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
Dipl.-Math. Michael Gabel
born in Heidelberg, Germany
Oral examination: 23.04.2018

Quantifying $CD8^+$ T cell dynamics in infectious diseases

Referees: Prof. Dr. Ursula Kummer
Dr. Frederik Graw

English summary

CD8$^+$ T cells are an important part of the adaptive immune system. They are not only able to efficiently kill infected cells, but can also provide protection against subsequent infections. Recently, T cell-based vaccines, which are able to elicit protection-mediating CD8$^+$ T cell responses, have shown their potential against various infectious diseases, including malaria or HIV. However, the efficient implementation of these approaches is currently hampered by a lack of knowledge regarding the dynamical processes generating the protective responses. In this thesis, we combined experimental data and mathematical modelling to analyse and quantify CD8+ T cell dynamics in different infectious diseases.

First, we study a specific immune responses elicited by cytomegalovirus (CMV) infection, known as 'memory inflation'. We used mathematical modelling to test different hypotheses regarding the processes that generate and maintain memory inflation and analysed how viral dynamics shape the corresponding CD8$^+$ T cell responses. Since CMV has already been used as a vaccine vector expressing foreign epitopes, our findings are relevant for improving the efficacy of these vector-based vaccination approaches.

To generate sufficient protection, vaccination strategies usually require the application of one or more booster injections. Here, factors such as dosage, frequency and timing of injections can influence the protective levels reached. To determine the impact of current vaccination strategies against malaria, we analysed the influence of vaccination regimens, differing in dosage and frequency of injections, on the generation of organ-specific CD8+ T cell responses. We identified the underlying cellular differentiation and migration pathway and determined the impact of different vaccination doses on the build-up and maintenance of protection-mediating liver-resident memory cells. Our results do not only provide a quantitative understanding of CD8$^+$ T cell responses elicited by immunisation, but can also be used to improve existing vaccination approaches.

Understanding the impact of T cell-based vaccination regimes on the immune system requires knowledge about the underlying cellular dynamics. While mathematical modelling allows the determination of differentiation pathways and the quantification of cellular turnover, the reliability of these analyses depends strongly on the quality of the available data. While the labelling of cells has been a useful method to increase the amount of information within cellular data, it has not been analysed so far how the design of applied labelling strategies affects the mathematical estimation of cellular turnover. To this end, we determined the robustness of different labelling strategies to infer cellular dynamics including data suffering from experimental limitations. Our findings can be used as a guideline to determine cellular dynamics more accurately in future experiments.

In summary, by combining experimental data and mathematical modelling our results do not provide a quantitative understanding of $CD8^+$ T cell responses in infectious diseases, but can also be used to improve the efficiency and efficacy of T cell-based vaccines.

Deutsche Zusammenfassung

$CD8^+$ T-Zellen sind ein wichtiger Teil des adaptiven Immunsystems. Sie sind nicht nur in der Lage, effizient infizierte Zellen zu töten, sondern können auch Schutz gegen Folgeinfektionen bieten. Vor Kurzem zeigten T-Zell-basierte Impfstoffe, welche schützende $CD8^+$ T-Zell-Antworten hervorrufen können, auch ihre Wirksamkeit gegen verschieden Infektionskrankheiten, wie Malaria oder HIV. Allerdings fehlt momentan noch genaueres Wissen über die zugrundeliegenden schutzbietenden Prozesse, um diese Ansätze gezielt anzuwenden. In dieser Dissertation kombinieren wir deswegen experimentelle Daten und mathematische Modellierung um $CD8^+$ T-Zell-Dynamiken in verschiedenen Infektionskrankheiten zu analysieren und quantifizieren.

Zunächst studierten wir eine spezielle Immunantwort, die sogenannte Gedächtnisinflation ,die durch Zytomegalovirusinfektion (ZMV) hervorgerufen wird. Hier benutzten wir mathematische Modellierung um verschiedene Hypothese hinsichtlich der Entstehung und Erhaltung der Gedächtnisinflation zu finden und analysierten wie virale Dynamiken die zugehörige $CD8^+$ T-Zell-Antwort beeinflussen. Da ZMV bereits als viraler Vektor benutzt wurde, welcher fremde Epitope präsentiert, sind unsere Resultate relevant um die Effektivität von vektorbasierten Impfmethoden zu verbessern.

Oftmals entsteht Schutz durch Vakzinierung erst nach der Zugabe einer oder mehrerer Boosterinjektionen. Hier können Faktoren wie Dosierung, Anzahl oder Zeitrahmen der Injektionen über den bereitgestellten Impfschutz entscheiden. Deswegen analysierten wir den Einfluss von verschiedenen Vakzinierungsstrategien, welche sich in der Anzahl ihrer Boosterinjektionen sowie der verabreichten Dosis unterscheiden, in Bezug auf die Generierung organabhängiger $CD8^+$ T-Zell-Antworten gegen Malaria. Wir identifizierten den zugrundeliegenden Differentiations- und Migrationspfad und modellierten den Effekt verschiedener Impfdosierungen auf die Entstehung und Erhaltung schutzbietender $CD8^+$ T-Zellen, welche in der Leber residieren. Unsere Resultate erlauben es nicht nur, ein quantitatives Verständnis von durch Immunisierung hervorgerufenen $CD8^+$ T-Zell-Antworten zu entwickeln, sondern ermöglichen es auch, bestehende Vakzinierungsstrategien zu verbessern.

Um den Effekt von T-Zell-basierten Vakzinierungsansätzen auf die Immunantwort zu verstehen, benötigt man Wissen über die zugrundeliegenden zellulären Dynamiken. Hier hängt die Zuverlässigkeit mathematischer Modellierung, welche es erlaubt Differentiationswege zu identifizieren und zelluläre Umsatzraten zu quantifizieren, stark von der Qualität der verfügbaren Daten ab. Hier stellt das Markieren von Zellen eine nützliche Methode bereit, um den Informationsgehalt von zellulären Daten zu erhöhen. Allerdings wurde bis jetzt noch nicht untersucht, wie das

Design von Markierungsstrategien die mathematische Schätzungen zellulärer Parameter beeinflusst. Deswegen untersuchten wir die Robustheit verschiedener Markierungsstrategien bezüglich der Schätzung zellulärer Dynamiken von simulierten Daten, inklusive Datensets, welche unter experimentellen Limitierungen leiden. Unsere Ergebnisse können als Leitfaden benutzt werden, um in Zukunft die Dynamiken zellulärer Systeme noch genauer zu schätzen.

Durch die Kombination experimenteller Daten und mathematischer Modellierung helfen unsere Resultate nicht nur dabei, unser quantitatives Verständnis von $CD8^+$ T-Zell-Antworten in Infektionskrankheiten zu erweitern, sondern können auch benutzt werden, um die Effizienz und Effektivität T-Zell-basierter Impfstoffe zu verbessern.

Contents

1 General introduction **1**

1.1 The innate and the adaptive immune system . . . 1

1.1.1 $CD8^+$ T cells and their role in immune responses . . . 2

1.2 The dynamics of $CD8^+$ T cell responses during infection . . . 2

1.3 $CD8^+$ T cell subsets . . . 4

1.3.1 The $CD8^+$ T cell differentiation pathway . . . 5

1.3.2 Mathematical analysis of cellular dynamics . . . 6

1.4 T cell-based vaccines . . . 8

1.5 The outline of this thesis . . . 9

2 Materials and Methods **11**

2.1 General methods . . . 11

2.1.1 Parameter estimation . . . 11

2.1.2 Model selection . . . 11

2.2 Materials and Methods of chapter 3 . . . 12

2.2.1 Experimental data . . . 12

2.2.2 Model fitting . . . 12

2.3 Materials and Methods of chapter 4 . . . 14

2.3.1 Experimental data . . . 14

2.3.2 Quantifying the interaction between spleen and liver . . . 14

2.3.3 Parameter estimation . . . 15

2.4 Materials and Methods of chapter 5 . . . 18

2.4.1 Parameter estimation . . . 18

2.4.2 Evaluating the quality of parameter estimates . . . 18

3 $CD8^+$ T cell dynamics in MCMV infection **21**

3.1 Introduction . . . 21

3.1.1 CMV infection induces $CD8^+$ T cell inflation . . . 21

3.1.2 Generation and maintenance of memory inflation . . . 22

3.1.3 Chapter overview . . . 23

3.2 Mathematical models of memory inflation . . . 23

3.2.1 Blood-based reactivation models . . . 24

3.2.2 Influx-dependent reactivation models . . . 26

3.3 Modelling the inflationary response . . . 26

3.4 Modelling the non-inflationary response . . . 28
3.5 The influence of sporadic virus reactivation on CD8$^+$ T cell dynamics . . . 30
3.5.1 Modelling sporadic viral reactivation . . . 30
3.5.2 Changes in viral burst patterns can lead to memory inflation . . . 30
3.6 Discussion . . . 32

4 CD8$^+$ T cell dynamics induced by variable malaria vaccination approaches 35
4.1 Introduction . . . 35
4.1.1 The infection cycle of malaria . . . 35
4.1.2 Vaccination approaches against malaria . . . 36
4.1.3 Chapter overview . . . 37
4.2 Modelling the CD8$^+$ T cell response after normal dose PbRAS vaccination . . . 38
4.2.1 The global model of cellular turnover . . . 38
4.2.2 Unbiased model selection algorithm . . . 39
4.2.3 Model selection and model ranking . . . 42
4.2.4 Parameter identifiability . . . 43
4.3 The influence of the vaccination dose on cellular dynamics . . . 47
4.3.1 The effect of a high dose on cellular dynamics . . . 47
4.3.2 The effect of a subprotective dose on the cellular dynamics . . . 47
4.3.3 Correlating model predictions with protection . . . 48
4.4 Discussion . . . 49

5 On the influence of labelling strategies to infer cellular dynamics 53
5.1 Introduction . . . 53
5.1.1 The application and use of different labelling approaches . . . 53
5.1.2 Appropriateness of labelling strategies for parameter estimation . . . 54
5.1.3 Chapter overview . . . 54
5.2 Mathematical models and simulation approaches . . . 56
5.2.1 The mathematical models of cellular dynamics . . . 56
5.2.2 Simulating labelling experiments . . . 57
5.3 The influence of experimental factors on parameter identification . . . 58
5.3.1 The influence of the labelling strategy . . . 58
5.3.2 The influence of the sampling time . . . 59
5.4 The influence of experimental limitations on the identification of parameters . . . 61
5.4.1 The influence of incomplete transfer . . . 61
5.4.2 The influence of incomplete sampling . . . 62
5.4.3 The influence of missing compartments . . . 64
5.5 Discussion . . . 67

6 General discussion 69
6.1 Implications for the design of T cell-based vaccines . . . 69
6.1.1 On the induction of protective CD8$^+$ T cell responses . . . 69
6.1.2 On the long-term efficacy of T cell-based vaccines . . . 70

6.2 Future work 71

Appendix A **73**

A.1 Model fits to the non-inflationary data 73
A.2 Individual parameter estimates 74
A.3 Subset dynamics of MCMV-specific T cells 75
A.4 The influence of periodic reactivation events on the corresponding $CD8^+$ T cell dynamics 76

Appendix B **77**

B.1 Cell numbers during expansion phases 77
B.2 Profile likelihoods 78
B.3 Akaike weights and evidence ratio 80
B.4 Parameter sets used for dose fitting 81
B.5 Best fit to the high dose data 82
B.6 Best fit to the sub dose data 83
B.7 Comparing different dosage dynamics 84

Appendix C **85**

C.1 The influence of the sampling time given the homoeostatic model 85
C.2 Solving the master equations 85
C.3 The influence of the labelling strategy on parameter estimation 88
C.4 Deriving estimates from non-uniformly distributed labelling strategies 90
C.5 The influence of the sampling time on parameter estimation 92
C.6 Adjusting for transfer loss 94
C.7 Adjusting for sampling loss 95
C.8 Failed parameter estimation due to negative correlation 96

List of abbreviations

APC	antigen-presenting cell
AICc	corrected Akaike information criterion
BI	biphasic influx model
CC	correlation coefficient
CD	cluster of differentiation
CI	constant influx model
CMV	cytomegalovirus
ENR	expanded non-haematopoietic reservoir model
FACS	fluorescence-activated cell sorting
FCR	false coverage rate
MCIL	mean confidence interval length
MCMV	murine cytomegalovirus
MHC	major histocompatibility complex
NLMEM	non-linear mixed effects model
NR	non-haematopoietic reservoir model
ODE	ordinary differential equation
PbRAS	*Plasmodium berghei* radiation-attenuated sporozoites
pfu	plaque-forming unit
p.i.	post infection
RMSE	root-mean-square error
SEM	standard error of the mean
TCR	T cell receptor
T_N	naïve $CD8^+$ T cell
T_{CM}	central memory $CD8^+$ T cell
T_{EM}	effector memory $CD8^+$ T cell
T_{EMS}	splenic effector/effector memory $CD8^+$ T cell
T_{EML}	hepatic effector/effector memory $CD8^+$ T cell
T_E	effector $CD8^+$ T cell
T_{RM}	tissue-resident memory $CD8^+$ T cell
VD	virus-dynamics model
VI	variable influx model

CHAPTER 1

General introduction

Getting sick is easy.

