenzieren zunächst zu Vorläufern von zentralen Gedächtniszellen (TCMp), differenzieren dann weiter zu Vorläufern von Effektor-Gedächtniszellen (TEMp) und schließlich zu Effektorzellen (TEF). Alle Differenzierungs- und Proliferationsraten des Modells können anhand der Daten identifiziert werden.

Zweitens konstruieren wir systematisch alle möglichen Differenzierungsmodelle, die naïve, TCMP-, TEMp- und TEF-Zellen über Ratenübergänge miteinander verbinden, und erlauben Zelltyp-abhängige Proliferationsraten. Wir identifizieren den linearen Differenzierungspfad $N \rightarrow \text{TCMP} \rightarrow \text{TEMp} \rightarrow \text{TEF}$ als den minimalen Pfad, den ein Differenzierungsmodell enthalten muss, um einen akzeptablen Fit an die Daten zu ermöglichen. Darüber hinaus sind zusätzliche Übergänge zwischen den Zelltypen stark durch die Daten eingeschränkt. Das $N \rightarrow \text{TCMP} \rightarrow \text{TEMp} \rightarrow \text{TEF}$ -Modell sagt einen progressiven Anstieg der Proliferationsraten voraus: Langsam proliferierende TCMP-Zellen differenzieren zu schnell proliferierenden TEMp- und sich noch schneller teilenden TEF-Zellen. Es zeigt sich, dass diese Hierarchie von Proliferationsraten ausreichend ist, um die überraschende Variabilität der Immunantworten zu erklären, die von individuellen naiven T-Zellen initiiert werden. Wichtige Vorhersagen des linearen Differenzierungsmodells wurden in Anschlussexperimenten validiert.

Drittens zeigen wir, dass sich Zelltod und asymmetrische Zellteilung ebenfalls in unsere Modelle einbauen lassen. Diese Erweiterungen ändern keine der Schlussfolgerungen aus den Modellvergleichen. Es zeigt sich, dass asymmetrische Zellteilung mit den Daten vereinbar ist. Allerdings ist die Vorhersage des Modells, dass nur ein kleiner Anteil der Zellteilungen asymmetrisch ist.

Viertens zeigen wir, dass unser Modell wichtige vorhergegangene Experimente zur T-Zell-Differenzierung erklären kann und, an manchen Stellen, alternative Interpretationen der Daten ermöglicht. Insbesondere zeigen wir, wie man mit einem einzigen Modell die Entwicklung von Effektor-T-Zellen während der Primär- und der Sekundärimmunantwort beschreiben kann.

Wir identifizieren das Differenzierungsprogramm von zytotoxischen T-Zellen während der Immunantwort durch datenbasierte Modellierung. Das bessere quantitative Verständnis dieses Prozesses wird Auswirkungen auf Immuntherapie und die Entwicklung von effektiven Impfstrategien haben. Wir denken, dass unser Ansatz breite Anwendungsmöglichkeiten für die Untersuchung von Zellschicksalen in proliferierenden Zellpopulationen bietet.

Summary

T cells provide a paradigm model system for the quantitative study of cell-fate decisions. Upon infection, naïve antigen-specific T cells expand vigorously and give rise to populations of short-lived effector cells that fight the invading pathogen and long-lived memory cells that protect against reinfection. Conflicting models have been proposed on how effector and memory cells develop out of the pool of activated naïve cells. A principle difficulty for identifying the cell-fate specification program is to distinguish inductive differentiation from selection through differential cell proliferation or death; this distinction cannot be made on the basis of population averages. In this thesis, we develop a novel theoretical approach that utilizes single-cell fate mapping to identify the differentiation program of cytotoxic T cells. To this end, we combine stochastic modeling of cell-fate decisions with large-scale statistical model discrimination.

First, we observe that a linear program with continuously declining potential of the cells to differentiate fits extensive in-vivo data. Naïve cells (N) initially differentiate into precursors of central memory cells ($TCMp$), and then pass through an intermediate phenotype of effector memory precursors ($TEMp$) into an effector stage (TEF). Importantly, all differentiation and proliferation parameters of the model are identified from the data.

Second, we construct systematically all possible differentiation models that link naïve, $TCMp$, $TEMp$ and effector cells by differentiation rate transitions and allow for differential proliferation. We identify the linear pathway $N \rightarrow TC Mp \rightarrow TEMp \rightarrow TEF$ to form the unique minimal core module of the differentiation scheme. None of the models that lack this core module produces an acceptable fit to the data, and additional conversions between the subsets are strongly constrained. The $N \rightarrow TC Mp \rightarrow TEMp \rightarrow TEF$ model predicts a progressive increase in the proliferation rates with slowly proliferating central memory precursors giving rise to rapidly expanding effector memory precursors and even faster dividing effector cells. This hierarchy of proliferation rates suffices to explain the surprising variability in the immune responses that are derived from individual naïve progenitor cells. Critical predictions of the unidirectional differentiation model have been validated in subsequent experiments.

Third, we show that our modeling framework is flexible to incorporate further biological detail, including cell death and asymmetric cell division. These extensions do not alter the principal conclusion of the model comparisons. Moreover, the data are compatible with the coexistence of symmetric and asymmetric cell divisions but, interestingly, predict only a small fraction of the latter.

Fourth, we show that our model fully accounts for important previous experiments on T cell differentiation and, in some cases, suggests an alternative, parsimonious ex-

planation for the data. In particular, we provide a unifying model for the generation of effector T cells during primary and recall immune responses.

In summary, we identify the differentiation program of cytotoxic T cells during an immune response by data-based modeling. The better quantitative understanding of this process will have implications for immunotherapy and the design of effective vaccines. We believe that our theoretical framework will be of wider applicability to the study of cell-fate decisions in proliferating cell populations.

1. Introduction

Clonal selection of lymphocytes by their cognate antigen has arguably been the most far-reaching theory for the understanding of adaptive immunity [11, 31]. Its achievements are even more intriguing considering that the supporting experimental evidence became available only after the formulation of the theory by Niels Jerne (1955) and Frank M. Burnet (1957).

One of the first quantitative models based on the clonal selection theory was formulated by George I. Bell in 1970 [6]. This model describes the expansion of antigen specific lymphocytes by means of a population-dynamic mathematical model. The use of deterministic population dynamics to model the response of T and B lymphocytes has served as a paradigm in quantitative immunology to the present day.

More recent insight into the nature of adaptive immunity has been the realization that antigen-specific T and B lymphocytes not only expand massively during the course of an immune response but also diversify into distinct subsets that can be distinguished using phenotypic markers [47, 51]. A fundamental aspect of this diversification is the development of short-lived effectors killing the invading pathogen and long-lived memory cells providing protection against reinfection by the same pathogen. The population of memory T cells further sub-divides into functionally distinct subsets [37, 50, 51]. Central memory T cells recirculate through secondary lymphoid organs and proliferate vigorously upon re-challenge while effector memory T cells reside in peripheral organs and can mount immediate effector function following stimulation.

The last decade has seen an enormous advance in the elucidation of the molecular details underlying B and T cell fate decisions [37]. However, fundamental questions about the mechanisms that govern the differentiation of lymphocytes into functionally distinct subsets have remained unanswered. One important question in the T cell field concerns the developmental order in which long-lived central and effector memory cells together with short-lived effectors emerge out of the pool of activated naïve cells.

The “linear differentiation model” proposes that memory T cells develop out of effector cells that have been rescued from death by specific signals [60]. In this biological model, following infection, effector cells differentiate into effector memory cells which then give rise to a population of central memory cells. An alternative model, termed “decreasing potential model”, suggests the opposite developmental

1. Introduction

order with central memory precursors being the direct descendants of activated naïve cells [2]. The observation that the first cell division of the activated naïve cell could be asymmetric [14] has led to yet another model of T cell diversification in which the decision for either memory or effector fate gets determined already after the first cell division of the naïve cell.

Perhaps the most important reason for the persistence of these fundamental questions is the inadequacy of most current experimental tools to distinguish between molecular differentiation and cellular selection (Fig. 1.1). Suppose some arbitrary inheritable marker allows to distinguish between cells belonging to two distinct subsets. A population consisting of cells of both subtypes is then observed at two consecutive time points. An increase in the frequency of marker positive cells can

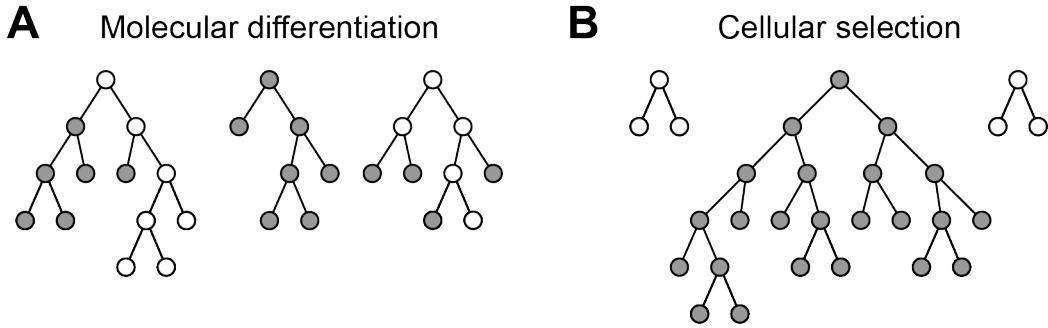


Figure 1.1.: Based on the population mean frequencies observed at two time points it is not possible to distinguish between molecular differentiation and cellular selection. (A) Molecular differentiation: During expansion, marker negative cells (white) differentiate into marker positive cells (gray). (B) Cellular Selection: No differentiation between the two subsets occurs, but marker positive cells (gray) proliferate faster than marker negative cells (white).

now have two mutually non-exclusive reasons. First, some of the cells in the dividing population up-regulate the respective marker and change their phenotype (“molecular differentiation”; Fig. 1.1A), or, second, the pre-existing marker positive cells divide faster than cells belonging to the other subset (“cellular selection”; Fig. 1.1B). Clearly, solely based on the mean number of marker positive and marker negative cells measured at different time points, it is not possible to distinguish between both cases. For deterministic population-dynamic models this implies inherent difficulties to estimate accurately both proliferation and differentiation rates.

In principal, monitoring the progeny of individual cells continuously over time would provide a means to distinguish between molecular differentiation and cellular decision [29]. However, such an approach is currently limited to *in vitro* experiments. One important exception to this is the usage of intravital two-photon microscopy to monitor T cell fates [7], but – compared to the time scale of the differentiation process – this technique is limited to rather short durations.

In this thesis we will show that an alternative, recently developed method yields rich information about T cell diversification [53]. In this method, a single genetically labelled naïve T cell is adoptively transferred into a host and cells belonging to its progeny population are recovered at a later time point during the immune response. The single cell-derived progeny populations can then be analyzed for phenotype and absolute cell numbers. In particular, the method allows to determine the absolute number of central memory and effector memory precursors as well as the absolute number of effector cells that emerged out of a single activated naïve T cell. We will develop a method to utilize such data to extract information about the T cell differentiation pathway as well as proliferation and differentiation rates of the emerging T cell subsets.

Importantly, this work has been done in intense collaboration with Dirk Busch and Veit Buchholz (TU Munich) who planned and conducted the experiments. In repeated iterations with experiments, model predictions were tested and new questions for the theoretical analysis were posed.

The second chapter provides an introduction to the experimental method and a statistical analysis of the single cell-progeny data. We place special emphasis on the unique features that distinguish these kind of data from data based on progenies derived from multiple precursors. The computational framework to efficiently estimate the parameters of rate-based stochastic differentiation models using the single cell progeny data is presented in chapter three. We then use this framework to study an unidirectional differentiation model in which, upon activation, the naïve cells differentiate into slowly proliferating central memory precursors (TCMp) which give rise to rapidly expanding effector memory precursors (TEMp) and effector (TEF) cells ($N \rightarrow \text{TCMp} \rightarrow \text{TEMp} \rightarrow \text{TEF}$ model). The fourth chapter contains a large scale model discrimination based on all possible differentiation models that link naïve, TCMP, TEMp and TEF cells by differentiation rate transitions. We further discuss asymmetric cell division models and present possible extensions of the $N \rightarrow \text{TCMp} \rightarrow \text{TEMp} \rightarrow \text{TEF}$ model. Various biological implications of the $N \rightarrow \text{TCMp} \rightarrow \text{TEMp} \rightarrow \text{TEF}$ model and important previous experiments on T cell differentiation are discussed in chapter five.

2. Statistical data analysis

Dirk Busch and colleagues have previously developed an experimental methodology to map the fate of individual naïve T cells during the immune response *in vivo* [53]. In this chapter we analyze a large data set that has been obtained through further improvements of this methodology by Veit Buchholz, Dirk Busch and colleagues [9].

The number of cells that develop from an individual naïve progenitor (“progeny size”) will turn out to be surprisingly variable. Moreover, the abundance of memory and effector phenotypes varies strongly from one progeny to another. We uncover interesting statistical features – correlations and coefficients of variation – in these single cell progeny data. Moreover, we show that this statistical information is no longer accessible when cell progenies are derived from more than one progenitor cell. Thus the single cell progeny data constitute a unique resource for dissecting the mechanisms that underlie T cell diversification.

2.1. Experimental fate mapping of individual T cells

Busch and coworkers developed an experimental approach that allows the transfer of a single T cell expressing a heritable congenic marker into a wild-type host [53]. Following infection, every descendant cell can be identified as being derived from one particular single ancestor. Recently, this technique was refined by the same laboratory (Dirk Busch, Veit Buchholz and colleagues) making it now possible to transfer and follow the progeny populations of up to eight distinguishable single cells in the same host. To this end, naïve CD8 T cells are obtained from eight C57BL/6 donor mice (OT-I SIINFEKL peptide-specific TCR transgenic) with different congenic backgrounds ranging from $CD90.1^{+/-} CD45.1^{+/-}$ to $CD90.2^{+/+} CD45.2^{+/+}$ (the congenic matrix; Fig. 2.1A).

In the usual setting, seven single OT-I SIINFEKL peptide-specific CD8⁺ T cells of matrix components A–G together with 100 cells of matrix component H are transferred to the $CD90.2^{+/+} CD45.2^{+/+}$ recipient mouse. Following infection with Listeria monocytogenes bacteria expressing the cognate antigen OVA (L.m.-OVA), the progeny populations of the transferred T cells are then recovered at day 6, day 8 or day 12 post infection (p.i.; Fig. 2.1B). If not indicated otherwise, the shown data are based on the recovered CD8⁺ T cells in the spleen.

2. Statistical data analysis

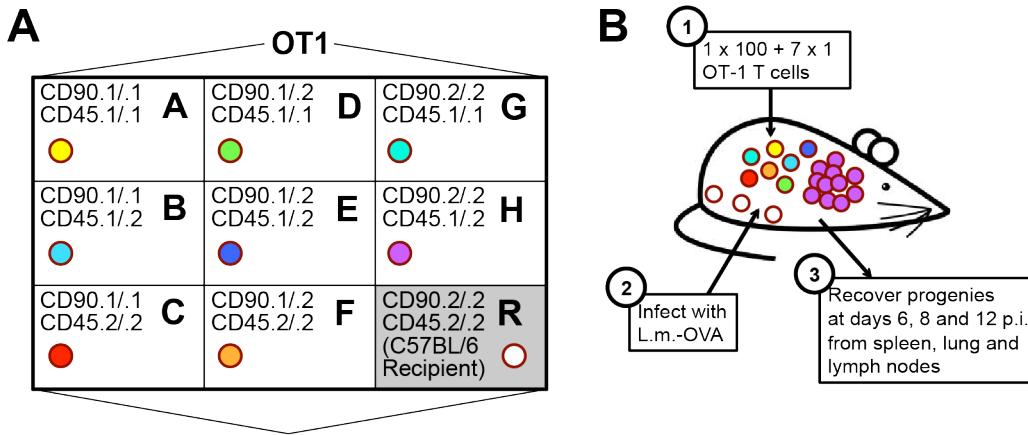


Figure 2.1.: Experimental setup. (A) The congenic matrix: The Matrix components A–H correspond to the eight different congenic backgrounds of the OT-I donor mice. Component R refers to the transgenic background of the recipient mouse. (B) Seven single cells of matrix components A - G and a control set of 100 cells of matrix component H are transferred to the recipient mouse. Subsequently, the recipient mice were intravenously infected with L.m.-OVA. At day 6, 8 or 12 post infection the progeny populations were recovered from spleen, lung and lymph nodes.

Importantly, the observed expansion and differentiation pattern of the 107 transferred OT-I cells closely resembles the endogenous SIINFEKL peptide-specific CD8⁺ T cells (not shown). Furthermore, the overall size of the responding endogenous SIINFEKL peptide-specific T cells at day 8 p.i. is not altered by the transferred OT-I cells thus excluding additional (unphysiological) competition between the T cells that could result from the artificial enlargement of the responding CD8⁺ T cell compartment. Adding to this argument, changing the transfer setting from “7x1 cell + 100 cells” to “8x1 cell” does not change the observed expansion and differentiation kinetics (not shown). The design of the experiment thus allows to study the response of individual CD8⁺ T cells upon activation *in vivo* under physiological conditions.

2.2. CD8⁺ T cell response is composed of highly variable single cell progenies

Single cell-derived progeny populations show a remarkable variability and asymmetry in both absolute cell number and phenotype (Fig. 2.2). At day 8 p.i., the absolute cell number varies over three orders of magnitude (Fig. 2.2A) with a coefficient of variation (CV; standard deviation divided by mean) of 2.2. In comparison, in a proliferation model with exponentially distributed cell division times and without cell-death the CV of the absolute size of single cell-derived progeny populations

2.2. CD8⁺ T cell response is composed of highly variable single cell progenies

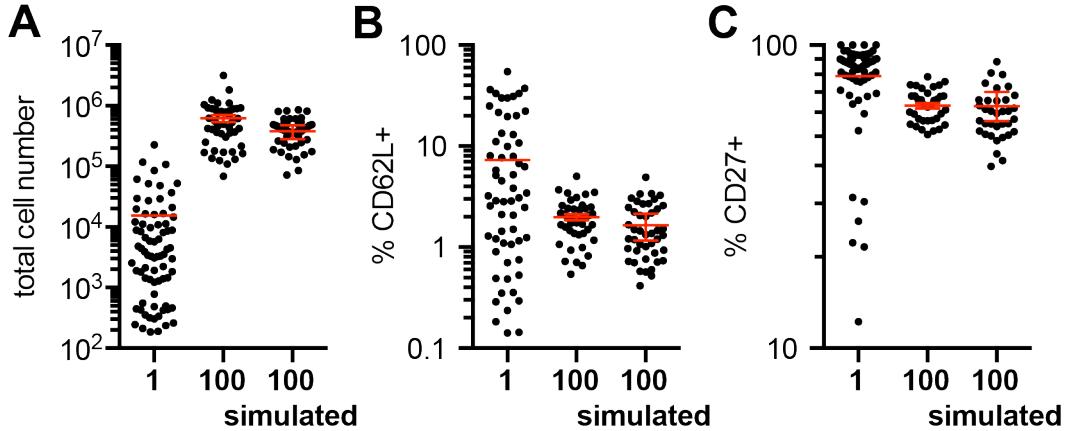


Figure 2.2.: Total cell numbers (A), CD27 (B) and CD62L (C) positive fractions of single and 100 cell-derived progeny populations (data and representative simulation of equal size) at day 8 p.i. The red line and error bars indicate mean and standard error of the mean (the bootstrap-based error of the predicted mean for “100 simulated” is based on many simulations). (Experiments by V. Buchholz and D. Busch)

would not exceed one [38].

The observed variability is accompanied by considerable asymmetry: Approximately 80% of the single cells lead to cell numbers well below the mean of ~15000 cells per spleen. This positive skewness leads to the fact that in an average response, half of the observed cells will be the descendants of only about 5% of the activated T cells (Fig. 2.3, black line). The resulting phenotype of a population of many activated CD8⁺ T cells will thus be dominated by the phenotype of the descendants of only a few heavily expanding single cells.

Importantly, the described variability and asymmetry is not due to differences in infection strength between the animals: Firstly, there is no evidence for differences in the median of the single cell-derived population sizes between different mice (Kruskal-Wallis test: $p = 0.55$). Secondly, normalizing each single cell progeny by the overall size of all OT-I matrix populations within the respective recipient does not reduce the observed asymmetry (Fig. 2.3, red dashed line).

Of particular interest is the composition of a population of responding T cells in terms of the percentage of terminal differentiated effectors and memory precursors. In this respect, two important surface molecules of CD8 T cells are the cell adhesion molecule CD62L (also known as L-selectin) and the tumor necrosis factor receptor CD27. In the present experimental setting, both markers together allow to define three different subsets of CD8⁺ T cells [34, 54]: CD62L⁺ CD27⁺ central memory T cell precursors (TCMp), CD62L⁻ CD27⁻ effector T cells (TEF) and CD62L⁻ CD27⁺ effector memory T cell precursors (TEMp). For the single cell-derived progenies, the positive fractions of both surface markers within the population show extreme vari-

2. Statistical data analysis

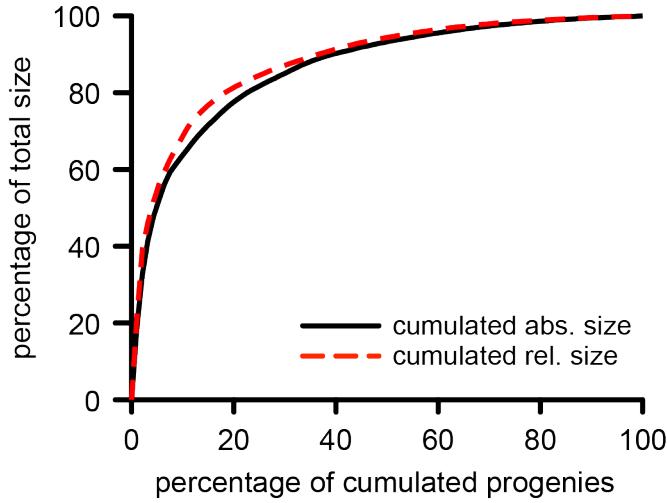


Figure 2.3.: Plot shows the percentage of total size plotted against the percentage of cumulated single cell-derived progenies (black line; ordered by size). Red dashed line: Same but after normalizing each single cell progeny by the overall size of all OT-I matrix populations within the respective recipient. (Experiments by V. Buchholz and D. Busch)

ability (Fig. 2.2B,C) and are inversely correlated with the population's total size (Fig. 2.4). Consequently, most of the single cell-derived progenies show a substantially higher fraction of memory precursors than observed on the population level (cf. Fig. 2.2B,C).

For later modeling, the pairwise correlations between the three subsets TCMP, TEMp and TEF will be of particular interest (Fig. 2.5). Although every population is derived from only one responding T cell, a substantial correlation between the subset sizes is only observed for TEMp vs. TEF (Fig. 2.5C). The red line in Fig. 2.5C represents a log-log regression line and, if cell death is negligible, can thus be interpreted as the relation between the number of divisions of both subsets. The slope of the regression line being not significantly different from one (95% confidence interval: (0.94, 1.24)) then tells us that both subsets have comparable proliferation rates. Extrapolating the regression line to the absolute size of one for the TEF compartment further provides an indication of the subset differentiation order being TEMp → TEF rather than vice versa. The size of both subsets on day 8 p.i. would then highly depend on the timing of the first TEMp emergence.

In chapter 3 we will formalize the above arguments and show that a slowly proliferating TCMP subset together with rare TCMP → TEMp transitions naturally generate the necessary temporal variability of the first TEMp differentiation event while at the same time explaining the low correlations between TCMP and the other subsets. Furthermore, we will see that it will prove appropriate to summarize this

2.3. Composing the 100 cell progenies out of single cell-derived progenies

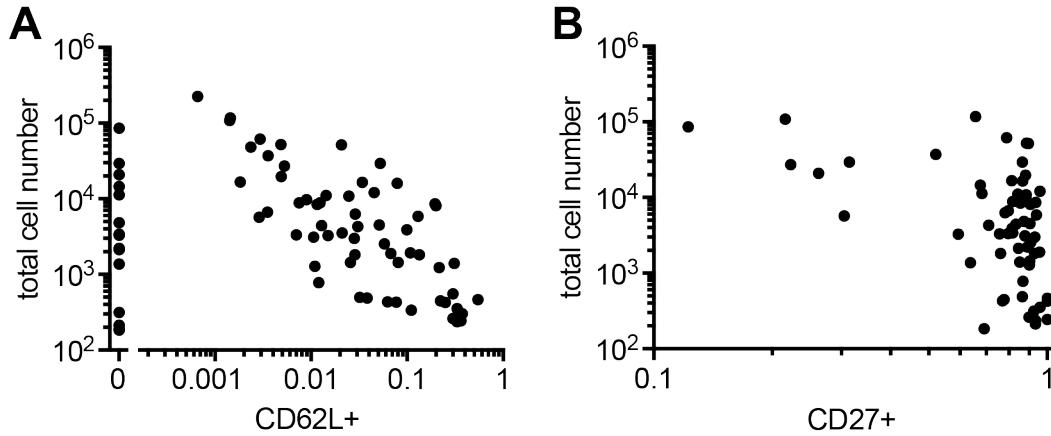


Figure 2.4.: Correlation between absolute cell number and fraction of (A) CD62L⁺ and (B) CD27⁺ of single cell-derived progenies at day 8 p.i. (spleen). Indicates that memory phenotype (CD62L⁺, CD27⁺) dominates in small progenies. (Experiments by V. Buchholz and D. Busch)

“statistical footprint” of the differentiation and proliferation histories in terms of the mean values, the CVs and the pairwise correlations of the three phenotypes (Fig. 2.6).

The statistical properties of the single cell-derived progenies at day 12 p.i. are very similar to the ones at day 8 p.i. (Table 2.1 and Fig. 2.7). However, the mean absolute sizes of all three subsets decline between day 8 and day 12 p.i. The peak of the response is expected to be around day 8 p.i. and afterwards cell death is expected to play a significant role [61].

2.3. Composing the 100 cell progenies out of single cell-derived progenies

At day 8 p.i., descendants were recovered from only 25% of the transferred single cells. We will now show that the 100 cell-derived progeny populations scale accordingly and are built up of about 25 responding single cells in average. This is to be expected since all 107 cells are transferred in one go.

In order to simulate the absolute numbers of descendants derived from 100 transferred cells we summed up the absolute cell numbers of randomly drawn single cell progenies (Fig. 2.2: “100 simulated”). The number of drawn single cell-derived populations was taken to be binomially distributed with $n = 100$ and $p = 0.25$ (sampled with replacement). In order to assess the predicted mean absolute cell number and its uncertainty for the “100 simulated”, the single cell measurements were resam-

2. Statistical data analysis

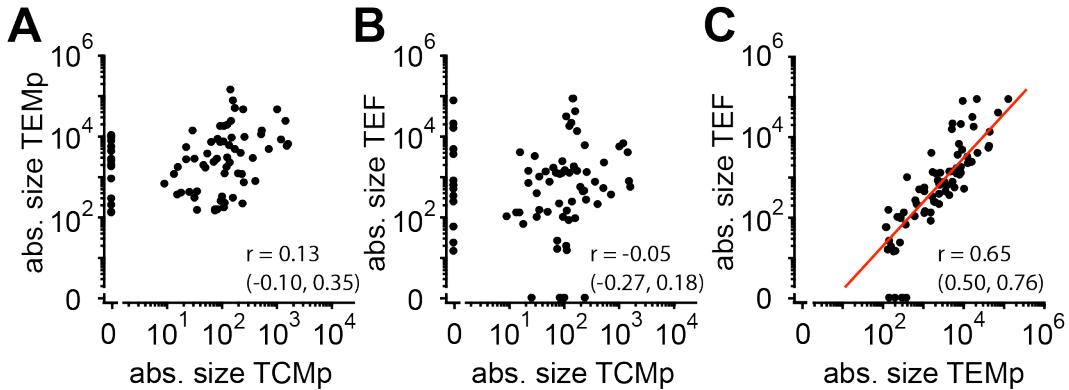


Figure 2.5.: (A-C) Pairwise correlations between the TCMp, TEMp and TEF subset sizes of the single cell-derived progenies at day 8 p.i. Linear correlation coefficients are indicated together with the 95% confidence intervals. Red line in C: log-log regression line. (Experiments by V. Buchholz and D. Busch)

pled using a non-parametric bootstrap approach ($n = 1000$) and each bootstrap sample was then used to generate a large number (10000) of 100 cell-derived progenies (cf. predicted mean and error shown in Fig. 2.2). A similar approach was used to simulate the percentage of CD62L and CD27 positive cells in the 100 cell-derived population: The randomly drawn single cell progeny population attributes (percentage of CD62L or CD27 positive cells) were weighted according to the corresponding absolute cell numbers (cf. Fig. 2.5) and then summed. The predicted mean values are found in good agreement with the measurements. Repeating this analysis with the data of day 12 p.i. leads to the same result (Fig. 2.7).

Interestingly, comparison with the endogenous pool of responding SIINFEKL-specific CD8⁺ T cells shows that the 100 cell-derived population size is in average about four fold smaller (data not shown), meaning that the endogenous repertoire of activated specific CD8⁺ T cells comprises around 100 cells.

2.4. Experimental variability in number of transferred cells

While the mean number of total cells and the mean fraction of CD62L and CD27 positive cells for the 100 cell-derived population is adequately modeled by assuming a constant probability (25%) of T cell recruitment (cf. Figs. 2.2 and 2.7), the variability in total cell number is under-estimated by the simulation: The measured coefficient of variation for total size of the 100 cell-derived progenies is $CV_{100} = (0.83 \pm 0.15)$ while the resulting variability of the simulation $CV_{\text{sim}} = (0.45 \pm 0.06)$ is much smaller. If the number of drawn and added single cells n in the simulation is constant, the CV of the simulation scales as $1/\sqrt{n}$ (upon adding n independent random

2.4. Experimental variability in number of transferred cells

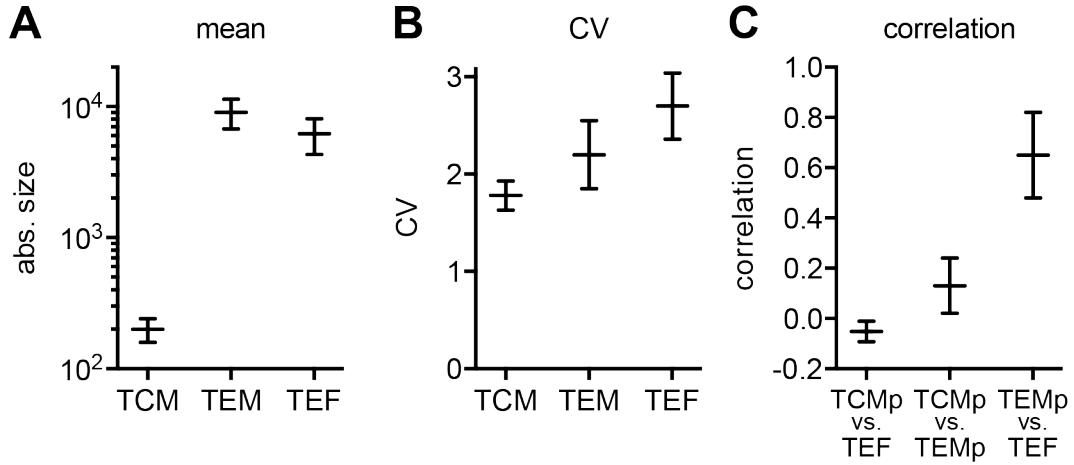


Figure 2.6.: Summary statistics of the single cell-derived progeny populations at day 8 p.i.: Mean values (A), coefficients of variations (B) and the pairwise correlations (C) of the three subsets (cf. Table 2.1). (Experiments by V. Buchholz and D. Busch)

variables both mean and variance of the sum are proportional to n). The coefficient of variation for the single cell progenies is $CV_{\text{single}} = (2.06 \pm 0.30)$. Hence, if a constant number of $n = 25$ of the transferred 100 T cells were recruited into the immune response, a coefficient of variation of $CV_{100}^{\text{exp}} = (0.41 \pm 0.06)$ would be expected for the total size of the 100 cell-derived population. Note that the small difference between CV_{100}^{exp} and CV_{sim} is due to the assumed binomial distribution of added single cells in the simulation (cf. chapter 2.3).

Obviously there is an additional source of variability. In this context note that the recruitment of successfully transferred T cells into the immune response has been shown to be almost complete [57] (cited study: 5000 transferred cells; same infection model). It is thus likely that the number of successfully transferred and surviving cells itself presents a major source of variability. In any case the number of T cells R that eventually become recruited (i.e. activated) upon transfer of 100 cells and subsequent infection represents a random variable with a mean of approximately 25 cells.

In the following we will first calculate the coefficient of variation of R . Then we will show that a non-constant R introduces correlations between the TCMp, TEMp and TEF subsets in the 100 cell-derived progenies that do not represent the true relations between the populations. It is primarily for this reason that the 100 cell progeny data cannot be used to draw conclusions on the individual cell proliferation and differentiation histories.

Let now S_i and H represent the random variables of the total cell number of single

2. Statistical data analysis

Quantity	day 8 p.i.	day 12 p.i.
mean total cell number	(15500 \pm 3800)	(3860 \pm 980)
mean number of TCMp cells	(200 \pm 41)	(100 \pm 14)
mean number of TEMp cells	(9100 \pm 2300)	(1830 \pm 410)
mean number of TEF cells	(6200 \pm 1900)	(1940 \pm 610)
CV of total cell number	(2.06 \pm 0.30)	(2.32 \pm 0.34)
CV of TCMp cells	(1.78 \pm 0.15)	(1.30 \pm 0.13)
CV of TEMp cells	(2.20 \pm 0.40)	(2.05 \pm 0.34)
CV of TEF cells	(2.72 \pm 0.34)	(2.97 \pm 0.40)
correlation cell number of TCMp vs. TEF	(-0.05 \pm 0.04)	(0.020 \pm 0.050)
correlation cell number of TCMp vs. TEMp	(0.13 \pm 0.11)	(0.063 \pm 0.060)
correlation cell number of TEMp vs. TEF	(0.65 \pm 0.17)	(0.77 \pm 0.20)
skewness of total cell number	(3.56 \pm 0.89)	(4.5 \pm 1.0)

Table 2.1.: Statistical quantities for the single cell-derived progenies at day 8 and day 12 p.i. and bootstrap estimates of the corresponding uncertainties.

cell-derived and 100 cell-derived populations, respectively. It then follows that

$$H = \sum_{i=1}^R S_i. \quad (2.1)$$

We use the law of iterated expectations

$$\mathbb{E}[X] = \sum_r p(r) \mathbb{E}[X | R = r] = \mathbb{E}[\mathbb{E}[X | R]] \quad (2.2)$$

and the law of total variance

$$\text{var}(X) = \mathbb{E}[\text{var}(X | R)] + \text{var}(\mathbb{E}[X | R]) \quad (2.3)$$

to find

$$\mathbb{E}[H] = \mathbb{E}[R] \mathbb{E}[S_i] \quad (2.4)$$

and

$$\text{var}(H) = \mathbb{E}[R] \text{var}(S_i) + \text{var}(R) \mathbb{E}[S_i]^2. \quad (2.5)$$

Using $\mathbb{E}[R] \approx 25$ and the sample means and variances estimated from the day 8 p.i. data then results in $\text{CV}_R = (1.36 \pm 0.57)$ (error estimation via non-parametric bootstrap).

Note that eqn. 2.4 implies that the relative mean subset sizes of the 100 cell-derived populations TCMp^H , TEMp^H and TEF^H are equal to the mean relative sizes in the

2.4. Experimental variability in number of transferred cells

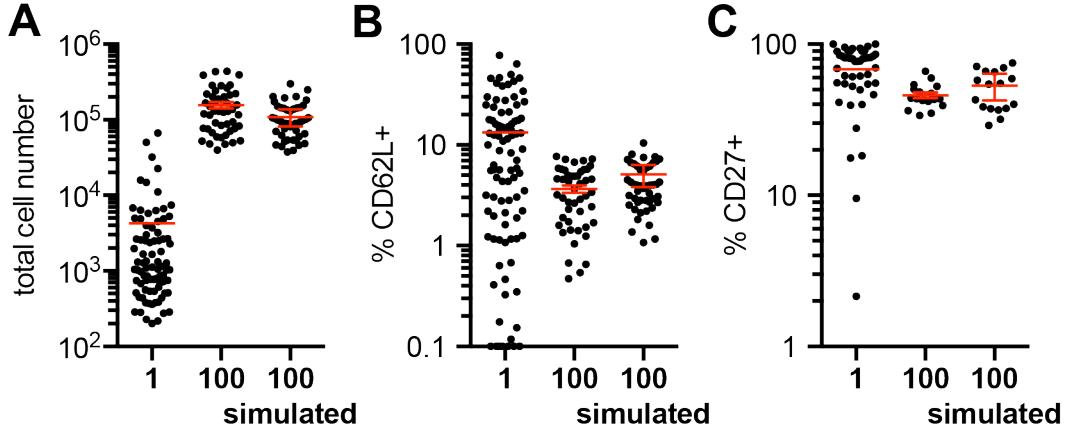


Figure 2.7.: Total cell numbers (A), CD27 (B) and CD62L (C) positive fractions of single and 100 cell-derived progeny populations (data and representative simulation of equal size) at day 12 p.i.. The red line and error bars indicate mean and error of the mean (the bootstrap-based error of the predicted mean for “100 simulated” is based on many simulations). (Experiments by V. Buchholz and D. Busch)

single cell-derived populations:

$$\frac{E[TCMp^H]}{E[H]} = \frac{E[TCMp]}{E[S_i]}, \quad \frac{E[TEMp^H]}{E[H]} = \frac{E[TEMp]}{E[S_i]}, \quad \frac{E[TEF^H]}{E[H]} = \frac{E[TEF]}{E[S_i]}. \quad (2.6)$$

Next we calculate the relation between the pairwise correlations $\text{corr}(S_i^1, S_i^2)$ of the three subsets of the single cell-derived progenies (indicated by S_i^1, S_i^2, S_i^3) and the 100 cell-derived progenies $\text{corr}(H^1, H^2)$ (subsets indicated by H^1, H^2, H^3). We note that

$$\begin{aligned} E\left[\sum_{i=1}^R S_i^1 \cdot \sum_{j=1}^R S_j^2\right] &= \sum_{r=1}^{100} p(r) E\left[\sum_{i=1}^R S_i^1 \cdot \sum_{j=1}^R S_j^2\right] \\ &= \sum_{r=1}^{100} p(r) \sum_{i=1}^r \sum_{j=1}^r E[S_i^1 S_j^2] \\ &= \sum_{r=1}^{100} p(r) \sum_{i=1}^r \sum_{j=1}^r (\text{cov}(S_i^1, S_j^2) + E[S_i^1] E[S_j^2]) \\ &= \sum_{r=1}^{100} r p(r) \text{cov}(S_i^1, S_i^2) + \sum_{r=1}^{100} r^2 p(r) E[S_i^1] E[S_i^2] \\ &= E[R] \text{cov}(S_i^1, S_i^2) + E[R^2] E[S_i^1] E[S_i^2], \end{aligned} \quad (2.7)$$

where we used eqn. 2.2 and the linearity of the expectation operator. With the help

2. Statistical data analysis

of this equation we now calculate

$$\begin{aligned}
 \text{corr}(H^1, H^2) &= \frac{1}{C} \text{cov} \left(\sum_{i=1}^R S_i^1, \sum_{j=1}^R S_j^2 \right) \\
 &= \frac{1}{C} \left(E \left[\sum_{i=1}^R S_i^1 \cdot \sum_{j=1}^R S_j^2 \right] - E \left[\sum_{i=1}^R S_i^1 \right] E \left[\sum_{j=1}^R S_j^2 \right] \right) \\
 &= \frac{1}{C} \left(E[R] \text{cov}(S_i^1, S_i^2) + E[R^2] E[S_i^1] E[S_i^2] - E[R]^2 E[S_i^1] E[S_i^2] \right) \\
 &= \frac{1}{C} \left(E[R] \sqrt{\text{var}(S_i^1) \text{var}(S_i^2)} \text{corr}(S_i^1, S_i^2) \right. \\
 &\quad \left. + E[S_i^1] E[S_i^2] \text{var}(R) \right), \tag{2.8}
 \end{aligned}$$

with

$$C = \sqrt{E[R] \text{var}(S_i^1) + \text{var}(R) E[S_i^1]^2} \sqrt{E[R] \text{var}(S_i^2) + \text{var}(R) E[S_i^2]^2}.$$

As expected, the above equation reduces to $\text{corr}(H^1, H^2) = \text{corr}(S_i^1, S_i^2)$ for $\text{var}(R) = 0$. Eqn. 2.8 enables us to calculate the expected values for the three pairwise correlations of the 100 cell-derived population which are found to be in agreement with the observed correlations (Fig. 2.8).

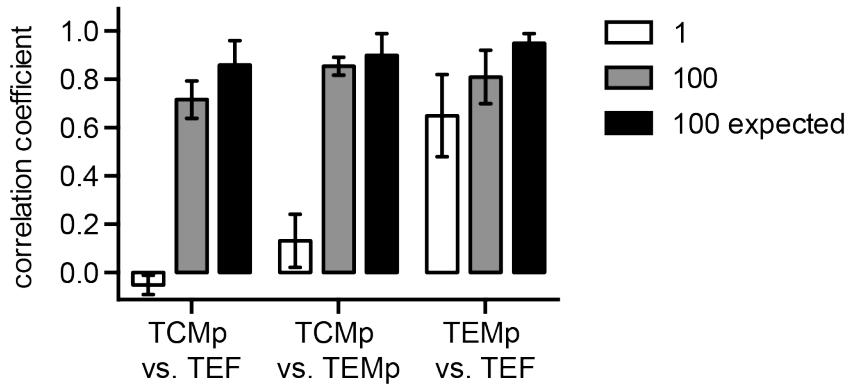


Figure 2.8.: Pairwise correlations between the absolute cell numbers of the subsets at day 8 p.i. for the single cell-derived progenies (white bars) and the 100 cell-derived progenies (grey bars: data; black bars: calculated based on single cell data and recruitment variability).

We thus identified non-constant recruitment of transferred cells into the mouse to be a confounding factor that leads to considerable information loss once a whole population of cells (e.g. 100 cells) is transferred. This finding most likely applies

2.4. Experimental variability in number of transferred cells

also to other experimental settings in which more than a single T cell is transferred and where the number of initially recruited T cells cannot be controlled.

To summarize, without making use of the single cell data, neither the mean number of recruited cells nor its variability can be inferred. This in turn precludes the estimation of both the true correlations between the total cell numbers of the subsets and their respective variability. In the next chapter we will see that these quantities are fundamental for identification of the underlying differentiation process. This makes the single cell-derived progeny data so valuable for the mathematical modeling that we will discuss in the next chapters.

3. The N→TCMp→TEMp→TEF model accounts for stochasticity of CD8⁺ T cell diversification

We will develop a computational framework that allows us to calculate efficiently the summary statistics of single cell-derived progenies from an underlying stochastic model of T cell diversification. We then use this framework to show that the single cell progenies data contain a “statistical footprint” of the individual proliferation and differentiation history. Specifically, we will find that these data suffice to identify the parameters of a model in which differentiation follows the pathway N→TCMp→TEMp→TEF. Based on the estimated parameters will then show that this unidirectional differentiation model accounts both for the observed variability and asymmetry as well as for the correlations in absolute cell number that are observed between the T cell subsets. Moreover, we will discuss various model predictions that were validated in subsequent experiments.

3.1. Computing moments of the cell number distributions

We model the proliferation and differentiation dynamics of the T cells by a standard Markov process, described in terms of the master equation [21]. In order to calculate the time evolution of the moments of the cell number distributions given such a differentiation model, we will first show how to translate the respective master equation into a partial differential equation (PDE) for the probability generating function. In general, for a system with k state variables n_1, \dots, n_k , the probability generating function F is defined as

$$F(z_1, \dots, z_k, t) = \sum_{n_1, \dots, n_k} z_1^{n_1} \cdot \dots \cdot z_k^{n_k} P(n_1, \dots, n_k, t), \quad (3.1)$$

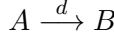
where $P(n_1, \dots, n_k, t)$ denotes the probability to find the system in state (n_1, \dots, n_k) at time t .

The master equation of stochastic differentiation models can be composed of elementary steps for cell proliferation and differentiation that we will now discuss and translate term by term into the corresponding part of the PDE for the probability generating function. Although later models will dependent on four states

3. The N→TCMp→TEMp→TEF model accounts for stochasticity of CD8⁺ T cell diversification

(the naïve, TCMP, TEMP and TEF subsets), it will be sufficient to consider the elementary steps in one- or two-state systems.

Cell differentiation: Differentiation from compartment A to compartment B (containing n_1 and n_2 cells, respectively) at a rate d corresponds to the reaction scheme



and leads to the following master equation

$$\dot{P}(n_1, n_2) = d [(n_1 + 1) P(n_1 + 1, n_2 - 1) - n_1 P(n_1, n_2)]. \quad (3.2)$$

The corresponding part of the PDE for the probability generating function is now obtained by multiplying both sides of this equation with $z_1^{n_1} z_2^{n_2}$ and subsequent summation. Using

$$\begin{aligned} \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} z_1^{n_1} z_2^{n_2} n_1 P(n_1, n_2) &= z_1 \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} n_1 z_1^{n_1-1} z_2^{n_2} P(n_1, n_2) \\ &= z_1 \partial_{z_1} F \end{aligned}$$

and

$$\begin{aligned} \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} z_1^{n_1} z_2^{n_2} (n_1 + 1) P(n_1 + 1, n_2 - 1) &= \frac{z_2}{z_1} \sum_{n'_1=0}^{\infty} \sum_{n'_2=0}^{\infty} z_1^{n'_1} z_2^{n'_2} n'_1 P(n'_1, n'_2) \\ &= \frac{z_2}{z_1} z_1 \partial_{z_1} F \\ &= z_2 \partial_{z_1} F \end{aligned}$$

then enables us to reformulate eqn. 3.2 as a PDE for F

$$\dot{F}(z_1, z_2) = d(z_2 - z_1) \partial_{z_1} F(z_1, z_2). \quad (3.3)$$

Similarly we find:

Cell death:

$$A \xrightarrow{\delta} \emptyset \Rightarrow \dot{F}(z_1) = \delta(1 - z_1) \partial_{z_1} F(z_1, z_2) \quad (3.4)$$

Symmetric cell division:

$$A \xrightarrow{\lambda} A + A \Rightarrow \dot{F}(z_1) = \lambda(z_1^2 - z_1) \partial_{z_1} F(z_1, z_2) \quad (3.5)$$

Asymmetric cell division:

$$A \xrightarrow{d} A + B \Rightarrow \dot{F}(z_1, z_2) = d(z_2 z_1 - z_1) \partial_{z_1} F(z_1, z_2) \quad (3.6)$$

3.1. Computing moments of the cell number distributions

Differentiation upon division:

$$A \xrightarrow{d} B + B \Rightarrow \dot{F}(z_1, z_2) = d(z_2^2 - z_1) \partial_{z_1} F(z_1, z_2). \quad (3.7)$$

Note that analogous expressions can also be used to quantify transcriptional and translational noise in gene regulatory networks [52, 56].

As can be seen from eqn. 3.1, expressions for moments of the probability distribution P can be obtained by evaluating the partial derivatives of F at $z_i = 1$, for all i . For the first three partial derivatives this results in

$$\begin{aligned} \partial_{z_i} F|_1 &= \langle n_i \rangle \\ \partial_{z_i}^2 F|_1 &= \langle n_i (n_i - 1) \rangle \\ \partial_{z_i} \partial_{z_j} F|_1 &= \langle n_i n_j \rangle, \text{ for } i \neq j, \\ \partial_{z_i}^3 F|_1 &= \langle n_i (n_i - 1)(n_i - 2) \rangle \\ \partial_{z_i}^2 \partial_{z_j} F|_1 &= \langle n_i (n_i - 1)n_j \rangle, \text{ for } i \neq j, \\ \partial_{z_i} \partial_{z_j} \partial_{z_k} F|_1 &= \langle n_i n_j n_k \rangle, \text{ for } i \neq j, j \neq k, i \neq k, \end{aligned} \quad (3.8)$$

where $\langle \cdot \rangle$ denotes the expectation operator.

From eqns. 3.8 we also obtain

$$\begin{aligned} \langle n_i^2 \rangle &= \partial_{z_i}^2 F|_1 + \partial_{z_i} F|_1 \\ \langle n_i^3 \rangle &= \partial_{z_i}^3 F|_1 + 3\partial_{z_i}^2 F|_1 + \partial_{z_i} F|_1 \\ \langle n_i^2 n_j \rangle &= \partial_{z_i}^2 \partial_{z_j} F|_1 + \partial_{z_i} \partial_{z_j} F|_1. \end{aligned} \quad (3.9)$$

We thus find the following expressions for the mean values (μ_i), variances ($\text{var}(n_i)$), coefficients of variation (CV_i) and the pairwise correlations ρ_{ij} in terms of the derivatives of the probability generating function:

$$\mu_i = \partial_{z_i} F|_1 \quad (3.10)$$

$$\text{var}(n_i) = \partial_{z_i}^2 F|_1 + \partial_{z_i} F|_1 - (\partial_{z_i} F|_1)^2 \quad (3.11)$$

$$\text{CV}_i = \frac{\sqrt{\partial_{z_i}^2 F|_1 + \partial_{z_i} F|_1 - (\partial_{z_i} F|_1)^2}}{\partial_{z_i} F|_1} \quad (3.12)$$

$$\rho_{ij} = \frac{\partial_{z_i} \partial_{z_j} F|_1 - \partial_{z_i} F|_1 \partial_{z_j} F|_1}{\sqrt{\text{var}(n_i)} \sqrt{\text{var}(n_j)}}, \text{ for } i \neq j. \quad (3.13)$$

Analogously, we can calculate the skewnesses γ_i (measure of asymmetry) which can also be defined in terms of the moments of the distribution

$$\gamma_i = \frac{\langle n_i^3 \rangle - 3\langle n_i \rangle \langle n_i^2 \rangle + 2\langle n_i \rangle^3}{\left(\langle n_i^2 \rangle - \langle n_i \rangle^2 \right)^{3/2}} \quad (3.14)$$

$$= \frac{\partial_{z_i} F|_1 + 3\partial_{z_i}^2 F|_1 + \partial_{z_i}^3 F|_1 - 3\langle n_i \rangle \text{var}(n_i) - \langle n_i \rangle^3}{\text{var}(n_i)^{3/2}}. \quad (3.15)$$

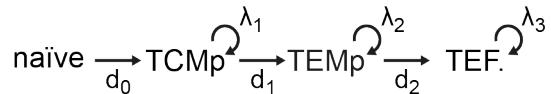
3. The $N \rightarrow TCMP \rightarrow TEMP \rightarrow TEF$ model accounts for stochasticity of CD8⁺ T cell diversification

The time evolution of these quantities can be computed as follows: Successively differentiating the PDE for the probability generating function (composed of the terms given in eqns. 3.3 to 3.7) with respect to the z_i and then setting $z_i = 1$ yields a system of linear ordinary differential equations (ODEs) of first order for the time evolution of the left hand sides of eqns. 3.9. For the four state systems that we will consider in the next sections, we will see that this results in a system of four coupled ODEs for the mean values (first moments), a 14-dimensional system of ODEs for the first two and a 34-dimensional ODE system for the first three moments. Finally, the initial conditions for the ODE system are obtained from the initial conditions of the master equation by evaluating the partial derivatives of eqn. 3.1 at $z_i = 1$ and setting $t = 0$.

This treatment allows us to compute the moments of the time-dependent distribution $P(n_1, \dots, n_k, t)$ much more efficiently than with stochastic simulations using Gillespie's algorithm – namely through solving systems of linear ODEs. In turn, this method enables us to fit diversification models to the data on a large scale. Utilizing the estimated parameters we can subsequently use stochastic simulations to compute model predictions that go beyond the moments.

3.2. The linear $N \rightarrow TCMP \rightarrow TEMP \rightarrow TEF$ model reproduces the observed correlations and variability

Using the framework introduced in the previous section we can immediately write down PDEs for the probability generating functions of possible differentiation models. Specifically, let us consider the unidirectional differentiation model that is given by the scheme



Here, the state vector (n_0, n_1, n_2, n_3) is formed by the number of naïve (n_0), TCMP (n_1), TEMP (n_2) and TEF (n_3) cells. In this model, the naïve T cell gets recruited into the immune response with the rate d_0 . Biologically, recruitment here subsumes migration of the naïve T cell into secondary lymphoid organs and the subsequent activation by an antigen presenting cell. The model makes no explicit assumptions about the stimuli (cytokine, T cell receptor stimuli etc.) that drive the differentiation (rates d_i) and division events (rates λ_i) and just assumes that both processes can be modeled by Markovian rate processes. Note that the model does allow for subset specific proliferation rates.

Using eqns. 3.3 and 3.5, the resulting PDE for the probability generating function

3.2. The linear N→TCMp→TEMp→TEF model reproduces the observed correlations and variability

is readily found to be

$$\dot{F}(z_0, z_1, z_2, z_3, t) = \sum_{i=1}^3 \lambda_i (z_i^2 - z_i) \partial_{z_i} F + \sum_{i=0}^2 d_i (z_{i+1} - z_i) \partial_{z_i} F. \quad (3.16)$$

Since we aim at modeling the T cell responses only at and before the peak of the immune response, we can neglect cell death in this model [22]. However, we will come back to this question in chapter 4.

To identify whether this model is able to account for the observed subset diversification pattern we focus on the single cell-derived progeny data at day 8 post infection (p.i.). It turns out that the mean values, CVs and the pairwise correlation coefficients (CCs) of the three subsets at this time point form appropriate summary statistics to quantify the “statistical footprint” (cf. chapter 2.2) that is left by the differentiation and proliferation history of the single cell progenies. As described in the previous section, in order to calculate the mean values, CVs and CCs in the model, we first successively differentiate eqn. 3.16 with respect to the z_i and then set $z_i = 1$. Using $\langle n_i \rangle = \partial_{z_i} F|_1$, this results in the following ODE system for the time dependence of the mean values

$$\begin{aligned} \dot{\langle n_0 \rangle} &= -d_0 \langle n_0 \rangle \\ \dot{\langle n_1 \rangle} &= (\lambda_1 - d_1) \langle n_1 \rangle + d_0 \langle n_0 \rangle \\ \dot{\langle n_2 \rangle} &= (\lambda_2 - d_2) \langle n_2 \rangle + d_1 \langle n_1 \rangle \\ \dot{\langle n_3 \rangle} &= d_2 \langle n_2 \rangle + \lambda_3 \langle n_3 \rangle, \end{aligned} \quad (3.17)$$

and the following ODEs for the time evolution of the second derivatives of F evaluated at $z_i = 1$

$$\begin{aligned} \dot{F}_{00} &= -2d_0 F_{00} \\ \dot{F}_{03} &= (\lambda_3 - d_0) F_{03} + d_2 F_{02} \\ \dot{F}_{33} &= 2(d_2 F_{23} + \lambda_3 F_{33} + \langle n_3 \rangle) \\ \dot{F}_{02} &= d_1 F_{01} - (d_0 + d_2 - \lambda_2) F_{02} \\ \dot{F}_{23} &= (-d_2 + \lambda_2 + \lambda_3) F_{23} + d_1 F_{13} + d_2 F_{22} \\ \dot{F}_{22} &= 2(d_1 F_{12} - d_2 F_{22} + \lambda_2 F_{22} + \langle n_2 \rangle) \\ \dot{F}_{01} &= d_0 F_{00} - (d_0 + d_1 - \lambda_1) F_{01} \\ \dot{F}_{13} &= (-d_1 + \lambda_1 + \lambda_3) F_{13} + d_0 F_{03} + d_2 F_{12} \\ \dot{F}_{12} &= (-d_1 - d_2 + \lambda_1 + \lambda_2) F_{12} + d_0 F_{02} + d_1 F_{11} \\ \dot{F}_{11} &= 2((\lambda_1 - d_1) F_{11} + d_0 F_{01} + \lambda_1 \langle n_1 \rangle), \end{aligned} \quad (3.18)$$

where we have used the short notation $F_{ij} = \partial_{z_i} \partial_{z_j} F|_1$. As initial conditions we find

$$\begin{aligned} \langle n_0 \rangle(0) &= 1 \\ \langle n_i \rangle(0) &= 0, \text{ for } i = 1, \dots, 3 \\ F_{ij}(0) &= 0, \text{ for } i, j = 0, \dots, 3. \end{aligned} \quad (3.19)$$

3. The N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model accounts for stochasticity of CD8 $^{+}$ T cell diversification

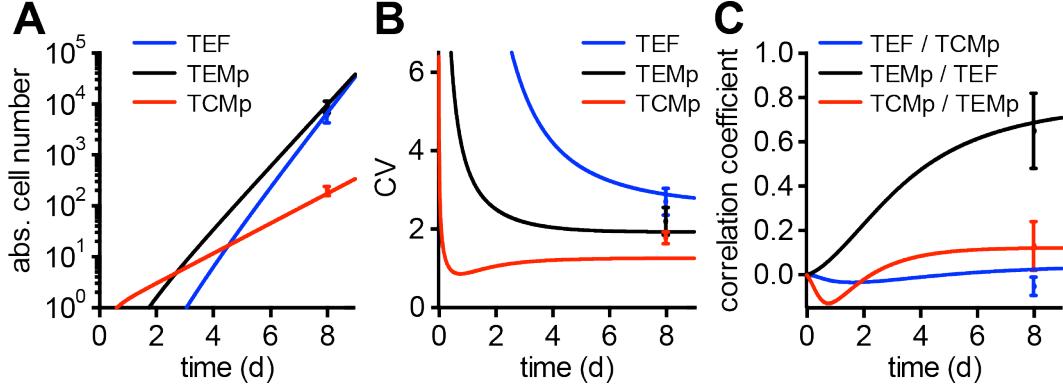


Figure 3.1.: Fit of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model to first and second moments of the abs. cell number distributions 8 d.p.i. (A-C) Best Fit of the model to the mean values (A), coefficients of variation (B) of the three subsets (TCMp, TEMp and TEF) and the three pairwise correlations (C). (Experiments by V. Buchholz and D. Busch)

We can now use the solution of this ODE system to get the time evolution of the mean values, CVs and CCs by making use of eqns. 3.10 to 3.13.

To implement parameter estimation we rely on standard χ^2 -minimization using the aforementioned summary statistics at day 8 p.i. for the objective function:

$$\chi^2 = \sum_{i=1}^3 \left(\frac{\langle n_i \rangle - \bar{n}_i}{\sigma_{\bar{n}_i}} \right)^2 + \sum_{i=1}^3 \left(\frac{CV_i - CV_i^D}{\sigma_{CV_i^D}} \right)^2 + \sum_{i=1}^3 \left(\frac{\rho_i - \rho_i^D}{\sigma_{\rho_i^D}} \right)^2. \quad (3.20)$$

Here the ρ_i denote the three pairwise correlation coefficients and CV_i^D , \bar{n}_i^D and ρ_i^D denote the sample estimates that correspond to the respective model quantities. We used bootstrapping to assess the uncertainties $\sigma_{\bar{n}_i}$, $\sigma_{CV_i^D}$ and $\sigma_{\rho_i^D}$ of these nine quantities [18]. Note that we chose the coefficients of variation instead of the variances to maximize independence between the different terms in the χ^2 -sum.

The resulting fit of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model is shown in Fig. 3.1. The model accurately reproduces the observed mean values, the variability and the correlations in the data. Table 3.1 lists the corresponding best fit parameter values (see chapter 3.6 for an assessment of the parameter uncertainties): Fast proliferating effector cells are derived from slowly expanding central memory precursors. In fact these differences in magnitude of the subset specific proliferation rates provide the key insight on how the observed variability and subset correlations are achieved in the model: On the one hand, the slow proliferation of the TCMP cells leads to a huge variability in the time point of the first TEMp differentiation event thereby generating strongly differing numbers of TEMp (and TEF) descendants at later time points (we will explore this further in chapter 5.1). On the other hand, rare differentiation events from slow to fast proliferating subsets when cell numbers are still small do not introduce a strong correlation of both subsets whereas bigger correlations are

3.3. Asymmetry of cell numbers in the N→TCMp→TEMp→TEF model

Parameter	value
TCMp proliferation rate (λ_1)	0.83 d ⁻¹
TEMp proliferation rate (λ_2)	1.44 d ⁻¹
TEF proliferation rate (λ_3)	1.59 d ⁻¹
naïve → TCMP differentiation rate (d_0)	2.4 d ⁻¹
TCMP → TEMp differentiation rate (d_1)	0.155 d ⁻¹
TEMp → TEF differentiation rate (d_2)	0.052 d ⁻¹

Table 3.1.: best fit parameters of the N→TCMp→TEMp→TEF model to the single cell-derived progenies data of day 8 p.i.

expected whenever differentiation to a subset with comparable or smaller turnover rate occurs.

The recruitment rate d_0 is found to be fast compared to the other rates. For this reason, the recruitment process contributes only little to the overall stochasticity. In fact, we would obtain a fit of comparable quality if we skipped the recruitment step in the model and directly started with one initial TCMP cell at $t = 0$. We will however keep the recruitment step for biological reasons.

3.3. Asymmetry of cell numbers in the N→TCMp→TEMp→TEF model

In addition to the high variability, the distribution of total cell number in the single cell-derived progenies also shows a considerable asymmetry (cf. chapter 2.2). We can quantify this asymmetry by estimating the sample skewness from the data

$$g = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{s} \right)^3, \quad (3.21)$$

where the sample of total cell number of the single cell progeny data is denoted by x_i , and \bar{x} and s are the corresponding sample estimates of mean and standard deviation, respectively. For the N→TCMp→TEMp→TEF model we can assess the skewness using

$$\gamma = \frac{\langle n_t^3 \rangle - 3 \langle n_t \rangle \langle n_t^2 \rangle + 2 \langle n_t \rangle^3}{\left(\langle n_t^2 \rangle - \langle n_t \rangle^2 \right)^{\frac{3}{2}}}, \quad (3.22)$$

with $n_t = n_1 + n_2 + n_3$ (cf. eqn. 3.14). The occurring moments of the sum $n_1 + n_2 + n_3$ can be evaluated straight forwardly and afterwards eqn. 3.9 can be used to calculate

3. The $N \rightarrow \text{TCMp} \rightarrow \text{TEMp} \rightarrow \text{TEF}$ model accounts for stochasticity of CD8^+ T cell diversification

the skewness γ as a function of the partial derivatives of the probability generating function. Since this expression for the skewness does depend on third order derivatives of the probability generating function F , we use the method described in chapter 3.1 to expand the ODE system given in eqns. 3.17 and 3.18 to also include the time evolution of the third order derivatives of F . The result is given in the appendix (eqn. A.5).

Fig. 3.2A shows the resulting time dependence of the skewness. Also shown is the sample skewness g with uncertainty estimated via non-parametric bootstrapping [18]. The skewness estimated via eqn. 3.21 is found to be considerably smaller than expected from the model calculation using eqn. 3.22.

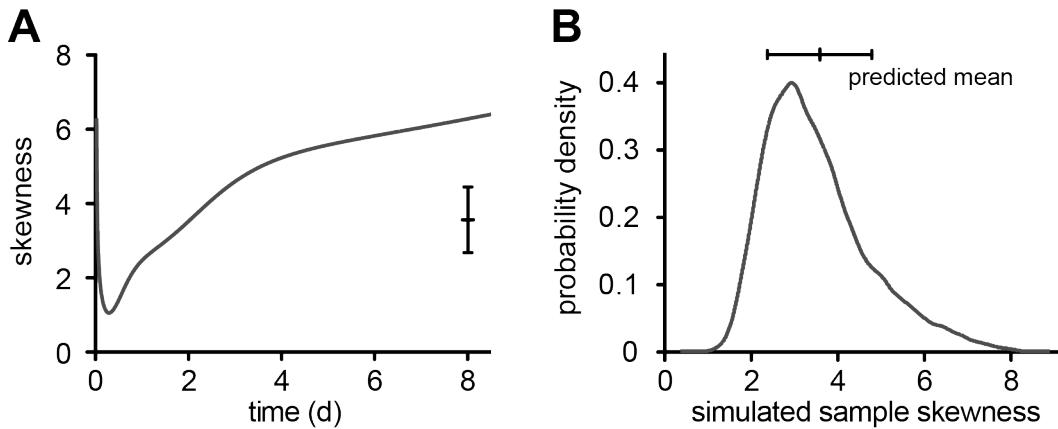


Figure 3.2.: (A) Predicted time evolution of the total cell number skewness in the $N \rightarrow \text{TCMp} \rightarrow \text{TEMp} \rightarrow \text{TEF}$ model and corresponding (biased) sample estimate with uncertainty based on the $N = 76$ single cell progeny populations of day 8 p.i. (B) Distribution of simulated skewness values ($N \rightarrow \text{TCMp} \rightarrow \text{TEMp} \rightarrow \text{TEF}$ model) for a sample size of $N = 76$ using the estimator given in eqn. 3.21. The shown mean of this distribution agrees with the observed sample skewness in A. The deviation between model and data in A is thus due to the small sample size and the resulting bias of the estimator for the skewness. (Experiments by V. Buchholz and D. Busch)

To understand this apparent discrepancy we simulated a large number (10^4) of samples of the same size as in the data ($N = 76$ single cell-derived progeny populations at day 8 p.i.) and then calculated the sample estimate for the skewness g for each sample. Fig. 3.2B shows a kernel estimate of the resulting distribution of the simulated sample skewness values. The mean simulated sample skewness for $N = 76$ is $\bar{g}_S^{N=76} = 3.6$, in agreement with the observed sample skewness $g = (3.56 \pm 0.89)$. Hence, for the total cell number distribution that results from the model, the estimate of g via eqn. 3.21 is highly biased. This is also the reason why we did not include the skewness in the summary statistics used for parameter estimation.

Thus the $N \rightarrow \text{TCMp} \rightarrow \text{TEMp} \rightarrow \text{TEF}$ model can reproduce the skewness in the total

3.4. Stochastic simulations of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

cell number distribution. The same analysis was also done for each subset separately showing that the model can also account for the skewness values of the individual subsets (not shown).

3.4. Stochastic simulations of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

Although only the linear correlations could be used for parameter estimation, the simulation using Gillespie's algorithm [28] shows that the general correlation structure of the data and specifically Spearman's rank correlation coefficients are correctly reproduced by the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model (Fig. 3.3A; all Spearman's rank correlation coefficients agree between model and data within 95% confidence level). In this figure, single cell progeny data (day 8 p.i.) are shown in black, simulated model points in red (same number of simulated single cell progenies as in the data) and estimates of the probability densities for the model are depicted in grey scale. In order to obtain the density estimates, a simulation of 10^5 cells was generated and

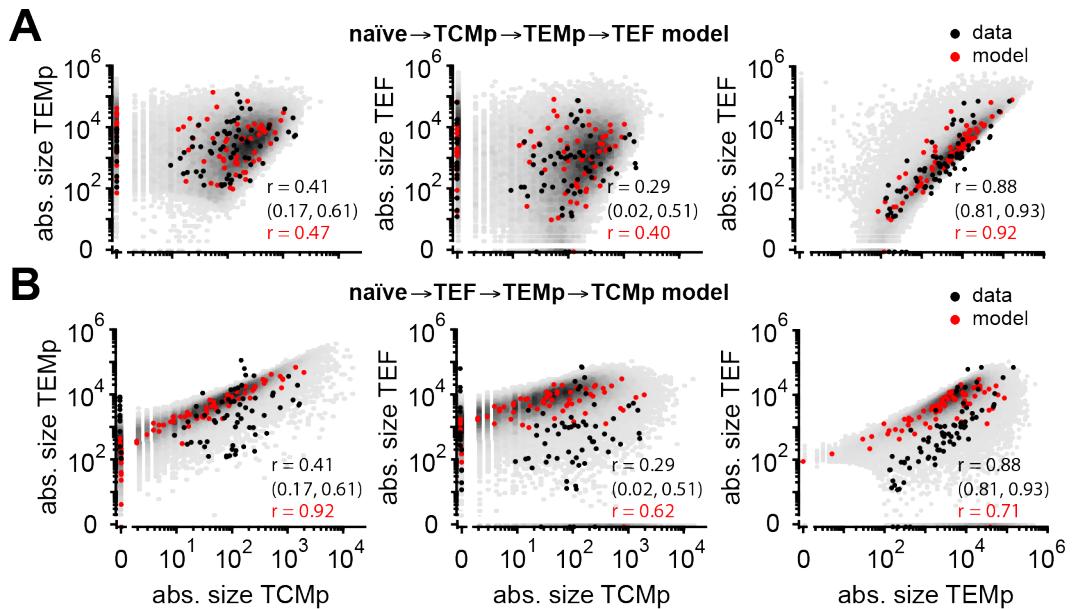


Figure 3.3.: Pairwise correlations of the three subsets TCMp, TEMp and TEF at day 8 p.i. (black dots) together with a representative simulation of equal size of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model (A; red dots) and the N \rightarrow TEF \rightarrow TEMp \rightarrow TCMp model (B; red dots). Also shown are density plots (grey scale) of the respective model simulations (simulating 10^5 cells). Spearman's rank correlation coefficients are shown in black for the data (with 95% confidence limits) and in red for the model. (Experiments by V. Buchholz and D. Busch)

3. The N→TCMp→TEMp→TEF model accounts for stochasticity of CD8⁺ T cell diversification

binned in log space. Spearman's rank correlation coefficients were also calculated based on this simulation. The simulations are based on the best fit parameters given in Table 3.1.

It is illustrative to consider the correlative structure that would result from the differentiation scheme with the reversed subset order N→TEF→TEMp→TCMp. To obtain reasonable best fit parameter values, we constrained the fitted subset mean values of the model to lie within the 95% confidence intervals of the mean values in the data and then optimized mean values, correlation coefficients and CVs under these constraints. The resulting model simulation is shown in Fig. 3.3B. None of the correlations in the data can be reproduced by this model: The correlations between TEMp-TCMp and TEF-TCMp are predicted as too strong, whereas the correlation TEF-TEMp is too weak in the model. This example illustrates that the order of subset emergence relates to the correlation structure in the data.

In chapter 2.2 we already mentioned the reversed correlation between total cell number and the fraction of CD62L⁺ and CD27⁺ cells in the single cell progenies. Since

$$\text{CD62L}^+ = \frac{n_0 + n_1}{\sum_i n_i} \quad (3.23)$$

$$\text{CD27}^+ = \frac{n_0 + n_1 + n_2}{\sum_i n_i}, \quad (3.24)$$

we can use the model to also simulate these correlations (Fig. 3.4). The model reproduces the observed relation between marker positive cells and the total size of the population. In the model, the inverse correlation between CD62L⁺ and total cell number is reflecting the fact that this marker efficiently divides the three subsets into the two groups of rapidly expanding TEMp and TEF cells and slowly proliferating TCMP cells.

For the other marker, CD27, the arising correlation is not as simple and expected to be smaller: Since the proliferation rates of the TEMp and TEF subsets are very comparable, populations can in principle get large irrespective whether the fraction of CD27⁺ cells is large or not. Nevertheless, a population with a high fraction of strongly expanding TEF cells (CD27⁻) is very unlikely to be small and the smallest populations are expected to contain a high fraction of slowly proliferating TCMP cells and are thus preferentially CD27⁺.