With bacteria, viruses and parasites seemingly lurking everywhere, waiting for a chance to exploit the human organism as a nutrition source or reproduction machinery, our environment presents a constant challenge to our health. Moreover, a constantly evolving and spatially heterogeneous pathogen distribution means that new threats are literally just around the corner.

Luckily, we do not face these challenges unarmed. Due to the constant evolutionary pressure over millions of years, we acquired a remarkably effective and efficient immune system, which is not only able to control and eradicate a broad range of different diseases, but does also provide protection against previously encountered pathogens, thereby increasing our resilience over the course of our lifetime.

To mediate all these different functions, however, our immune system needs to be immensely complex.

1.1 The innate and the adaptive immune system

Our immune system is remarkably competent in handling a wide range of different diseases, which are caused by various pathogens, including viruses and bacteria. This efficacy is a result of the tightly woven interplay of many different cell types, which exert a multitude of immunological functions. The cells involved in immune responses are usually classified by belonging either to the innate or the adaptive immune system. The innate immune system serves as first line of defence and provides general protection against many different pathogens, e.g. by inducing inflammation, phagocytosing foreign pathogens or killing infected cells. The adaptive immune system, on the other hand, provides specifically tailored responses against infectious agents, such as the production of pathogen-specific antibodies, the cytotoxic clearance of infected cells, or the generation of immunological memory, which provides long-term protection against previously encountered diseases [Kuby 1997]. Despite their classification into separate genera, the innate and adaptive immune responses are intimately connected and cells of both domains are able to stimulate to each other.

1.1.1 CD8$^+$ T cells and their role in immune responses

CD8$^+$ T cells (also known as cytotoxic T lymphocytes), are a major part of the adaptive immune system. Their main function lies in the killing of infected cells and therefore the eradication of intracellular pathogens, such as viruses or parasites (a sketch of the role of CD8$^+$ T cells, their functions and their interactions with other immune cells is shown in Fig. 1.1 **A**).

Every CD8$^+$ T cell expresses a specific T cell receptor (TCR) on its surface, which can only recognise a single antigen. CD8$^+$ T cells are triggered by the presentation of pathogen-specific epitopes on the surface of infected cells, which leads to the release of cytotoxic granules, consisting mainly of perforin and granzymes, that induce cell death by apoptosis. These cytotoxic granules are specifically released towards the direction of the target cell to minimise damage to neighbouring tissue [Kuby 1997]. Because of the specificity of the TCR, every CD8$^+$ T cell targets only cells which present a matching antigen on their MHC I molecules.

Besides their cytotoxic effector function, CD8$^+$ T cells also release interferon-γ (IFN-γ) and tumour necrosis factor α (TNF-α), which leads to attraction, activation and stimulation of different immune cells, such macrophages and neutrophils [Kambayashi et al. 2003; Ratner et al. 1993]. Likewise, the efficacy of CD8$^+$ T cells can be increased by stimulatory cytokines secreted by other immune cells, e.g. by interleukin-2 (IL-2) production from CD4$^+$ T cells [Boyman et al. 2010].

Since the effector functions exerted by CD8$^+$ T cells present powerful immunological weaponry, their use needs to be tightly regulated. To acquire cytotoxicity, CD8$^+$ T cells first need to be activated by professional antigen-presenting cells (APCs), like dendritic cells, macrophages or B cells. Most of the initial priming takes place in the T cell zones of lymph nodes and is mediated by cross-presentation of antigens on the surface of matured dendritic cells [Haan et al. 2000; Kurts et al. 2010]. However, to successfully induce CD8$^+$ effector responses and to allow for the formation of immunological memory, dendritic cells have to be licensed by CD4$^+$ T cells first, otherwise the CD8$^+$ T cell responses are terminated early. [Allan et al. 2006; Smith et al. 2004]. All these factors guarantee that potent CD8$^+$ T cell functions are only provided when necessary.

1.2 The dynamics of CD8$^+$ T cell responses during infection

After egressing the thymus, where CD8$^+$ T cells are tested for functionality and self-tolerance, naïve, i.e. antigen-inexperienced, CD8$^+$ T cells continuously recirculate throughout the vasculature and secondary lymphoid organs until they encounter their cognate antigen [Kuby 1997]. Naïve T cells are thought to be long-lived and to divide in a homoeostatic fashion [Surh et al. 2008]. When naïve T cells are activated by the presentation of their cognate antigen on APCs, they start to proliferate extensively, which leads to a substantial increase in the number of pathogen-specific CD8$^+$ cells. It has been found that a single naïve cell can generate up to hundreds of thousands daughter cells [Gerlach et al. 2013].

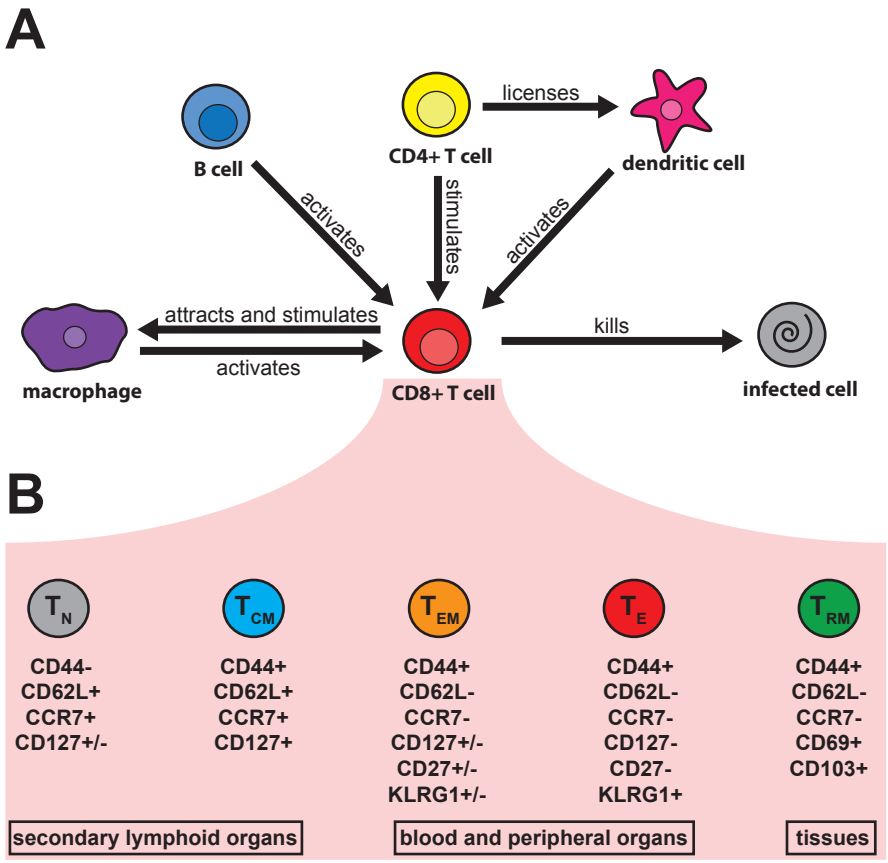

Figure 1.1: Immunological properties and phenotypes of CD8$^+$ T cells: (**A**) The sketch shows selected interactions of CD8$^+$ T cells and other immune cells: CD8$^+$ T cells are activated and stimulated by APCs (i.a. dendritic cells, macrophages and B cells) which leads to the exertion of cytotoxic effector functions and therefore to the death of infected cells. CD8$^+$ T cells also secrete cytokines that attract, activate and stimulate different immune cells. Likewise, they react to cytokine secretion from other cells types. (**B**) A selection of CD8$^+$ T cells subsets with their corresponding surface marker expression and corresponding anatomical location are shown. The following subsets are depicted (from left to right): naïve (T_N), central memory (T_{CM}), effector memory (T_{EM}), effector (T_E) and tissue-resident memory T cells (T_{RM}).

During clonal expansion, activated CD8$^+$ T cells give rise to memory- and effector-like CD8$^+$ T cell subsets by acquiring different phenotypes and functions. All these subsets mediate different actions during or after an infection (see section 1.3).

At the height of an induced CD8$^+$ T cell immune response, most cells are belonging to the class of effector cells, meaning they are able to release cytotoxic granules. Their high numbers allows them to closely patrol the body tissues and to efficiently eradicate intracellular pathogens. Once the infection is resolved, most effector CD8$^+$ T cells die by the activation of apoptosis-inducing pathways, leaving only long-lived memory-like cells behind [Krammer et al. 2007]. The establishment of immunological memory provides an important contribution towards future protection against previously encountered pathogens, as memory T cells are able to mount robust

and rapid immune responses upon reinfection [Bouneaud et al. 2005; Bachmann et al. 2005a; Woodland et al. 2009].

1.3 CD8$^+$ T cell subsets

Since their first classification in the 1970s, CD8$^+$ T cells have turned out to be quite heterogeneous. While early classifications only distinguished between naïve, effector and memory cells, the research in the last decades has lead to further fragmentation and the proposition of numerous subsets [Masopust et al. 2007; Jameson et al. 2009; Sallusto et al. 2004; Mahnke et al. 2013]. Regarding the subset classification, however, there is some inconsistency in the literature, as identical subset names have been used to define cells based on different properties such as surface marker expression, effector functions, short- and long-term maintenance or anatomical localisation [Masopust et al. 2013; Rocha et al. 2006; Ahmed et al. 2009]. To account for this varying definitions, the following descriptions of the CD8$^+$ T cell subsets[1] summarise the phenotypes based on various characterisations found in literature. A sketch of the subsets, their surface marker expression and their preferred anatomical localisation can be found in Fig. 1.1 **B**.

Naïve T cells ($\mathbf{T_N}$) Naïve T cells are cells that have not encountered their cognate antigen yet. They recirculate between the secondary lymphoid organs, such as lymph nodes or spleen, and proliferate homoeostatically [Surh et al. 2008]. The surface marker expression of T_N is usually defined as CD44-/CD62L+/CCR7+, and they show a variable expression of CD127 [Farber et al. 2014; Park et al. 2004]. In mice, it was recently estimated that for every antigen around $80 - 1200$ specific naïve cells exist [Obar et al. 2008].