Another important model validation is to compare the binned single cell progeny's absolute subset sizes with the expected cell number distributions in the model (Fig. 3.5). Remarkably, in addition to the overall good agreement between the binned log-transformed subset sizes and the predicted model distributions, also the number of TCMP and TEF populations of zero absolute size in the data are captured well by the model. Note that in the data there is no TEMp subset of zero absolute size and the probability for this in the model is also negligible. In the model, populations

3.4. Stochastic simulations of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

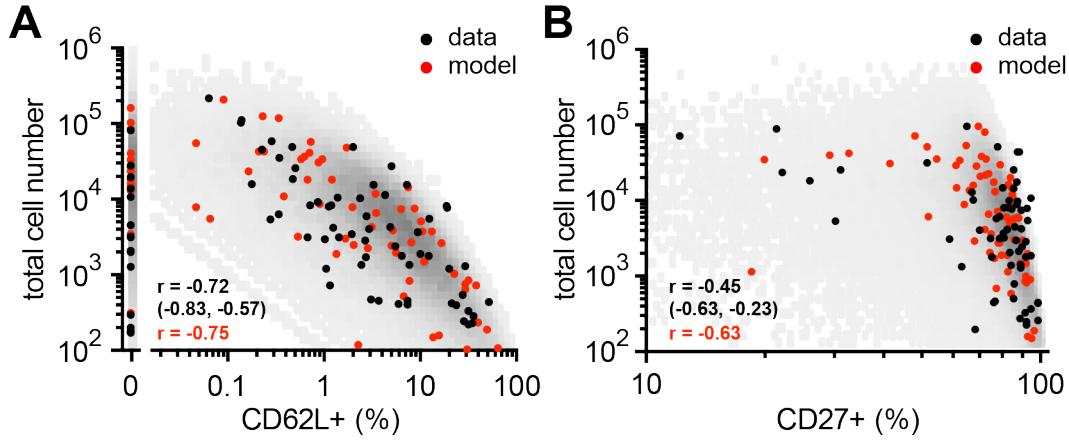


Figure 3.4.: CD62L $^{+}$ (A) and CD27 $^{+}$ (B) versus absolute cell number at day 8 p.i. (black dots) together with a representative simulation of equal size of the N \rightarrow TEF \rightarrow TEMp \rightarrow TCMp model (red dots) and a density plot (grey) of a simulation of 10^5 cells. Spearman's rank correlation coefficients are shown in black for the data (with 95% confidence limits) and in red for the model. (Experiments by V. Buchholz and D. Busch)

that lack TEF cells arise due to the small probability that even until day 8 p.i. no TEMp cell has yet differentiated into the TEF compartment. Accordingly, these populations show a small absolute size also for the TEMp compartment both in the model and in the data (cf. Fig. 3.3A: TEMp vs. TEF). We will discuss the matching model's prediction for the fraction of TCMP depleted populations in chapter 5.5.

All model simulations in this section were solely based on the best fit parameter values obtained by fitting the mean values, CVs and correlations of the single cell data of day 8 (Table 3.1). Thus, these nine quantities prove appropriate to efficiently summarize the single cell-derived progeny data.

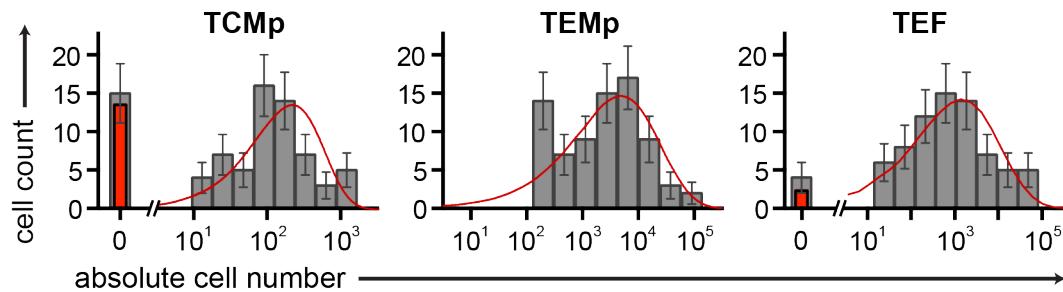


Figure 3.5.: Binned abs. size of the single cell progenies of the three subpopulations at day 8 p.i. (grey) and the expected distributions (red curves and bars). (Experiments by V. Buchholz and D. Busch)

3. The $N \rightarrow TCMp \rightarrow TEMP \rightarrow TEF$ model accounts for stochasticity of CD8⁺ T cell diversification

3.5. Validated predictions of the $N \rightarrow TCMp \rightarrow TEMP \rightarrow TEF$ model

In the previous section we showed various validations of the $N \rightarrow TCMp \rightarrow TEMP \rightarrow TEF$ model that were all based on the single cell-derived progenies data of day 8 p.i. We then asked whether the model was able to predict the dynamics of CD8⁺ T cell diversification also at earlier time points solely based on the parameter values estimated from the data of day 8 (Table 3.1). To test this, further experiments were conducted in the lab of Dirk Busch.

Predicting the day 6 cell number distributions

We first explored model predictions for the single cell-derived progeny populations at various time points prior to day 8 p.i. One complication here is the experimental cut-off: In the spleen, populations with total absolute sizes under 100 cells are likely to be missed in the analysis of the organ. We thus reasoned 6 days after infection to be the earliest time point where the expected absolute cell numbers would be experimentally feasible.

Based on the single cell-derived progenies data at day 6 p.i. we then tested various predictions of the model. First, we first compare the predicted absolute cell size distributions for the three subsets with the binned single cell progenies data (Fig. 3.6). The model predictions (red line and bars) take into account the estimated experimental cut-off of 100 absolute cells: The distributions shown in Fig. 3.6 are based on simulated populations with a total size of at least 100 cells. The general shape of the observed distributions (grey bars) is captured well by the model prediction. Moreover, consistency with the model prediction is also found for the percentage of populations that lack TCMp and TEF cells. Note that the model prediction for zero

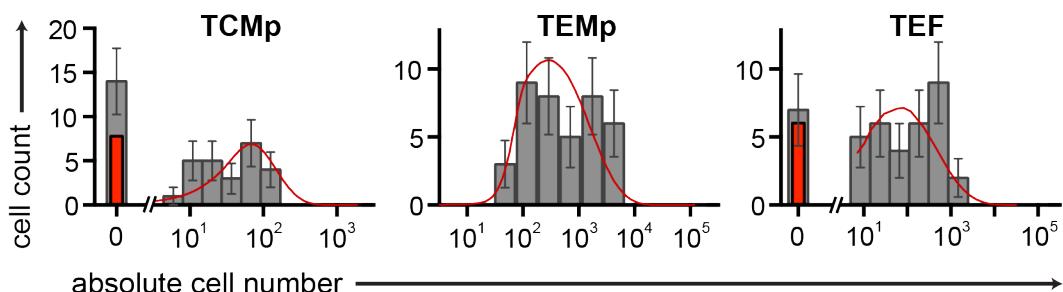


Figure 3.6.: Binned abs. size of the single cell progenies of the three subpopulations at day 6 post infection (grey) and the predicted distributions (red curves and bars) simulated using the best fit parameters of the $N \rightarrow TCMp \rightarrow TEMP \rightarrow TEF$ model to the data of day 8 p.i. (Experiments by V. Buchholz and D. Busch)

3.5. Validated predictions of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

absolute size of the TEF compartment assumes that the corresponding count in the data includes TEF subsets with sizes smaller than 10.

Predicting the day 6 correlations

We then tested whether the model correctly predicts the non-linear correlations between total cell number and the fraction of CD62L $^{+}$ and CD27 $^{+}$ cells (Fig. 3.7). Indeed, we find agreement between the data and the predicted relation between both markers and the total sizes. Notably, both Spearman's rank correlation coefficients agree between model prediction and data within 95% confidence level.

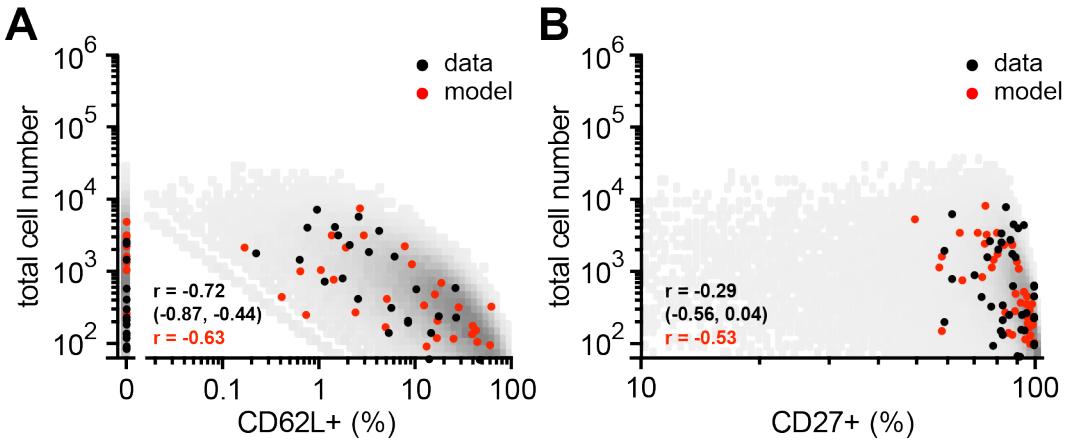


Figure 3.7.: CD62L $^{+}$ (A) and CD27 $^{+}$ (B) versus absolute cell number at day 6 p.i. (black dots) together with a simulation of equal size of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model (red dots) and a density plot (grey) of a simulation of 10^5 cells (prediction made by using the best fit parameters to the data of day 8 p.i.). (Experiments by V. Buchholz and D. Busch)

We can also calculate summary statistics of the single-cell derived progenies at day 6 and compare them with the corresponding model predictions (Fig. 3.8). Again we used non-parametric bootstrapping [18] to assess the uncertainties in the data.

The 95% confidence intervals for the model predictions shown in Fig. 3.8 were calculated as follows. First, the single cell-derived progenies data of day 8 p.i. were resampled to get bootstrap samples (about 10000 bootstrap sample were generated). Then the summary statistics used for the parameter estimation procedure (as described in chapter 3.2) were calculated based on the bootstrap samples. For every bootstrap sample, the model was then refitted based on these summary statistics. This leads to a set of resulting best fit parameters. We subsequently simulated large populations of single cell-derived progenies at day 6 p.i. for each of the obtained best fit parameter vectors (300,000 populations were generated for every best fit

3. The N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model accounts for stochasticity of CD8 $^{+}$ T cell diversification

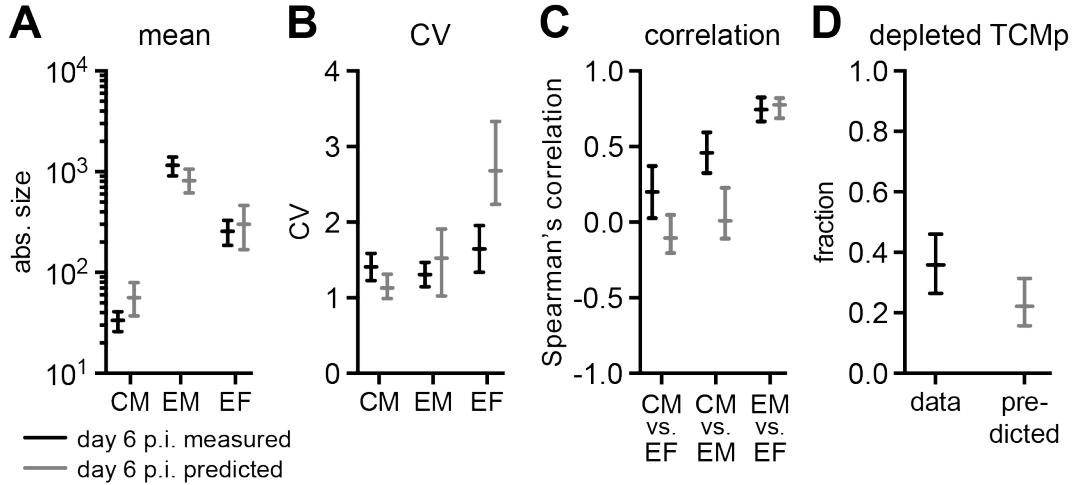


Figure 3.8.: Comparison of mean values (A), CVs (B), Spearman's rank correlation coefficients (C) and the fraction of TCMP depleted populations (D) for the single cell-derived data of day 6 p.i. (black data points with bootstrap-estimated standard errors) with the corresponding model predictions (mean prediction and 95% confidence region depicted in grey). (Experiments by V. Buchholz and D. Busch)

parameter vector). Population means for the sought quantities were then calculated for each of these simulated populations and in the end used to determine the mean and percentiles of the predicted quantities. To take into account the experimental cut-off of around 100 cells, progenies with a total size of less than 100 cells were not considered for this calculation.

The predicted values for the mean absolute subset sizes (Fig. 3.8A) agree very well with the observed mean values. It should be noted that these mean values change considerably between day 6 and day 8 (cf. Fig. 3.1). The CVs for the TCMP and the TEMp subset sizes are also correctly predicted (Fig. 3.8B). The CV for the TEF population on the other hand is found to be smaller than predicted. However, both data and model prediction for this quantity are subject to large uncertainties. Also, the rigorous cut-off of 100 absolute cells most probably represents an over-simplification of the real situation. Especially statistical quantities that are dependent on second or higher order moments of the distribution might be affected by this. In accordance with this, a small discrepancy between model prediction and data is also found for the TCMP-TEMp correlation (Fig. 3.8C; using Spearman's rank correlation coefficient). The remaining two correlations are predicted correctly. Also the expected fraction of TCMP depleted populations is found to be in agreement with the observed value (Fig. 3.8D).

3.5. Validated predictions of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

Predicting the relative subset size kinetics

As we have seen, for the single cell-derived progeny populations already day 6 p.i. is problematic due to the experimental cut-off in absolute cell size. To explore also earlier time points we thus had to rely on experiments with a higher number of transferred cells. In chapter 2.4 we discussed how variability in the number of T cells that get recruited into the immune response precludes reliable estimation of the subset's mean values, CVs and correlations. However, we also showed that the mean relative subset sizes are not affected by this recruitment variability (eqn. 2.6). Hence, for early time points, we focused on measuring the kinetics of the relative TCMP, TEMp and TEF subset sizes.

In a first experiment, 10^5 OT-I cells were transferred and relative subset sizes were evaluated at day 1 to 4 p.i. (Fig. 3.9A). For day 6 and 8 p.i. the relative subset sizes in Fig. 3.9A are based on the 100 cell-derived progeny populations. The calculation of the shown 95% confidence bands was done as described above, but the set of obtained best fit parameter vectors from the bootstrap procedure were used to compute a set of kinetics for each of the three subsets. At every predicted time point, the 95% confidence region for each subset was then based on the corresponding percentiles.

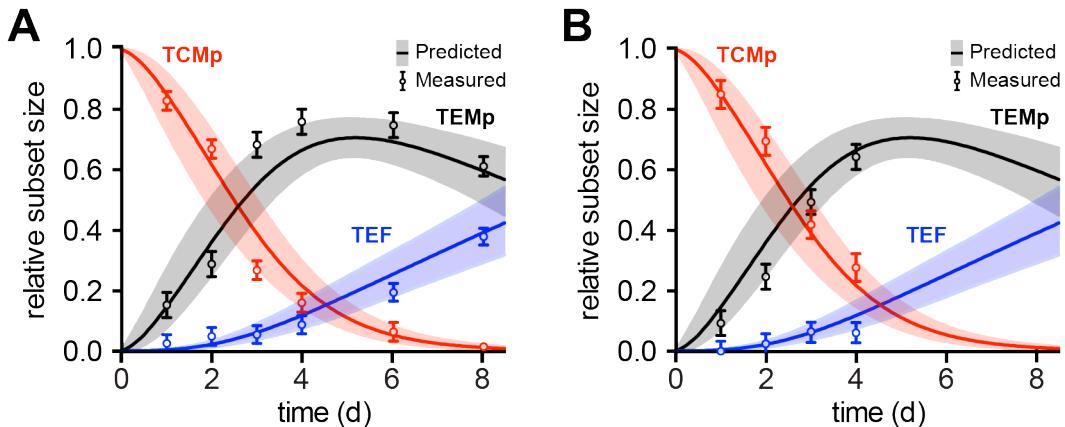


Figure 3.9.: Relative subset sizes (circles) of day 1 to 4 following transfer of 100,000 (A) or 10,000 OT-I cells (B) and after transfer of 100 cells at day 6 and day 8 p.i. (A). The filled regions show the 95% confidence prediction bands of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model computed from the day 8 single cell progeny data. (Experiments by V. Buchholz and D. Busch)

All measured relative subset sizes fall into the 95% prediction bands that were calculated based solely on the single cell progeny data of day 8. Remarkably, the time point after which the TEMp subset becomes dominating and the time point after which there are more TEF than TCMP cells are accurately predicted by the model. Also, the model correctly describes the occurrence of a maximum in the relative size of the TEMp compartment around day 5 p.i.

3. The N→TCMp→TEMp→TEF model accounts for stochasticity of CD8⁺ T cell diversification

It has been reported that upon adoptive transfer of a high number of T cells (as compared to the endogenous size of specific T cells), T cell diversification during infection becomes dependent on the number of transferred transgenic precursors [4, 40]. However, these studies focused on time points later than day 4 p.i. Indeed, no change in the dynamics of the relative subset sizes could be observed in a second experiment where 10⁴ OT-I cells were transferred into the mice (Fig. 3.9B). Also in this experiment, the data points fall within the 95% prediction bands.

We thus demonstrated the predictive power of the N→TCMp→TEMp→TEF model. Remarkably, the validated predictions of days 1 to 6 p.i. were based solely on the single cell-progeny data of day 8. Apparently, the used framework for parameter estimation efficiently utilizes the information on the differentiation and proliferation histories that is contained in these data (“statistical footprint”).

3.6. Identifiability of model parameters

In the previous section we showed that the relative subset size kinetics of the TCMP, TEMp and TEF cells can be accurately predicted based solely on the single cell-derived progeny data of day 8 p.i. In the following we will use these kinetics to further constrain the model parameters and compute confidence regions for the estimated parameters based on the profile likelihood method [59]. Nevertheless, we will also see that all model parameters except for the recruitment rate d_0 can be identified already by just using the day 8 data.

Since the relative subset sizes sum to one, we can only use the kinetics of two of the subsets for parameter estimation. The TEF and the TCMP subsets best reflect the dynamics of the two measured markers CD27 and CD62L. We have thus chosen to base the fitting on the relative size kinetics of those two subsets (cf. Fig. 3.9A). The single cell progeny data of day 8 was included into the χ^2 -sum as before (eqn. 3.20). Note that we do not include the single cell-derived progeny data from day 6 p.i. in the parameter estimation due to the discussed complications that arise because of the experimental cut-off in the absolute cell number (cf. chapter 3.5).

Fig. 3.10 shows the best fit of the N→TCMP→TEMp→TEF model to the complete data set; best fit parameters are given in Table 3.1.

In order to assess the uncertainties of the parameter estimates, we will utilize the profile likelihood method [59]. In this method, the $(1 - \alpha)$ -confidence region Θ_i for parameter θ_i is obtained as the region $\theta_i \in \Theta$ for which

$$\chi_{\min}^2(\theta_i) - \chi_{\min}^2 \leq \chi_{1-\alpha}^2. \quad (3.25)$$

Here χ_{\min}^2 stands for the minimized χ^2 -value of the fit, $\chi_{1-\alpha}^2$ stands for the $(1 - \alpha)$ quantile of the χ^2 -distribution with one degree of freedom and $\chi_{\min}^2(\theta_i)$ denotes the

3.6. Identifiability of model parameters

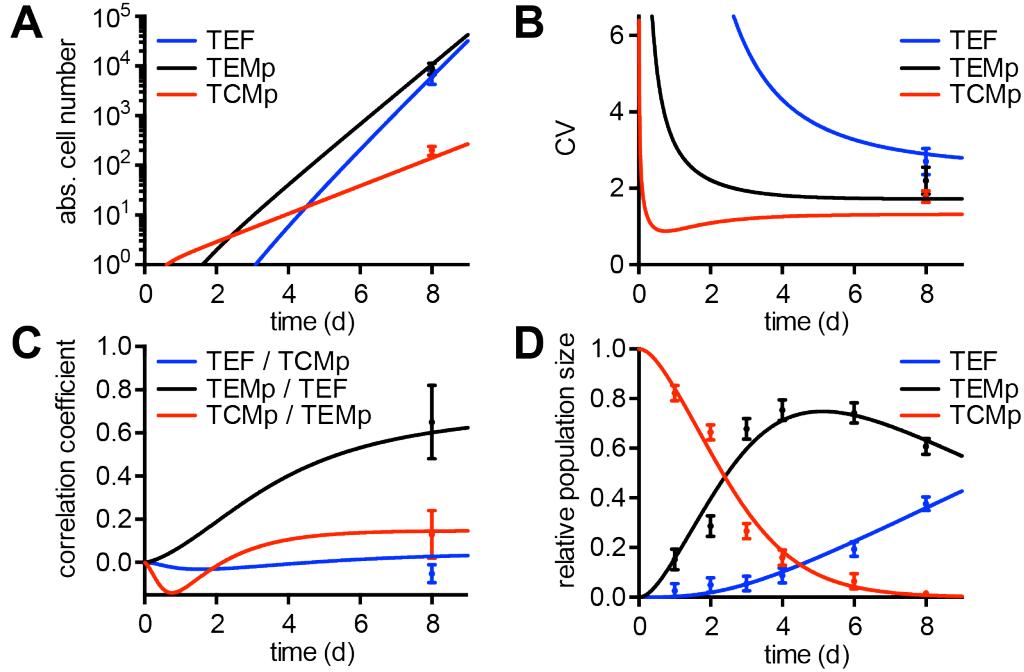


Figure 3.10.: Fit of the $N \rightarrow TCMP \rightarrow TEMp \rightarrow TEF$ model to mean values (A), CVs (B) and pairwise correlations (C) of the single cell progeny data of day 8 p.i. and to the relative subset size kinetics (D; day 1 to 4: transfer of 10^5 OT-I cells, day 6 and 8: transfer of 100 OT-I cells). (Experiments by V. Buchholz and D. Busch)

minimal χ^2 -value that is obtained upon fixing θ_i and minimizing over the remaining parameters.

In practice, parameter bounds can be found by plotting $\Delta\chi^2(\theta_i) = \chi^2_{\min}(\theta_i) - \chi^2_{\min}$ and checking for which parameter values of θ_i the function $\Delta\chi^2(\theta_i)$ exceeds certain thresholds. In Fig. 3.11 two functions $\Delta\chi^2(\theta_i)$ are plotted for every model parameter: The dashed lines show $\Delta\chi^2(\theta_i)$ for the respective parameter if parameter estimation is solely based on the day 8 single cell progeny data and the straight lines result if the relative subset size kinetics are incorporated into the fitting as well.

Using $\chi^2_{68\%} = 1.0$ and $\chi^2_{95\%} = 3.8$, we can read off the resulting 68% and 95% confidence bounds for every parameter directly from Fig. 3.11. The confidence intervals obtained from fitting the complete data set are depicted in Fig. 3.11 and listed in Table 3.2. As can be seen from Fig. 3.11, including the relative subset size kinetics in the parameter estimation does not lead to very different estimates for the parameter values. The main change is that the differentiation rates can now be estimated more accurately.

Note that since all our estimates are based on absolute cell numbers in the spleen, the proliferation rates will be slightly under-estimated. However, estimates at day

3. The N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model accounts for stochasticity of CD8 $^{+}$ T cell diversification

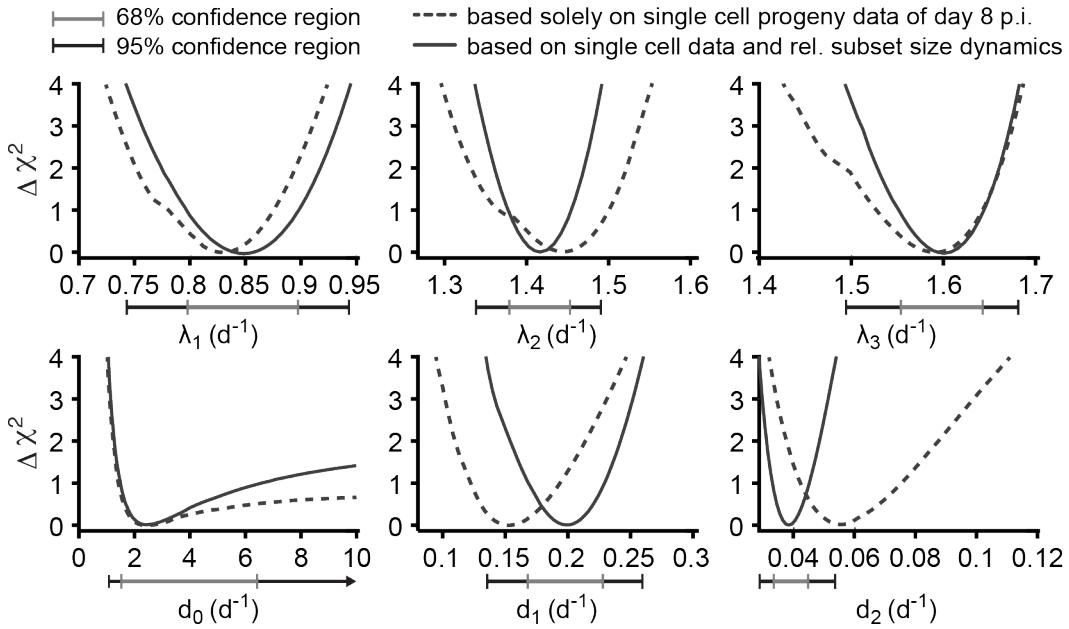


Figure 3.11.: Change in χ^2 -function obtained upon fixing the indicated parameter and minimizing over the remaining ones using just the day 8 single cell progeny data (dashed line) or using the rel. subset size kinetics of the TEF and TCMP cells in addition (straight line). Also indicated are the resulting 68% and 95% confidence regions of the respective parameters (profile likelihood method; based on rel. size kinetics and day 8 data). (Experiments by V. Buchholz and D. Busch)

12 p.i. show that the overall total T cell number (spleen, lung and lymph nodes) is around twice as big as the number of T cells that is recovered in the spleen. Assuming a proliferation phase of around 8 days this amounts to a change of all proliferation rates of only around 0.1 d^{-1} .

Remarkably, the single cell data of day 8 already sufficiently constrain all parameter

Parameter	best fit	68% conf.	95% conf.
TCMP proliferation rate (λ_1)	0.83 d^{-1}	$(0.79, 0.90) \text{ d}^{-1}$	$(0.74, 0.95) \text{ d}^{-1}$
TEMp proliferation rate (λ_2)	1.42 d^{-1}	$(1.37, 1.46) \text{ d}^{-1}$	$(1.32, 1.49) \text{ d}^{-1}$
TEF proliferation rate (λ_3)	1.59 d^{-1}	$(1.55, 1.68) \text{ d}^{-1}$	$(1.49, 1.81) \text{ d}^{-1}$
naïve \rightarrow TCMP rate (d_0)	2.8 d^{-1}	$(1.4, 7.4) \text{ d}^{-1}$	$(1.5, \infty) \text{ d}^{-1}$
TCMP \rightarrow TEMp rate (d_1)	0.192 d^{-1}	$(0.168, 0.231) \text{ d}^{-1}$	$(0.130, 0.262) \text{ d}^{-1}$
TEMp \rightarrow TEF rate (d_2)	0.039 d^{-1}	$(0.032, 0.045) \text{ d}^{-1}$	$(0.028, 0.054) \text{ d}^{-1}$

Table 3.2.: best fit parameters and confidence intervals (profile likelihood method) of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model based on the fit to the single cell-derived progenies data of day 8 p.i. and the relative subset size kinetics.

3.6. Identifiability of model parameters

values except for the recruitment rate d_0 . This shows again that the single cell progeny data contain a statistical footprint of the proliferation and differentiation history and that this footprint can be appropriately summarized by the mean values, CVs and pairwise correlations coefficients of the absolute subset sizes. We will explore this further in the next chapter where among other things we will show that these data can also be utilized to effectively rule out alternative models that are obtained upon subset permutations.

4. Systematic comparison of possible differentiation models

In chapter 3 we demonstrated that the N→TCMp→TEMp→TEF model can accurately reproduce both the stochasticity of the single cell-derived progenies and the kinetics of the relative T cell subset sizes. In the present chapter, we use the same data set to probe also various alternative differentiation models. Among the tested model topologies, we find that only the N→TCMp→TEMp→TEF model (and a closely related variant thereof) provides a good fit to data. Furthermore, we will show that this result still holds if we include cell death into the models. In addition, we can base the analysis solely upon the data of day 6 and day 8 post infection, where physiological numbers of T cells could be used in the adoptive transfers. At the end of this chapter we discuss models in which differentiation is initiated by an asymmetrically dividing naïve cell.

4.1. Alternative unidirectional models fail to reproduce the single cell progeny data

In addition to the N→TCMp→TEMp→TEF model, there are five other unidirectional differentiation models with alternative subset sequences (Fig. 4.1A). We will now show that these models cannot reproduce the summary statistics for the single cell progeny data of day 8 p.i.

Since all five alternative models can be obtained from the N→TCMp→TEMp→TEF model upon subset permutations, we can again define the χ^2 -sum as in eqn. 3.20 and simply reuse the framework introduced in chapter 3.2 to fit each model to the day 8 data. The best fit of each model was obtained by using both simulated annealing and a local optimization algorithm (trust-region-reflective algorithm using Matlab's optimization toolbox) with at least 300 different random initial values.

All five models fail to reproduce the summary statistics. Fig. 4.1B illustrates this by ranking the six models in terms of the corrected Akaike information criterion (AICc) [12],

$$\text{AICc} = 2k - 2 \log L + \frac{2k(k+1)}{n-k-1}. \quad (4.1)$$

4. Systematic comparison of possible differentiation models

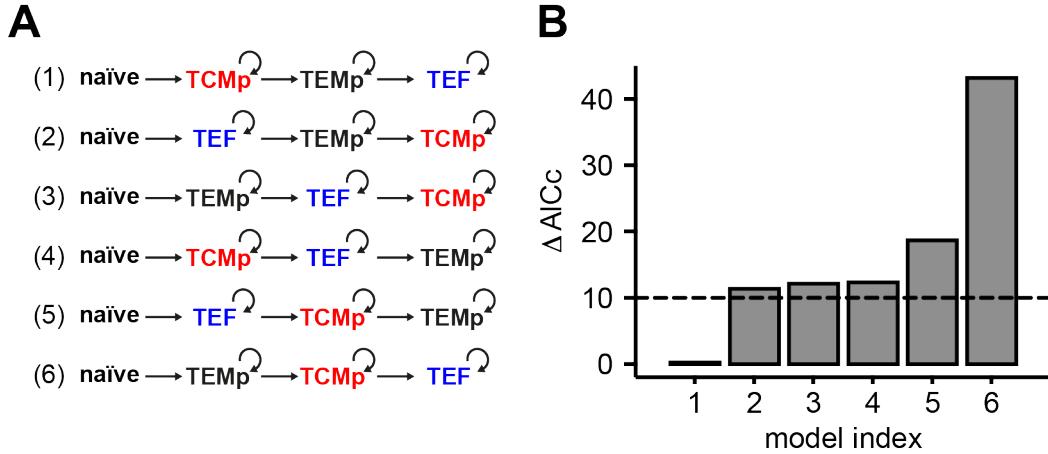


Figure 4.1.: (A) Schemes of the six unidirectional differentiation models. (B) Ranking of the best fit of each model in A to the single cell progeny data of day 8 in terms of the corrected Akaike information criterion (AICc). Model index as in A.

Here, k denotes the number of model parameters (for the models given in Fig. 4.1A: $k = 6$), n is the sample size ($n = 9$ for the summary statistics at day 8 p.i.) and L is the maximized value of the likelihood that the considered model would generate the data. For standard χ^2 -fitting we can use $\log L = -\frac{1}{2}\chi^2_{\min} + \text{const.}$ resulting in

$$\text{AICc} = \chi^2_{\min} + 2k + \frac{2k(k+1)}{n-k-1} + \text{const.} \quad (4.2)$$

We then evaluate the performance of each model by calculating the difference in AICc to the best fitting model (Fig. 4.1B). An AICc difference between two models of 10 and above indicates that the worse-fitting model is essentially unsupported by the data [12]. Since in this case all models have the same number of parameters, ranking based upon χ^2_{\min} would have led to the same order of models.

Thus, the single cell progeny data of day 8 suffices to identify the order of subset emergence in unidirectional differentiation models.

4.2. Probing a general class of differentiation models

We will now fit more general classes of differentiation models to the data. Consider the model scheme in Fig. 4.2: If all rates are taken to be non-zero, then an activated naïve cell can differentiate into all three subsets at a rate d_0 . Subsequently, at every given time point, cells of each subset can differentiate into every other subset at rates d_i and d_{-i} . By setting a subset of the rates to zero, more specific differentiation models can be obtained. For instance, the N→TCMp→TEMP→TEF model is retrieved upon setting d_{02} , d_{03} , d_{-1} , d_{-2} , d_3 and d_{-3} to zero. However, not each of

4.2. Probing a general class of differentiation models

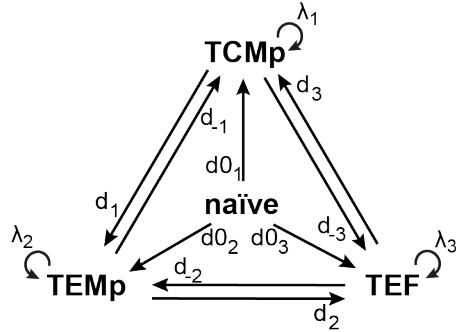


Figure 4.2.: Model scheme allowing for all possible diversification pathways from naïve to TCMP, TEMP and TEF as well as subset-specific proliferation rates.

the 512 combinatorially possible topologies generates an acceptable differentiation model. For a model to be valid, each of the three states TCMP, TEMP and TEF has to be reachable from the naïve state.

In order to enumerate the number of valid models that can be derived from this model scheme, we first consider the case where $d0_1 > 0$, $d0_2 = 0$ and $d0_3 = 0$. If both, d_1 and d_{-3} were zero, the TEMP and the TEF subset could not be reached. If both rates are non-zero, 16 valid models can be generated. On the other hand, if only one of the rates is non-zero, e.g. d_1 , then also d_2 has to be non-zero in order to be able to reach also the TEF state. This way we can generate another eight valid models. Similarly, we find eight more models in the case when $d_1 = 0$ and $d_{-3} > 0$. Altogether we thus find 32 valid models if the naïve cells can only differentiate into the TCMP subset. It follows that there are 96 valid models if only exactly one of the $d0_i$ is non-zero. If two of those rates are non-zero, say $d0_1$ and $d0_2$, then invalid models are generated only if d_2 and d_{-3} are both zero. On the other hand, 16 valid models each can be found if either one of the rates is non-zero or both are non-zero. This leaves us with $3 \times 48 = 144$ valid models for the case that exactly two of the $d0_i$ are non-zero. Finally, if all $d0_i$ are non-zero, then all 64 combinations of the remaining differentiation steps will generate valid model structures. With the scheme depicted in Fig. 4.2 we can thus altogether generate $96 + 144 + 64 = 304$ models, with the property that, starting from the naïve state, all three other subsets can be reached.

We can use eqns. 3.3 and 3.5 to find the PDE for the probability generating function of the full model scheme

$$\begin{aligned} \partial_t F(t) = \sum_{i=1}^3 & \left\{ \lambda_i (z_i^2 - z_i) \partial_{z_i} F + d0_i (z_i - z_0) \partial_{z_0} F + d_i (z_{i+1} - z_i) \partial_{z_i} F \right. \\ & \left. + d_{-i} (z_i - z_{i+1}) \partial_{z_{i+1}} F \right\}, \end{aligned} \quad (4.3)$$

where we have set $z_4 = z_1$ for notational convenience.

4. Systematic comparison of possible differentiation models

We can also parametrize the above model in terms of the recruitment rate d_0 and the ratios of N→TCMp (r_{CM}) and N→TEF (r_{EF}) transitions:

$$d_0 = \sum_i d_{0i}, \quad r_{CM} = \frac{d_{01}}{d_0}, \quad r_{EF} = \frac{d_{03}}{d_0}. \quad (4.4)$$

The mean values, CVs and correlation coefficients as a function of time are obtained analogously to chapter 3.2 by solving the ODE system of the first and second derivatives of F evaluated at $z_i = 1$ and then using eqns. 3.10 to 3.13 (the ODE system is given in appendix A.1).

Model ranking based on complete data set

For parameter estimation, we define the χ^2 -sum as in chapter 3.6 and include the measured kinetics of the relative TEF and TCMP subset sizes as well as the summary statistics (mean values, CVs and the correlation coefficients of the three subsets) of the single cell-progeny data of day 8. As in the previous section, the best fit for each of the 304 models was obtained by using both simulated annealing and a local optimization algorithm (trust-region-reflective algorithm using Matlab's optimization toolbox) with at least 300 different random initial values.

A substantial number of the best fits of the 304 models evaluated one or several of the fitted rate constants as zero, so that only 72 topologically different models remained after the fitting. Fig. 4.3A shows a ranking of these models in terms of the corrected Akaike information criterion (cf. eqn. 4.2). The corresponding goodness of fit for each model can be quantified by the ratio χ^2_{\min} over degrees of freedom ν (equal to the number of data points minus the number of fitted parameters; Fig. 4.3B). This ratio should be of the order of one since the expectation value for a χ^2 -distribution with ν degrees of freedom is ν .

Based on these rankings we can now make the following conclusions. Firstly, two minimal model structures are in accordance with the data: The already discussed N→TCMP→TEMp→TEF model (ranked second) and a model with the same scheme, but with an additional differentiation step from the naïve to the TEMp compartment (ranked first; Fig. 4.3C). Secondly, all other model structures either reduce to one of these topologies for their best fit parameter values (with one or several rates being zero in the best fit) or cannot account for the data.

The difference in AICc between the models ranked first and second is $\Delta\text{AICc} = 2.7$. As a rule of thumb, models with $\Delta\text{AICc} < 2$ have “substantial support”, those with $4 \leq \Delta\text{AICc} \leq 7$ have “considerably less support” and models with $\Delta\text{AICc} > 10$ have “essentially no support” [12]. Based on the data at hand we can thus not distinguish between the two best-fitting models. However, according to the best fit parameters,

4.2. Probing a general class of differentiation models

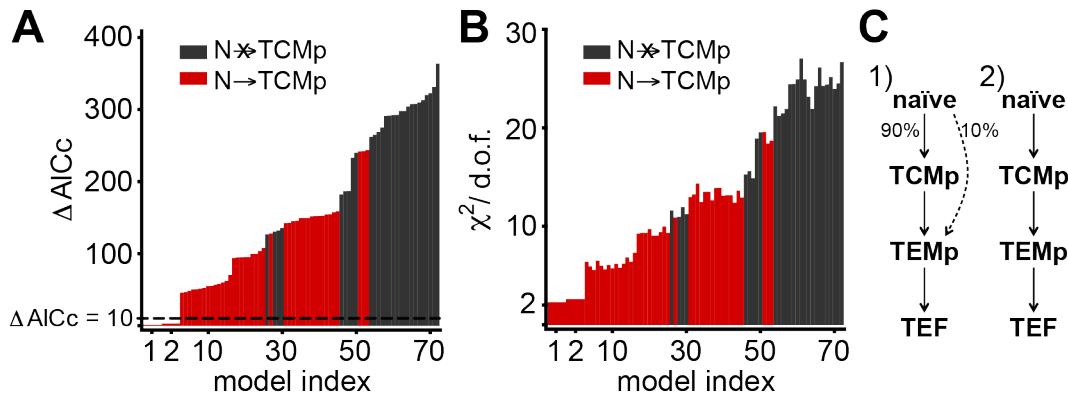


Figure 4.3.: Model comparison based on fitting the relative subset size kinetics and the single cell progeny data of day 8 p.i. (A) Model ranking with corrected Akaike information criterion (histogram shows the difference in AICc between the respective model and the best model). (B) Same ranking but showing the corresponding goodness of fit using χ^2_{min} over degrees of freedom. Red bars: models including a naïve \rightarrow TCMp differentiation step; black bars: other models. (C) Scheme of the two models that are in accordance with the data.

also in the model with the additional naïve \rightarrow TEMP differentiation step, 90% of the activated naïve cells generate progeny populations that differentiate along the pathway N \rightarrow TCMp \rightarrow TEMP \rightarrow TEF (Fig. 4.3C).

A much discussed model of CD8⁺ T cell diversification is the linear differentiation model which proposes differentiation along the pathway N \rightarrow TEF \rightarrow TEM \rightarrow TCM [60]. It is thus an interesting finding that in our case models in which the naïve cells cannot directly differentiate into central memory precursors yield the worst fits (grey bars in Fig. 4.3A and B).

Model ranking based solely on the data of the single cell and 100 cell-derived progenies

We already mentioned in chapter 3.5 that the adoptive transfer of an unphysiological high number of cells has been reported to influence T cell differentiation [4, 40]. We also showed that early after infection (until day 4 p.i.; the cited studies focused on later time points), no change in the kinetics of the relative subset sizes could be observed between two experiments in which 10^5 (cf. Fig. 3.9A) or 10^4 OT-I cells (cf. Fig. 3.9B) were adoptively transferred. Nevertheless, we also repeated the above analysis using just the relative subset sizes of day 6 and 8 (based on the 100 cell-progeny populations) in addition to the single cell progeny data of day 8 (Fig. 4.4). In this case, 69 topological different models remained after the parameter estimation procedure. As can be seen from Fig. 4.4A, five of these models have $\Delta\text{AICc} < 10$. Furthermore, the difference in ΔAICc between the best model and the models ranked

4. Systematic comparison of possible differentiation models

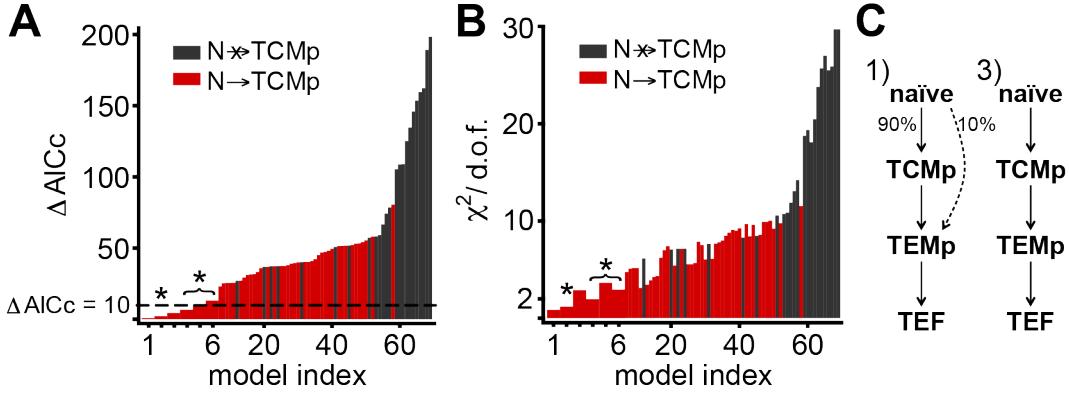


Figure 4.4.: As in Fig. 4.3 but based solely on the relative subset sizes of day 6 and 8 (using the 100 cell-derived progenies) and the single cell progeny data of day 8 p.i. (A) Model ranking based on $AICc$. Models marked with asterisks are also excluded by the data (see main text). (B) Same ranking but showing the corresponding goodness of fit using χ^2_{min} over degrees of freedom. Red bars: models including a naïve \rightarrow TCMP differentiation step; black bars: other models. (C) Scheme of the two models that are in accordance with the data.

six and seven is $\Delta AICc \approx 13$ and $\Delta AICc \approx 24$, respectively. We will therefore look more closely on the models ranked one to six. These models are also ranked best when comparing the goodness of fit (Fig. 4.4B). Fig. 4.5 shows the schemes of these models. In models 2, 4, 5 and 6, the fraction $d0_1/(d0_1 + d0_2 + d0_3)$ of the naïve cells that differentiate into the TCMP subset would generate progeny populations that cannot differentiate into the TEMp subset. More specifically, as can be inferred from the best fit parameters (Table 4.1), models 2, 4, 5 and 6 (marked with asterisks in Figs. 4.5 and 4.4) would predict the absence of the TEMp subset in 42%, 50%, 52% and 56% of the single-cell progenies, respectively. In the data, TEMp cells are observed in every single cell-derived progeny population (cf. Fig. 3.5). Therefore those models are also not in accordance with the data and we are again left with the same two models (Fig. 4.4C) that we found in the previous ranking.

model #	$d0_1 (d^{-1})$	$d0_2 (d^{-1})$	$d0_3 (d^{-1})$	w/o TEMp
2	0.43	0.59	0	42%
4	0.50	0.51	0	50%
5	0.50	0.46	0	52%
6	0.57	0.45	0	56%

Table 4.1.: best fit values of the recruitment rates $d0_i$ and the resulting predicted fraction of progeny populations without TEMp cells for the models marked with asterisks in Figs. 4.5 and 4.4.

4.3. Identifiability of the generalized differentiation model

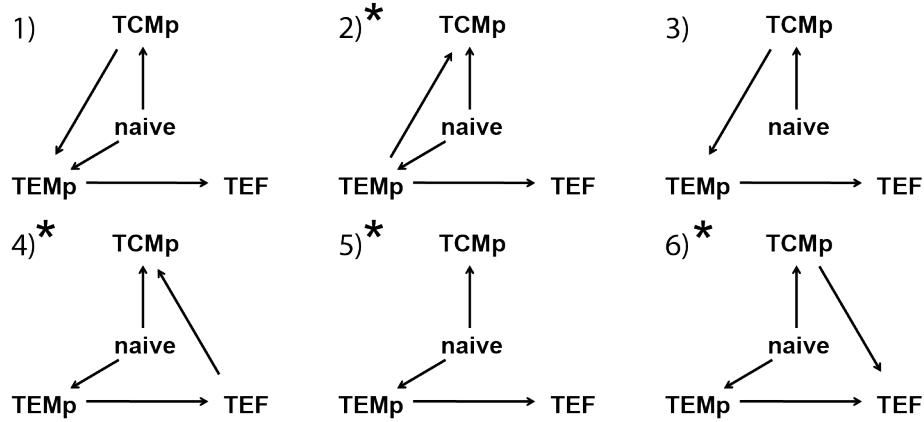


Figure 4.5.: Schemes of the models ranked one to six using just the data of day 6 and 8 (Fig. 4.4). The best fit parameters of models 2, 4, 5 and 6 (marked with asterisks) would predict the absence of the TEMp subset in 42%, 50%, 52% and 56% of the single-cell progenies, respectively, and are therefore also excluded by the data.

4.3. Identifiability of the generalized differentiation model

In the previous section we introduced a generalized model of CD8⁺ T cell differentiation (Fig. 4.2). We showed that all 304 sub-models that can be generated from this scheme either cannot account for the data, or, due to one or several best fit values for the rate constants being zero, reduce to one of two closely related models. One of these models is the N→TCMp→TEMp→TEF model. In the other model, differentiation follows the same scheme except that about 10% of the naïve cells differentiate directly into the TEMp compartment. The schemes of these two models thus form the minimal necessary core modules of the differentiation pathway.

It also follows from the analysis of the previous section that the best fit estimates for the rates of additional differentiation steps must be zero. However, based on this analysis, no conclusions can be made about the maximum size of these rates that would still be in agreement with the data. We will now again use the profile likelihood method to show that except for the TEF→TEMp differentiation rate, the rate constants of all other additional differentiation steps are essentially zero due to their very low upper bounds.

Fig. 4.6 shows the best fit of the full model (scheme depicted in Fig. 4.2) to the data. The corresponding parameter estimates are given in Table 4.2. Because the best fit estimate for the rate constants d_{-1} , d_{-2} , d_3 and d_{-1} is zero, the best fit curves depicted in Fig. 4.6 correspond to the N→TCMp→TEMp→TEF model with the additional N→TEMp differentiation step. If we compare the best fit of this model with the best fit of the N→TCMp→TEMp→TEF model (Fig. 3.10), a slight improvement for the estimate of the day 8 coefficient of variation of the TCMP subset

4. Systematic comparison of possible differentiation models

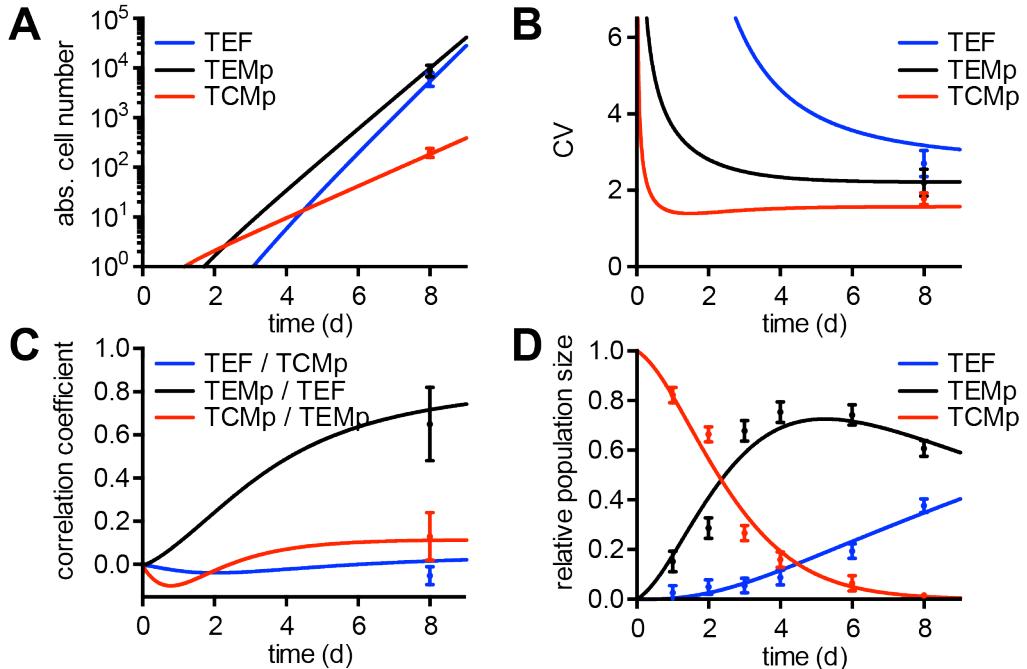


Figure 4.6.: Fit of the full model depicted in Fig. 4.2 to mean values (A), CVs (B) and pairwise correlations (C) of the single cell progeny data of day 8 p.i. and to the relative subset size kinetics (D; day 1 to 4: transfer of 10^5 OT-I cells, day 6 and 8: transfer of 100 OT-I cells). (Experiments by V. Buchholz and D. Busch)

is noticeable. The reason behind this is that the additional N \rightarrow TEMp differentiation step allows for a smaller overall recruitment rate d_0 which in turn leads to an increase in the variability of the TCMP absolute cell number.

To assess the parameter confidence bounds using the profile likelihood method [59] (cf. chapter 3.6)), we again plot the change in the χ^2 -function $\Delta\chi^2$ that is obtained for each parameter upon fixing the respective parameter at varying values while minimizing over the remaining ones (Fig. 4.7). All parameters are identifiable within narrow bounds. Note that also the confidence interval for the recruitment rate d_0 is identifiable in this model (cf. 3.11): A very fast recruitment rate d_0 would result in an estimate for the CV of the absolute size of the TCMP subset that would be too low.

The resulting confidence intervals are given in Table 4.2. Thus, according to the 95% confidence interval estimates, upon recruitment less than one percent of the naïve cells would develop directly into TEF cells, and more than about 80% of the naïve cells would differentiate into the TCMP compartment. The rate constants for TEMp \rightarrow TCMP (d_{-1}), TCMP \rightarrow TEF (d_{-3}) and TEF \rightarrow TCMP (d_3) have all very low upper bounds ($d_{-1} < 4 \cdot 10^{-4} \text{ d}^{-1}$, $d_{-3} < 8 \cdot 10^{-3} \text{ d}^{-1}$, $d_3 < 6 \cdot 10^{-4} \text{ d}^{-1}$). Those differentiation events are thus very rare and do not play a significant role in this

4.4. Incorporating cell death into the model

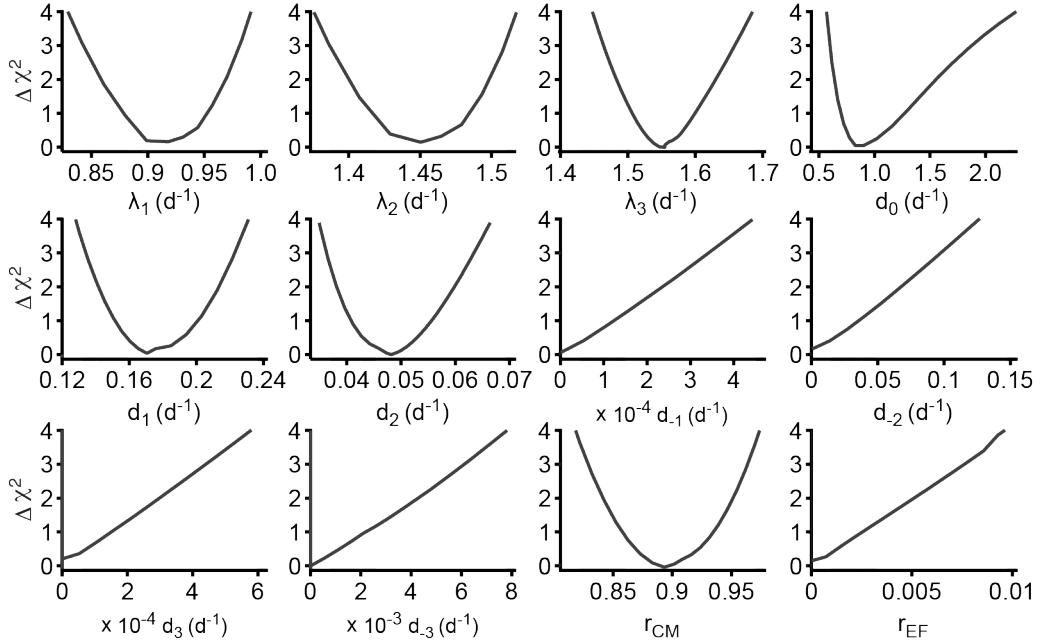


Figure 4.7.: Profile likelihood method for the full model depicted in Fig. 4.2: Change in χ^2 -function obtained upon fixing each parameter at the abscissa values and minimizing over the remaining ones.

type of T cell diversification models. On the other hand, the rate constant for TEF \rightarrow TEMp differentiation events (with $d_{-2} < 0.12 \text{ d}^{-1}$) does not necessarily have to be small. So even though the best fit estimate for the corresponding rate is zero, with the data at hand we cannot exclude that TEF cells occasionally differentiate back to the TEMp compartment.

4.4. Incorporating cell death into the model

So far we have neglected cell death in our modeling approaches. We will now introduce constant subset-dependent death rates and show that the two main conclusions from the last sections still hold: Firstly, the core modules of the differentiation pathway (N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF with or without direct N \rightarrow TEMp differentiation) are unchanged. Secondly, the only non-rare additional differentiation step in accordance with the data is TEF \rightarrow TEMp.

Incorporating subset-specific death rates δ_i into the general differentiation model results in the scheme depicted in Fig. 4.8. To derive the corresponding PDE of the probability generating function we use eqn. 3.4 to extend the PDE of the probability

4. Systematic comparison of possible differentiation models

Parameter	best fit	68% conf.	95% conf.
TCMp proliferation rate (λ_1)	0.92 d ⁻¹	(0.88, 0.953) d ⁻¹	(0.83, 0.99) d ⁻¹
TEMp proliferation rate (λ_2)	1.45 d ⁻¹	(1.42, 1.48) d ⁻¹	(1.39, 1.52) d ⁻¹
TEF proliferation rate (λ_3)	1.55 d ⁻¹	(1.51, 1.60) d ⁻¹	(1.45, 1.68) d ⁻¹
recruitment rate (d_0)	0.9 d ⁻¹	(0.7, 1.7) d ⁻¹	(0.6, 2.2) d ⁻¹
TCMp \rightarrow TEMp rate (d_1)	0.17 d ⁻¹	(0.15, 0.20) d ⁻¹	(0.13, 0.23) d ⁻¹
TEMp \rightarrow TEF rate (d_2)	0.048 d ⁻¹	(0.041, 0.056) d ⁻¹	(0.035, 0.066) d ⁻¹
TEF \rightarrow TEMp rate (d_{-2})	0 d ⁻¹	(0, 0.035) d ⁻¹	(0, 0.122) d ⁻¹
TEMp \rightarrow TCMP rate (d_{-1})	0 d ⁻¹	(0, 1.3) 10 ⁻⁴ d ⁻¹	(0, 4.3) 10 ⁻⁴ d ⁻¹
TEF \rightarrow TCMP rate (d_3)	0 d ⁻¹	(0, 1.6) 10 ⁻⁴ d ⁻¹	(0, 56) 10 ⁻⁴ d ⁻¹
TCMP \rightarrow TEF rate (d_{-3})	0 d ⁻¹	(0, 2.0) 10 ⁻³ d ⁻¹	(0, 7.4) 10 ⁻⁴ d ⁻¹
ratio N \rightarrow TCMP (r_{CM})	0.90	(0.86, 0.94)	(0.82, 0.97)
ratio N \rightarrow TEF (r_{EF})	0	(0, 3.0) 10 ⁻³	(0, 9.3) 10 ⁻³

Table 4.2.: best fit parameters and confidence intervals (profile likelihood method) of the full differentiation model depicted in Fig. 4.2 based on the fit to the single cell-derived progenies data of day 8 p.i. and the relative subset size kinetics.

generating function of the general model without cell death (eqn. 4.3). This yields

$$\partial_t F(t) = \sum_{i=1}^3 \left\{ \lambda_i (z_i^2 - z_i) \partial_{z_i} F + \delta_i (1 - z_i) \partial_{z_i} F + d_0 i (z_i - z_0) \partial_{z_0} F + d_i (z_{i+1} - z_i) \partial_{z_i} F + d_{-i} (z_i - z_{i+1}) \partial_{z_{i+1}} F \right\}, \quad (4.5)$$

where we again have set $z_4 = z_1$ for notational convenience. For the ODE system of the first and second derivatives of F evaluated at $z_i = 1$ see appendix A.1.

To systematically rank the best fits of the 304 sub-models that result from setting a subset of the differentiation rates to zero, we follow the same steps as chapter

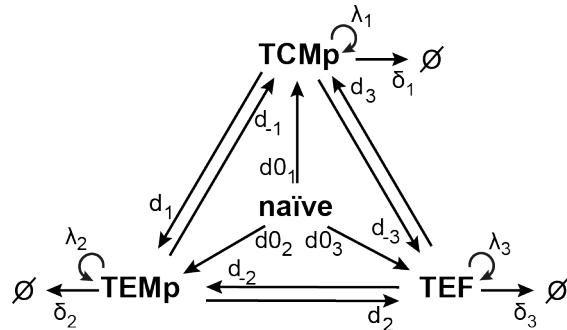


Figure 4.8.: Model scheme allowing for all possible diversification pathways from naïve to TCMP, TEMp and TEF as well as subset-specific proliferation and death rates.

4.4. Incorporating cell death into the model

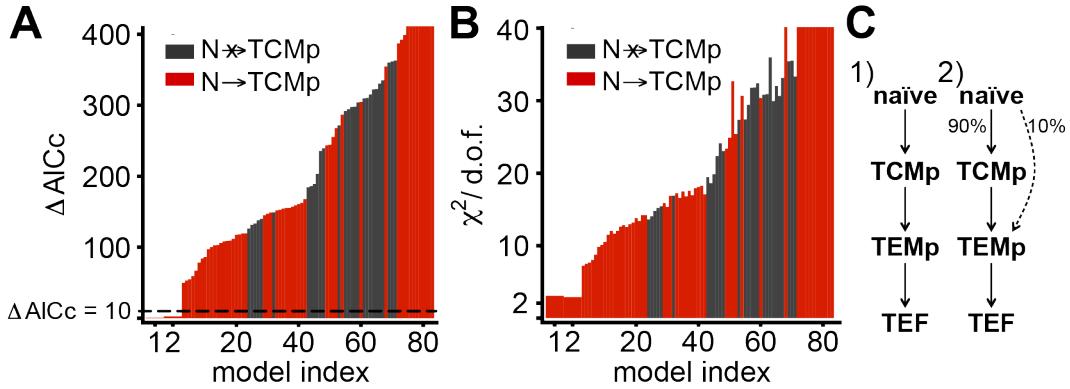


Figure 4.9.: As in Fig. 4.3 but incorporating cell death. (A) Model ranking based on ΔAICc . (B) Same ranking but showing the corresponding goodness of fit using χ^2_{\min} over degrees of freedom. Red bars: models including a naïve \rightarrow TCMP differentiation step; black bars: other models. (C) Scheme of the two models that are in accordance with the data.

4.2. In this case, 83 topologically different models are left over after the parameter estimation procedure (again for many models one or more best fit estimates for the rates are zero). The ranking of these models is shown in Fig. 4.9. Both in terms of the Akaike information criterion (Fig. 4.9A) and χ^2 over degrees of freedom (Fig. 4.9B) we find the two already discussed core modules of the differentiation pathway (Fig. 4.9C).

Interestingly, here, the $N \rightarrow \text{TCMP} \rightarrow \text{TEMP} \rightarrow \text{TEF}$ model is ranked before the model that additionally allows for direct $N \rightarrow \text{TEMP}$ differentiation. The reason for this is that the death rate δ_1 leads to an increase of the CV of the TCMP absolute cell number without having to lower the recruitment rate d_0 . The simpler model can thereby produce an equally good fit of the observed CVs. We can see this as follows. A very simple expression for the CV of both models can be obtained if the $N \rightarrow \text{TCMP}$ step is ignored (corresponding to the limit of a very fast recruitment rate d_0). In this case we can use eqn. 3.11 to derive (for sufficiently late times t)

$$CV_1 = \sqrt{\frac{\lambda_1 + d_1 + \delta_1}{\lambda_1 - d_1 - \delta_1}}. \quad (4.6)$$

So, a non-zero death rate δ_1 for the TCMP cells leads to an increase of CV_1 . This also means that the effect of a fast recruitment rate on CV_1 can be balanced by an increase in δ_1 . Consequently, as we will see next, the recruitment rate d_0 is no longer identifiable if non-zero death rates are introduced.

As expected, the profile likelihood method applied to the full model with cell death reveals that the proliferation and death rates can only be estimated with great uncertainty (Fig. 4.10 and Table 4.3; see chapter 3.6 for a description of the method). Nevertheless, only the recruitment rate d_0 is non-identifiable (for the reason dis-

4. Systematic comparison of possible differentiation models

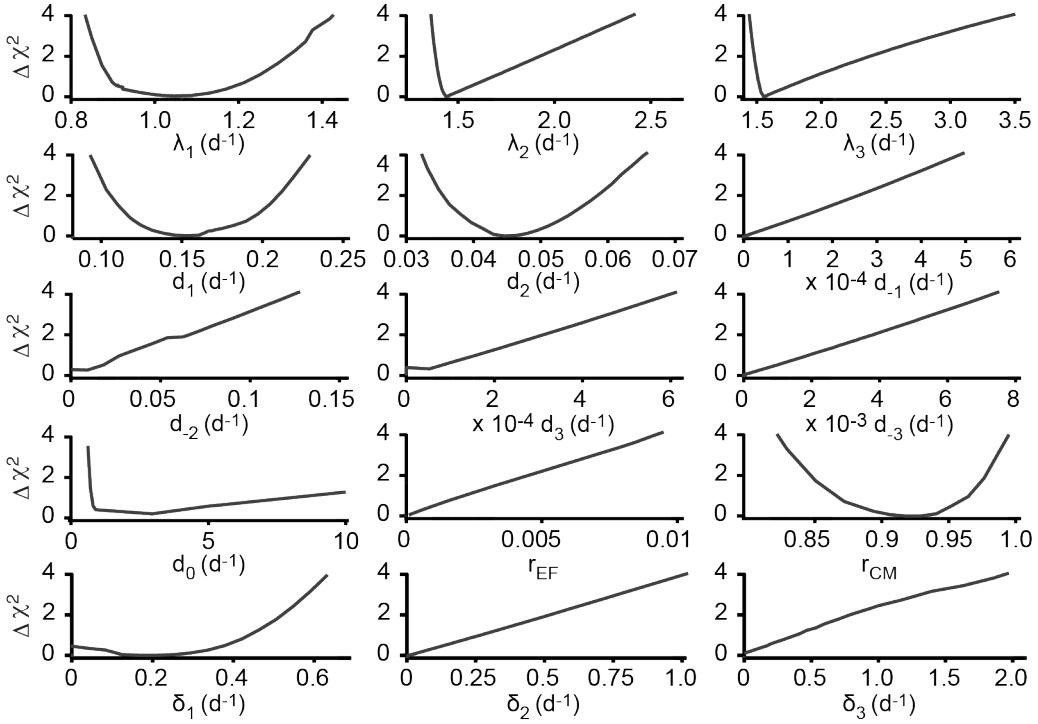


Figure 4.10.: Profile likelihood method for the full model with cell death depicted in Fig. 4.8: Change in χ^2 -function obtained upon fixing each parameter at the abscissa values and minimizing over the remaining ones.

cussed above). The best fit parameter values for all other differentiation rates as well as their accuracies are almost unchanged compared to the fit of the same model without cell death (cf. Fig. 4.7). This is also true for the estimates of the ratios of N \rightarrow TCMp (r_{CM}) and N \rightarrow TEF (r_{EF}) transitions. The best fit estimates for the death rates of the TEMp and TEF subset are zero. Also the death rate of the TCMP cells δ_1 is not significantly different from zero. However, its best fit estimate is not zero due to the reason discussed above (cf. eqn. 4.6).

Most importantly, we find the same small upper bounds for TEMp \rightarrow TCMp (d_{-1}), TCMP \rightarrow TEF (d_{-3}) and TEF \rightarrow TCMP (d_3) differentiation steps. Thus, also if cell death is explicitly taken into account we can still conclude that in addition to differentiation along the pathway N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF, this model would only allow N \rightarrow TEMp and TEF \rightarrow TEMp differentiation events to occur at significant rates.

4.5. Incorporating asymmetric cell division into the model

Parameter	best fit	68% conf.	95% conf.
TCMp proliferation rate (λ_1)	1.05 d^{-1}	$(0.89, 1.24) \text{ d}^{-1}$	$(0.84, 1.42) \text{ d}^{-1}$
TEMp proliferation rate (λ_2)	1.43 d^{-1}	$(1.40, 1.7) \text{ d}^{-1}$	$(1.36, 2.4) \text{ d}^{-1}$
TEF proliferation rate (λ_3)	1.55 d^{-1}	$(1.49, 1.91) \text{ d}^{-1}$	$(1.44, 3.3) \text{ d}^{-1}$
recruitment rate (d_0)	3 d^{-1}	$(0.8, 8) \text{ d}^{-1}$	$(0.5, \infty) \text{ d}^{-1}$
TCMp \rightarrow TEMp rate (d_1)	0.15 d^{-1}	$(0.12, 0.20) \text{ d}^{-1}$	$(0.09, 0.23) \text{ d}^{-1}$
TEMp \rightarrow TEF rate (d_2)	0.045 d^{-1}	$(0.038, 0.054) \text{ d}^{-1}$	$(0.033, 0.065) \text{ d}^{-1}$
TEF \rightarrow TEMp rate (d_{-2})	0 d^{-1}	$(0, 0.028) \text{ d}^{-1}$	$(0, 0.12) \text{ d}^{-1}$
TEMp \rightarrow TCMP rate (d_{-1})	0 d^{-1}	$(0, 1.6) 10^{-4} \text{ d}^{-1}$	$(0, 4.7) 10^{-4} \text{ d}^{-1}$
TEF \rightarrow TCMP rate (d_3)	0 d^{-1}	$(0, 1.6) 10^{-4} \text{ d}^{-1}$	$(0, 5.8) 10^{-4} \text{ d}^{-1}$
TCMP \rightarrow TEF rate (d_{-3})	0 d^{-1}	$(0, 1.8) 10^{-3} \text{ d}^{-1}$	$(0, 7.0) 10^{-3} \text{ d}^{-1}$
ratio N \rightarrow TCMP (r_{CM})	0.92	$(0.87, 0.96)$	$(0.82, 0.99)$
ratio N \rightarrow TEF (r_{EF})	0	$(0, 0.002)$	$(0, 0.009)$
TCMP death rate (δ_1)	0.2 d^{-1}	$(0, 0.43) \text{ d}^{-1}$	$(0, 0.63) \text{ d}^{-1}$
TEMp death rate (δ_2)	0 d^{-1}	$(0, 0.24) \text{ d}^{-1}$	$(0, 0.95) \text{ d}^{-1}$
TEF death rate (δ_3)	0 d^{-1}	$(0, 0.35) \text{ d}^{-1}$	$(0, 1.89) \text{ d}^{-1}$

Table 4.3.: best fit parameters and confidence intervals (profile likelihood method) of the full differentiation model with cell death depicted in Fig. 4.8 based on the fit to the single cell-derived progenies data of day 8 p.i. and the relative subset size kinetics.

4.5. Incorporating asymmetric cell division into the model

It has been suggested by Chang et al. [14] that the first cell division of the naïve cell following activation could be asymmetric, with one daughter cell being prone to generate descendants of memory phenotype whereas the other one would generate a population of effectors.

This motivates the following extension of the class of differentiation models that we have considered so far. Consider the scheme in Fig. 4.11: Upon recruitment (rate d_0), the naïve cell divides asymmetrically into two daughter cells of differing subtypes, e.g. give rise to one TCMP and one TEF cell. In this scheme, each of the three states A, B and C can stand for either TCMP, TEMp or TEF.

Based on eqns. 3.3, 3.5 and 3.6 we readily find for the PDE of the moment generating function F

$$\begin{aligned} \partial_t F = d_0 (z_1 z_2 - z_0) \partial_{z_0} F + \sum_{i=1}^3 \lambda_i (z_i^2 - z_i) \partial_{z_i} F + \sum_{i=1}^2 d_i (z_{i+1} - z_i) \partial_{z_i} F \\ + \sum_{i=1}^2 d_{-i} (z_i - z_{i+1}) \partial_{z_{i+1}} F. \end{aligned} \quad (4.7)$$

4. Systematic comparison of possible differentiation models

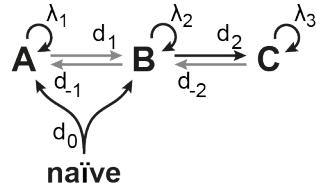


Figure 4.11.: Scheme of a model that allows the first division of the naïve cell to be asymmetric, giving rise to one cell of type A and one cell of type B. Each of the three pairwise different states A, B and C can stand for either TCMP, TEMP or TEF. Grey arrows indicate optional differentiation steps.

See appendix A.2 for the ODE system of the first and second derivatives of F evaluated at $z_i = 1$.

For this class of models, having $d_2 > 0$ is sufficient to generate models in which each of the three states TCMP, TEMP and TEF is reachable from the naïve state. The other three differentiation steps (with rates d_1 , d_{-1} and d_{-2}) are optional. Taking into account the six possible permutations among the three states therefore leads to $6 \times 8 = 48$ different sub-models that can be generated from the scheme in Fig. 4.11.

As described in chapter 4.2, we now fit each of these models to the single cell progeny data of day 8 p.i. and the relative subset size kinetics of the TCMP and TEF subsets (using simulated annealing and local optimization with at least 300 different random initial values for each model). Subsequently, we rank the best fit of the models based on the Akaike information criterion together with the class of models that we discussed previously (Fig. 4.12; cf. chapter 4.2). The first asymmetric cell division model (naïve \rightarrow TCMP + TEMP, TCMP \rightarrow TEMP \rightarrow TEF) is ranked third,

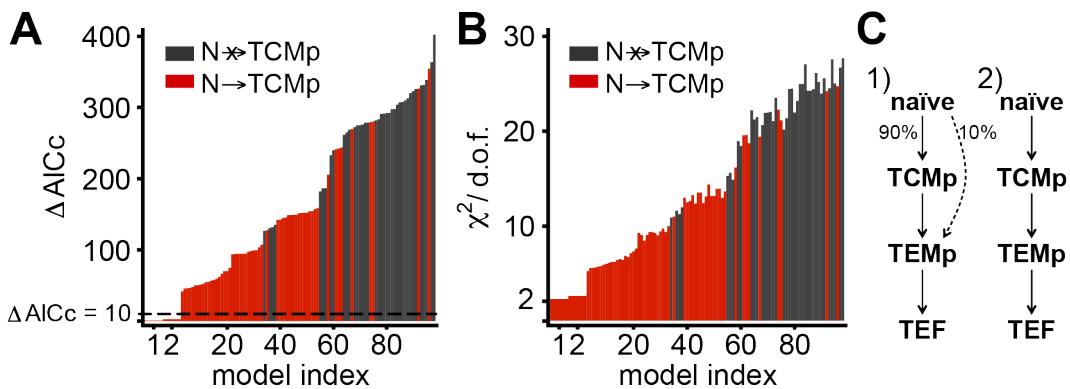


Figure 4.12.: As in Fig. 4.3 but including asymmetric cell division models (Fig. 4.11). (A) Model ranking based on AICc. (B) Same ranking but showing the corresponding goodness of fit using χ^2_{\min} over degrees of freedom. Red bars: models including a naïve \rightarrow TCMP differentiation step; black bars: other models. (C) Scheme of the two models that are in accordance with the data.