Central memory T cells ($\mathbf{T_{CM}}$) Central memory cells are long-lived T cells that arise after an infection and provide lasting protection against the encountered pathogen [Farber et al. 2014; Sallusto et al. 2004]. In case of reinfection, central memory cells show a superior expansion capacity compared to naïve or effector memory cells, as they are able mount rapid and robust T cell responses [Bachmann et al. 2005b; Huster et al. 2004; Wherry et al. 2003; Bouneaud et al. 2005; Roberts et al. 2005]. T_{CM} reside mostly in secondary lymphoid organs and are characterised by the surface marker expression of CD44+/CD62L+/CCR7+/CD127+ [Arsenio et al. 2014; Bachmann et al. 2005a; Kaech et al. 2007].

Effector memory T cells ($\mathbf{T_{EM}}$) As the name suggests, effector memory T cells combine traits associated with both effector and memory compartments. They are mostly found in blood and peripheral tissues where they provide immediate protection against intruding pathogens by mounting rapid effector functions such as the secretion of cytokines [Sallusto et al. 2004; Stemberger et al. 2007]. High levels of T_{EM} were shown to protect against diseases like HIV or malaria [Hansen et al. 2011; Schmidt et al. 2008]. Therefore, their efficient induction is one of the major goals of T cell-based vaccine research (see section 1.4). Effector memory cells are

[1] Only the CD8$^+$ T cell subsets relevant in this thesis are characterised.

mostly classified as being CD44+/CD62L-/CCR7-, and, depending on the publication, either by the lack or the expression of CD127, CD27 and KLRG1 [Arsenio et al. 2014; Bachmann et al. 2005a; Bassett et al. 2012; Bolinger et al. 2013; Buchholz et al. 2013; Farber et al. 2014].

Effector T cells ($\mathbf{T_E}$) Effector T cells, also known as short-lived effector cells, are cytotoxic T cells that are responsible for the detection and eradication of infected cells and are mostly present during acute infections. Effector cells are defined by the surface marker expression of CD44+/CD62L-/CCR7-/CD127-/CD27-/KLRG1+ and the production of IFN-γ upon antigen stimulation [Arsenio et al. 2014; Bassett et al. 2012; Buchholz et al. 2013; Stemberger et al. 2007].

Tissue resident memory T cells ($\mathbf{T_{RM}}$) Tissue-resident memory cells are a relatively recent addition to the CD8$^+$ T cell subsets and therefore still under a lot of investigation [Gebhardt et al. 2009]. While it is known that they are patrolling organ-specific tissues with seemingly no exposure to the blood [Fernandez-Ruiz et al. 2016], little is known about their genesis and maintenance [Farber et al. 2014]. It is assumed that T_{RM} provide immediate *in situ* protection against intruding pathogens, as a high number of T_{RM} correlates with protection against tissue-specific disease [Fernandez-Ruiz et al. 2016; Thom et al. 2016]. T_{RM} are characterised by the expression of CD44+/CD62L-/CCR7-/CD69+/CD103+.

1.3.1 The CD8$^+$ T cell differentiation pathway

Ever since the identification of different subsets, there has been an ongoing debate about the underlying CD8$^+$ T cell differentiation pathway. While early work focused mostly on the question at which time during infection the memory compartment arises, current approaches face the challenge to identify the differentiation pathways of the various CD8$^+$ T cell subsets, which might include spatial and temporal aspects as well.

Many hypotheses have been proposed over the last decades concerning the developmental path of CD8$^+$ T cells. Originally, the only distinction in the activated T cell pool was with respect to memory and effector cells. As memory cells were simply defined as cells that are present after an infection, the general assumption about the underlying CD8$^+$ T cell differentiation pathway was that naïve cells turn into effector cells upon antigen encounter and a fraction of these would differentiate into long-lived memory cells during or after the resolution of the acute infection [Althaus et al. 2007; Antia et al. 2003; De Boer et al. 2001] (see Figure 1.2 **A**). The identification of heterogeneity in the memory population [Sallusto et al. 2004] resulted in the revision of this pathway and questions about the generation and connection of central memory and effector memory cells needed to be addressed [Bouneaud et al. 2005; Gerlach et al. 2010; Kaech et al. 2007; Wherry et al. 2003; Crauste et al. 2017]. This eventually led to hypotheses of memory stages preceding the effector stages [Kohler 2007; Buchholz et al. 2013], challenging the previously established naïve-effector-memory CD8$^+$ T cell differentiation pathways. As of today, the ongoing identification of new CD8$^+$ T cell subsets over the recent years [Jameson et al. 2009; Farber et al. 2014] made it clear that our current understanding of CD8$^+$ T cell differentiation

dynamics is far from being complete and that more data and modelling approaches are needed to uncover the whole picture.

Challenges regarding the pathway identification The challenges with respect to the identification of the true differentiation pathway for $CD8^+$ T cells, as well as the reporting of seemingly contradictory differentiation patterns, can be partially attributed to some fundamental issues concerning the complexity of the underlying systems. One of these problems is the inconsistency in terms of $CD8^+$ T cell subset classification (see section 1.3). In previous studies, for example, the distinction of T_{CM} and T_{EM} in mice was based on the expression of surface markers CD127/CD62L [Bachmann et al. 2005a; Bassett et al. 2012], CCR7/CD62L [Wherry et al. 2003; Bouneaud et al. 2005] or CD27/CD62L [Buchholz et al. 2013]. These differences in subset classifications do not only leave room for ambiguity regarding the lineage identification, but can potentially result in contradicting predictions about the differentiation dynamics (see Fig. 1.2 **B1**).

Another problem is the open question, whether $CD8^+$ T cell subsets represent fates or states. A fate would imply that a cell moves along a predefined differentiation pathway, while a state would allow some plasticity with respect to differentiation dynamics, as e.g. cells react to the presence of immunological stimuli (see Fig. 1.2 **B2**). While there is some indication about programmed proliferation, immunological factors are also likely to play a role in determining the shaping of the $CD8^+$ T cell response [Raue et al. 2013; Redeker et al. 2014].

Another obstacle arises from the anatomical distribution of different $CD8^+$ T cell subsets: As many of the expressed surface markers of the T cell subsets correspond to a specific anatomical location [Masopust et al. 2013; Jung et al. 2010; Steinert et al. 2015], a lack of organ-specific $CD8^+$ T cell data might potentially bias the identification of the differentiation pathway (see Fig. 1.2 **B3**). For example, while measurements of $CD8^+$ T cell responses in the blood will probably represent the general dynamics of T_{EM} or T_E , they are less likely to capture the turnover of T_{CM} and T_{RM} pools, which preferentially localise in secondary lymphoid organs and organ-tissues, respectively.

In summary, the identification of the underlying $CD8^+$ T cell differentiation pathway is facing several structural challenges that need to be properly addressed in order to reach conclusive results.

1.3.2 Mathematical analysis of cellular dynamics

The application of mathematical methods has proven to be very useful to determine underlying cellular differentiation pathways as it allows the comparison of different hypotheses by determining the likeliness of each hypothesis (e.g. by using the Akaike or Bayesian information criterion [Burnham et al. 2003]). These methods do not only take the ability of each mathematical model to reproduce the experimental data into account, but also penalise for complexity with respect to the number of model parameters. Hence, mathematical analysis makes it possible to distinguish between likely and unlikely cellular interactions while simultaneously help to select the simplest model which is able to explain the observed cellular dynamics. This kind of model selection

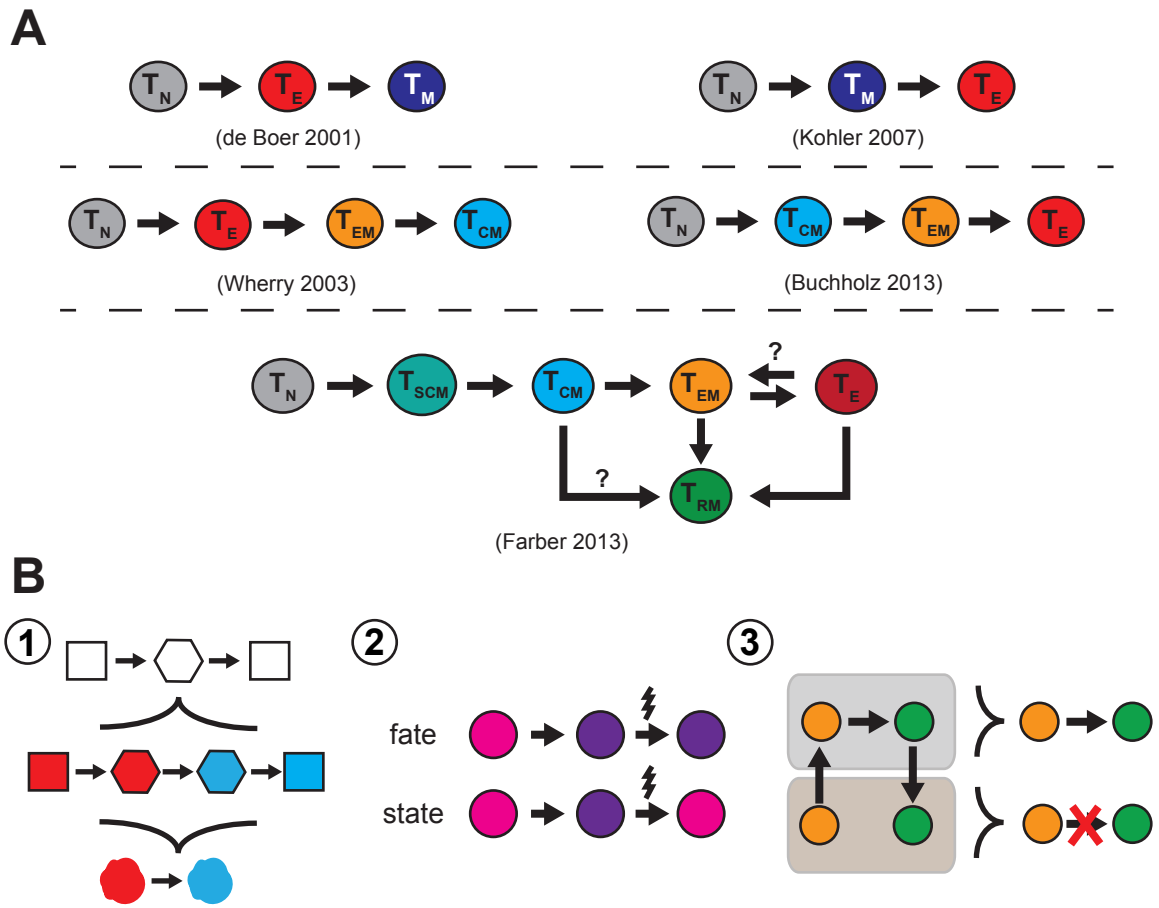

Figure 1.2: Previously proposed CD8$^+$ T cell differentiation pathways and challenges regarding the identification of cellular lineages: (**A**) A selection of previously proposed pathways for CD8$^+$ T cell differentiation. Originally, it was assumed that naïve CD8$^+$ T cells (T_N) give rise to effector cells (T_E), which can turn into memory cells (T_M) during or after an acute infection [De Boer et al. 2001]. This general trend was also found, as the subsets became more diverse and central and effector memory cells (T_{CM} and T_{EM}) were included in the models [Wherry et al. 2003]. However, this pathway has been recently challenged by models that assume the memory stage precedes the effector stage [Kohler 2007; Buchholz et al. 2013]. Additionally, the identification of new cellular subsets over the last years, such as memory stem cells (T_{SCM}) or tissue-resident memory cells (T_{RM}) asks for a further revision of the cellular differentiation dynamics [Farber et al. 2014]. (**B**) Potential pitfalls regarding the identification of cellular lineages: 1) Focussing on different properties can result in contradicting results about the underlying dynamics: While the shape returns to its original form in our example, the colour stays permanently changed. 2) The problem of cell fate vs. cell state: The cellular differentiation dynamics might depend on a predefined pathway (fate) or external stimuli (state). 3) Neglecting spatial compartments can bias the predicted dynamics: Conversion from orange to green circles is only found in the upper compartment in our example and would not be predicted if only the lower compartment was to be analysed.

has also been extensively used for determining the dynamics of CD8$^+$ T cells. For example, previous studies have analysed possible differentiation pathways by testing and comparing up

to several hundred different models [Buchholz et al. 2013; Crauste et al. 2017]. In the future, increasingly available computational resources will make it possible to test even more models, therefore allowing to analyse the underlying pathways of very complex cellular systems.