4.5. Incorporating asymmetric cell division into the model

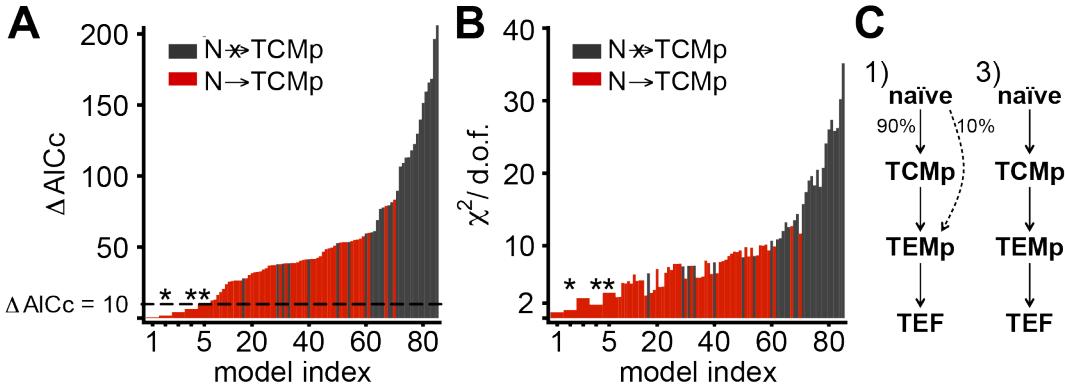


Figure 4.13.: As in Fig. 4.4 but including asymmetric cell division models (Fig. 4.11). (A) Model ranking based on AICc . (B) Same ranking but showing the corresponding goodness of fit using χ^2_{\min} over degrees of freedom. Red bars: models including a naïve $\rightarrow \text{TCMp}$ differentiation step; black bars: other models. (C) Scheme of the two models that are in accordance with the data.

but with an AICc difference to the best fitting model of almost 50 (Fig. 4.12A) and $\chi^2_{\min}/\text{d.o.f.} \approx 5$. Obviously, none of the asymmetric cell division models produces a satisfying fit to the data.

As in chapter 4.2, we also repeated the ranking based solely on the data of day 6 and day 8 where we do not have to rely on the experiments that use a high number of adoptively transferred cells. The resulting ranking is shown in Fig. 4.13. Also in this case there is no asymmetric cell division model among the models with $\Delta \text{AICc} < 10$ (five models fulfill this criterion). Note that the models ranked second, fourth and fifth (marked with asterisks in Fig. 4.13) are also excluded by the data since their respective best fit estimates for the rates d_0 would predict the absence of the TEMP subset in 42%, 50% and 52% of the single-cell progenies, respectively (cf. Fig. 4.5; this was already discussed in chapter 4.2).

The first asymmetric cell division model is found at rank six with an AICc difference of 12 to the best fitting model. However, with $\chi^2_{\min}/\text{d.o.f.} \approx 3$, this model still provides a mediocre fit to the data of day 6 and 8. In this model, the naïve cell divides asymmetrically into one TCMp and one TEMP cell and then differentiation follows $\text{TCMp} \rightarrow \text{TEMP} \leftrightarrow \text{TEF}$. Both, this model and the model found at rank three in the previous ranking (cf. Fig. 4.12) above are very closely related to the linear differentiation models that we have discussed in the previous sections. However, the model assumption that every naïve cell divides asymmetrically is obviously too strong. In the next section we will therefore present a model that allows for both symmetric and asymmetric divisions of the naïve cells.

4. Systematic comparison of possible differentiation models

4.6. More refined asymmetric cell division models

None of the asymmetric models that we analyzed in the last section provided a satisfying fit to the whole data set. However, a mediocre fit to the data of day 6 and 8 p.i. could be obtained by a model in which differentiation along the linear pathway $\text{TCMp} \rightarrow \text{TEMp} \leftrightarrow \text{TEF}$ is induced by a naïve cell that asymmetrically divides into one TCMp and one TEMp cell. On the other hand, we already know that a good fit to the whole data set can be obtained by a model in which differentiation follows $\text{TCMp} \rightarrow \text{TEMp} \rightarrow \text{TEF}$ after the activated naïve cell differentiated either into one TCMp or one TEMp cell.

Asymmetrically and symmetrically dividing naïve cells

In an attempt to combine both models, we considered the model scheme depicted in Fig. 4.14. In this model, following recruitment (rate d_0), the activated naïve cell can either divide symmetrically into two TCMp (ratio rS_1) or two TEMp cells (ratio rS_2) or it can divide asymmetrically into cells of both phenotypes (ratio rA). After that, differentiation follows the linear pathway $\text{TCMp} \rightarrow \text{TEMp} \rightarrow \text{TEF}$.

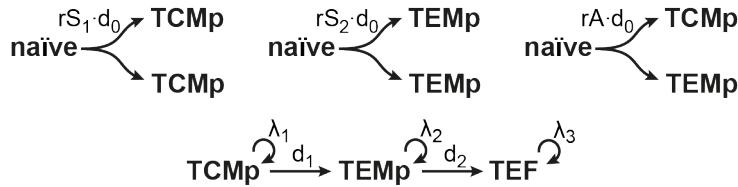


Figure 4.14.: Scheme of an asymmetric cell division model in which the activated naïve cell (recruitment rate d_0) can either divide symmetrically into two TCMp (ratio rS_1) or two TEMp cells (ratio rS_2). Alternatively, the naïve cell can divide asymmetrically into two cells of both phenotypes (ratio rA). Otherwise differentiation follows $\text{TCMp} \rightarrow \text{TEMp} \rightarrow \text{TEF}$.

The PDE for the moment generating function is constructed in the usual way using eqns. 3.3, 3.5 and 3.6. The result is

$$\begin{aligned} \partial_t F = & d_0 (rS_1 z_1^2 + rS_2 z_2^2 + rA z_1 z_2 - z_0) \partial_{z_0} F \\ & + \sum_{i=1}^3 \lambda_i (z_i^2 - z_i) \partial_{z_i} F + \sum_{i=1}^2 d_i (z_{i+1} - z_i) \partial_{z_i} F, \end{aligned} \quad (4.8)$$

with

$$rA + rS_1 + rS_2 = 1. \quad (4.9)$$

Analogously to chapter 3.2, in order to obtain an ODE system for the time dependence of the mean values, we successively differentiate this equation with respect to

the z_i and then set $z_i = 1$. This results in

$$\begin{aligned}\dot{\langle n_0 \rangle} &= -d_0 \langle n_0 \rangle \\ \dot{\langle n_1 \rangle} &= (\lambda_1 - d_1) \langle n_1 \rangle + d_0 (rA + 2rS_1) \langle n_0 \rangle \\ \dot{\langle n_2 \rangle} &= (\lambda_2 - d_2) \langle n_2 \rangle + d_0 (rA + 2rS_2) \langle n_0 \rangle + d_1 \langle n_1 \rangle \\ \dot{\langle n_3 \rangle} &= d_2 \langle n_2 \rangle + \lambda_3 \langle n_3 \rangle,\end{aligned}\tag{4.10}$$

where $\langle n_i \rangle = \partial_{z_i} F|_1$. We further obtain the following ODEs for the time evolution of the second derivatives of F evaluated at $z_i = 1$

$$\begin{aligned}\dot{F}_{00} &= -2d_0 F_{00} \\ \dot{F}_{03} &= d_2 F_{02} - F_{03} (d_0 - \lambda_3) \\ \dot{F}_{33} &= 2(d_2 F_{23} + \lambda_3 (F_{33} + \langle n_3 \rangle)) \\ \dot{F}_{02} &= d_1 F_{01} - F_{02} (d_0 + d_2 - \lambda_2) + d_0 F_{00} (rA + 2rS_2) \\ \dot{F}_{23} &= d_1 F_{13} + F_{23} (-d_2 + \lambda_2 + \lambda_3) + d_0 F_{03} (rA + 2rS_2) + d_2 F_{22} \\ \dot{F}_{22} &= 2(d_1 F_{12} + d_0 (F_{02} (rA + 2rS_2) + rS_2 \langle n_0 \rangle) - d_2 F_{22} + \lambda_2 (F_{22} + \langle n_2 \rangle)) \quad (4.11) \\ \dot{F}_{01} &= d_0 F_{00} (rA + 2rS_1) - F_{01} (d_0 + d_1 - \lambda_1) \\ \dot{F}_{13} &= (-d_1 + \lambda_1 + \lambda_3) F_{13} + d_2 F_{12} + d_0 F_{03} (rA + 2rS_1) \\ \dot{F}_{12} &= d_0 ((rA + 2rS_2) F_{01} + rA (F_{02} + \langle n_0 \rangle) + 2rS_1 F_{02}) + d_1 F_{11} \\ &\quad + (-d_1 - d_2 + \lambda_1 + \lambda_2) F_{12} \\ \dot{F}_{11} &= 2((\lambda_1 - d_1) F_{11} + d_0 ((rA + 2rS_1) F_{01} + rS_1 \langle n_0 \rangle) + \lambda_1 \langle n_1 \rangle),\end{aligned}$$

where $F_{ij} = \partial_{z_i} \partial_{z_j} F|_1$ (see eqn. 3.19 for the initial conditions). The fit of this model to the data is then obtained as in chapter 3.2 (Fig. 4.15). Thus, the model provides a good fit to the relative subset size kinetics as well as to the summary statistics of the single cell progeny data of day 8 p.i. But, it turns out that the best fit estimate for the ratio of asymmetric cell divisions (rA) is zero (Table 4.4).

To get also an upper bound on rA , we again apply the profile likelihood method (cf. chapter 3.6). The resulting deviance plots are shown in Fig. 4.16. All model parameters are identifiable within narrow bands (Table 4.4). Except for the ratio of asymmetric divisions rA , all parameter estimates are similar to the corresponding estimates of the general differentiation model that we discussed in chapter 4.3.

At first glance, it might be surprising that the deviance plot for rA is not more similar to the one for rS_2 . After all, both parameters have the same effect (up to a factor of two) on the mean values of the subsets. This can be seen by rewriting eqn.

4. Systematic comparison of possible differentiation models

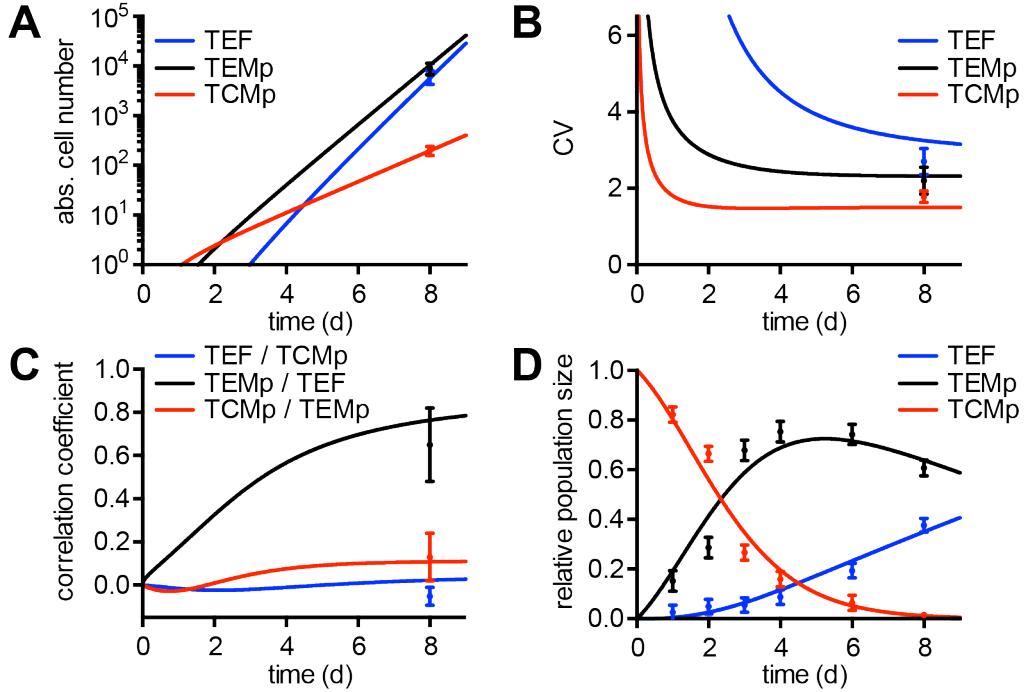


Figure 4.15.: Fit of the asymmetric cell division model depicted in Fig. 4.14 to mean values (A), CVs (B) and pairwise correlations (C) of the single cell progeny data of day 8 p.i. and to the relative subset size kinetics (D; day 1 to 4: transfer of 10⁵ OT-I cells, day 6 and 8: transfer of 100 OT-I cells). (Experiments by V. Buchholz and D. Busch)

Parameter	best fit	68% conf.	95% conf.
TCMP proliferation rate (λ_1)	0.87 d^{-1}	$(0.83, 0.90) \text{ d}^{-1}$	$(0.79, 0.93) \text{ d}^{-1}$
TEMP proliferation rate (λ_2)	1.4 d^{-1}	$(1.37, 1.44) \text{ d}^{-1}$	$(1.34, 1.47) \text{ d}^{-1}$
TEF proliferation rate (λ_3)	1.53 d^{-1}	$(1.47, 1.57) \text{ d}^{-1}$	$(1.41, 1.61) \text{ d}^{-1}$
recruitment rate (d_0)	0.44 d^{-1}	$(0.38, 0.56) \text{ d}^{-1}$	$(0.34, 0.71) \text{ d}^{-1}$
TCMP → TEMP rate (d_1)	0.15 d^{-1}	$(0.13, 0.18) \text{ d}^{-1}$	$(0.11, 0.21) \text{ d}^{-1}$
TEMP → TEF rate (d_2)	0.045 d^{-1}	$(0.038, 0.053) \text{ d}^{-1}$	$(0.033, 0.064) \text{ d}^{-1}$
ratio asym. div. (r_A)	0	$(0, 0.04)$	$(0, 0.14)$
ratio sym. div. to TEMP (r_{S2})	0.17	$(0.13, 0.20)$	$(0.10, 0.23)$

Table 4.4.: best fit parameters and confidence intervals (profile likelihood method) of the asymmetric cell division model depicted in Fig. 4.14 based on the fit to the single cell-derived progenies data of day 8 p.i. and the relative subset size kinetics.

4.10 using $r_{S1} = 1 - r_{S2} - r_A$:

$$\begin{aligned}\dot{\langle n_0 \rangle} &= -d_0 \langle n_0 \rangle \\ \dot{\langle n_1 \rangle} &= (\lambda_1 - d_1) \langle n_1 \rangle - d_0 (r_A + 2r_{S2} - 2) \langle n_0 \rangle \\ \dot{\langle n_2 \rangle} &= (\lambda_2 - d_2) \langle n_2 \rangle + d_0 (r_A + 2r_{S2}) \langle n_0 \rangle + d_1 \langle n_1 \rangle \\ \dot{\langle n_3 \rangle} &= d_2 \langle n_2 \rangle + \lambda_3 \langle n_3 \rangle.\end{aligned}\tag{4.12}$$

4.6. More refined asymmetric cell division models

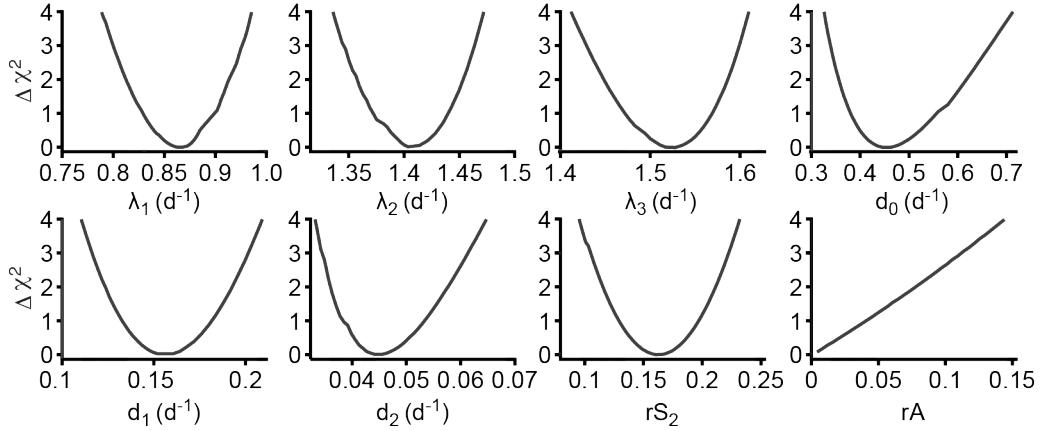


Figure 4.16.: Profile likelihood method for the asymmetric cell division model depicted in Fig. 4.14: Change in χ^2 -function obtained upon fixing each parameter at the abscissa values and minimizing over the remaining ones.

So why is the best fit estimate for rA zero while we find a lower bound for rS_2 ? One reason for this is that the two parameters have an opposing effect on the correlation between the absolute size of TCMP and the two other subsets. The observed correlations for TCMP-TEMP and TCMP-TEF are small (cf. Fig. 4.15). Increasing the ratio of naïve \rightarrow 2 TEMP transitions rS_2 naturally leads to a decrease in the correlation between TCMP and the two other subsets. In contrast, increasing the ratio of asymmetric divisions naïve \rightarrow TEMP + TCMP will also increase the correlations between TCMP and the TEMP and TEF subset (early / late recruitment increases the probability for both TCMP and the other two subsets to become bigger / smaller). The predicted low percentage of asymmetric divisions thus has to do with the low correlation that is observed between TCMP and the other subsets.

It should also be noted that it is evident from the ODE system for the first moments (eqn. 4.12) that basically no inference about any of the ratios rA , rS_1 and rS_2 could be made by basing the parameter estimation solely on the mean values: Since $\langle n_0 \rangle$ is not directly measured, d_0 could not be estimated and therefore only the whole term ($d_0 rA + 2d_0 rS_2$) would be accessible for parameter estimation.

The upper bound on the percentage of asymmetric cell divisions rA is found to be approximately 15%. This low percentage of allowed asymmetric divisions is an interesting model prediction that could be checked in future experiments.

Asymmetrically and symmetrically dividing naïve and TCMP cells

Initiation of asymmetric cell division in T cells is thought to require interactions with dendritic cells (DCs) that provide the necessary polarity cue [45]. It is known that

4. Systematic comparison of possible differentiation models

the early progenies of activated T cells still engage in brief interactions with DCs [42]. If T cell differentiation follows the linear pathway N→TCMp→TEMp→TEF, it is therefore conceivable that asymmetric cell divisions also occur in the population of the TCMP cells.

Fig. 4.17 shows the scheme of a model where in addition to the naïve cells also TCMP cells can divide asymmetrically. In this model, TCMP cells and naïve cells divide either symmetrically into two TCMP (ratio rS_1) or two TEMp cells (ratio rS_2) or divide asymmetrically into two TEMp cells (ratio rA). The ratios of the three differentiation events are supposed to be the same for TCMP and naïve cells.

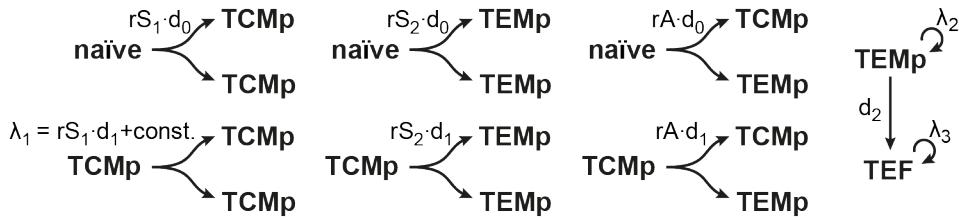


Figure 4.17.: Scheme of an asymmetric cell division model in which both, the activated naïve cell (recruitment rate d_0) and each TCMP cell can either divide symmetrically into two TCMP (ratio rS_1) or two TEMp cells (ratio rS_2) or can divide asymmetrically into two cells of the TCMP and TEMp phenotype (ratio rA). TEMp cells further differentiate into TEF cells at a rate d_2 .

The PDE for the moment generating function can be constructed analogously to the previously discussed model

$$\begin{aligned} \partial_t F = & d_0 (rS_1 z_1^2 + rS_2 z_2^2 + rA z_1 z_2 - z_0) \partial_{z_0} F \\ & + d_1 (rS_2 z_2^2 + rA z_1 z_2 - z_1) \partial_{z_1} F + d_2 (z_3 - z_2) \partial_{z_2} F \\ & + \sum_{i=1}^3 \lambda_i (z_i^2 - z_i) \partial_{z_i} F. \end{aligned} \quad (4.13)$$

For the time evolution of the mean values we obtain

$$\begin{aligned} \dot{\langle n_0 \rangle} &= -d_0 \langle n_0 \rangle \\ \dot{\langle n_1 \rangle} &= d_0 (rA + 2rS_1) \langle n_0 \rangle + (\lambda_1 - d_1 rS_2) \langle n_1 \rangle \\ \dot{\langle n_2 \rangle} &= (\lambda_2 - d_2) \langle n_2 \rangle + (rA + 2rS_2) (d_0 \langle n_0 \rangle + d_1 \langle n_1 \rangle) \\ \dot{\langle n_3 \rangle} &= d_2 \langle n_2 \rangle + \lambda_3 \langle n_3 \rangle. \end{aligned} \quad (4.14)$$

It is again worthwhile to explicitly rewrite these equation using $rA + rS_1 + rS_2 = 1$:

$$\begin{aligned} \dot{\langle n_0 \rangle} &= -d_0 \langle n_0 \rangle \\ \dot{\langle n_1 \rangle} &= (\lambda_1 - d_1 rS_2) \langle n_1 \rangle - d_0 (rA + 2rS_2 - 2) \langle n_0 \rangle \\ \dot{\langle n_2 \rangle} &= (\lambda_2 - d_2) \langle n_2 \rangle + (rA + 2rS_2) (d_0 \langle n_0 \rangle + d_1 \langle n_1 \rangle) \\ \dot{\langle n_3 \rangle} &= d_2 \langle n_2 \rangle + \lambda_3 \langle n_3 \rangle. \end{aligned} \quad (4.15)$$

4.6. More refined asymmetric cell division models

As expected, also in this model, information about the ratios of asymmetric and symmetric cell divisions can therefore just be obtained by taking into account the second order moments in the parameter estimation.

The ODEs for the time evolution of the second derivatives of F evaluated at $z_i = 1$ read

$$\begin{aligned}
\dot{F}_{00} &= -2d_0 F_{00} \\
\dot{F}_{03} &= d_2 F_{02} - F_{03} (d_0 - \lambda_3) \\
\dot{F}_{33} &= 2(d_2 F_{23} + \lambda_3 (F_{33} + \langle n_3 \rangle)) \\
\dot{F}_{02} &= d_1 (rA + 2rS_2) F_{01} + (\lambda_2 - d_2) F_{02} + d_0 (F_{00} (rA + 2rS_2) - F_{02}) \\
\dot{F}_{23} &= (rA + 2rS_2) (d_1 F_{13} + d_0 F_{03}) + F_{23} (-d_2 + \lambda_2 + \lambda_3) + d_2 F_{22} \\
\dot{F}_{22} &= 2d_1 (F_{12} (rA + 2rS_2) + rS_2 \langle n_1 \rangle) + 2F_{22} (\lambda_2 - d_2) \\
&\quad + 2d_0 (F_{02} (rA + 2rS_2) + rS_2 \langle n_0 \rangle) + \lambda_2 \langle n_2 \rangle \\
\dot{F}_{01} &= d_0 F_{00} (rA + 2rS_1) - F_{01} (d_0 (rA + rS_1 + rS_2) + d_1 rS_2 - \lambda_1) \\
\dot{F}_{13} &= F_{13} (-d_1 rS_2 + \lambda_1 + \lambda_3) + d_2 F_{12} + d_0 F_{03} (rA + 2rS_1) \\
\dot{F}_{12} &= (-d_2 + \lambda_1 + \lambda_2) F_{12} + d_1 (F_{11} (rA + 2rS_2) - rS_2 F_{12} + rA \langle n_1 \rangle) \\
&\quad + d_0 ((rA + 2rS_2) F_{01} + rA (F_{02} + \langle n_0 \rangle) + 2rS_1 F_{02}) \\
\dot{F}_{11} &= 2(d_0 ((rA + 2rS_1) F_{01} + rS_1 \langle n_0 \rangle) + F_{11} (\lambda_1 - d_1 rS_2) + \lambda_1 \langle n_1 \rangle).
\end{aligned} \tag{4.16}$$

This model variant provides a good fit to the data (Fig. 4.18), too. The best fit estimates for all parameters are summarized in Table 4.5. As expected, the ratio of asymmetric cell divisions rA is again zero in the best fit. Moreover, an even stricter upper bound for rA is found (Table 4.5 and Fig. 4.19): According to this model, less than 10% of the divisions of the naïve and TCMP cells are predicted to be asymmetric. In the last section we already discussed that too many asymmetrically dividing naïve cells contradict the small TCMP-TEMp and TCMP-TEF correlations. This effect is enhanced by the additionally asymmetrically dividing TCMP cells, hence, the decrease in the upper bound is to be expected.

Taken together with the analysis of last section, we find strong data constraints on the occurrence of asymmetric cell divisions during T cell diversification in all discussed model variants. These constraints are encoded in the second order moments of the single cell progeny data of day 8. With the presented framework we were able to unravel this information and use it for model inference.

4. Systematic comparison of possible differentiation models

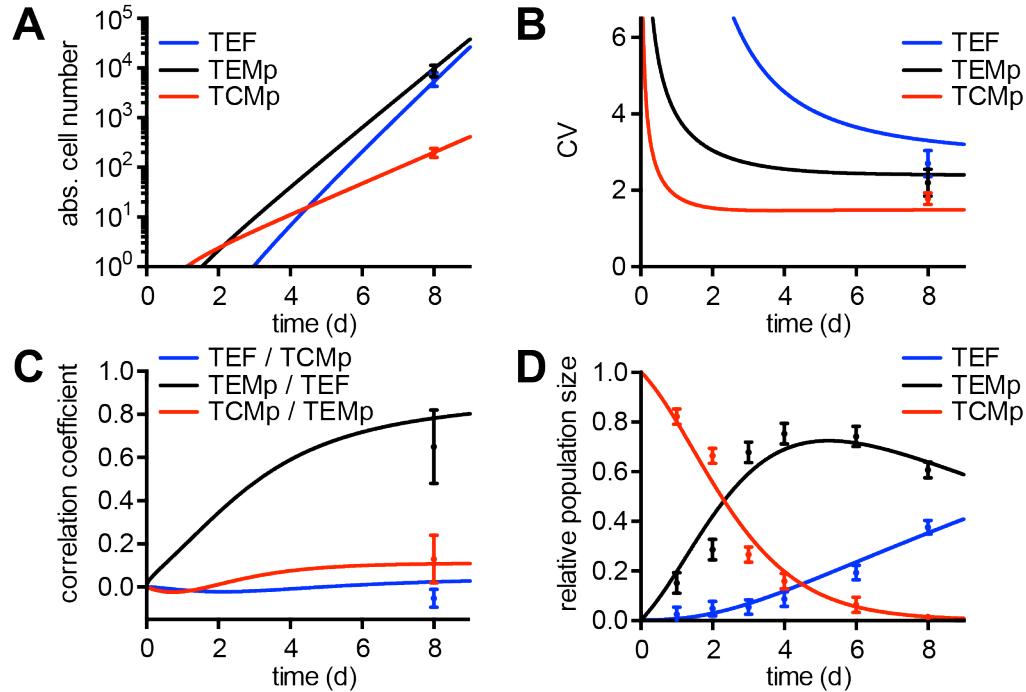


Figure 4.18.: Fit of the asymmetric cell division model depicted in Fig. 4.17 to mean values (A), CVs (B) and pairwise correlations (C) of the single cell progeny data of day 8 p.i. and to the relative subset size kinetics (D; day 1 to 4: transfer of 10⁵ OT-I cells, day 6 and 8: transfer of 100 OT-I cells). (Experiments by V. Buchholz and D. Busch)

Parameter	best fit	68% conf.	95% conf.
TCMP proliferation rate (λ_1)	0.80 d^{-1}	$(0.77, 0.83) \text{ d}^{-1}$	$(0.73, 0.86) \text{ d}^{-1}$
TEMP proliferation rate (λ_2)	1.4 d^{-1}	$(1.37, 1.44) \text{ d}^{-1}$	$(1.33, 1.47) \text{ d}^{-1}$
TEF proliferation rate (λ_3)	1.52 d^{-1}	$(1.47, 1.57) \text{ d}^{-1}$	$(1.41, 1.61) \text{ d}^{-1}$
recruitment rate (d_0)	0.42 d^{-1}	$(0.36, 0.51) \text{ d}^{-1}$	$(0.32, 0.64) \text{ d}^{-1}$
TCMP differentiation rate (d_1)	0.48 d^{-1}	$(0.35, 0.65) \text{ d}^{-1}$	$(0.26, 0.94) \text{ d}^{-1}$
TEMP \rightarrow TEF rate (d_2)	0.045 d^{-1}	$(0.038, 0.054) \text{ d}^{-1}$	$(0.033, 0.065) \text{ d}^{-1}$
ratio asym. div. (r_A)	0	$(0, 0.026)$	$(0, 0.098)$
ratio sym. div. to TEMP (r_{S2})	0.17	$(0.14, 0.21)$	$(0.10, 0.24)$

Table 4.5.: best fit parameters and confidence intervals (profile likelihood method) of the asymmetric cell division model depicted in Fig. 4.17 based on the fit to the single cell-derived progenies of day 8 p.i. and the relative subset size kinetics.

4.6. More refined asymmetric cell division models

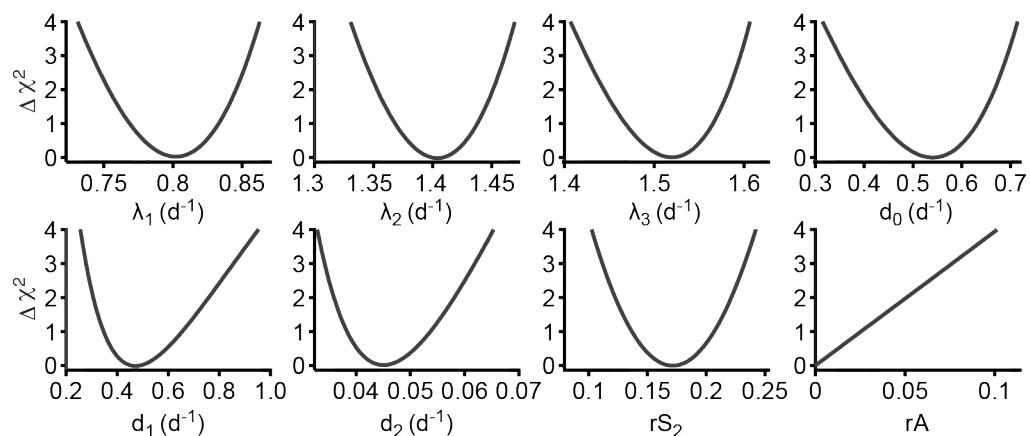


Figure 4.19.: Profile likelihood method for the asymmetric cell division model depicted in Fig. 4.17: Change in χ^2 -function obtained upon fixing each parameter at the abscissa values and minimizing over the remaining ones.

5. Implications of the $N \rightarrow TCMP \rightarrow TEMP \rightarrow TEF$ model

In this chapter we will analyze important implications of the $N \rightarrow TCMP \rightarrow TEMP \rightarrow TEF$ model and show that this model can account for several important previous experiments on T cell differentiation. Moreover, we will explore the implications of the $N \rightarrow TCMP \rightarrow TEMP \rightarrow TEF$ model for recall responses.

5.1. Timing of early division and differentiation events determines progeny population size and phenotype

In chapter 3 we demonstrated that the $N \rightarrow TCMP \rightarrow TEMP \rightarrow TEF$ model accurately accounts for the asymmetry and variability in absolute cell number as well as for the heterogeneity in phenotype that is observed during T cell diversification. To better understand how this is achieved in the model, we will now analyze how the timing of the first differentiation and division events influence the progeny's phenotype and total cell number at the peak of the response.

For this purpose it is illustrative to look at a few single cell trajectories (Fig. 5.1; simulated using Gillespie's algorithm [28] based on the best fit parameter set given in Table 3.2). In this figure, trajectory one represents the progeny of a TCMP cell that starts proliferation prior to day one, giving rise to the first TEMP cell shortly after. The TEF subset emerges with about two days delay and does not significantly contribute to the progeny's total size that is dominated by the TEMP cells and is about average size.

Trajectories two and three are examples first cell divisions that started late (around day 3 p.i.): The corresponding ancestor TCMP cells start proliferation about the same time, but while in one case the first TEMP differentiation occurs at day 4 (trajectory 3), in the other case (trajectory 2) no fast proliferating TEMP cell emerges until almost day 7. Due to the late first division of the TCMP ancestor, both progeny populations are small. On the other hand, the later occurrence of the first TCMP \rightarrow TEMP differentiation event causes a much smaller total population size in trajectory two. Obviously, both the timing of the first division and the first differentiation event have a strong influence on the progeny's total size at the peak of the response.

5. Implications of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

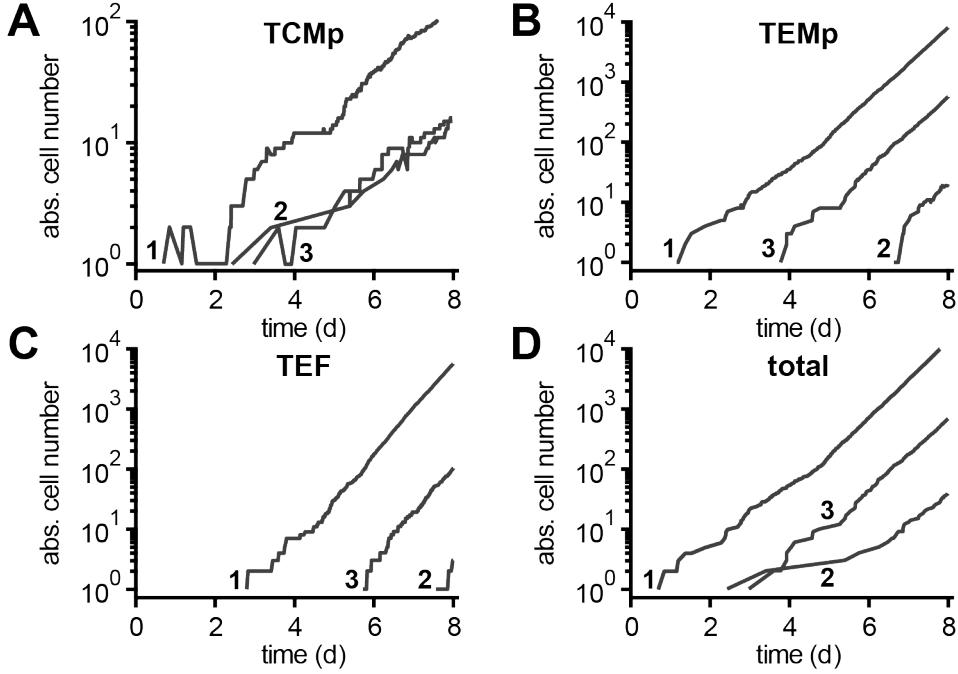


Figure 5.1.: Three simulated trajectories of single cell-derived progenies for the TCMP (A), TEMp (B) and TEF subset (C) and the sum of all three populations (D) using the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model.

We first look more closely at the time to first division. In order to calculate its distribution, we first compute the general waiting time for two poisson processes to occur. In a system with three states A , B and C and transition rates r_1 and r_2 that is described by the master equation

$$\begin{aligned}\dot{P}(A, t) &= -r_1 P(A, t) \\ \dot{P}(B, t) &= r_1 P(A, t) - r_2 P(B, t) \\ \dot{P}(C, t) &= r_2 P(B, t)\end{aligned}\tag{5.1}$$

with

$$P(A, 0) = 1, \quad P(B, 0) = P(C, 0) = 0,\tag{5.2}$$

we find the following probability density function (p.d.f.) of the waiting time until state C is reached for the first time

$$P(\rightarrow C, t) = \dot{P}(C, t) = r_2 P(B, t) = \frac{r_1 r_2}{r_1 - r_2} (e^{-r_2 t} - e^{-r_1 t}).\tag{5.3}$$

Given that the first dividing cell belongs to the TCMP compartment, we can directly use this formula to calculate the p.d.f. of the waiting time until the first division in

5.1. Timing of early division and differentiation events determines progeny population size and phenotype

the $N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF$ model

$$P(TCMp \text{ div.}, t) = \frac{d_0 \lambda_1}{d_0 - \lambda_1} \left(e^{-\lambda_1 t} - e^{-d_0 t} \right), \quad (5.4)$$

where d_0 is the recruitment rate and λ_1 the TCMp proliferation rate. The waiting time under the condition that the first dividing cell is a TEMp cell corresponds to the waiting time for three poisson processes and can be calculated analogously. The overall waiting time is the weighted sum of the waiting time conditioned on either of the two subsets being the first to divide (we can safely neglect the very small probability that a TEF cell divides prior to the other two subsets). The weights are formed by the probability that the first dividing cell is of TCMp phenotype ($\lambda_1 / (d_1 + \lambda_1) \approx 0.8$) and the much smaller probability that it is of TEMp phenotype ($d_1 / (d_1 + \lambda_1) \approx 0.2$). Eqn. 5.4 thus gives a good approximation to the overall distribution of time to first division. The resulting p.d.f. is depicted in Fig. 5.2 (dashed line) together with the distribution that results from the simulation of 10^5 progeny populations (solid line). The time to first division distribution is very broad and thus constitutes a major source of variability among the progenitors.

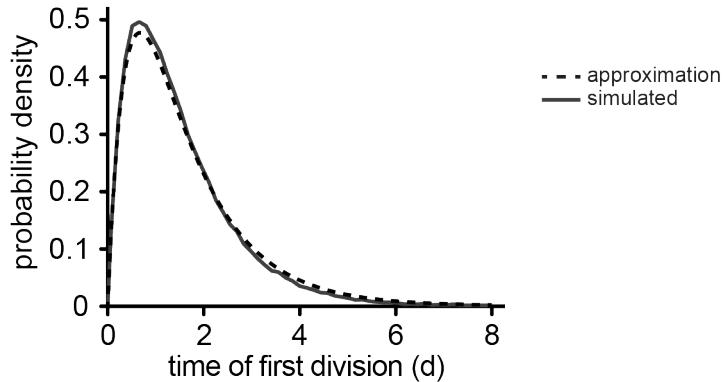


Figure 5.2.: Probability density distribution of the time of first cell division using the approximate formula given in Eqn. 5.4 (dashed line) and based on a simulation of $3 \cdot 10^5$ progeny populations.

Next we analyzed how the timing of the first division influences the progeny's total size at day 8 p.i. (Fig. 5.3A). Indeed, we found a rather strong negative correlation (Spearman's rank correlation coefficient $r = -0.66$) between both quantities: As expected, the biggest progeny populations are derived from progenitors that become activated and initiate cell division very early after infection. However, it is evident from Fig. 5.3A that the early onset of proliferation is not sufficient to yield a progeny population of large total size. This is because of the already mentioned second event that plays a crucial role here: The time point of the first differentiation event of the TCMp cells into the compartment of fast proliferating TEMp cells. Fig. 5.3B shows the strong inverse correlation (Spearman's rank correlation coefficient $r = -0.83$) of

5. Implications of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

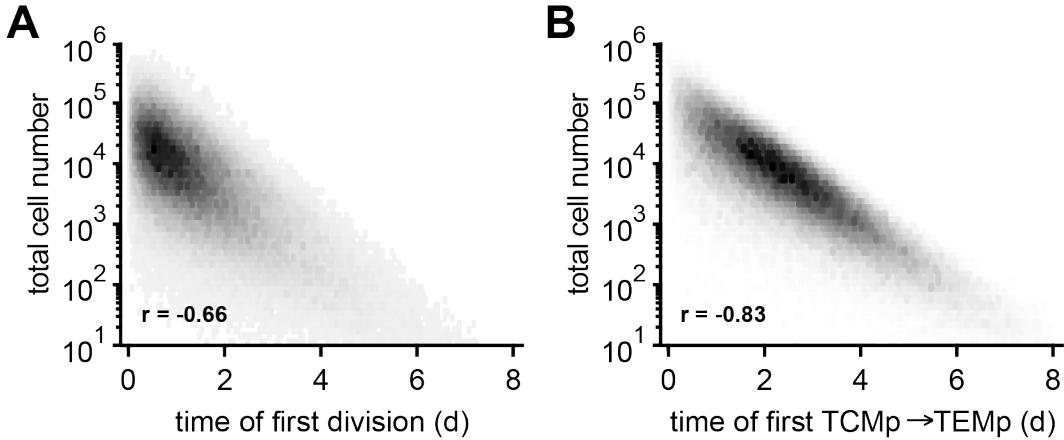


Figure 5.3.: Model simulation showing inverse correlations between total cell number (day 8 p.i.) and time of first cell division (A) and time of first TEMp emergence (B). Spearman's rank correlation coefficients are indicated. The simulations are based on $3 \cdot 10^5$ progeny populations.

the timing of this differentiation event with the progeny's total size at day 8 p.i. In the biggest single cell-derived populations, fast proliferating TEMp cells emerge very early after infection. Conversely, very early differentiation almost certainly leads to a big total population size.

Finally, the phenotypes of the progeny populations depend on the timing of the first TCMP \rightarrow TEMp differentiation event as well (Fig. 5.4). The relative fraction of CD62L $^+$ (Fig. 5.4A) and CD27 $^+$ cells (Fig. 5.4B) is augmented if this differentiation occurs late. However, this is not true for the absolute number of CD62L $^+$ and CD27 $^+$ cells: In the model the absolute number of TCMP cells (which are CD62L $^+$) is inversely correlated with the time point of TCMP \rightarrow TEMp differentiation since an early first division of the TCMP cells (beneficial for the resulting absolute number of TCMP cells) makes it more likely for one TCMP cell of the resulting progeny population to differentiate early into the TEMp compartment. On the other hand, the majority of the cells at day 8 p.i. are CD27 $^+$ (corresponding to the sum of TCMP and TEMp cells in the model). Thus, the absolute number of CD27 $^+$ cells is inversely correlated with the TCMP \rightarrow TEMp differentiation time point in the same way as the total cell number (Fig. 5.3B).

5.2. Probability for TCMP depletion

A sufficient number of both central memory and effector memory T cells is crucial to efficiently generate immunological memory and to ensure protection against reinfection [61]. It is therefore an interesting finding that in about 15% of the sin-

5.2. Probability for TCMP depletion

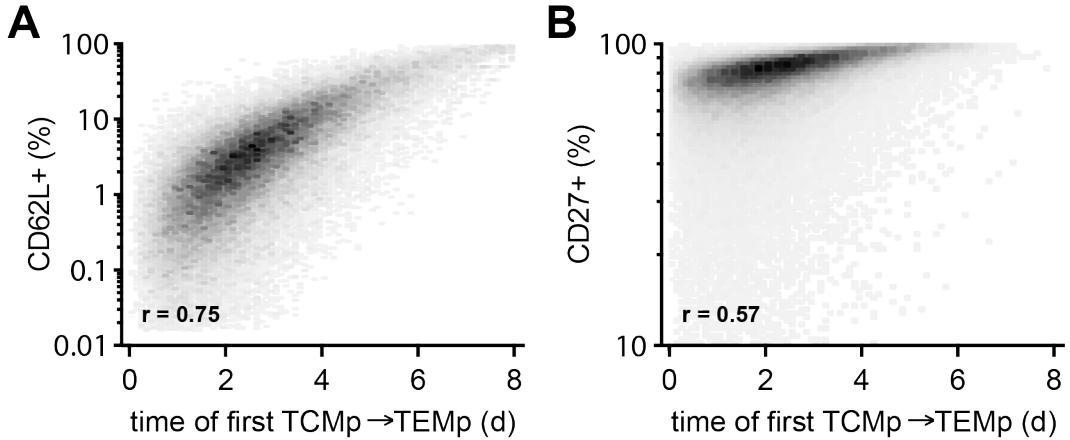


Figure 5.4.: Model simulation showing positive correlations between time of first TEMP emergence and the fraction of CD62L⁺ cells (A) and the fraction of CD27⁺ cells (B). Spearman's rank correlation coefficients are indicated. The simulations are based on $3 \cdot 10^5$ progeny populations.

gle cell-derived progeny populations no central memory precursor cells are found (cf. Fig. 3.5). Due to the experimental cutoff this estimation of the percentage of depleted TCMP populations should be regarded as an upper limit. Interestingly, depleted TCMP populations arise naturally in the N→TCMP→TEMP→TEF model due to the possibility of a differentiation event (instead of a proliferation event) when there is only a single TCMP cell.

N→TCMP→TEMP→TEF model

In chapter 3.4 (Fig. 3.5) we saw that the simulation for the number of depleted TCMP populations at day 8 p.i. in the model agrees with the observed value. In order to obtain an analytical solution, we will first assume that the transition N→TCMP is fast. The model scheme then effectively simplifies to TCMP→TEMP→TEF with initial condition $n_1(t=0)=1$. In this case we can use the probability of population extinction in a birth-and-death process [38] yielding

$$P(n_1=0,t) = \frac{d_1(e^{(\lambda_1-d_1)t}-1)}{\lambda_1 e^{(\lambda_1-d_1)t} - d_1}.$$

We can readily generalize this expression to account also for the N→TCMP transition. The probability for depletion given that the N→TCMP transition happens at $t=t'$ is given by

$$P(n_1=0,t|N \rightarrow \text{TCMP at } t') = \frac{d_1(e^{(\lambda_1-d_1)(t-t')} - 1)}{\lambda_1 e^{(\lambda_1-d_1)(t-t')} - d_1}.$$

5. Implications of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

Thus,

$$P(n_1 = 0, t) = \int_0^t dt' d_0 e^{-d_0 t'} \frac{d_1 (e^{(\lambda_1 - d_1)(t-t')} - 1)}{\lambda_1 e^{(\lambda_1 - d_1)(t-t')} - d_1}. \quad (5.5)$$

Fig. 5.5 shows the resulting time dependence of $P(n_1 = 0, t)$ using the best fit parameter values based on the fit to the day 8 data (cf. Table 3.1). For late time points, $P(n_1 = 0, t)$ approaches d_1/λ_1 . Hence, the observed percentage of depleted TCMP populations provide an estimate for the ratio of the TCMP differentiation to proliferation rate (if the observed TCMP depleted populations are in part due to the experimental cutoff, then we still have an upper bound for d_1/λ_1).

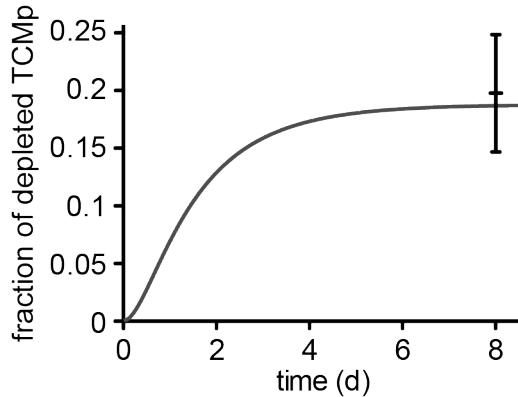


Figure 5.5.: Fraction of single cell progeny populations with depleted TCMP subsets as predicted by the model (solid line) and as measured at day 8 p.i. (black data point). (Experiments by V. Buchholz and D. Busch)

Asymmetric cell division models

We will now derive the probability for TCMP depletion in the two asymmetric models that we introduced in chapter 4.6. In these two models, naïve cells get recruited at the rate d_0 , and subsequently divide either symmetrically into two TCMP cells (ratio rS_1) or two TCMP cells (ratio rS_2) or the naïve cell divides asymmetrically into one TCMP cell and one TEMp cell (ratio rA_1 ; $rS_1 + rS_2 + rA = 1$). In model variant one, differentiation otherwise follows the linear pathway TCMP \rightarrow TEMp \rightarrow TEF with the TCMP \rightarrow TEMp differentiation rate d_1 . In model variant two, the TCMP cells were also allowed to divide symmetrically into two TEMp cells (at the effective rate $d_1 \cdot rS_2$) and asymmetrically into cells of both phenotypes (at the effective rate $d_1 \cdot rA$).

5.3. Robustness of T cell response

Defining

$$f(\lambda_1, d_1, t, t') = \frac{d_1 (e^{(\lambda_1 - d_1)(t-t')} - 1)}{\lambda_1 e^{(\lambda_1 - d_1)(t-t')} - d_1} \quad (5.6)$$

we find analogously to Eqn. 5.5 for the probability of TCMP depletion in model variant one

$$P(n_1 = 0, t) = \int_0^t dt' d_0 e^{-d_0 t'} \left(rS_1 f(\lambda_1, d_1, t, t')^2 + rA f(\lambda_1, d_1, t, t') + rS_2 \right). \quad (5.7)$$

The corresponding equation for model variant two is obtained by substituting all occurrences of d_1 in this formula with $d_1 \cdot rS_2$ since only symmetric division into two TEMp cells leads to a reduction in the number of TCMP cells. In these models the predicted fraction of TCMP depleted populations coincides with the measured value.

It has been estimated that there are around 50 - 200 naïve CD8⁺ T cells sharing specificity for the same epitope [61]. According to our previous discussion in chapter 2.3, the pool of endogenous SIINFEKL-specific CD8⁺ T cells comprises about 100 naïve cells. Depletion of all central memory precursor cells in any given immune response will thus most likely be an extremely rare event.

5.3. Robustness of T cell response

We saw that single cell-derived progeny populations show enormous variability in total cell number. On the other hand, T cell responses to any given antigen need to be robust. Above all, this requires a certain minimum size of the response that results from all activated progenitor cells. Thus, robustness in this sense will critically depend on the number of T cells that get activated by a particular antigen.

For a given minimum size n_{\min} of the overall response and a given probability value p that quantifies how certain this threshold should be reached, we can use the N→TCMP→TEMp→TEF model to compute the required minimum number N of responding naïve progenitors. To accomplish this, we simulated a sufficient number of progeny populations that are derived from N initial ancestors and determined the $(100-p \cdot 100)$ percentile of the resulting absolute size distribution. Fig. 5.6 shows the resulting minimum number N of activated naïve precursors that are needed to reach a certain minimum progeny size per progenitor (n_{\min}/N) at day 8 p.i. for three different certainties p (the three solid lines correspond to $p = 90\%$, $p = 99\%$ and $p = 99.9\%$). As the number of activated naïve precursors increases, the three curves first show a steep increase and then flatten out as they approach the mean progeny size of a single progenitor (dashed line in Fig. 5.6). We can interpret this as follows.

5. Implications of the N→TCMp→TEMp→TEF model

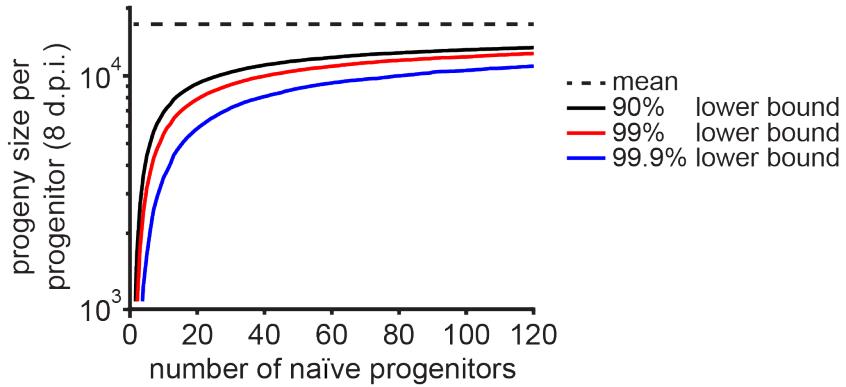


Figure 5.6.: Robust T cell responses require the recruitment of multiple precursors. The plot shows the minimum number of activated naïve precursors that are needed to reach a certain minimum progeny size per progenitor (8 d.p.i.) for three different certainties p (90%, 99% and 99.9%; solid lines). Also shown is the mean progeny size of a single progenitor (dashed line) that is asymptotically approached by the three solid lines.

A huge gain in robustness is achieved upon increasing the number of activated T cells to a few dozens. Further increase only leads to a minor gain in robustness.

Robustness of the T cell response in the above sense directly depends on the variability of the total cell number of a single cell-derived population which can be quantified by calculating the CV of the total cell number. In the N→TCMp→TEMp→TEF model, this CV is strongly dependent on the value of the TCMP→TEMp differentiation rate d_1 . Fig. 5.7A shows the CV of the total cell number at day 8 p.i. for different values of d_1 (all other parameters are held fixed at their corresponding best fit values). Interestingly, decreasing the value of d_1 from its best fit estimate quickly leads to a steep increase in the CV of the total cell number (the decline of the CV for d_1 close to zero is because in this region the model effectively reduces to N→TCMp). On the other hand, the CV only decreases marginally for values of d_1 that lie above its best fit estimate. However, the rate d_1 has a second important effect: As we derived in the previous section, the probability for depletion of the TCMP subset also critically depends on the value of d_1 (cf. eqn. 5.5). The chance of TCMP depletion as a function of d_1 is plotted in Fig. 5.7B: Every increase of d_1 is accompanied by an increase in the probability of TCMP depletion.

The TCMP→TEMp differentiation rate d_1 thus plays an interesting dual role: Decreasing its value leads to less robust T cell responses and an increase implies a higher fraction of progeny populations with a depleted TCMP population.

5.4. Distribution of the number of cell divisions over time

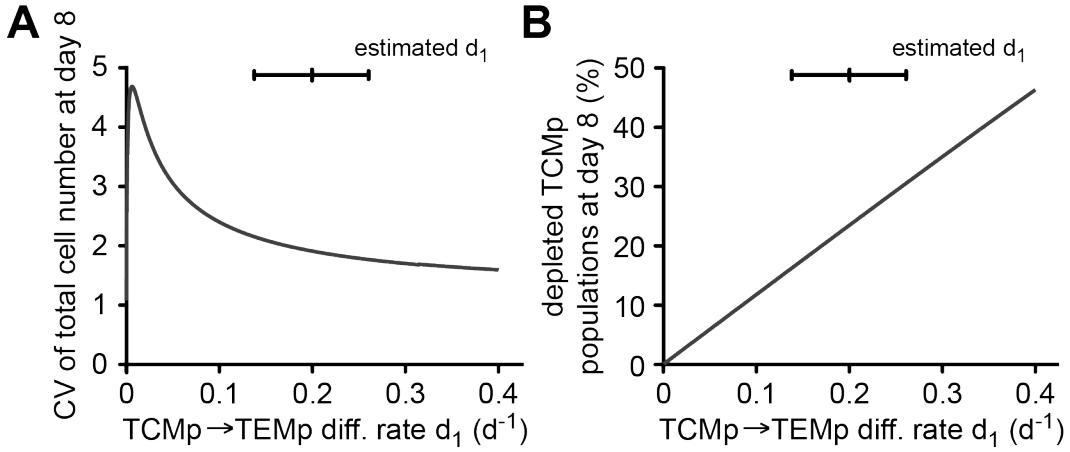


Figure 5.7.: The TCMP \rightarrow TEMP differentiation rate d_1 plays a dual role. (A) CV of total cell number as a function of d_1 . Decreasing the value of d_1 from its best fit estimate leads to an increase in the CV (and hence less robustness). (B) Percentage of depleted TCMP populations plotted against the probability for TCMP depletion. For larger values of d_1 , this probability increases. All parameters except for d_1 are held fixed at their best fit estimates.

5.4. Distribution of the number of cell divisions over time

In this section we will compute the time evolution of the distribution of the number of cell divisions that is predicted by the N \rightarrow TCMP \rightarrow TEMP \rightarrow TEF model and compare it to experimental data.

Let n^0 , n^1 , n^2 and n^3 be the mean number of naïve, TCMP, TEMP and TEF cells, respectively. Further, for the respective subset, let n_k^1 , n_k^2 and n_k^3 be the mean number of cells that divided k times. We are interested in the relative number of cells in generation k that is given by

$$r_k = \frac{\sum_i n_k^i}{\sum_i n^i}. \quad (5.8)$$

In a simple model with just one population of cells and a generation-independent proliferation rate λ , the relative number of cells per generation follows a Poisson distribution [20]

$$r_k(t) = \frac{(2\lambda t)^k}{k!} e^{-2\lambda t}. \quad (5.9)$$

In the N \rightarrow TCMP \rightarrow TEMP \rightarrow TEF model we have to consider subset-dependent proliferation rates λ_i and differentiation between the subsets. To incorporate this, we again write down the ODE system for the mean number of cells in the N \rightarrow TCMP \rightarrow TEMP \rightarrow TEF model (this was already given in chapter 3.2, eqn. 3.17 and we use the same notation

5. Implications of the N→TCMp→TEMp→TEF model

here)

$$\begin{aligned}\dot{n}^0 &= -d_0 n^0 \\ \dot{n}^1 &= (\lambda_1 - d_1) n^1 + d_0 n^0 \\ \dot{n}^2 &= (\lambda_2 - d_2) n^2 + d_1 n^1 \\ \dot{n}^3 &= d_2 n^2 + \lambda_3 n^3,\end{aligned}\tag{5.10}$$

with

$$n^0(0) = N_0 \text{ and } n^i(0) = 0, \text{ for } i = 1, \dots, 3,\tag{5.11}$$

where N_0 denotes the initial number of naïve cells. We can readily generalize eqn. 5.10 to get the corresponding ODE system for the time evolution of the mean number of cells n_k^1 , n_k^2 and n_k^3 that have undergone k divisions

$$\begin{aligned}\dot{n}_k^0 &= -d_0 n_k^0 \\ \dot{n}_k^1 &= 2\lambda_1 n_{k-1}^1 - \lambda_1 n_k^1 - d_1 n_k^1 + d_0 n_k^0 \\ \dot{n}_k^2 &= 2\lambda_2 n_{k-1}^2 - \lambda_2 n_k^2 - d_2 n_k^2 + d_1 n_k^1 \\ \dot{n}_k^3 &= 2\lambda_3 n_{k-1}^3 - \lambda_3 n_k^3 + d_2 n_k^2,\end{aligned}\tag{5.12}$$

where $k = 0, 1, 2, \dots$, and

$$\begin{aligned}n_{-1}^i &= 0, \text{ for } i = 1, \dots, 3, \\ n^0(0) &= N_0, \\ n_0^i(0) &= 0, \text{ for } i = 1, \dots, 3.\end{aligned}\tag{5.13}$$

We can now numerically solve the ODE systems of eqns. 5.10 and 5.12 and compute the relative number r_k of cells in generation k using eqn. 5.8.

If T cell numbers are sufficiently high, the distribution r_k can be assessed experimentally using the fluorescent cell staining dye CFSE (Carboxyfluorescein succinimidyl ester): Upon cell division, the CFSE molecules of the labelled T cells are partitioned equally between the daughter cells, resulting in a twofold reduction of the intensities of cellular fluorescence between mother and daughter cells. The fluorescence intensity for each cell in a population of dividing T cells can then be measured using flow cytometry. This allows for a direct estimation of the relative number r_k of cells per generation.

Figs. 5.8A and 5.8B show the result of two experiments where an initial number of 10^4 (Fig. 5.8A) or 10^5 (Fig. 5.8B) CFSE labelled naïve OT-I T cells have been transferred into the mice. Otherwise the experimental setting was as described in chapter 2.1. The CFSE fluorescence intensity profiles have then been analyzed 2, 3 and 4 days post infection. In this experiments, only the first five cell generations could be resolved separately, the remaining fraction of cells that divided more

5.4. Distribution of the number of cell divisions over time

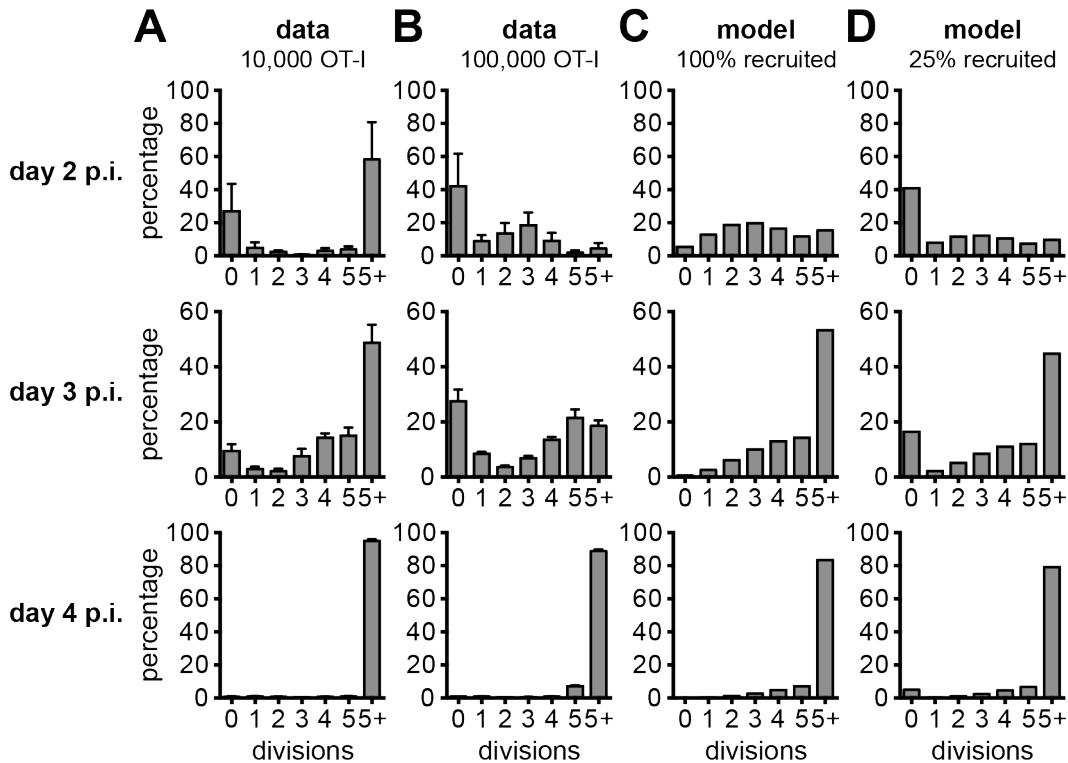


Figure 5.8.: (A-B) Histograms showing the relative number of cells that underwent the indicated number of divisions until day 2, 3 and 4 p.i. based on CFSE dilution. In the experiments either 10^5 (A) or 10^4 OT-I cells (B) were adoptively transferred. (C) Corresponding prediction of the N \rightarrow TCMP \rightarrow TEMP \rightarrow TEF model. (D) Same prediction, but assuming recruitment of only 25% of the naïve T cells within the first three days after infection. (Experiments by V. Buchholz and D. Busch)

than five times is denoted “5+” in Fig. 5.8. The corresponding predictions of the N \rightarrow TCMP \rightarrow TEMP \rightarrow TEF model are depicted in Fig. 5.8C. At day 2 and 3 p.i. the observed relative numbers of cells per generation differ between both experiments and at day 2 p.i. neither of both coincides with the model prediction. It has been reported before that the progeny population size derived from a high number (as compared to the size of the endogenous pool of responding T cells) of adoptively transferred T cells is no longer proportional to the number of transferred cells [4]. It is thus suggestive that competition between the T cells leads to the observed differences between both experiments.

Competition would be expected to play a less important role in the 10^4 transfer experiment. If anything, the model prediction should thus be compared to this experiment. But, in this experiment, the T cell frequencies at day 2 are so low that the CFSE profiles are hard to evaluate. This is also reflected in the large error bars that belong to the measurements of the fraction of undivided cells (generation 0) and the fraction of cells that divided more than five times (generation “5+”). However,

5. Implications of the N→TCMp→TEMp→TEF model

based on this experiment, it seems that the model underestimates the proliferation at day 2 p.i. This could have to do with an underestimation of the proliferation rates due to the neglect of cell death in this model: As we showed in chapter 4.4, the incorporation of cell death into the model leads to a significant increase in the upper bounds of all proliferation rates.

Another discrepancy between model prediction and data concerns the number of undivided cells. One possible explanation for this would be that competition between the naïve T cells (e.g. for antigen presenting cells) leads to delayed recruitment of a considerable fraction of naïve cells. Consistent with this explanation would be the fact that even more cells are undivided in the experiment where 10^5 OT-I cells were transferred.

To estimate the quantitative effect of delayed recruitment on the CFSE profiles, we assumed that a fraction of the transferred cells would not get activated during the first three days and recalculated the model prediction accordingly. As an example, Fig. 5.8D shows how the model prediction would change if 75% of the naïve cells could not get recruited within the first three days of the infection. Indeed, it seems that this explanation suffices to explain the discrepancies between model and data for the undivided cells.

5.5. Latecomer T cells in the N→TCMp→TEMp→TEF model

It has been proposed that TCMP cells are preferentially derived from naïve precursors that are recruited late in the response [16, 13]. The experimental evidence in both studies is based on the comparison of the response of naïve T cells that have been transferred prior to infection with the response of naïve T cells that have been transferred 2-4 days post infection (so-called “latecomer” T cells). However, high cell numbers were used in all conducted adoptive transfer experiments (an approach that was later shown to alter the kinetics of memory cell formation [4, 40]).

Nevertheless, the increase in the fraction of central memory precursors has been interpreted as evidence for the so-called “decreasing potential model” of T cell diversification [2]. In this model, the central memory precursors are derived directly from the naïve cells and subsequently differentiate into effector memory precursors which finally give rise to terminally differentiated effector cells. Progression through this differentiation pathway was proposed to be driven by the cumulative stimulation history experienced by the cells. Accordingly, the latecomer T cells would be expected to accumulate less stimulative signals. Thus, the decreasing potential model would predict a higher fraction of memory precursors among the latecomer T cells. We will now investigate how we can account for the different differentiation

5.5. Latecomer T cells in the N→TCMp→TEMp→TEF model

dynamics of these late transferred cells using the N→TCMp→TEMp→TEF model.

Dirk Busch, Veit Buchholz and coworkers experimentally investigated the characteristics of latecomer T cells using their technique of single cell transfers [9]. To this end, single naïve OT-I cells were transferred into an ongoing infection at day 3 p.i. The progeny populations of these latecomers were recovered in the spleen at day 11 p.i.

Table 5.1 summarizes some important statistical quantities of the outcome of this experiment: Compared to the (so far considered) single cell-derived population size

	latecomer	early transfer
mean total cell number	(960 ± 230)	(15500 ± 3800)
CV of total cell number	(1.6 ± 0.2)	(2.1 ± 0.3)
percentage of CD27 ⁺ cells	(78 ± 4)	(62 ± 3)

Table 5.1.: Statistical quantities for the single cell-derived latecomer progenies at day 11 p.i. (transfer at day 3 p.i.). For comparison also the corresponding quantities for the single cell-derived progenies at day 8 p.i. (transfer at day 0) are given.

after eight days of proliferation, the total cell number of the latecomer T cells is about 15 times reduced. The CV of the total cell number on the other hand is only slightly reduced for the latecomer populations. Note that we cannot analyze this experiment in terms of the TCMP, TEMp and TEF subsets, because CD62L has not been measured. The measured number of CD27⁺ cells however does correspond to the sum of the TCMP and TEMp subsets in the model.

After 8 days of proliferation, the latecomers reach a mean total cell number that is much smaller than the one of the single cell-progenies that were transferred prior to infection, but proliferated for the same period of time. It is therefore evident that the latecomers cannot be described by the N→TCMP→TEMp→TEF model using the same set of parameters. To be able to simultaneously describe both data sets using the same model we have to expand the model. One possibility is to introduce time dependent rates: If the initiation of cell division or differentiation is dependent on environmental factors that change during the time course of the infection (like a decreasing strength of antigen or cytokine stimulus), then the corresponding rates would also show a time dependence. Thus we can use the latecomer data to test for which of the rates the assumption of time dependence enables a simultaneous fit to both data sets.

Fitting the latecomer data using time dependent rates

We first tested whether it is sufficient to only allow for time dependent proliferation rates while keeping the differentiation rates constant. We assumed the following

5. Implications of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

time dependence of the proliferation rates

$$\lambda(t) = \lambda_i^0 f(t), \quad (5.14)$$

with

$$f(t) = \begin{cases} 1, & \text{if } t < t_c, \\ e^{-\alpha(t-t_c)}, & \text{if } t \geq t_c. \end{cases} \quad (5.15)$$

Here λ_1^0 , λ_2^0 and λ_3^0 denote the TCMP, TEMP and TEF proliferation rates for $t < t_c$ and α denotes the decay rate of the exponential decline for $t \geq t_c$. The parameters λ_i^0 , t_c and α were estimated together with the remaining three differentiation rates by simultaneously fitting the data of the early transferred T cells and the latecomer data.

The resulting best fit estimate for the time dependence of the proliferation rates is depicted in Fig. 5.9A. All best fit parameter estimates are given in Table 5.2. For the latecomer data, Fig. 5.9B-D shows the best fit to the mean (Fig. 5.9B)

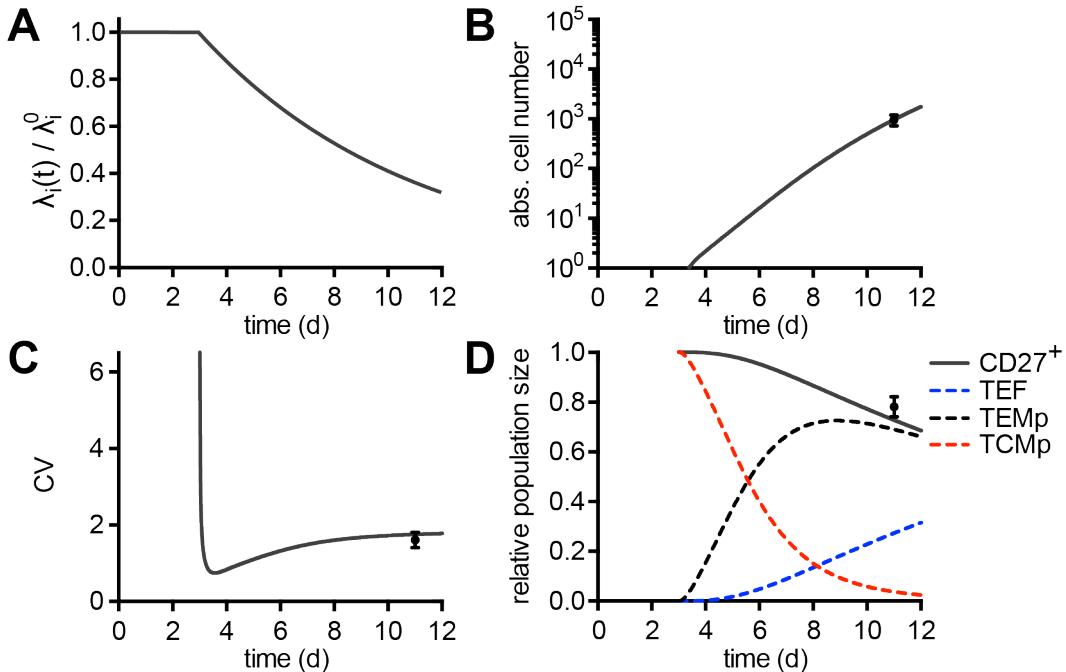


Figure 5.9.: Fit of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model to the latecomer data assuming time dependent proliferation rates (fitted simultaneously with the data of the early transferred T cells shown in Fig. 5.10). (A) Best fit estimate of the time dependence of the proliferation rates. (B-D) Best fit to the mean (B) and CV (C) of the total cell number as well as to the fraction of CD27 $^{+}$ cells (D). Data is shown in black. The dashed lines in (C) show the predicted time courses for the TCMP, TEMP and TEF subset of the latecomers. (Experiments by V. Buchholz and D. Busch)

5.5. Latecomer T cells in the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

and CV (Fig. 5.9C) of the total cell number as well as to the fraction of CD27⁺ cells (Fig. 5.9D). The best fit to the data of the early transferred T cells is shown in Fig. 5.10. Parameter estimation was performed as described in chapter 3.2 and chapter 3.6. Note that the formalism that we introduced in chapter 3 to calculate the second moments of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model can also be used with time dependent rates.