Determining the cellular lineage differentiation pathway is only one side of the coin, quantifying the cellular dynamics is the other. Here, mathematical modelling has proven to be an essential tool as well. For example, the mathematical studies analysing cellular lineages depicted in Fig. 1.2 **A** [De Boer et al. 2001; Kohler 2007; Buchholz et al. 2013] also led to the estimation of differentiation, proliferation and death rates, and therefore gave insight into cellular half-lives, doubling times and long-term maintenance. Additionally, mathematical modelling was used to quantify cellular migration [Ganusov et al. 2014], the influence of cytokines on cellular turnover [Arias et al. 2014] or the division-dependent cellular expansion dynamics [Yates et al. 2007], thereby broadening our understanding of cellular dynamics in many different ways. Most of these studies used ordinary differential equations to model the cellular kinetics, which allows to determine the mean dynamics observed in cellular systems. Other approaches include partial differential equations to incorporate spatial aspects [Stromberg et al. 2012; Bouchnita et al. 2017] or stochastic modelling to determine the contribution of individual variations [Luciani et al. 2001; Currie et al. 2012]. Here, the broadness of the mathematical toolbox allows to quantify cellular processes from almost all kinds of experimental data.

Retrieving information from cellular data However, the amount of information obtainable by mathematical modelling is limited to the amount of information which is present within the biological data. For example, measuring only the numbers of specific type of cell allows for the quantification of the corresponding mean population dynamics, but contains no information about individual cell behaviour. Therefore, biological experiments frequently use cellular markers to make individual processes visible. Common labelling methods include the use of genetic barcodes for lineage determination [Schumacher et al. 2010; Gerlach et al. 2013], the introduction of congenic markers for subset classification [Shen et al. 1985; Kearney et al. 1994; Buchholz et al. 2013] or the application of proliferation markers, such as bromodeoxyuridine (BrDU) or deuterated glucose [Mohri et al. 1998; Borghans et al. 2017]. Mathematical analysis based on biological data using cellular markers allows for a more accurate determination of rates describing cellular turnover [Kaiser et al. 2013; Buchholz et al. 2013; Borghans et al. 2017], thus improving the reliability of model predictions. As not all labelling strategies give the same amount of information the set-up of the biological experiments has to be carefully planned. Here, mathematical modelling can be used to determine the reliability of experimental designs and therefore also serves as a tool to measure the amount and quality of information which is obtainable from cellular experiments (see chapter 5).

1.4 T cell-based vaccines

Successful vaccination programs against pathogens like smallpox, polio or measles have helped to control or even eradicate these infectious diseases, thus effectively preventing millions of deaths per year. The development of vaccines is still a major part of medical research and its relevance

today is highlighted by a recent WHO announcement that declared our decade as the "decade of vaccines".

The basic idea behind every vaccination approach is to generate immune responses that will result in effective and long-lasting protection against a specific pathogen. To this end, vaccines stimulate the adaptive immune system as long-term protection requires the generation of immunological memory. Pre-existing memory cells allow for a robust and quick expansion of pathogen-specific immune cells in case the disease is actually encountered. Many vaccines induce a pronounced B cell response as various infections can be prevented by the release of pathogen-specific antibodies [Clem 2011].

However, for some diseases, such as malaria or HIV, the production of antibodies alone is not sufficient to efficiently clear an infection. In these cases, a promising approach lies in the generation of vaccines that induce pronounced $CD8^+$ T cell responses, which could specifically target intracellularly replicating pathogens.

As of today, there has been a lot of research on the topic of T cell-based (also known as T cell-inducing) vaccines [Gilbert 2012; Koup et al. 2011; Korber et al. 2009]. Previous studies have shown that the induction of high numbers of effector memory $CD8^+$ T cells resulted in protection of rhesus macaques against SIV infection [Liu et al. 2009; Hansen et al. 2011; Hansen et al. 2013b]. In these studies, the use of recombinant adeno- and cytomegalovirus vectors expressing SIV proteins elicited pronounced and protective immune responses. The importance of readily available $CD8^+$ T cell responses has also been shown for malaria vaccinations, where high levels of $CD8^+$ T cells correlated with protection against the disease [Schmidt et al. 2008; Schmidt et al. 2010; Seder et al. 2013; Fernandez-Ruiz et al. 2016].

Hence, an aim for effective T cell-based vaccines in humans lies in the induction of high levels of pathogen-specific $CD8^+$ T cell responses, that can exhibit immediate effector functions and are preferable located at the main site of infection. However, tailoring those vaccination approaches is currently hampered by a lack of knowledge regarding the underlying cellular dynamics. Thus, identifying and quantifying the cellular responses represent important steps for the creation of effective T cell-based vaccines [Korber et al. 2009].

1.5 The outline of this thesis

The creation of efficient and effective T cell-based vaccines is a promising approach to tackle many of today's most dangerous infectious diseases. However, to achieve this goal, more research is necessary, especially in terms of the identification of the underlying $CD8^+$ T cell differentiation pathway as well as the quantification of $CD8^+$ T cell turnover *in vivo*. In this thesis, we combine experimental data and mathematical modelling to address some of these topics, which allows us to gain insight into cellular processes that are difficult to be determined from biological experiments alone.

A major goal with respect to protective T cell-based vaccination approaches lies in the generation of sufficient readily available disease-specific cytotoxic cells. As it is known that cytomegalovirus elicits high levels of effector-like $CD8^+$ T cells over long periods of times, understanding the factors influencing this particular immune response is of special interest in medical

research. In chapter 3 we therefore analyse this atypical $CD8^+$ T cell response mathematically. We test different biological hypotheses regarding the generation of this 'memory-inflation' response and quantify the underlying cellular processes. Additionally, we analyse the influence of viral reactivation patterns on the shaping of the inflationary $CD8^+$ T cell response. Since cytomegalovirus can serve as a recombinant vector expressing viral epitopes, our findings are relevant for determining the efficacy of vector-based vaccination approaches.

In many cases, vaccination-mediated protection requires the application of one or more booster injections. Here, factors such as dosage, frequency and timing of injections can determine the protective levels reached by vaccination. To this end, we assess the influence of different malaria vaccination strategies on the generation of organ-specific $CD8^+$ T cell responses in chapter 4. To determine the underlying cellular differentiation pathway we established an unbiased model selection algorithm, which allows to efficiently search for the best model describing the experimental data given a huge variety of possible models. By successfully applying this algorithm to the malaria data, we can identify and quantify the $CD8^+$ T cell responses in spleen and liver, which are elicited by different whole sporozoite vaccination approaches. Furthermore, we model the impact of different vaccination doses on the build-up and maintenance of protection-mediating liver-resident memory cells. Our results can be used to improve existing vaccination strategies in terms of dosage and frequency of immunisations.

Understanding the impact of T cell-based vaccination regimes requires knowledge about the underlying cellular dynamics. While mathematical modelling allows the determination of differentiation pathways and the quantification of cellular turnover, the reliability of these analyses depends strongly on the provided biological data. Therefore, we study in chapter 5 how the labelling of cells affects the estimation of cellular turnover, such as proliferation and differentiation dynamics. Here, we establish suitable labelling strategies and determine the robustness of mathematical estimates derived from data suffering from experimental limitations, such as incomplete transfer or sampling of cells. Our theoretical findings can help to improve the quantification of cellular dynamics in the future.

Taken together, the analyses performed in this thesis help to broaden our understanding of $CD8^+$ T cell dynamics in infectious diseases and have potential implications for the design of T cell-based vaccines.

CHAPTER 2

Materials and Methods

2.1 General methods

2.1.1 Parameter estimation

Unless specified otherwise, all simulations were carried out using the `R`-language of statistical computing [R Core Team 2016]. Data fitting and parameter estimation was done based on the built-in `optim`-function using the default 'Nelder-Mead' method [Avriel 1976]. Each fitting routine was run with at least five different randomly selected initial conditions. Additionally, to guarantee convergence, the fitted output of each run was again used as an initial condition for a new run. This re-fitting was done until no further change in the fitness measurement[1] was observed.

2.1.2 Model selection

Model comparison was based on the corrected Akaike information criterion (AICc, see [Burnham et al. 2003]). The AICc is defined as

$$\mathrm{AICc} := -2LL + 2k + \frac{2k(k+1)}{n-k-1} \ ,$$

where $-2LL$ stands for the negative log-likelihood of the model describing the data, k denotes the number of model parameters and n is the number of data points.

Given a number of n different models, their goodness is assessed by calculating the differences in their AICcs:

$$\Delta\mathrm{AICc}_i := \mathrm{AICc}_i - \min_{i=1,..,n}\{\mathrm{AICc}_i\} \ .$$

[1] We used both maximum likelihood approaches and χ^2-minimization approaches, see the following sections.

These values can be used to measure the support each model has given all other models by calculating Akaike weights, which are defined as

$$w_i := \frac{e^{-\frac{\Delta \mathrm{AICc}_i}{2}}}{\sum_{j=1}^{n} e^{-\frac{\Delta \mathrm{AICc}_j}{2}}}$$

Alternatively, evidence ratios measuring the support of different model types are calculate by

$$\frac{\sum_{i \in A} w_i}{\sum_{j \in B} w_j} ,$$

where A and B are sets containing models with specific properties. Here, the normalised probability can be calculated by taking B as the set that contains all tested models.

According to [Burnham et al. 2003], the performance of different models can be quickly assessed by looking at the ΔAICc: Models with a $\Delta\mathrm{AICc} < 2$ are performing almost equally well as the best model, while models having a $\Delta\mathrm{AICc} > 10$ are basically not supported.

2.2 Materials and Methods of chapter 3

2.2.1 Experimental data

Naïve C57BL/6 mice were infected with intravenously injected 5×10^6 pfu MCMV-Δ157 and blood was sampled at several time points between day 7 and 70 p.i. Cells were analysed by flow cytometry with respect to their surface marker expression, including markers for CD8, CD44, CD127, CD62L and KLRG1 and their specificity for MCMV epitopes M38 (inflationary) and M45 (non-inflationary). Cell numbers per millilitre blood were calculated based on extrapolation with a given number of added fluorescently-labelled PE+ beads. The data is shown in Fig. 2.1. Measurements having a living leukocyte percentage lower than 90% or a measured PE+ bead number higher than 10^4 were excluded from the mathematical analysis, as these values indicated corrupt data. All experiments were performed by the lab of Prof. Dr. A. Oxenius, Institute of Microbiology, ETH Zürich.

2.2.2 Model fitting

All models were fitted to the data based on a nonlinear mixed effects model (NLMEM) approach [Pinheiro et al. 1995]. This method assumes that data from different individuals can be described based on fixed (i.e. population) and random (i.e. individual) effects. Having data from n individuals, the aim of an NLMEM approach is to find parameter sets $\vec{\theta}_i,\ i = 1, ..., n$, which maximise the likelihood to observe the experimental data. Before fitting, the underlying distribution of the parameters needs to specified. In our case, we assumed that the parameters are log-normally distributed and can be expressed in the form of

$$\log(\vec{\theta}_i) = \log(\vec{\theta}_{\mathrm{pop}}) + \vec{\psi}_i \ .$$

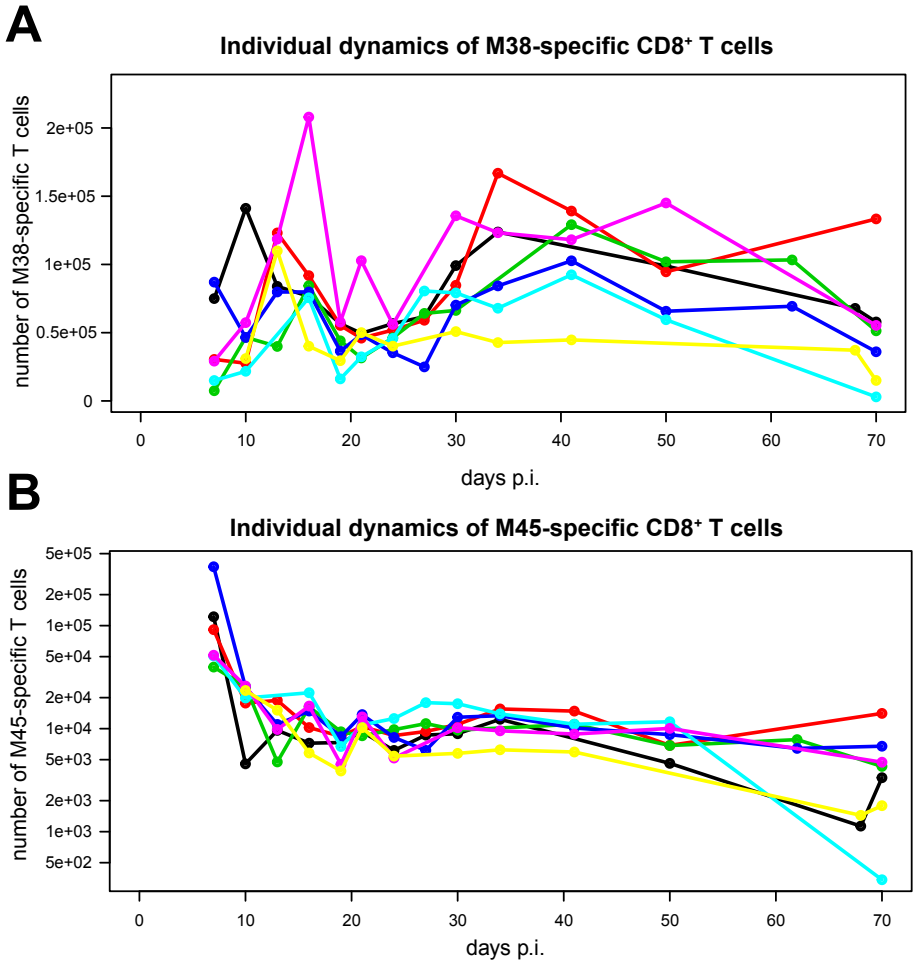

Figure 2.1: Inflationary and non-inflationary CD8$^+$ T cell dynamics during MCMV infection: (**A**) Longitudinal data of inflationary M38-specific CD8$^+$ T cells in the blood for seven different mice. Measurements were obtained over the course of 70 days after infection with MCMV. (**B**) The corresponding data of non-inflationary M45-specific CD8$^+$ T cells. All experiments were performed by the lab of Prof. Dr. A. Oxenius, Institute of Microbiology, ETH Zürich. For details see Materials and Methods 2.2.1.

with $\vec{\theta}_{\text{pop}}$ being the vector of the fixed population effects and $\vec{\psi}_i \sim \mathcal{N}(0, \sigma^2\Omega)$ being normally distributed individual effects. The variance-covariance matrix, Ω, was chosen to be the identity matrix ($\Omega = \mathbb{1}$) in our simulations. In order to find the population parameters $(\vec{\theta}_{\text{pop}}, \sigma^2)$, it is necessary to maximise the following likelihood

$$\mathbb{P}(Y|\vec{\theta}_{\text{pop}}; \sigma^2) = \int \mathbb{P}(Y|\vec{\theta}_{\text{pop}}; \sigma^2; \vec{\psi})\mathbb{P}(\vec{\psi})d\vec{\psi} \ ,$$

where Y is the experimental data. Because of the non-linearity in the underlying models (see section 3.2), this value needs to be approximated numerically. Here, our results are based on calculations performed by the software Monolix (Version 2016R1, see http://lixoft.com/products/monolix/)

using the stochastic approximation expectation-maximisation algorithm [Delyon et al. 1999] and assuming a proportional error in the measurements. Once the population parameters are estimated, the software determines the individual parameter sets $\vec{\theta}_i,\ i = 1, ..., n$ by maximising the corresponding likelihood:

$$\mathbb{P}(\vec{\theta}_i | Y_i; \vec{\theta}_{\text{pop}}; \sigma^2) \ ,$$

where Y_i denotes the data belonging to the ith individual.

2.3 Materials and Methods of chapter 4

2.3.1 Experimental data

In order to test the influence of dosage and frequency of vaccination regimens, groups of C57BL/6 mice received one, two or three intravenous injections with different doses of *Plasmodium berghei* radiation-attenuated sporozoites (PbRAS) or salivary gland debris (mock control). The irradiation of sporozoites is assumed to prevent the development of liver-stage schizonts, therefore inducing a pre-erythrocytic termination of *Plasmodium* infection [Hoffman et al. 2002]. Three different doses were applied: subprotective (1×10^3 PbRAS), normal (1×10^4 PbRAS) and high (1×10^5 PbRAS). Organs were harvested and the number of $CD8^+$ T cells in spleen and liver was analysed at the time points depicted in Fig. 2.2. Additionally, groups of mice were challenged with 1×10^4 infectious *P. berghei* sporozoites to assess protective efficacy.

The gathered samples were analysed by FACS and cells were gated for the surface markers CD8, CD44, CD62L and CD69. $CD8^+$ T cell populations were classified by their expression of markers: central memory ($CD44^+/CD62L^+$), effector memory/effector ($CD44^+/CD62L^-/CD69^-$) and tissue-resident memory cells ($CD44^+/CD62L^-/CD69^+$).

To assess the pathogen-specific response, the cell numbers were normalised by subtracting the corresponding subset levels measured in naïve mice. The adjusted data is shown in Fig. 2.3.

All experiments were performed by the lab of Prof. Dr. A.-K. Mueller, Centre for Infectious Diseases, Parasitology Unit, University Hospital Heidelberg.

C_{N2}	C_{H2}	C_{S2}	C_{S3}	C_{N2,d_118}	C_{H2,d_118}
3/6	7/7	0/3	0/3	6/8	7/8

Table 2.1: Number of mice that showed sterile protection upon challenge (C) with different dosages of infectious sporozoites. For the corresponding groups see Fig. 2.3. All experiments were performed by the lab of Prof. Dr. A.-K. Mueller, Centre for Infectious Diseases, Parasitology Unit, University Hospital Heidelberg.

2.3.2 Quantifying the interaction between spleen and liver

To quantify the effect of boost and dosing, Dr. F. Graw developed a boost-effect model, which incorporates a negative feedback from spleen to liver (R. Frank, M. Gabel, K. Heiss, A.-K.

Mueller, F. Graw, manuscript in preparation). The model is given as

$$\mathrm{T}_{\mathrm{EMS}}(b,d) = \left[1 + \lambda_d\Big(1 - \frac{\mathrm{T}^2_{\mathrm{EML}}}{\mathrm{T}^2_{\mathrm{EML}} + \Omega^2_{\mathrm{EML}}}\Big)^{b-1}\right] \mathrm{T}_{\mathrm{EMS}}(b-1,d) \ ,$$

where b is the number of boosts, d the dose (being either subprotective, normal or high), $T_{\cdot}$ the number of effector memory T cells in either spleen or liver, λ_d the dose-dependent expansion factor and Ω_{EML} the level of effector memory cells in the liver at which the saturation reaches half of its maximal value. The parameters $\lambda_{\cdot}$ and Ω_{EML} were estimated based on the data given in section 2.3.1. The 95% confidence interval for Ω_{EML} was estimated to range from $195,000$ to $269,000$ cells, with the best estimate being $236,000$ cells, which we used in our model parametrisation in chapter 4.

2.3.3 Parameter estimation

Models were fitted based on maximum likelihood estimation approach and for each model the negative log-likelihood

$$-2LL = \min_{\theta} \sum_{i=1}^{n} \left(\frac{f(\theta, t_i) - \mu_i}{\sigma_i}\right)^2$$

was minimised. Here, $t_i,\ i = 1, ..., n$ denotes the measurement times of the experimental data, μ_i the corresponding data mean, σ_i the standard error of the mean and $f(\cdot)$ the simulated prediction given a set of model parameters θ.

The 95% confidence intervals for the parameters were calculate based on a profile likelihood approach [Raue et al. 2009]. Given a parameter $\psi \in \theta$, the corresponding confidence interval is defined as

$$\mathrm{CI}_{\psi} = \Big\{\psi | -2LL_{\psi} + 2LL < \chi^2_{\{0.95,1\}}\Big\} \ ,$$

where $-2LL_{\psi}$ is the negative log-likelihood estimate obtained by fitting the model with a fixed value of ψ to the data and $\chi^2_{\{0.95,1\}} (\approx 3.84)$ is the 95%-quantile of the χ^2-distribution with one degree of freedom.

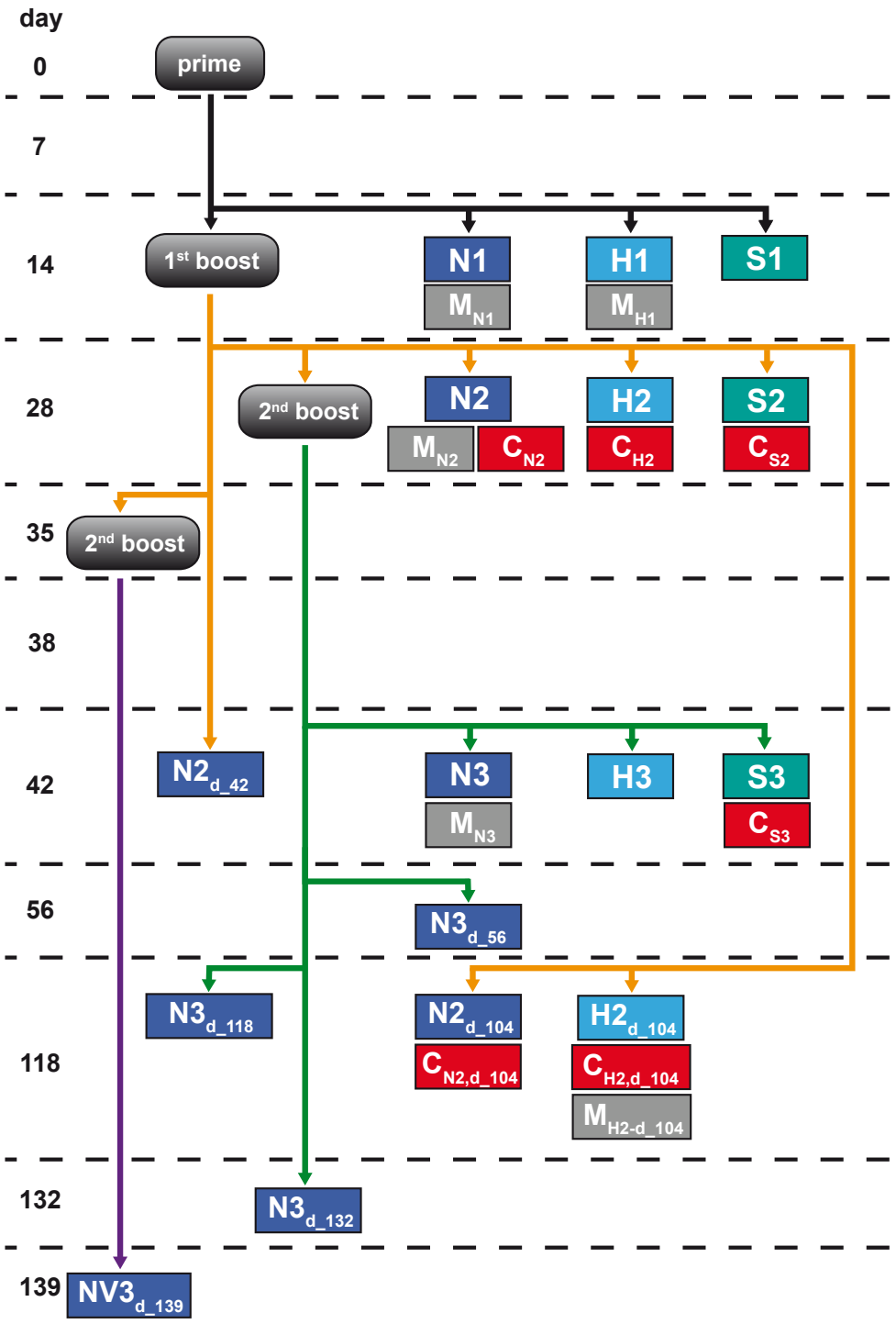

Figure 2.2: Sketch of the experimental protocol: Boxes indicate the respective time points of vaccination shots (black) and harvests (coloured). Mice were either vaccinated with mosquito salivary gland debris (mock control (M) - grey boxes) or radiation attenuated sporozoites of different dosages (normal dose (N) - blue, high dose (H) - light blue, subprotective dose (S) - turquoise). Additionally, challenge experiments were performed by administering live *P.berghei* parasites (red boxes). Timing between shots was either 14 days (black, orange and green arrows), or first 14, then 21 days (purple arrows, group $NV3_{d\_139}$). Experiments were performed by the lab of Prof. Dr. A.-K. Mueller, Centre for Infectious Diseases, Parasitology Unit, University Hospital Heidelberg.