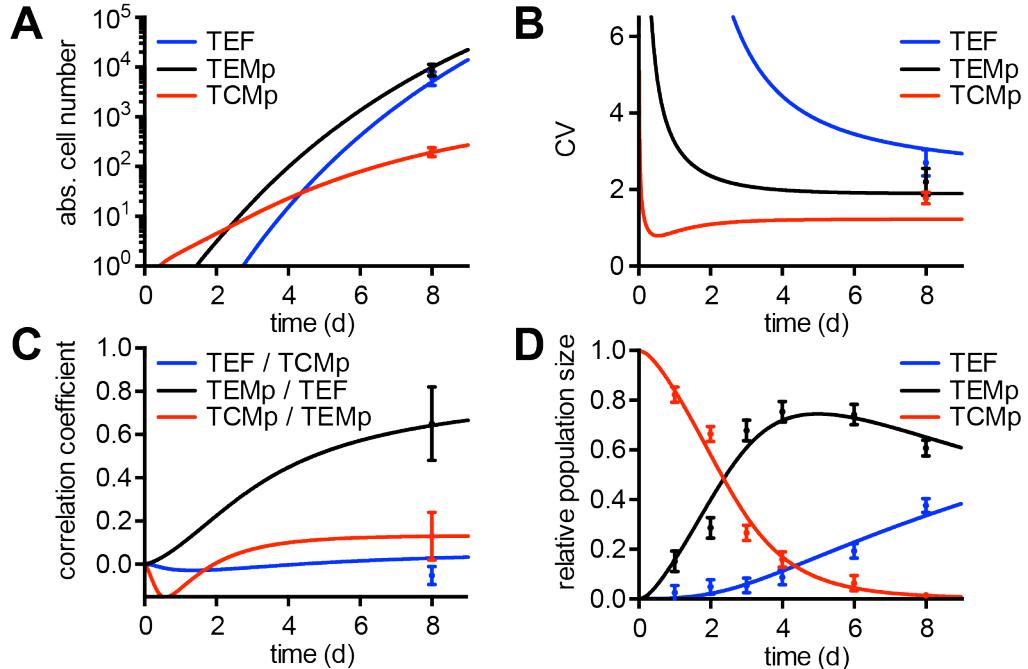


Figure 5.10.: Fit of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model to the data of the early transferred cells assuming time dependent proliferation rates (fitted simultaneously with the latecomer data shown in Fig. 5.9B-D). (A-D) Best fit to the mean values (A), CVs (B) and pairwise correlations (C) of the single cell progeny data of day 8 p.i. and to the relative subset size kinetics (D; day 1 to 4: transfer of 10⁵ OT-I cells, day 6 and 8: transfer of 100 OT-I cells). (Experiments by V. Buchholz and D. Busch)

Interestingly, the estimated value of the time point t_c after which the proliferation rates start declining is 3.0 d. This corresponds quite well to the time point of the peak of the infection [61]. After this time point, the pathogen load will start to decrease.

Comparing early and late transferred T cell progenies at the same time point after infection, the two already mentioned studies [16, 13] found an increased percentage of central memory precursor cells in the progeny populations of the latecomer T cells. This would also be the model prediction based on the above fit: At day 8 p.i., the latecomer progeny populations are predicted to comprise about 15% TCMP cells (cf. red dashed line in Fig. 5.9C). In contrast, only about 1.6% TCMP cells

5. Implications of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

(Fig. 5.10D) were observed in the progeny populations of the early transferred T cells. Nevertheless, this is just true for the relative number of central memory precursors. In absolute numbers, much more central memory precursors are found in the progenies of the early transferred T cells.

Simulating the latecomer progeny populations

So far we summarized the latecomer data by mean and CV of the total cell number as well as the percentage of CD27 $^{+}$ cells. Based on the obtained best fit parameters we then simulated single cell-derived latecomer progenies and compared them with the observed progenies. For this we looked at the correlation of the percentage of CD27 $^{+}$ cells with the total cell number. The result of the simulation together with the latecomer data at day 11 p.i. is shown in Fig. 5.11A. Indeed, the simulated progeny populations (red) resemble the observed ones. In chapter 3.4 we already showed that the N \rightarrow TEF \rightarrow TEMp \rightarrow TCMp model can reproduce the observed relation between both CD27 $^{+}$ and CD62L $^{+}$ positive cells and the total size of the population. As shown in Fig. 5.11B, this finding also holds in the case of time dependent proliferation rates.

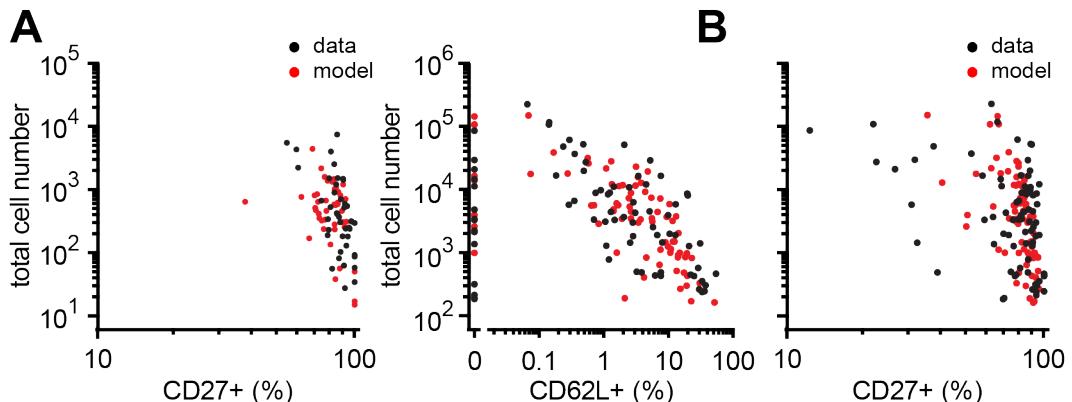


Figure 5.11.: (A) Percentage of CD27 $^{+}$ cells versus total cell number at day 11 p.i. for the single cell-derived latecomer cells (black dots) together with a representative simulation of equal size of the N \rightarrow TEF \rightarrow TEMp \rightarrow TCMp model with time dependent rates (red dots). **(B)** As in A, but for the early transferred T cells at day 8 p.i. and additionally showing the percentage of CD62L $^{+}$ cells versus total cell number. (Experiments by V. Buchholz and D. Busch)

So, the introduced time dependence of the proliferation rates suffices to reproduce the data of both early and late transferred T cells. On the other hand, it turns out that no acceptable fit to the whole data set can be obtained if the same time dependency (eqn. 5.14) is imposed on any of the three differentiation rates (d_0 , d_1 and d_2) or any combination of the differentiation rates. However, the data can still

5.5. Latecomer T cells in the N→TCMp→TEMp→TEF model

be described if the proliferation rates and one or more differentiation rates share the time dependency that is given by eqn. 5.14.

Alternative approach to model the latecomer cells

Another way to approach the modeling of the latecomers would be to assume that early and late transferred T cells proliferate and / or differentiate at different rates. In the so-called “autopilot” model of T cell diversification [8], this assumption would be justified. The “autopilot” model proposes that upon activation of the naïve T cell, a program of cell division and differentiation is initiated that does not depend on further environmental stimuli. However, also in this model, during the activation of the T cell, different stimuli could still lead to different outcomes in terms of cell division and differentiation.

Following this approach, we assumed that the proliferation rates of the latecomers are reduced by a constant factor c (relative to the proliferation rates of the early transferred cells). We then fitted this model to the data of the early and late transferred T cells and estimated the factor c together with the six remaining parameters of the N→TEF→TEMp→TCMp model (Fig. A.1). Also in this case a good fit to the data can be obtained (not shown). The best fit estimate for c is 0.7. Again no acceptable fit to the whole data set can be obtained if only the differentiation rates of the latecomers are reduced by a constant factor.

Based on the N→TEF→TEMp→TCMp model we showed that the increase in the relative number of memory precursors in the latecomer population does not necessarily imply a change in the differentiation rates for this population. Based on the N→TEF→TEMp→TCMp model, the latecomer experiment could thus not be interpreted as evidence for the decreasing potential model.

Parameter	best fit
TCMp proliferation rate for $t < t_c$ (λ_1^0)	1.0 d^{-1}
TEMp proliferation rate for $t < t_c$ (λ_2^0)	1.7 d^{-1}
TEF proliferation rate for $t < t_c$ (λ_3^0)	1.9 d^{-1}
recruitment rate (d_0)	3.8 d^{-1}
TCMp → TEMp rate (d_1)	0.17 d^{-1}
TEMp → TEF rate (d_2)	0.038 d^{-1}
time until which proliferation rates are constant (t_c)	3.0 d
decay rate of proliferation rates (α)	0.13 d^{-1}

Table 5.2.: best fit parameters of the N→TCMp→TEMp→TEF model with time dependent proliferation rates based on the fit to the latecomer data and the data of the early transferred T cells.

5. Implications of the N→TCMp→TEMp→TEF model

5.6. Simulating antigen stimulus withdrawal

Bevan and colleagues [48] established a very elegant experimental system which allows to control the time period in which T cells can receive antigen stimulus *in vivo*. This is achieved by the diphtheria toxin-induced killing of the dendritic cells (carrying a diphtheria toxin receptor transgene) that present the antigen for which the reacting T cells are specific. One interesting result of this study is that even very early truncation of the antigen signal has only a minor effect on the observed fraction of CD62L positive cells on day 5 p.i. Another major finding is that the withdrawal of the antigen stimulus after 2-3 days leads to a T cell response at day 5 p.i. that is approximately of the same size as in the unperturbed system. Based on these findings the authors put forward the following conclusions: First, less than a day of antigen stimulus is sufficient to program differentiation of the T cells. Second, to equip the T cells with optimal proliferative potential, around 2-3 days of antigen stimulation is needed.

We will now use the N→TCMp→TEMp→TEF model and show that we can also explain the observed phenomena by making the following converse assumptions: Firstly, the withdrawal of antigen stimulus does not affect the proliferation rates of the activated T cells. Secondly, upon antigen withdrawal, T cells are no longer recruited and do no longer differentiate from the central memory precursor to the effector memory precursor phenotype. Note that since recruitment is fast ($d_0 \approx 3 \text{ d}^{-1}$) in the N→TCMp→TEMp→TEF model, antigen withdrawal after day 1 p.i. would nevertheless barely affect T cell recruitment. Subsequently we will further consider the case where in addition to the TCMp→TEMp differentiation rate, the TEMp→TEF rate is strictly antigen dependent as well.

Assuming antigen withdrawal precludes further recruitment and TCMp→TEMp differentiation

To simulate antigen stimulus withdrawal at $t = t_0$, we used Gillespie's algorithm [28] and split up the simulation into two parts: For $t < t_0$ the simulation was done as before. For $t \geq t_0$ the resulting T cell numbers from the first part were used as initial values for a second Gillespie simulation in which the TCMp→TEMp rate d_1 and the recruitment rate d_0 were set to zero. Fig. 5.12A shows the resulting predicted mean population size per progenitor at day 8 p.i. as a function of the time point of antigen withdrawal t_0 . Indeed, as observed by Bevan and colleagues [48], the predicted mean population size per progenitor in the simulation barely changes for t_0 bigger than 2-3 days. A similar result is found for the fraction of CD62L and CD27 positive cells (Fig. 5.12B): The main effect of the antigen withdrawal is obtained if it is enforced prior to day 1-2 p.i.

To better understand this result it is illustrative to consider also the effect of the

5.6. Simulating antigen stimulus withdrawal

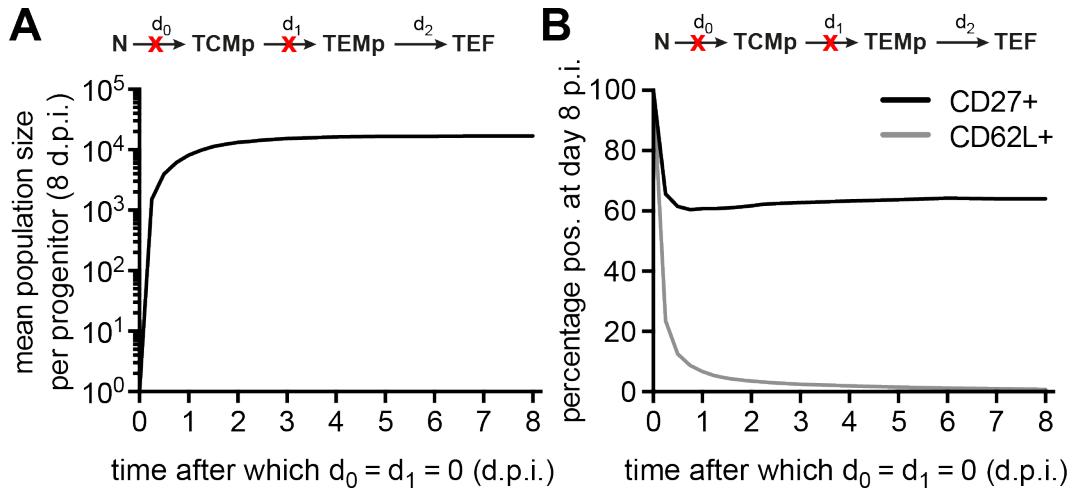


Figure 5.12.: Simulation of antigen withdrawal using the $N \rightarrow \text{TCMp} \rightarrow \text{TEMP} \rightarrow \text{TEF}$ model. Upon antigen withdrawal, d_0 and d_1 are set to zero. (A-B) Predicted mean population size per progenitor (A) and the predicted mean fractions of CD62L and CD27 positive cells (B) at day 8 p.i. as a function of the time point of antigen withdrawal.

antigen withdrawal on the single cell-derived progenies. Fig. 5.13 shows the percentage of CD62L and CD27 positive cells vs. total cell number for single cell-derived progeny populations assuming antigen withdrawal at day 2 p.i. Although many progeny populations are entirely positive both for CD62L and for CD27 (these are the progenies where no $\text{TCMp} \rightarrow \text{TEMP}$ differentiation happened prior to day 2), the biggest populations show a considerably smaller fraction of CD62L and CD27 positive cells. These populations are dominating the CD62L and CD27 phenotype in progenies derived from many precursors (which is the case in the cited study of Bevan and colleagues).

Assuming antigen withdrawal precludes further recruitment as well as $\text{TCMp} \rightarrow \text{TEMP}$ and $\text{TEMP} \rightarrow \text{TEF}$ differentiation

We also repeated the analysis under the assumption that antigen withdrawal additionally precludes further $\text{TEMP} \rightarrow \text{TEF}$ differentiation events (Fig. 5.14). For the prediction of the mean population size per progenitor (Fig. 5.14A) and the mean fraction of CD62L positive cells (Fig. 5.14B) the resulting dependence on the time point of antigen withdrawal is almost unchanged compared to the previously discussed case. However in this case, for the fraction of CD27 positive cells (Fig. 5.14B), even late truncation of the antigen stimulus would still have an effect. This prediction could be tested experimentally.

So the $N \rightarrow \text{TCMp} \rightarrow \text{TEMP} \rightarrow \text{TEF}$ model would predict that even under the extreme

5. Implications of the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model

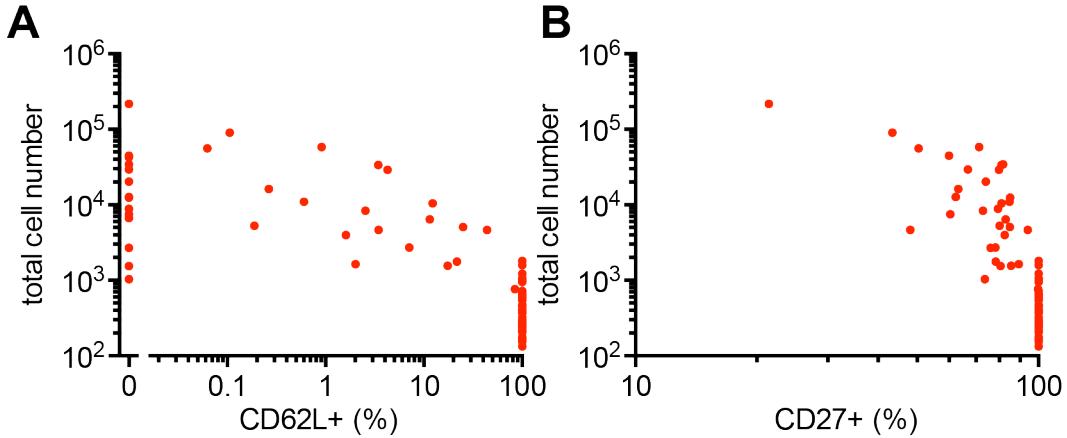


Figure 5.13.: Simulation of single cell-derived progenies assuming antigen withdrawal at day 2 p.i. using the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model. Upon antigen withdrawal, d_0 and d_1 are set to zero. (A-B) Percentage of CD62L $^{+}$ (A) and CD27 $^{+}$ cells (B) versus total cell number.

assumption of total antigen stimulus dependence of the differentiation events, the effect of antigen withdrawal on T cell differentiation markers (e.g. CD62L and CD27) could still be small if assessed in populations that are derived from many progenitor cells. In addition, we were able to explain the observed effect of antigen withdrawal on the absolute population size without having to change the proliferation rates. This is due to the reason that in the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model, the proliferative capacity is coupled to the differentiation status of the T cell. Overall, this analysis demonstrates the necessity to repeat such an experiment using single cell transfers.

5.7. Simulating recall responses

So far we just considered T cell responses derived from naïve progenitor cells (primary response). We will now use the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model to simulate recall responses. To this end we will take advantage of the fact that in this model the precursors of memory cells develop prior to the effector population. This allows for one obvious way to model recall responses: Upon re-challenge, the memory cells that are derived from the memory precursor cells could lead to the generation of effector cells following the same dynamics of differentiation as in the primary response.

Among the population of memory cells, the subset of central memory cells is known to have the greatest proliferative capacity upon re-challenge (recall response) [24, 10]. In the data and also in the N \rightarrow TCMp \rightarrow TEMp \rightarrow TEF model, the absolute number of the central memory precursor cells are only weakly correlated with the total sin-

5.7. Simulating recall responses

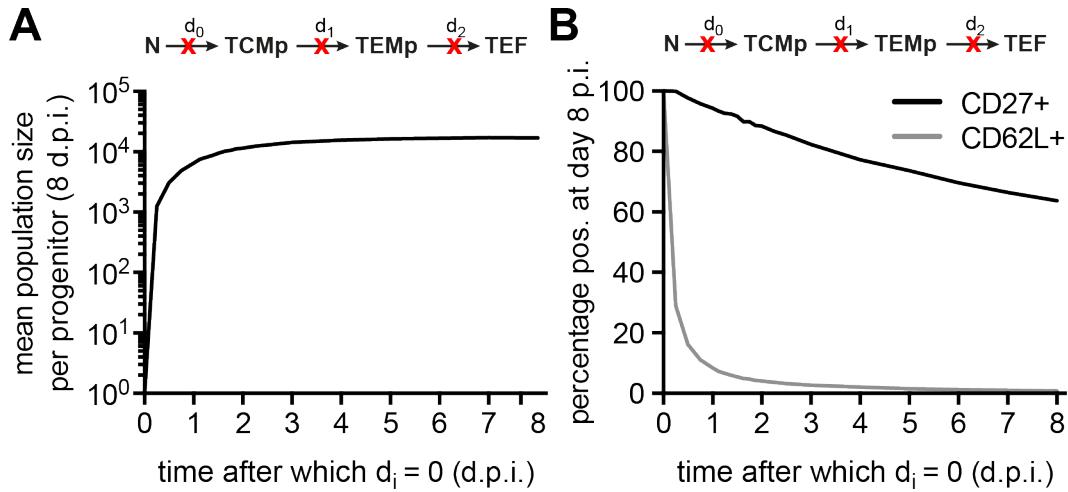


Figure 5.14.: Simulation of antigen withdrawal using the $N \rightarrow \text{TCMp} \rightarrow \text{TEMP} \rightarrow \text{TEF}$ model. Upon antigen withdrawal, d_0 , d_1 and d_2 are set to zero. (A-B) mean population size per progenitor (A) and the predicted mean fractions of CD62L and CD27 positive cells (B) at day 8 p.i. as a function of the time point of antigen withdrawal.

gle cell-derived progeny size (Fig. 5.15A). This would imply that a progeny's total size in the primary infection would be a poor predictor for its total size in a recall response. This prediction was experimentally tested in the group of Dirk Busch at TU Munich (experiments were conducted by Veit Buchholz): Indeed, primary

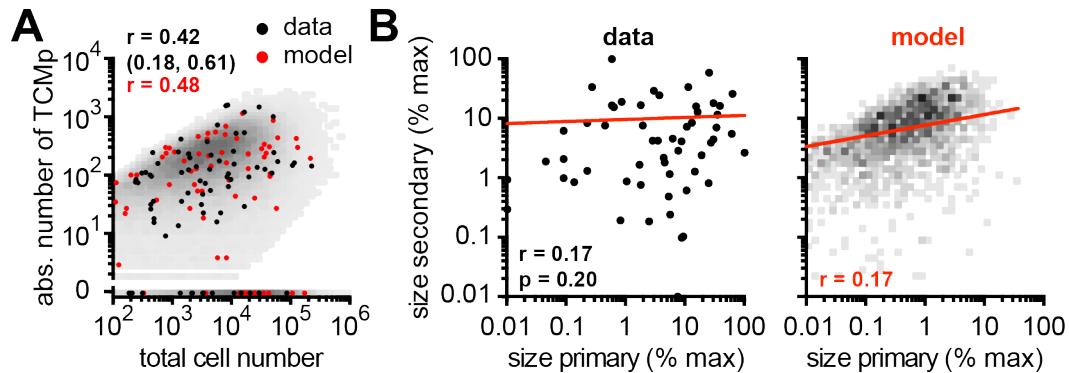


Figure 5.15.: (A) Correlation between total cell number and absolute number of TCMp cells for the single cell-derived progenies at day 8 p.i. (black dots: data; red dots: prediction of the $N \rightarrow \text{TCMp} \rightarrow \text{TEMP} \rightarrow \text{TEF}$ model.) (B) Data (left) and model prediction (right) for the relation between total cell number in primary and secondary response for single cell-derived progenies (largest progeny normalized to one; progenies recovered from peripheral blood at day 8 after primary and day 6 after secondary infection). Red line: log-log regression line. Also indicated in A and B are Spearman's rank correlation coefficients for both model (red) and data (black; together with the 95% confidence interval (A) or p-value (B)). (Experiments by V. Buchholz and D. Busch)

5. Implications of the N→TCMp→TEMp→TEF model

and secondary progeny population sizes are barely correlated (Fig. 5.15B). To make more quantitative statements about the expected correlation between primary and secondary progeny size, we used the N→TCMp→TEMp→TEF model to simulate the recall response in the following way. The TCMP subset sizes on day 8 p.i. of the simulation of the primary response were used as initial sizes of the TCMP compartment for the subsequent simulations (another 6 days). For both simulations we used the parameter value estimates that were obtained by fitting the primary response (cf. Table 3.2). The resulting correlation between primary and secondary size is shown in Fig. 5.15B. The predicted correlation based on Spearman's rank correlation coefficient ($r = 0.17$) coincides with the observed rank correlation coefficient in the data ($r = 0.17$).

Next, we analyzed the relation between secondary and tertiary response. We simulated the tertiary response again based on the number of TCMP cells that resulted from the secondary response simulation (Fig. 5.16). The model predicts a strong correlation between both response sizes. This is due to the fact that the tertiary response is no longer derived from a single expanding T cell, but from a whole population of ancestor TCMP cells.

A strong correlation between secondary and tertiary response is observed in the data (Fig. 5.16; same experiment as mentioned above) as well. However, the agreement between data and model is only qualitative. Clearly, the assumptions that went into the modeling are oversimplifying reality. First of all, the dynamics of secondary and tertiary responses could be different from the primary response. Specifically, the sizes of the secondary and tertiary responses are unlikely to scale with the number of available central memory cells due to competition between the T cells (e.g. for antigen presenting cells or cytokine stimuli). Furthermore, not every memory

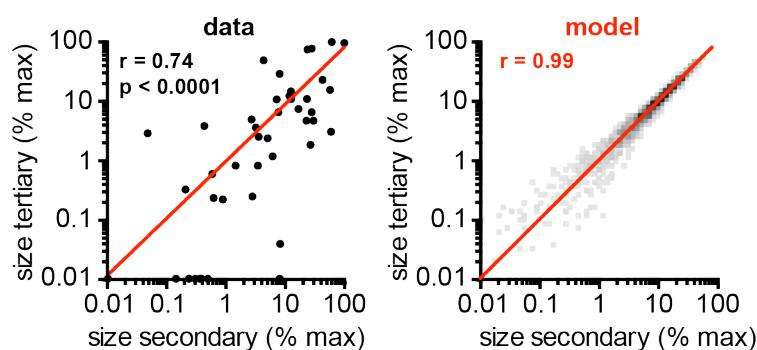


Figure 5.16.: Data (left) and model prediction (right) for relation between secondary and tertiary response for single cell-derived progenies (largest progeny normalized to one; progenies recovered from peripheral blood at day 6 after secondary and day 6 after tertiary infection). Red line: log-log regression line. Spearman's rank correlation coefficients are given for both model (red) and data (black; together with p-value). (Experiments by V. Buchholz and D. Busch)

5.7. Simulating recall responses

precursor cell develops into a functional memory cell [36]. These neglected processes might cause additional variability that could explain the smaller correlation that is observed.

6. Discussion

In this thesis we combined the stochastic modeling of T cell proliferation and differentiation with large-scale model discrimination to gain insight into the mechanisms that underlie CD8⁺ T cell differentiation. Arguably the most significant general finding is that the application of our theoretical framework to single-cell derived fate mapping data yields identifiable models. Thus we can reliably discriminate the contributions of differentiation and selective proliferation to the temporal evolution of the different phenotypic subpopulations. To our knowledge, this has here been achieved for the first time for the *in vivo* dynamics of lymphocytes. In turn, this has allowed us to identify a ‘core motif’ of the differentiation program that is required to account for the experimental data.

Biological questions and mathematical answers. Two scientific questions initially motivated the mathematical modeling: Which mechanisms are responsible for the large variations observed between the total population sizes of the single cell-derived progenies? Is there a differentiation pathway through which memory precursors and effector cells emerge out of a single naïve T cell? First, we found that the unidirectional N→TCMp→TEMp→TEF model fully accounts for the observed variability and in addition also suffices to explain the correlations between the subset sizes. Based on a large scale model comparison allowing for all possible conversions between the TCMP, TEMp and TEF subsets, we furthermore identified the pathway N→TCMp→TEMp→TEF to form the unique minimal core module of the differentiation scheme. None of the models that lack this core module produces a sufficient fit to the data. Moreover, the data strongly constrain additional differentiation steps resulting in only two optional model extensions: Less than 20% of the naïve cells could also differentiate directly into the TEMp phenotype, and, although the best fit estimate of this rate is zero, a differentiation of the TEF cells back into the TEMp compartment cannot be excluded by the data. We also found that these conclusions still hold if we incorporate cell death into the models. It must be stressed, however, that a recent careful study has shown that cell death is minimal until the peak of the immune response – when our data are observed [22]. Therefore, the neglect of cell death is an appropriate simplifying assumption.

In the N→TCMp→TEMp→TEF model, recruited naïve cells give rise to slowly proliferating central memory precursors that eventually differentiate into the first effector memory precursor. Since the TCMP proliferation and differentiation rates are small, both corresponding events happen at very variable time points for in-

6. Discussion

dividual single cell progenies. On the other hand, we showed that the timing of the first cell division and the first TCMP \rightarrow TEMp differentiation event strongly correlate with the later total size of the progeny population. Using the model, the observed large variation in total cell number can thus be attributed to the small TCMP proliferation and differentiation rates.

Markov-process framework. In the models that we introduced in this thesis, cell proliferation and differentiation are modeled by constant rates, implying exponentially distributed waiting times between subsequent division and differentiation events. This assumption led to a straightforward formulation of the models in terms of Markov processes. In turn, we used available theory to efficiently compute summary statistics from the first two moments of the ‘phenotype distribution’ by solving systems of ordinary differential equations. Of note, this approach made large-scale model comparison and model-based predictions (utilizing statistical tools like profile likelihoods and bootstrap) feasible in the first place, because it kept the computation time for the parameter estimation small and allowed us to carry out more than 10^5 individual fits to compute the results in this thesis. By comparison, using stochastic simulations to generate time-dependent distributions of the subpopulation frequencies would have caused prohibitively large computation times for parameter estimation. Of course, we subsequently validated our results with stochastic simulations, observing that our best-fitting model and its variants not only reproduced the measured summary statistics, but also the shapes of measured distributions. This finding suggests that the use of summary statistics to construct objective functions for parameter estimation is potentially a very efficient approach for stochastic systems.

Previously, more general waiting time distributions than the simple exponential have also been used to model lymphocyte expansion, including log-normal, gamma and Weibull distributions [62]. The most prominent example of such models is the “cyton model” put forward by Philip Hodgkin and his co-workers [29, 30]. A generalized form of the cyton model has recently been successfully used to describe B-cell fate decisions [17]. In principal, such a ‘refined’ model could also be appropriate to model T cell diversification. However, the B-cell fate model already introduces fifteen parameters to describe the fate decisions of a single species of B cells (division, death, differentiation and isotype switching) [17]. Parameter estimation in this model was done using measurements of the division time distributions of extensive long-term microscopic imaging *in vitro*. Clearly, for our case where parameter estimation relies on comparatively more sparse *in vivo* data, it would have not been possible to reliably estimate the resulting large parameter set for a model with three T cell subsets. Without the possibility to directly measure division time distributions – which is currently not feasible *in vivo*, a rate-based modeling approach was thus more appropriate.

Interestingly, we originally implemented a modified distribution for the first division

of a cell (TCMp), because it is thought that there is a large delay for the cell leaving the G0 phase of the cell cycle and enter a multiple cycling state, which is indeed seen in cultured T cells [29, 30]. Thus we included a fixed time delay for this step but found that it did not improve the fit to the data (in fact, it made the fit slightly worse). One explanation could be that *in vivo* and *in vitro* proliferation dynamics cannot directly be compared.

The Markov property of the rate-based models also implies that each cell decision to either divide or to differentiate depends only on its current state (naïve, TCMP, TEMp or TEF) and not on the individual cell's division history. It is important to note, however, that this does not imply independence between the number of cell divisions a cell has taken and its differentiation state because the proliferation rates are subset-specific. For example, it has been reported that in CD4⁺ T cells the ability to secrete cytokines correlates with cell division [27]. This phenomenon would be seen also in our model for cytokines produced preferentially by TCMP cells.

Stochastic program of CD8⁺ T cell diversification. Long-lived central and effector memory cells are derived from (a fraction of) their respective precursors following an additional maturation process after infection [35, 49, 51]. The modeling presented here focuses on differentiation pathways between these memory precursors and the subset of pure effector cells (defined via a combination of the surface markers CD62L and CD27). Our results contribute to the current intense discussion on how these memory precursor and effector cells arise out of the pool of activated naïve cells [1].

Interestingly, our results provide no evidence for the currently very popular model stating that memory cells emerge from effector cells late in the immune response, the so-called “linear differentiation model” [46, 60]. In this model, a small fraction of the fully differentiated effector cells down-regulate effector function and acquire effector memory phenotype. A further fraction of the latter continue to develop into central memory cells [60]. Although a later study has been found that TEM→TCM conversion is only observed after adoptive transfer of an unphysiological high number of cells [40], the linear model has received further support by a study showing that a large fraction of the later memory cells did express the effector molecule Granzyme B during primary infection [5]. We cannot formally exclude that effectors become TCM cells at later time points than we observe here, but until the peak of the infection, this conversion is incompatible with the data. Moreover, the N→TCMP→TEMp→TEF model makes no statement about possible effector functions such as cytokine secretion for cells of the TCMP and TEMp phenotype. In fact, cells of the memory precursor phenotype expressed the effector cytokine IFN- γ to a large extent also in the data set that we used for the modeling [9]. Furthermore, we find that the central memory precursor (TCMP) number can be used to predict the extent of a later recall response, thus indicating that these cells develop into func-

6. Discussion

tional memory cells. By contrast, memory responses are essentially uncorrelated with the number of effector cells at the peak of the infection, which is incompatible with a ‘memory-out-of-effectors’ model.

The results of this thesis are compatible with a biological model known as “decreasing potential model” [2] or “progressive differentiation model” [50] that proposes differentiation following the pathway N→TCMp→TEMp→TEF. Cumulative stimuli (comprised of T cell receptor stimulus, co-stimulation and cytokine signals) progressively lead to the loss of memory potential and drive T cell differentiation towards the terminal effector state in which the cells are destined to die during the contraction phase. The findings of this thesis provide strong quantitative evidence for this model. Moreover, we add to this model a progressive increase in the proliferation rates with slowly proliferating central memory precursors giving rise to rapidly expanding effector memory precursors and even faster dividing effector cells. Interestingly, principal component analysis of multiple surface markers on CD8⁺ T cell subtypes leads to a grouping of the subsets that is identical to the order in the dynamic pathway identified by us [44]. Since central memory cells also seem to possess some stem cell-like qualities [23], a recent finding for the stem cell based maintenance of tissue homeostasis is very interesting: A similar hierarchy of proliferation rates, with slow-cycling stem cells giving rise to committed progenitor cells has been found to be essential for the maintenance of homeostasis in the skin [41].

Contrary to the role of the environment in the decreasing potential model, the so-called “autopilot model” [8] proposes that a brief encounter with antigen suffices to program the naïve T cell to commit to multiple rounds of cell divisions leading to a progeny population that is comprised of both effectors and memory precursor cells [33, 43, 58]. In the strictest version of the autopilot model, in which the initial antigen encounter leads to deterministic division and differentiation times (that might be distributed according to the strength of the initial stimulus), stochastic rate models would need to be modified to provide the appropriate framework to model T cell diversification, e.g. by including ‘age-structure’, with the age being the time since recruitment [3]. However, initial recruitment of the cells and cell-cycle progression would still have inevitable stochastic components. Given that live-cell imaging of lymphocyte proliferation and differentiation in culture shows strong stochastic effects even within individual cell clones, a strict autopilot is unlikely [17].

Time-dependent stimuli. For the modeling, we do not have to make any assumptions about environmental cues that might influence cell differentiation and proliferation as long as these cues are approximately constant during the expansion phase. However, the change in pathogen load during the time course of the infection is accompanied by the temporal change of many environmental factors [54] that could have an impact on cell differentiation and proliferation (e.g. cytokines like interleukin-12 and type I interferon [32, 39], T cell receptor stimulus [26]). In this case, the corresponding rates would become time-dependent. We explicitly incor-

porated time-dependent rates into the model to study the response of the so-called “latecomer” T cells: These are naïve T cells that have been transferred 2-4 days post infection. Such late transferred T cells have been used to study the influence of the environment on T cell diversification [13, 16]. These latecomer cells might be exposed to mainly weak stimuli and thus, according to the decreasing potential model, differentiate preferentially into memory precursor cells. Indeed, such an increase in the percentage of memory precursors has been observed for the latecomer T cells [13, 16]. However, using the $N \rightarrow TCMp \rightarrow TEMP \rightarrow TEF$ model we could show that this experimental result does not necessarily imply a change in the differentiation rates of the latecomer T cells. Rather, we found that the lower total cell number of the latecomer cells requires a time-dependency of the proliferation rates that is already sufficient to also explain the increased percentage of memory precursors.

Experimental tests and extensions of the model. For model discrimination in this thesis, we exclusively considered data generated during the expansion phase of the T cell response (until day 8 p.i.). Hence, based on the modeling, we cannot exclude that a fraction of the effector cells does convert into memory cells in the contraction phase following infection. However, the $N \rightarrow TCMp \rightarrow TEMP \rightarrow TEF$ model would predict that memory cells can also be derived from precursors that are direct descendants of naïve cells and that did not go through a fast proliferating effector stage. Moreover, in the model comparisons we found particularly bad fits to the data for those models that did not allow for $N \rightarrow TCMp$ conversion.