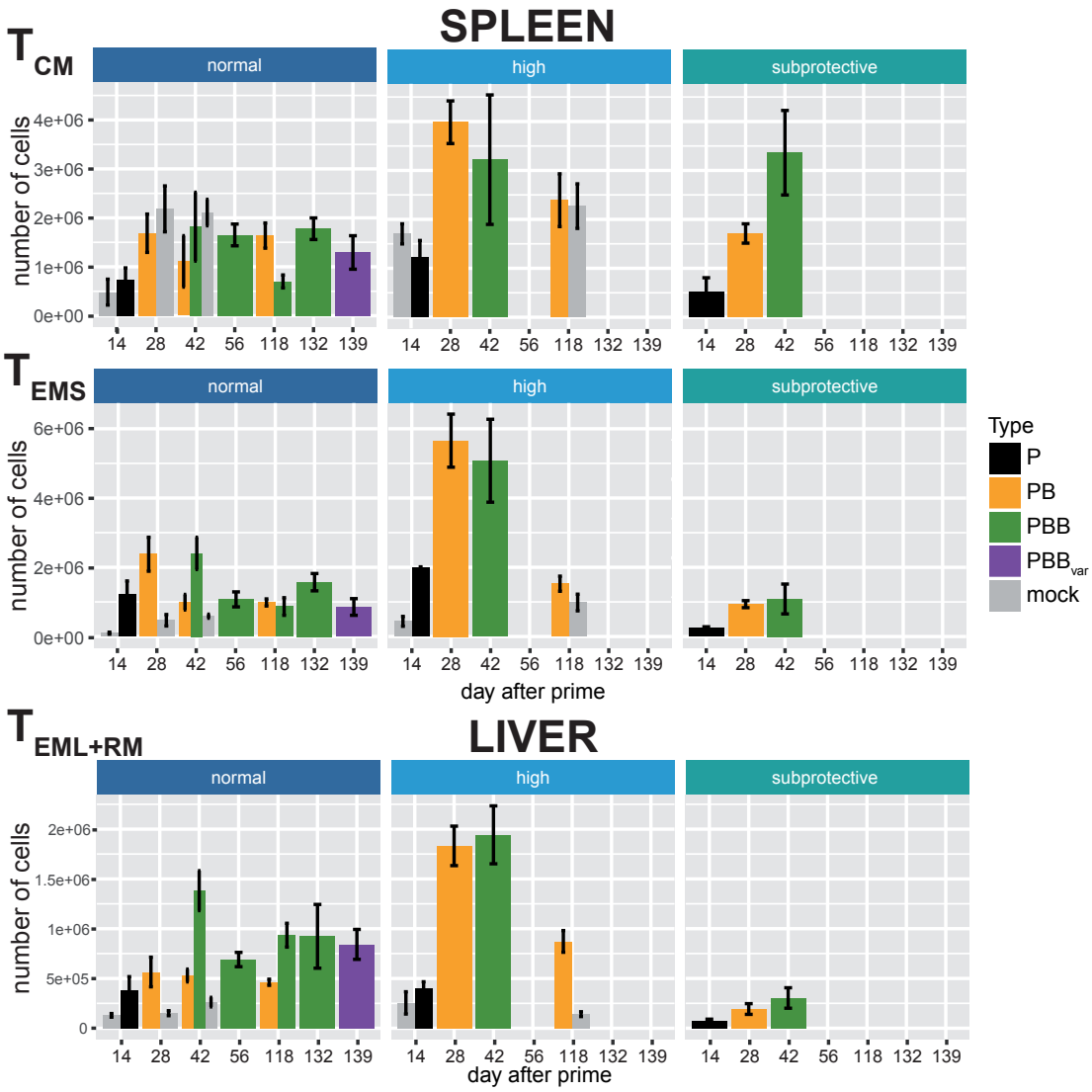

Figure 2.3: The number of CD8$^+$ T cell subset in different organs: Barplots show the mean ± SEM of the different groups depicted in Fig. 2.2 for spleen and liver according to prime (P, black), prime-boost (PB, orange), prime-boost-boost (PBB, green), varied prime-boost-boost (PBB_{var}, purple) and mock vaccination (grey). Panels show the the number of central memory cells in spleen (T_{CM}), the number of effector memory/effector cells (T_{EMS}) and the combined number of effector/effector memory and tissue-resident memory cells (T_{EML+RM}) in the liver. All experiments were performed by the lab of Prof. Dr. A.-K. Mueller, Centre for Infectious Diseases, Parasitology Unit, University Hospital Heidelberg.

2.4 Materials and Methods of chapter 5

2.4.1 Parameter estimation

Estimates for the model parameters were obtained by fitting the predicted summary statistics for each cell population to the stochastically generated data. The predicted summary statistics, i.e. the expected mean ($\mathbb{E}$), the coefficients of variation (CV) and - depending on the model - the correlation coefficients (CC), are obtained by solving the corresponding master equations of the systems specified in section 5.2.1 (see Appendix C.2 for a detailed description of the calculations). The fitting procedure is based on Pearson's χ^2-minimisation [Berkson 1980], which was chosen to avoid assumptions about the underlying distributions of the summary statistics. Let $\mathcal{O} = (\mathcal{O}_{\mathbb{E}}, \mathcal{O}_{\mathrm{CV}}, \mathcal{O}_{\mathrm{CC}}))$ be the observed and $\mathcal{P}(\vec{\theta}) = (\mathcal{P}_{\mathbb{E}}(\vec{\theta}), \mathcal{P}_{\mathrm{CV}}(\theta), \mathcal{P}_{\mathrm{CC}}(\theta)))$ the predicted summary statistics, where θ denotes a set of model parameters. The best parameter set is then calculated by minimising

$$\min_{\theta} \left| \frac{(\mathcal{P}(\theta) - \mathcal{O})^2}{\mathcal{P}(\theta)} \right| .$$

Confidence intervals for parameter estimates were obtained by bootstrapping the data using the built-in `R`-package `boot`, which are calculated based on Efron's non-parametric and accelerated bootstrap (BCa) method [Carpenter et al. 2000] with 999 repeats and a significance level of $\alpha = 0.05$ (see also Fig. 5.3.1 **B**).

2.4.2 Evaluating the quality of parameter estimates

The appropriateness of different labelling strategies to retrieve the underlying cellular dynamics was determined by calculating and comparing the following statistical quantities derived from Burton et al. [Burton et al. 2006].

Bias: The bias indicates how much the mean parameter estimate deviates from the true value. Let $\hat{\theta}_i,\ i = 1, ..., m$ be estimates of the true parameter θ, with $\hat{\theta} := \frac{1}{m}\sum_{i=1}^{m} \hat{\theta}_i$ defining the empirical mean. The bias is then calculated by

$$\mathrm{Bias} := \hat{\theta} - \theta \ . \tag{2.1}$$

Percentage bias: The percentage bias shows on a relative scale how much the mean parameter estimate differs from the true value and therefore allows the comparison of values based on different scales.
The percentage bias is defined by

$$\mathrm{pBias} := \frac{\hat{\theta} - \theta}{\theta} \ . \tag{2.2}$$

Root-mean-square error (RMSE): The root-mean-square error serves as a measure to describe the deviation of the estimators and is defined as

$$\text{RMSE} := \sqrt{(\hat{\theta} - \theta)^2 + \mathbf{Var}(\hat{\theta})} \ , \tag{2.3}$$

with $\mathbf{Var}(\hat{\theta}) := \frac{1}{m-1} \sum_{i=1}^{m} (\theta_i - \hat{\theta})^2$ being the sample variance.

Mean confidence interval length (MCIL): The MCIL serves as a measure of uncertainty for the parameter estimate. If $\text{CI}_i = [a_i, b_i]$ is the estimated confidence interval for parameter θ in run i, with $l(\text{CI}_i) = b_i - a_i$ defining the length of the confidence interval, the mean confidence interval length calculates as

$$\text{MCIL} := \frac{1}{m} \sum_{i=1}^{m} l(\text{CI}_i) \ , \tag{2.4}$$

where m denotes the total number of simulation runs.
In some cases, the MCIL cannot be calculated because at least one of the calculated confidence intervals does not have an upper bound (meaning all values above a certain threshold are equally likely). In the plots, this is indicated by a grey coloured box for the respective parameter combination.

False coverage rate (FCR): The false coverage rate is defined as the fraction of simulation runs in which the estimated confidence interval does not contain the true parameter.

CHAPTER 3

CD8$^+$ T cell dynamics in MCMV infection

3.1 Introduction

Cytomegaloviruses (CMVs) are a group of species-specific doubled-stranded DNA viruses, which belong to the family of herpesviruses (*Herpesviridae*) [Davison et al. 2009]. They are known to cause life-long persistent infections in different mammal species, including mice (MCMV[1]), rhesus monkeys (RhCMV) and humans (HCMV). It is estimated that the prevalence rate of CMV infections can reach up to 90% in certain human subpopulations [Colugnati et al. 2007]. Although CMV infections are usually clinically asymptomatic in immunocompetent hosts, they can be a major health issue in immunocompromised individuals [Krmpotic et al. 2002].

The course of an infection with CMV can be divided into an acute and a persistent phase: During acute infection, the virus replicates rapidly and spreads throughout different host organs such as spleen, lungs and liver [Zhang et al. 2016; Kurz et al. 1999]. The ongoing viral spread activates the immune system, which starts clearing the virus, although it is hampered by a number of viral immune-evasive strategies, such as the down-regulation of MHC I molecules or the expression of decoy receptors [Reddehase 2002; Hengel et al. 1998]. After a couple of days, CMV enters a stage of latency, which leads to the establishment of a persistent infection in the host [Brune et al. 2017; Reddehase et al. 2002]. Preferential sites of latency include the salivary gland and endothelial cells [Seckert et al. 2009; Thom et al. 2015; Loewendorf et al. 2011]. During latency, the virus is assumed to reactivate sporadically to maintain the status of a persistent infection and to allow for horizontal transmission to other susceptible hosts [Bolinger et al. 2013; Campbell et al. 2008; Seckert et al. 2012].