In this thesis we discussed data [9] showing that no correlation between primary and secondary expansion size of the single cell-derived progenies can be observed. We found this to be in accordance with the prediction of the $N \rightarrow TCMp \rightarrow TEMP \rightarrow TEF$ model. On the other hand, if most of the central memory cells that play a dominant role in the recall response (TCM cells proliferate vigorously upon re-challenge [24]) developed out of those cells that dominated the primary response (as predicted by the linear model), a certain correlation between primary and secondary size would be expected.

In the “one cell-one fate” [10] or “separate precursor model” [34], the fate of an activated naïve cell is determined before the first cell division (depending only on the stimuli received during activation), leading to a progeny that homogeneously consists of either memory precursors or terminally differentiated effector cells. In this form, the model is directly challenged by the data presented in this thesis [9] and also previous studies [25, 53]: The vast majority of the single cell-derived progenies are comprised of all three subsets. However, the class of models considered in this thesis also include generalizations of such a model in which the naïve T cell can give rise to all three subsets, but which additionally allow for subsequent transitions between the subsets. All of these models that do not include the minimal core module $N \rightarrow TCMp \rightarrow TEMP \rightarrow TEF$ fail to reproduce the data.

It has been suggested that the first cell division of the activated progenitor cell could

6. Discussion

be asymmetric [14]. In this “asymmetric differentiation model”, one of the daughter cells of the naïve cell would generate a progeny of effectors and the other one would give rise to a population of memory precursors. We systematically tested models that incorporate asymmetric first cell divisions. None of the these models produced an acceptable fit to the data. In an extension of the $N \rightarrow TCMp \rightarrow TEMP \rightarrow TEF$ model, we allowed for both symmetric and asymmetric divisions of the naïve cell ($N \rightarrow 2 TCMp$, $N \rightarrow TCMp + TEMP$ and $N \rightarrow 2 TEMP$). We found that the upper bound on the fraction of asymmetric cell divisions would be approximately 15%. This surprisingly low fraction is an interesting model prediction that could be tested experimentally.

To our knowledge, there has only been one study to date that used mathematical modeling to discriminate between alternative diversification models [19]. In this study, two different versions of the decreasing potential model have been compared to the linear differentiation model. All three (deterministic) models were fitted solely to the total number of $CD8^+$ T cells (during expansion and contraction phase). The author found that only the linear differentiation model fitted the data at biological reasonable parameter values. However, all three models were constrained to allow only one of the subsets to have a non-zero proliferation rate.

In contrast, the model discrimination presented in this thesis relies on a much richer data set and can thus consider more general types of models. It is remarkable that the result of this model comparison is relatively clear-cut – with the $N \rightarrow TCMp \rightarrow TEMP \rightarrow TEF$ pathway as the minimal core module and only two optional additional conversions ($N \rightarrow TEMP$ and $TEF \rightarrow TEMP$).

Future work has to be done to extend the models to describe the single cell progenies also during the contraction phase following day 8 p.i. Traditional approaches have used time thresholds after which death rates are set to a non-zero value [15]. A more elegant way would be to limit, for example, the overall number of divisions that a clone can undergo [55] and allow for constant non-zero death rates already during the expansion phase. It would also be interesting to incorporate competition between the cells into the model. Different transfer settings in which the single progenitor cells are transferred together with progenitor populations of varying sizes could provide the experimental basis for such models.

Bibliography

- [1] R. Ahmed, M. J. Bevan, S. L. Reiner, and D. T. Fearon. The precursors of memory: models and controversies. *Nat. Rev. Immunol.*, 9(9):662–668, Sep 2009.
- [2] R. Ahmed and D. Gray. Immunological memory and protective immunity: understanding their relation. *Science*, 272(5258):54–60, Apr 1996.
- [3] R. Antia, C. T. Bergstrom, S. S. Pilyugin, S. M. Kaech, and R. Ahmed. Models of CD8+ responses: 1. What is the antigen-independent proliferation program. *J. Theor. Biol.*, 221(4):585–598, Apr 2003.
- [4] V. P. Badovinac, J. S. Haring, and J. T. Harty. Initial T cell receptor transgenic cell precursor frequency dictates critical aspects of the CD8(+) T cell response to infection. *Immunity*, 26(6):827–841, Jun 2007.
- [5] O. Bannard, M. Kraman, and D. T. Fearon. Secondary replicative function of CD8+ T cells that had developed an effector phenotype. *Science*, 323(5913):505–509, Jan 2009.
- [6] G. I. Bell. Mathematical model of clonal selection and antibody production. *Nature*, 228(5273):739–744, Nov 1970.
- [7] H. Beuneu, F. Lemaitre, J. Deguine, H. D. Moreau, I. Bouvier, Z. Garcia, M. L. Albert, and P. Bousso. Visualizing the functional diversification of CD8+ T cell responses in lymph nodes. *Immunity*, 33(3):412–423, Sep 2010.
- [8] M. J. Bevan and P. J. Fink. The CD8 response on autopilot. *Nat. Immunol.*, 2(5):381–382, May 2001.
- [9] V. R. Buchholz, M. Flossdorf, Hensel I., L. Kretschmer, Weissbrich B., Gräf P., M. Schiemann, T. Höfer, and D. H. Busch. Disparate individual fates compose robust CD8+ T cell immunity. (in revision), 2013.
- [10] V. R. Buchholz, P. Graf, and D. H. Busch. The origin of diversity: studying the evolution of multi-faceted CD8+ T cell responses. *Cell. Mol. Life Sci.*, 69(10):1585–1595, May 2012.
- [11] F. M. Burnet. A modification of Jerne’s theory of antibody production using

Bibliography

- the concept of clonal selection. *CA Cancer J Clin*, 26(2):119–121, 1976.
- [12] K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, 2 edition, 2002.
 - [13] D. M. Catron, L. K. Rusch, J. Hataye, A. A. Itano, and M. K. Jenkins. CD4+ T cells that enter the draining lymph nodes after antigen injection participate in the primary response and become central-memory cells. *J. Exp. Med.*, 203(4):1045–1054, Apr 2006.
 - [14] J. T. Chang, V. R. Palanivel, I. Kinjyo, F. Schambach, A. M. Intlekofer, A. Banerjee, S. A. Longworth, K. E. Vinup, P. Mrass, J. Oliaro, N. Killeen, J. S. Orange, S. M. Russell, W. Weninger, and S. L. Reiner. Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science*, 315(5819):1687–1691, Mar 2007.
 - [15] R. J. De Boer, D. Homann, and A. S. Perelson. Different dynamics of CD4+ and CD8+ T cell responses during and after acute lymphocytic choriomeningitis virus infection. *J. Immunol.*, 171(8):3928–3935, Oct 2003.
 - [16] W. N. D’Souza and S. M. Hedrick. Cutting edge: latecomer CD8 T cells are imprinted with a unique differentiation program. *J. Immunol.*, 177(2):777–781, Jul 2006.
 - [17] K. R. Duffy, C. J. Wellard, J. F. Markham, J. H. Zhou, R. Holmberg, E. D. Hawkins, J. Hasbold, M. R. Dowling, and P. D. Hodgkin. Activation-induced B cell fates are selected by intracellular stochastic competition. *Science*, 335(6066):338–341, Jan 2012.
 - [18] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.
 - [19] V. V. Ganusov. Discriminating between different pathways of memory CD8+ T cell differentiation. *J. Immunol.*, 179(8):5006–5013, Oct 2007.
 - [20] V. V. Ganusov, S. S. Pilyugin, R. J. de Boer, K. Murali-Krishna, R. Ahmed, and R. Antia. Quantifying cell turnover using CFSE data. *J. Immunol. Methods*, 298(1-2):183–200, Mar 2005.
 - [21] C. Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences (Springer Series in Synergetics)*. Springer, 4th ed. edition, January 2009.
 - [22] K. R. Garrod, H. D. Moreau, Z. Garcia, F. Lemaitre, I. Bouvier, M. L. Albert, and P. Bousso. Dissecting T cell contraction in vivo using a genetically encoded reporter of apoptosis. *Cell Rep*, 2(5):1438–1447, Nov 2012.
 - [23] L. Gattinoni, E. Lugli, Y. Ji, Z. Pos, C. M. Paulos, M. F. Quigley, J. R. Almeida,

Bibliography

- E. Gostick, Z. Yu, C. Carpenito, E. Wang, D. C. Douek, D. A. Price, C. H. June, F. M. Marincola, M. Roederer, and N. P. Restifo. A human memory T cell subset with stem cell-like properties. *Nat. Med.*, 17(10):1290–1297, Oct 2011.
- [24] J. Geginat, A. Lanzavecchia, and F. Sallusto. Proliferation and differentiation potential of human CD8+ memory T-cell subsets in response to antigen or homeostatic cytokines. *Blood*, 101(11):4260–4266, Jun 2003.
- [25] C. Gerlach, J. W. van Heijst, E. Swart, D. Sie, N. Armstrong, R. M. Kerkhoven, D. Zehn, M. J. Bevan, K. Schepers, and T. N. Schumacher. One naive T cell, multiple fates in CD8+ T cell differentiation. *J. Exp. Med.*, 207(6):1235–1246, Jun 2010.
- [26] A. V. Gett, F. Sallusto, A. Lanzavecchia, and J. Geginat. T cell fitness determined by signal strength. *Nat. Immunol.*, 4(4):355–360, Apr 2003.
- [27] A.V. Gett and P.D. Hodgkin. Cell division regulates the t cell cytokine repertoire, revealing a mechanism underlying immune class regulation. *Proc Natl Acad Sci U S A*, 95(16):9488–93, 1998.
- [28] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, December 1977.
- [29] E. D. Hawkins, J. F. Markham, L. P. McGuinness, and P. D. Hodgkin. A single-cell pedigree analysis of alternative stochastic lymphocyte fates. *Proc. Natl. Acad. Sci. U.S.A.*, 106(32):13457–13462, Aug 2009.
- [30] E. D. Hawkins, M. L. Turner, M. R. Dowling, C. van Gend, and P. D. Hodgkin. A model of immune regulation as a consequence of randomized lymphocyte division and death times. *Proc. Natl. Acad. Sci. U.S.A.*, 104(12):5032–5037, Mar 2007.
- [31] N. K. Jerne. THE NATURAL-SELECTION THEORY OF ANTIBODY FORMATION. *Proc. Natl. Acad. Sci. U.S.A.*, 41(11):849–857, Nov 1955.
- [32] N. S. Joshi, W. Cui, A. Chandele, H. K. Lee, D. R. Urso, J. Hagman, L. Gapin, and S. M. Kaech. Inflammation directs memory precursor and short-lived effector CD8(+) T cell fates via the graded expression of T-bet transcription factor. *Immunity*, 27(2):281–295, Aug 2007.
- [33] S. M. Kaech and R. Ahmed. Memory CD8+ T cell differentiation: initial antigen encounter triggers a developmental program in naïve cells. *Nat. Immunol.*, 2(5):415–422, May 2001.
- [34] S. M. Kaech and W. Cui. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat. Rev. Immunol.*, 12(11):749–761, Nov 2012.

Bibliography

- [35] S. M. Kaech, S. Hemby, E. Kersh, and R. Ahmed. Molecular and functional profiling of memory CD8 T cell differentiation. *Cell*, 111(6):837–851, Dec 2002.
- [36] S. M. Kaech, J. T. Tan, E. J. Wherry, B. T. Konieczny, C. D. Surh, and R. Ahmed. Selective expression of the interleukin 7 receptor identifies effector CD8 T cells that give rise to long-lived memory cells. *Nat. Immunol.*, 4(12):1191–1198, Dec 2003.
- [37] V. Kalia, S. Sarkar, T. S. Gourley, B. T. Rouse, and R. Ahmed. Differentiation of memory B and T cells. *Current opinion in immunology*, 18(3):255–264, June 2006.
- [38] D. G. Kendall. On the Generalized "Birth-and-Death" Process. *The Annals of Mathematical Statistics*, 19(1):1–15, March 1948.
- [39] G. A. Kolumam, S. Thomas, L. J. Thompson, J. Sprent, and K. Murali-Krishna. Type I interferons act directly on CD8 T cells to allow clonal expansion and memory formation in response to viral infection. *J. Exp. Med.*, 202(5):637–650, Sep 2005.
- [40] A. L. Marzo, K. D. Klonowski, A. Le Bon, P. Borrow, D. F. Tough, and L. Lefrancois. Initial T cell frequency dictates memory CD8+ T cell lineage commitment. *Nat. Immunol.*, 6(8):793–799, Aug 2005.
- [41] G. Mascre, S. Dekoninck, B. Drogat, K. K. Youssef, S. Brohee, P. A. Sotiropoulou, B. D. Simons, and C. Blanpain. Distinct contribution of stem and progenitor cells to epidermal maintenance. *Nature*, 489(7415):257–262, Sep 2012.
- [42] T. R. Mempel, S. E. Henrickson, and U. H. Von Andrian. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature*, 427(6970):154–159, Jan 2004.
- [43] R. Mercado, S. Vijh, S. E. Allen, K. Kerksiek, I. M. Pilip, and E. G. Pamer. Early programming of T cell populations responding to bacterial infection. *J. Immunol.*, 165(12):6833–6839, Dec 2000.
- [44] E. W. Newell, N. Sigal, S. C. Bendall, G. P. Nolan, and M. M. Davis. Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of CD8+ T cell phenotypes. *Immunity*, 36(1):142–152, Jan 2012.
- [45] J. Oliaro, V. Van Ham, F. Sacirbegovic, A. Pasam, Z. Bomzon, K. Pham, M. J. Ludford-Menting, N. J. Waterhouse, M. Bots, E. D. Hawkins, S. V. Watt, L. A. Cluse, C. J. Clarke, D. J. Izon, J. T. Chang, N. Thompson, M. Gu, R. W. Johnstone, M. J. Smyth, P. O. Humbert, S. L. Reiner, and S. M. Russell.

Bibliography

- Asymmetric cell division of T cells upon antigen presentation uses multiple conserved mechanisms. *J. Immunol.*, 185(1):367–375, Jul 2010.
- [46] J. T. Opferman, B. T. Ober, and P. G. Ashton-Rickardt. Linear differentiation of cytotoxic effectors into memory T lymphocytes. *Science*, 283(5408):1745–1748, Mar 1999.
 - [47] K. A. Pape, J. J. Taylor, R. W. Maul, P. J. Gearhart, and M. K. Jenkins. Different B cell populations mediate early and late memory during an endogenous immune response. *Science*, 331(6021):1203–1207, Mar 2011.
 - [48] M. Prlic, G. Hernandez-Hoyos, and M. J. Bevan. Duration of the initial TCR stimulus controls the magnitude but not functionality of the CD8+ T cell response. *J. Exp. Med.*, 203(9):2135–2143, Sep 2006.
 - [49] R. L. Rutishauser and S. M. Kaech. Generating diversity: transcriptional regulation of effector and memory CD8 T-cell differentiation. *Immunol. Rev.*, 235(1):219–233, May 2010.
 - [50] F. Sallusto, J. Geginat, and A. Lanzavecchia. Central memory and effector memory T cell subsets: function, generation, and maintenance. *Annu. Rev. Immunol.*, 22:745–763, 2004.
 - [51] F. Sallusto, D. Lenig, R. Forster, M. Lipp, and A. Lanzavecchia. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature*, 401(6754):708–712, Oct 1999.
 - [52] V. Shahrezaei and P. S. Swain. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 105(45):17256–17261, Nov 2008.
 - [53] C. Stemberger, K. M. Huster, M. Koffler, F. Anderl, M. Schiemann, H. Wagner, and D. H. Busch. A single naive CD8+ T cell precursor can develop into diverse effector and memory subsets. *Immunity*, 27(6):985–997, Dec 2007.
 - [54] C. Stemberger, M. Neuenhahn, V. R. Buchholz, and D. H. Busch. Origin of CD8+ effector and memory T cell subsets. *Cell. Mol. Immunol.*, 4(6):399–405, Dec 2007.
 - [55] V. G. Subramanian, K. R. Duffy, M. L. Turner, and P. D. Hodgkin. Determining the expected variability of immune responses using the cyton model. *J Math Biol.*, 56(6):861–892, Jun 2008.
 - [56] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.*, 98(15):8614–8619, Jul 2001.
 - [57] J. W. van Heijst, C. Gerlach, E. Swart, D. Sie, C. Nunes-Alves, R. M. Kerkhoven, R. Arens, M. Correia-Neves, K. Schepers, and T. N. Schumacher.

Bibliography

- Recruitment of antigen-specific CD8+ T cells in response to infection is markedly efficient. *Science*, 325(5945):1265–1269, Sep 2009.
- [58] M. J. van Stipdonk, E. E. Lemmens, and S. P. Schoenberger. Naïve CTLs require a single brief period of antigenic stimulation for clonal expansion and differentiation. *Nat. Immunol.*, 2(5):423–429, May 2001.
 - [59] D. J. Venzon and S. H. Moolgavkar. A Method for Computing Profile-Likelihood-Based Confidence Intervals. *Applied Statistics*, 37(1):87–94, 1988.
 - [60] E. J. Wherry, V. Teichgraber, T. C. Becker, D. Masopust, S. M. Kaech, R. Antia, U. H. von Andrian, and R. Ahmed. Lineage relationship and protective immunity of memory CD8 T cell subsets. *Nat. Immunol.*, 4(3):225–234, Mar 2003.
 - [61] M. A. Williams and M. J. Bevan. Effector and memory CTL differentiation. *Annu. Rev. Immunol.*, 25:171–192, 2007.
 - [62] A. Zilman, V. V. Ganusov, and A. S. Perelson. Stochastic models of lymphocyte proliferation and death. *PLoS ONE*, 5(9), 2010.

A. Appendix

A.1. ODE system for first and second moments of the general model

We here give the ODE system for the time evolution of the first and second moments of the general model (without cell death: depicted in Fig. 4.2, $\delta_i = 0$; with cell death: Fig. 4.8). As described in chapter 3.2, in order to obtain an ODE system for the time dependence of the mean values, we successively differentiate the probability generating function F from eqn. 4.5 with respect to the z_i and then set $z_i = 1$. This results in

$$\begin{aligned}\dot{\langle n_0 \rangle} &= -(\text{d}0_1 + \text{d}0_2 + \text{d}0_3) \langle n_0 \rangle \\ \dot{\langle n_1 \rangle} &= (\lambda_1 - d_{-3} - d_1 - \delta_1) \langle n_1 \rangle + d_{-1} \langle n_2 \rangle + d_3 \langle n_3 \rangle + \text{d}0_1 \langle n_0 \rangle \\ \dot{\langle n_2 \rangle} &= (\lambda_2 - d_{-1} - d_2 - \delta_2) \langle n_2 \rangle + d_{-2} \langle n_3 \rangle + d_1 \langle n_1 \rangle + \text{d}0_2 \langle n_0 \rangle \\ \dot{\langle n_3 \rangle} &= (\lambda_3 - d_{-2} - d_3 - \delta_3) \langle n_3 \rangle + d_{-3} \langle n_1 \rangle + d_2 \langle n_2 \rangle + \text{d}0_3 \langle n_0 \rangle,\end{aligned}\tag{A.1}$$

where $\langle n_i \rangle = \partial_{z_i} F|_1$. We further obtain the following ODE system for the time evolution of the second derivatives of F evaluated at $z_i = 1$

$$\begin{aligned}\dot{F}_{00} &= -2(\text{d}0_1 + \text{d}0_2 + \text{d}0_3) F_{00} \\ \dot{F}_{03} &= -F_{03} (d_{-2} + d_3 + \delta_3 + \text{d}0_1 + \text{d}0_2 - \lambda_3) + d_{-3} F_{01} + d_2 F_{02} + \text{d}0_3 (F_{00} - F_{03}) \\ \dot{F}_{33} &= 2(d_{-3} F_{13} - F_{33} (d_{-2} + d_3 + \delta_3) + d_2 F_{23} + \text{d}0_3 F_{03} + \lambda_3 (F_{33} + \langle n_3 \rangle)) \\ \dot{F}_{02} &= -F_{02} (d_{-1} + d_2 + \delta_2 + \text{d}0_1 + \text{d}0_3 - \lambda_2) + d_1 F_{01} + d_{-2} F_{03} + \text{d}0_2 (F_{00} - F_{02}) \\ \dot{F}_{23} &= d_{-3} F_{12} + d_1 F_{13} - F_{23} (d_{-1} + d_2 + d_3 + \delta_2 + \delta_3 - \lambda_2 - \lambda_3) \\ &\quad + d_{-2} (F_{33} - F_{23}) + d_2 F_{22} + \text{d}0_2 F_{03} + \text{d}0_3 F_{02} \\ \dot{F}_{22} &= 2(d_1 F_{12} - F_{22} (d_{-1} + d_2 + \delta_2) + d_{-2} F_{23} + \text{d}0_2 F_{02} + \lambda_2 (F_{22} + \langle n_2 \rangle)) \\ \dot{F}_{01} &= -F_{01} (d_{-3} + d_1 + \delta_1 + \text{d}0_1 + \text{d}0_2 + \text{d}0_3 - \lambda_1) + d_{-1} F_{02} + d_3 F_{03} + \text{d}0_1 F_{00} \\ \dot{F}_{13} &= -F_{13} (d_{-3} + d_{-2} + d_1 + \delta_1 + \delta_3 - \lambda_1 - \lambda_3) + d_{-3} F_{11} \\ &\quad + d_2 F_{12} + d_3 (F_{33} - F_{13}) + d_{-1} F_{23} + \text{d}0_3 F_{01} + \text{d}0_1 F_{03} \\ \dot{F}_{12} &= -F_{12} (d_{-3} + d_1 + d_2 + \delta_1 + \delta_2 - \lambda_1 - \lambda_2) + d_{-2} F_{13} \\ &\quad + d_{-1} (F_{22} - F_{12}) + d_1 F_{11} + d_3 F_{23} + \text{d}0_2 F_{01} + \text{d}0_1 F_{02} \\ \dot{F}_{11} &= 2(-F_{11} (d_{-3} + d_1 + \delta_1 - \lambda_1) + d_{-1} F_{12} + d_3 F_{13} + \text{d}0_1 F_{01} + \lambda_1 \langle n_1 \rangle),\end{aligned}\tag{A.2}$$

A. Appendix

where $F_{ij} = \partial_{z_i} \partial_{z_j} F|_1$ (see eqn. 3.19 for the initial conditions).

A.2. ODE system for first and second moments of general asymmetric cell division model

We here give the ODE system for the time evolution of the first and second moments of the general asymmetric cell division model depicted in Fig. 4.11.

As described in chapter 3.2, in order to obtain an ODE system for the time dependence of the mean values, we successively differentiate the probability generating function F from eqn. 4.7 with respect to the z_i and then set $z_i = 1$. This results in

$$\begin{aligned}\dot{\langle n_0 \rangle} &= -d_0 \langle n_0 \rangle \\ \dot{\langle n_1 \rangle} &= (\lambda_1 - d_1) \langle n_1 \rangle + d_{-1} \langle n_2 \rangle + d_0 \langle n_0 \rangle \\ \dot{\langle n_2 \rangle} &= -(d_{-1} + d_2 - \lambda_2) \langle n_2 \rangle + d_{-2} \langle n_3 \rangle + d_0 \langle n_0 \rangle + d_1 \langle n_1 \rangle \\ \dot{\langle n_3 \rangle} &= (\lambda_3 - d_{-2}) \langle n_3 \rangle + d_2 \langle n_2 \rangle,\end{aligned}\tag{A.3}$$

where $\langle n_i \rangle = \partial_{z_i} F|_1$. We further obtain the following ODE system for the time evolution of the second derivatives of F evaluated at $z_i = 1$

$$\begin{aligned}\dot{F}_{00} &= -2d_0 F_{00} \\ \dot{F}_{03} &= d_2 F_{02} - (d_{-2} + d_0 - \lambda_3) F_{03} \\ \dot{F}_{33} &= 2(d_{-2}(-F_{33}) + d_2 F_{23} + \lambda_3(F_{33} + \langle n_3 \rangle)) \\ \dot{F}_{02} &= d_1 F_{01} - (d_{-1} + d_2 - \lambda_2) F_{02} + d_{-2} F_{03} + d_0(F_{00} - F_{02}) \\ \dot{F}_{23} &= d_1 F_{13} + F_{23}(-d_{-1} - d_2 + \lambda_2 + \lambda_3) \\ &\quad + d_0 F_{03} + d_{-2}(F_{33} - F_{23}) + d_2 F_{22} \\ \dot{F}_{22} &= 2(d_1 F_{12} + d_0 F_{02} + d_{-2} F_{23} - (d_{-1} + d_2) F_{22} + \lambda_2(F_{22} + \langle n_2 \rangle)) \\ \dot{F}_{01} &= -(d_0 + d_1 - \lambda_1) F_{01} + d_{-1} F_{02} + d_0 F_{00} \\ \dot{F}_{13} &= (-d_{-2} - d_1 + \lambda_1 + \lambda_3) F_{13} + d_2 F_{12} + d_0 F_{03} + d_{-1} F_{23} \\ \dot{F}_{12} &= (-d_1 - d_2 + \lambda_1 + \lambda_2) F_{12} + d_0(F_{01} + F_{02} + \langle n_0 \rangle) \\ &\quad + d_{-2} F_{13} + d_{-1}(F_{22} - F_{12}) + d_1 F_{11} \\ \dot{F}_{11} &= 2((\lambda_1 - d_1) F_{11} + d_0 F_{01} + d_{-1} F_{12} + \lambda_1 \langle n_1 \rangle),\end{aligned}\tag{A.4}$$

where $F_{ij} = \partial_{z_i} \partial_{z_j} F|_1$ (see eqn. 3.19 for the initial conditions).

A.3. ODE system for third moments of N→TCMp→TEMp→TEF model

A.3. ODE system for third moments of N→TCMp→TEMp→TEF model

We here give the ODE system for the time evolution of the third moments of the N→TCMp→TEMp→TEF model.

As described in chapter 3.2, in order to obtain an ODE system for the time dependence of the third order moments, we successively differentiate the probability generating function from eqn. 3.16 three times with respect to the z_i and then set $z_i = 1$. This results in

$$\begin{aligned}
\dot{F}_{000} &= -3d_0 F_{000} \\
\dot{F}_{333} &= 3(d_2 F_{233} + \lambda_3(2F_{33} + F_{333})) \\
\dot{F}_{222} &= 3(d_1 F_{122} - d_2 F_{222} + \lambda_2(2F_{22} + F_{222})) \\
\dot{F}_{111} &= 3((\lambda_1 - d_1)F_{111} + d_0 F_{011} + 2\lambda_1 F_{11}) \\
\dot{F}_{033} &= -d_0 F_{033} + 2d_2 F_{023} + 2\lambda_3(F_{03} + F_{033}) \\
\dot{F}_{022} &= 2d_1 F_{012} - (d_0 + 2d_2)F_{022} + 2\lambda_2(F_{02} + F_{022}) \\
\dot{F}_{011} &= 2(d_0 F_{001} + \lambda_1 F_{01}) - (d_0 + 2d_1 - 2\lambda_1)F_{011} \\
\dot{F}_{003} &= (\lambda_3 - 2d_0)F_{003} + d_2 F_{002} \\
\dot{F}_{002} &= d_1 F_{001} + (-2d_0 - d_2 + \lambda_2)F_{002} \\
\dot{F}_{001} &= (-2d_0 - d_1 + \lambda_1)F_{001} + d_0 F_{000} \\
\dot{F}_{223} &= 2d_1 F_{123} + (\lambda_3 - 2d_2)F_{223} + d_2 F_{222} + 2\lambda_2(F_{23} + F_{223}) \\
\dot{F}_{113} &= (\lambda_3 - 2d_1)F_{113} + 2d_0 F_{013} + d_2 F_{112} + 2\lambda_1(F_{13} + F_{113}) \\
\dot{F}_{233} &= d_1 F_{133} + (\lambda_2 - d_2)F_{233} + 2d_2 F_{223} + 2\lambda_3(F_{23} + F_{233}) \\
\dot{F}_{133} &= (\lambda_1 - d_1)F_{133} + 2d_2 F_{123} + d_0 F_{033} + 2\lambda_3(F_{13} + F_{133}) \\
\dot{F}_{122} &= (-d_1 - 2d_2 + \lambda_1)F_{122} + 2d_1 F_{112} + d_0 F_{022} + 2\lambda_2(F_{12} + F_{122}) \\
\dot{F}_{112} &= (-2d_1 - d_2 + \lambda_2)F_{112} + 2d_0 F_{012} + d_1 F_{111} + 2\lambda_1(F_{12} + F_{112}) \\
\dot{F}_{023} &= d_1 F_{013} + (-d_0 - d_2 + \lambda_2 + \lambda_3)F_{023} + d_2 F_{022} \\
\dot{F}_{013} &= (-d_0 - d_1 + \lambda_1 + \lambda_3)F_{013} + d_2 F_{012} + d_0 F_{003} \\
\dot{F}_{012} &= (-d_1 - d_2 + \lambda_1 + \lambda_2)F_{012} + d_0(F_{002} - F_{012}) + d_1 F_{011} \\
\dot{F}_{123} &= (-d_2 + \lambda_1 + \lambda_2 + \lambda_3)F_{123} + d_1(F_{113} - F_{123}) + d_2 F_{122} + d_0 F_{023},
\end{aligned} \tag{A.5}$$

where $F_{ijk} = \partial_{z_i} \partial_{z_j} \partial_{z_k} F|_1$ (initial conditions: $F_{ijk}(0) = 0$). The ODE system for the first and second moments is given in chapter 3.2. The third moments are needed to compute the skewness that we discussed in chapter 3.3.

A. Appendix

A.4. Modeling the latecomer cells by assuming reduced proliferation rates

As discussed in chapter 5.5, an alternative approach to model the latecomer T cells is to assume that early and late transferred T cells proliferate at different rates. Specifically, we assumed that the proliferation rates of the latecomers are reduced by a constant factor c (relative to the proliferation rates of the early transferred cells). We then fitted this model to the data of the early and late transferred T cells and estimated the factor c together with the six remaining parameters of the N→TEF→TEMP→TCMP model. A good fit to the data can be obtained (Fig. A.1). The best fit estimate for c is 0.7.

A.4. Modeling the latecomer cells by assuming reduced proliferation rates

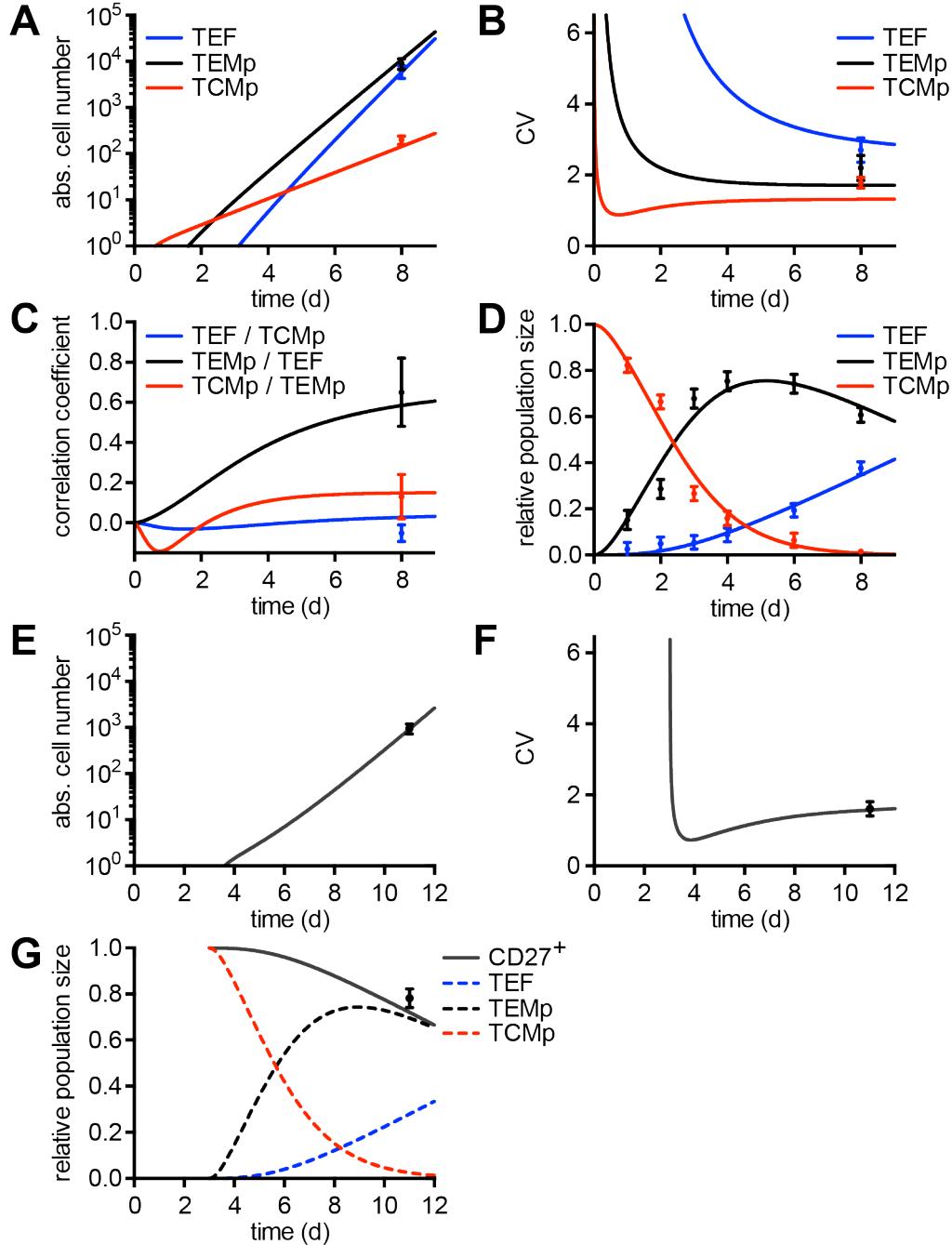


Figure A.1.: Fit of the $N \rightarrow TCMP \rightarrow TEMp \rightarrow TEF$ model to the data of the early and late transferred cells simultaneously (assuming just the proliferation rates of the latecomers get reduced by a (fitted) factor of $c = 0.72$). (A-D) Early transferred cells. Best fit to the mean values (A), CVs (B) and pairwise correlations (C) of the single cell progeny data of day 8 p.i. and to the relative subset size kinetics (D). (E-G) Latecomer cells. Best fit to the mean (E) and CV (F) of the total cell number as well as to the fraction of CD27⁺ cells (G). Data is shown in black. The dashed lines in (C) show the predicted time courses for the TCMP, TEMp and TEF subsets of the latecomers. (Experiments by V. Buchholz and D. Busch)

Abbreviations

AICc	corrected Akaike information criterion
CC	correlation coefficient
CFSE	Carboxyfluorescein succinimidyl ester
CV	coefficient of variation
N	naïve
ODE	ordinary differential equation
PDE	partial differential equation
p.d.f.	probability density function
PGF	probability generating function
p.i.	post infection
TCM	central memory T cells
TCMp	central memory precursor T cells
TEF	effector T cells
TEM	effector memory T cells
TEMp	effector memory precursor T cells

Acknowledgments

Above all, I would like to express my deepest gratitude to my supervisor Thomas Höfer, whose support I could always count on and from whom I learned a lot. His scientific intuition, passion and expertise has been an inspiration and an invaluable source for ideas.

I thank Prof. Roland Eils for being thesis referee and Chairperson of the examining committee. I would also like to thank Prof. Ursula Klingmüller and Prof. Max Löhning for being in the examining committee.

I am very grateful to my parents and my whole family for all their unconditional support and understanding.

I would very much like to thank my good friend Florian for all the great time we have spent together both at and after work. I cannot imagine how my time as a PhD student would have been without the numerous coffees and beers that we had together.

I very much thank my good friend and office mate Anna – one of the very few persons that I don't mind seeing eight hours a day – for brightening up everyday office-life and sharing the obligatory 5-pm-cookies with me.

I am very thankful to Veit Buchholz and Dirk Busch for trusting me with their data. This work would not have been possible without Veit's brilliant ideas and devotion to this project. It is hard to imagine a more intense and efficient collaboration. I really enjoyed the uncountable discussions we had and look forward to many more.

I thank the whole Höfer group for the pleasant working atmosphere and especially Diana who often offered kind help with last minute organizational necessities.

Special thanks go to Anna Schulze, Florian Lamprecht, Elan Gin and Sandip Kar for critical reading of this thesis.