3.1.1 CMV infection induces CD8$^+$ T cell inflation

Contrary to most other diseases, CMV infection triggers two different kinds of CD8$^+$ T cell responses, termed “inflationary” and “non-inflationary” response [Sierro et al. 2005; Munks et al. 2006; Kim et al. 2015]. The dynamics of non-inflationary T cells follows the commonly observed ’expansion-contraction-memory maintenance’ dynamics: During the initial phase of

[1] According to a recent taxiconomical change proposed by the International Comitee on Taxonomy of Viruses [Aiaenssens et al. 2017], MCMV is now reclassified as murid betaherpesvirus-1 belonging to the newly established group of Muromegaloviruses instead of Cytomegaloviruses. However, we will stick with the old terminology as it is still widely used in the present literature.

infection, non-inflationary CD8$^+$ T cells increase massively in numbers, but once the acute infection is cleared, they die off rapidly and only a small pool of memory T cells is maintained long-term (see section 1.2 and Fig. 3.1 **A**). Non-inflationary responses are observed for CD8$^+$ T cells specific for the CMV epitopes M45 or M57 [Munks et al. 2006]. On the other hand, inflationary T cells are also expanding during the acute phase, but instead of contracting at the onset of the persistent infection, high levels of T cells are maintained over time (see Fig. 3.1 **A**). These inflated levels of T cells also show a markedly different composition of T cell phenotypes compared to their non-inflationary counterparts, as most of the maintained inflationary T cells do not exhibit a central memory but rather an effector memory or effector phenotype [Snyder et al. 2008; Hertoghs et al. 2010; Vieira Braga et al. 2015]. Inflationary responses are found for T cells specific for viral epitopes M38 or IE3 [Munks et al. 2006].

The generation and maintenance of elevated CD8$^+$ T cell responses over the course of a persistent CMV infection is known as "memory inflation" [Karrer et al. 2003]. The feature of CMV-induced memory inflation has garnered a lot of interest from immunologists: As CMV can be used as a vaccine vector expressing pathogen-specific epitopes, it can generate high levels of protection-mediating CD8$^+$ T cell responses against diseases such as HIV and malaria [Hansen et al. 2013a; O'Hara et al. 2012; Klenerman 2016; Fruh et al. 2017]. As HCMV and MCMV share many characteristics, studies analysing the phenomenon of memory inflation are often based on experiments in mice [Reddehase et al. 2002].

3.1.2 Generation and maintenance of memory inflation

Although several studies have broadened our understanding of the causes of memory inflation over the recent years, many questions about the generation and maintenance of the inflationary responses are still left unanswered.

Much progress has been made in determining why memory inflation occurs for certain but not all CMV-specific epitopes. Recent studies indicate a critical involvement of peptide-processing by the immunoproteasome and the localisation of the respective epitope in the CMV genome, as it was shown that recombination of the genome results in the inflation of different epitopes [Dekhtiarenko et al. 2016; Bolinger et al. 2013; Hutchinson et al. 2011]. Furthermore, it was found that antigen presentation by non-haematopoietic cells is critical to sustain memory inflation [Torti et al. 2011b; Seckert et al. 2011], which can be induced even if the initial priming of CD8$^+$ T cells is hampered [Torti et al. 2011a].

However, the exact location of the cells that drive memory inflation has not been determined yet. One study found that central memory T cells in lymph nodes are dividing rapidly during MCMV infection and therefore suggested that memory inflation is driven by proliferating T cells in the lymph nodes [Torti et al. 2011b]. This hypothesis was tested in another study, in which they blocked the egress of T cells from the lymph nodes [Smith et al. 2014]. However, they found that memory inflation was not abrogated. Instead they observed that most inflationary cells have access to the blood supply and therefore argued that memory inflation is dependent on antigen presentation by cells with access to the vasculature, as e.g. endothelial cells.

Another currently open question addresses the role of viral dynamics in the shaping and maintenance of memory inflation. It was found that inflationary T cells have a rather short half-life, which indicates the need for a constant replacement of lost cells by a source compartment [Snyder et al. 2008]. However, no viral load is detectable experimentally during the late stages of infection [Snyder et al. 2011]. Besides the possibility of a continuous viral antigen presentation with viral loads below the detection limit [Seckert et al. 2012], this has led to the hypothesis of memory inflation being dependent on sporadic reactivation of CMV from latency [Sims et al. 2015]. It is assumed that viral reactivation can be initiated by local inflammation or a decreased immune pressure [Reddehase et al. 2002]. However, how a fluctuating antigen stimulus would affect the long-term stability of the inflationary T cell pool over time (see Fig. 3.1 **B**) has not been determined yet.

3.1.3 Chapter overview

Due to its intriguing $CD8^+$ T cell responses and its potential use as a vaccine vector, there have been lots of studies about T cell dynamics in CMV infection. However, up to now, the factors involved in the generation and maintenance of memory inflation remain poorly understood.

To gain new insights into the cellular dynamics during CMV infection, we combined mathematical modelling with experimental data measuring $CD8^+$ T cell dynamics in MCMV-infected mices, which allowed us to not only distinguish between different modes of memory inflation maintenance, but also to quantify the underlying cellular dynamics and to elucidate the potential influence of sporadically reactivating virus on the shaping of inflationary $CD8^+$ T cell responses. To our knowledge, the present work also provides the first mathematical analysis of MCMV-induced memory inflation.

This chapter is structured as follows: In sections 3.2 - 3.4, we use mathematical modelling to determine in which way the inflationary and non-inflationary $CD8^+$ T cell response are maintained. To this end, we set up different mathematical models, describing memory inflation either by blood-based restimulation or by fuelling from external compartments (section 3.2), and fit them to the inflationary (section 3.3) and non-inflationary data (section 3.4) obtained from longitudinal blood measurements of MCMV-infected mice (see Fig. 2.1). In section 3.5, we analyse how a sporadically reactivating virus would shape the corresponding T cell response and which factors determine the long-term levels of inflationary $CD8^+$ T cells.

In summary, our results provide a quantitative understanding of the factors determining the creation and maintenance of memory inflation.

3.2 Mathematical models of memory inflation

To test whether inflationary T cells are maintained primarily by stimulation in the blood periphery or are rather fuelled by an external compartment, e.g. by T cells residing in lymph nodes, we developed several models of different complexity for each of the two hypotheses.

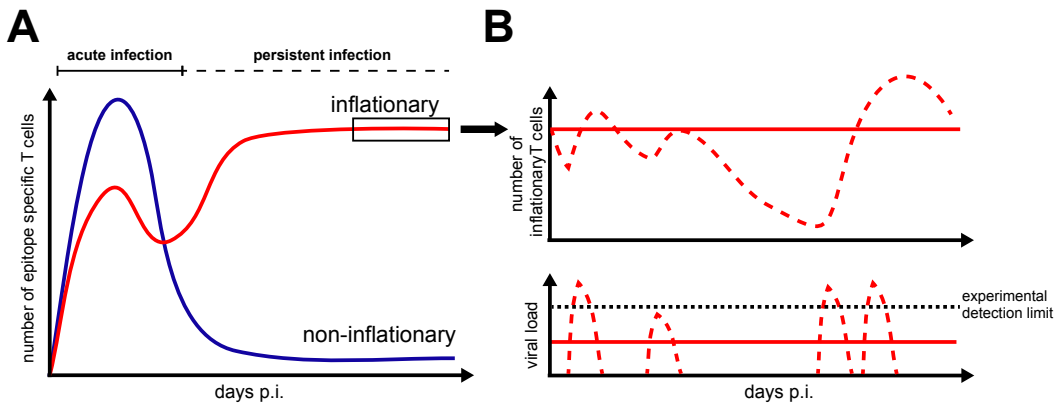

Figure 3.1: CD8$^+$ T cell dynamics during CMV infection: (**A**) CMV induces inflationary (red) and non-inflationary T cell responses (blue). Inflationary responses are maintaining high numbers of effector and effector memory T cells over the course of the persistent infection, while non-inflationary responses decline in numbers and maintain mostly central memory cells once the acute infection is resolved. (**B**) Possible virus and T cell dynamics during latency. Constant virus replication below the detection limit would result in a stable inflationary response over time (red solid lines). On the other hand, a sporadic reactivation of virus could trigger oscillating or highly variable T cell dynamics (red dotted lines).

3.2.1 Blood-based reactivation models

According to the theory proposed by [Smith et al. 2014], the maintenance of the inflationary pool is dependent on the presentation of viral antigen by cells with access to the blood supply. We will use the following models to describe the possible interactions of viral load and CD8$^+$ T cells in the vasculature.

Virus-dynamics model (VD) Describing the dependency of MCMV and CD8$^+$ T cells in a simple manner, we set up a virus-dynamics model, which is based on Lotka–Volterra equations. The system of ODEs is given as:

$$\begin{aligned} \frac{dV}{dt} &= \rho_{\mathrm{V}} V - kVT_{\mathrm{E}} \\ \frac{dT_{\mathrm{E}}}{dt} &= (\rho_{\mathrm{T}} V - \delta) T_{\mathrm{E}} \ . \end{aligned} \tag{3.1}$$

Here, V denotes the viral load, which triggers the immune system, and T_{E} the number of CD8$^+$ T cells. The replication and proliferation rates of virus and T cells are denoted by ρ_V and ρ_{T}, respectively. Virus is assumed to be cleared by T cells at rate k and CD8$^+$ T cells die over time based on the death rate δ. Including the initial value for the CD8$^+$ T cells at day 7 p.i. (T_7), the VD model has five different parameters. As no data for the viral load was available we set the initial value for the viral load to $V_7 = 1/\mathrm{ml}$.

Non-haematopoietic reservoir model (NR) To test the hypothesis that inflationary restimulation is mainly triggered by antigen presentation on non-haematopoietic cells [Torti et

al. 2011b; Smith et al. 2014], we expanded the virus-dynamics model by splitting the viral load into two different compartments, representing the stimulus during acute infection and the non-haematopoietic stimulus during persistent infection. Here, we assume that the non-haematopoietic viral reservoir builds up over the course of the acute infection and reaches a constant level shortly after. Additionally, we assume that $CD8^+$ T cells are not able to deplete the non-haematopoietic viral reservoir[2].

The system of ordinary differential equations describing these dynamics are given as

$$\begin{aligned} \frac{dV}{dt} &= \rho_{\mathrm{V}} V - kVT_{\mathrm{E}} \\ \frac{dR}{dt} &= (\beta V + \rho_{\mathrm{R}} R)(1-R) \\ \frac{dT_{\mathrm{E}}}{dt} &= \big(\rho_{\mathrm{T}}(V+R) - \delta\big)T_{\mathrm{E}} \ , \end{aligned} \tag{3.2}$$

where V denotes the viral stimulus during acute infection, R the viral load in the latent non-haematopoietic reservoir and T_{E} the number of $CD8^+$ T cells. The expansion rates of virus (acute and latent) and T cells are denoted by ρ_{V}, ρ_{R} and ρ_{T}, respectively. The viral load during acute infection is assumed to be cleared by $CD8^+$ T cells at rate k and infects non-haematopoietic cells at rate β. Since no data about the viral load is available, we set the limit of the non-haematopoietic reservoir equal to one. $CD8^+$ T cells are assumed to proliferate proportionally to the overall (i.e. acute + latent) viral load and die with rate δ. Together with the initial conditions V_7 and T_7 (the initial viral load in the non-haematopoietic reservoir is set to zero), the NR model has eight different parameters.

Expanded non-haematopoietic reservoir model (ENR) The expanded non-haematopoietic reservoir model is almost identical to the NR model. The only difference is that the viral transmission from acute to latent stages is not subject to the limit of the non-haematopoietic reservoir. Therefore, the $CD8^+$ T cells can initially receive a stronger stimulus before the non-haematopoietic steady state is reached.

The expanded non-haematopoietic reservoir model is defined as:

$$\begin{aligned} \frac{dV}{dt} &= \rho_{\mathrm{V}} V - kVT_{\mathrm{E}} \\ \frac{dR}{dt} &= \beta V + \rho_{\mathrm{R}} R(1-R) \\ \frac{dT_{\mathrm{E}}}{dt} &= \big(\rho_{\mathrm{T}}(V+R) - \delta\big)T_{\mathrm{E}} \ . \end{aligned} \tag{3.3}$$

The ENR model has eight different parameters.

[2]This restriction was imposed to guarantee a long-term viral steady state and does not mean that $CD8^+$ T cells are not able to clear infected cells. Biologically, it rather corresponds to a balance with respect to new infections and lysing of infected cells.

3.2.2 Influx-dependent reactivation models

According to Torti et al., memory inflation is fuelled by rapidly dividing T cells in the secondary lymphoid organs, especially the lymph nodes [Torti et al. 2011b]. Therefore, the number of CD8$^+$ T cells in the blood is supplied by an external influx. We set up models describing this influx with different levels of complexity.

Constant influx model (CI) Assuming the influx of CD8$^+$ T cells doesn't change over time, the constant influx model is given as:

$$\frac{dT_{\mathrm{E}}}{dt} = \Lambda - \delta T_{\mathrm{E}} \ , \tag{3.4}$$

where Λ models the constant influx and δ_T is the loss rate of blood-based CD8$^+$ T cells. The CI model has three parameters.

Biphasic influx model (BI) Assuming the input of cells changes over time, we considered a biphasic description of the cellular dynamics which is given as:

$$\begin{aligned} \Lambda &= \begin{cases} L_1, \ t < T \\ L_2, \ t \geq T \end{cases} \\ \frac{dT_{\mathrm{E}}}{dt} &= \Lambda - \delta T_{\mathrm{E}} \end{aligned} \tag{3.5}$$

Here, L_1 and L_2 are two different influx rates, T is the time the influx changes and δ is the loss rate of CD8$^+$ T cells. Together with the initial conditions for T_7^E, the BI model has five different parameters.

Variable influx model (VI) Expanding the BI model to allow for a more realistic description of the influx over time, we set up a new model that describes the external supply by a continuous function:

$$\begin{aligned} \frac{d\Lambda}{dt} &= \begin{cases} \alpha(L_1 - \Lambda), \ t < T \\ \alpha(L_2 - \Lambda), \ t \geq T \end{cases} \\ \frac{dT_{\mathrm{E}}}{dt} &= \Lambda - \delta T_{\mathrm{E}} \end{aligned} \tag{3.6}$$

Here, L_1 and L_2 are the maximal influx limits, α is a scaling factor that regulates how quickly the limits are reached, T is the time at which the influx changes and δ is the loss rate of CD8$^+$ T cells. Together with the initial conditions for Λ_7 and T_7^E, the influx model has seven different parameters.

3.3 Modelling the inflationary response

Fitting the mathematical models from section 3.2 to the M38 data based on a non-linear mixed effects model approach (see Materials and Methods 2.2.2) revealed that the expanded non-

	stimulation in blood			external influx of T cells		
model	VD	NR	ENR	CI	BI	VI
M38	2092	2084	2077	2091	2097	2084
M45	1896	1794	1812	1769	1777	1785

Table 3.1: Model comparison: The table shows the AICc for the tested models based on either the inflationary M38 or the non-inflationary M45 data set. The best models are highlighted.

haematopoietic reservoir model (ENR) is the most appropriate to describe the overall cellular dynamics and performs much better than all the other models in terms of AICc (see Table 3.1). The individual and population fits are shown in Fig. 3.2. The ENR model is able to recreate the secondary expansion dynamics and predicts a long-term decline in cell numbers.

The ENR model estimates that non-latent virus is present up to approximately day 40 p.i. (see Fig. 3.2 **B**, orange lines). This is in line with findings by Torti et al. [Torti et al. 2011b], in which MCMV was still detectable in the lungs and the salivary gland 28 days p.i. The individual and population dynamics of non-latent virus indicate a change in the viral clearance around 20 days p.i., with some even showing a small increase in acute viral load (Fig. 3.2 **B**, orange lines). The slowing in viral clearance coincides with the build-up of the non-haematopoietic viral reservoir, which starts increasing around day 15 p.i. and reaches its steady-state approximately ten days later (see Fig. 3.2 **B**, blue lines). The combined viral load and the corresponding T cell dynamics are shown in Fig. 3.2 **C** and **D** and resemble a damped oscillation.

The population parameter estimates obtained from fitting the ENR model are found in Table 3.2. Our model estimates that the half-life of inflationary $CD8^+$ T cells is around 2 days. However, since there is a continuous restimulation, the T cell pool itself is estimated to have a total half-life of approximately 26 days.

Comparing the modelled expansion dynamics with the individual parameter sets (see Table A.1), we find a very strong correlation between the peak cell numbers during the acute phase (before day 20 p.i.) and the parameters determining the viral replication, ρ_V, ($CC = 0.97$) and the T cell-mediated clearance, k, ($CC = -0.96$). Similar values[3] are observed when correlating these parameters to the increase in cell numbers during the secondary expansion phase, i.e. the second peak minus the lowest cell number between first and second peak. Therefore, we also find a strong correlation between the cell numbers at the first peak and the increase during the secondary expansion ($CC = 0.95$, neglecting mouse 7).

In summary, our model fits supports the hypothesis proposed by [Smith et al. 2014] that the inflationary $CD8^+$ T cell response is mainly dependent on viral antigen presence connected to the blood supply. Additionally, our findings highlight the need for different viral stimuli in the shaping of the overall inflationary response, corroborating the importance of antigen presentation on non-haematopoietic cells during persistent infection [Torti et al. 2011b; Smith et al. 2014].

[3]Here, the model prediction from mouse 7 formed an outlier and was therefore neglected in the following calculations.

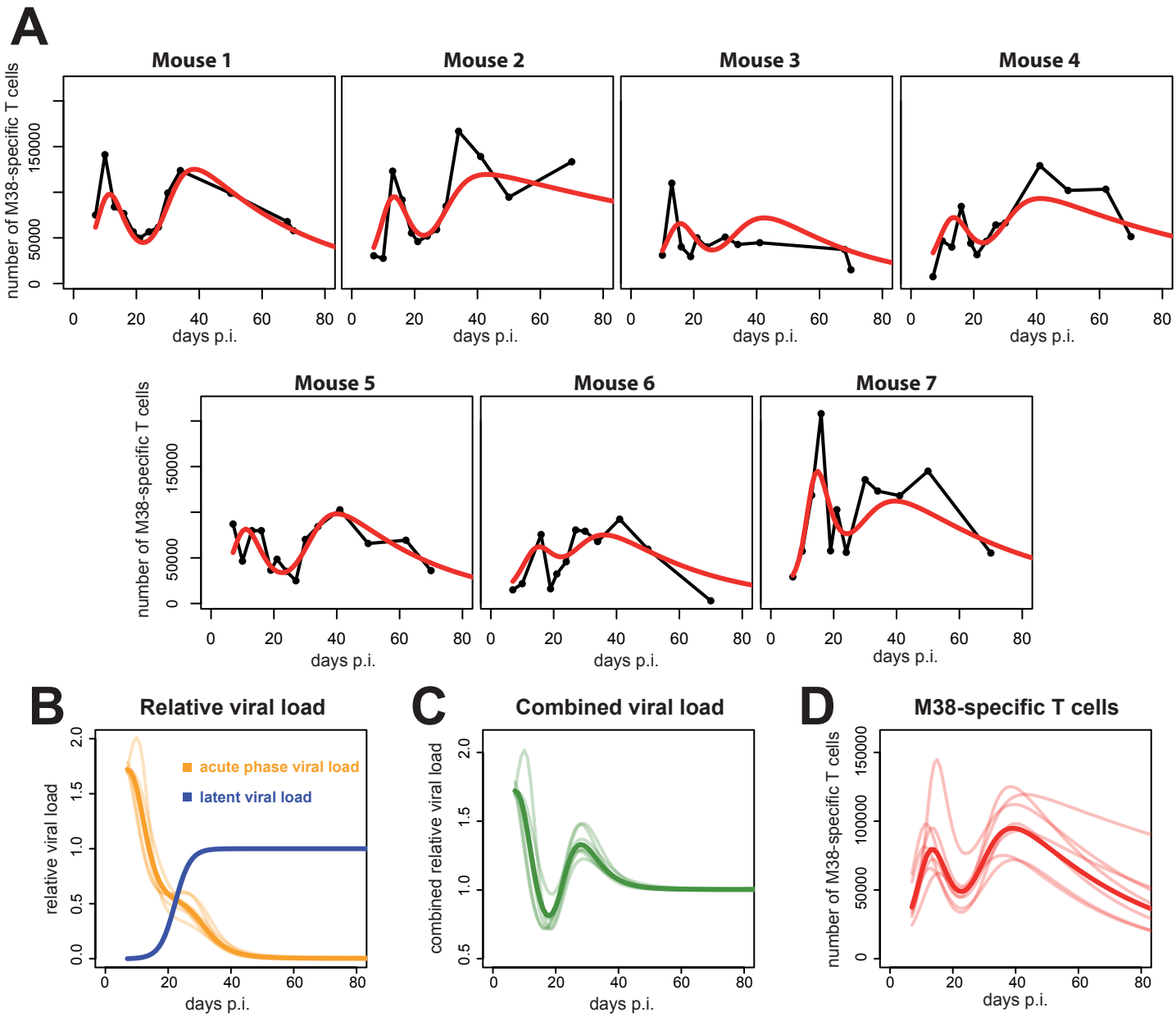

Figure 3.2: Individual fits to the inflationary M38-specific CD8$^+$ T cell response: (**A**) Panels show the model prediction (red) compared to the individual experimental data (black). (**B-D**) Population dynamics of viral load and CD8$^+$ T cells.(**B**) The relative viral load over time. Depletable viral load, V, is shown in orange, latent reservoir viral load, R, in blue. (**C**) The combined load of depletable and latent virus. (**D**) The population and individual dynamics of the M38-specific T cells. Solid lines indicate the population dynamics, opaque lines the individual fits. All experiments were performed by the lab of Prof. Dr. A. Oxenius, Institute of Microbiology, ETH Zürich.

3.4 Modelling the non-inflationary response

The dynamics of non-inflationary M45-specific CD8$^+$ T cells are characterised by a sharp decline in cell numbers after day 7 p.i., which is followed by the maintenance of a low number of T cells

parameter	unit	population parameters	population s.e.	relative s.e. (%)
V_7	viral load/ml	1.72	0.06	3
T_7	#cells/ml	3.6×10^4	6.6×10^3	18
ρ_V	day^{-1}	0.114	0.036	32
k	day^{-1}	3×10^{-6}	5.5×10^{-7}	18
β	day^{-1}	4.9×10^{-4}	2.2×10^{-4}	45
ρ_R	day^{-1}	0.414	0.035	8
ρ_T	day^{-1}	0.296	0.03	10
δ	day^{-1}	0.322	0.033	10

Table 3.2: Population parameter estimates for the ENR model: Population parameters, standard error (s.e.) and relative standard error are given. For model description see Eqs. 3.3.

over time (see Fig. 2.1 **B**). Fitting our models to the experimental data we find that the constant influx model (CI) is not only sufficient to describe the experimental data, but also yields the lowest AICc (see Table 3.1 and Appendix A.1). All other models perform much worse. Interestingly, all influx models perform better than any of the blood-based reactivation model, indicating a different long-term maintenance of the non-inflationary compared to the inflationary pool.

Based on our parameter estimations, the CI model predicts a half-life of roughly one day (see Table 3.3), which, together with our estimated influx rate, means that around 60% of the non-inflationary T cells are replaced every day. However, the analysis of the $CD8^+$ T cell subset distribution reveals that around one third of the non-inflationary pool consists of long-lived central memory cells (see Appendix A.3). Therefore it is possible that the estimated loss rate only describes the death of cells due to apoptosis-inducing pathways after resolution of the acute infection [Krammer et al. 2007]. To correctly determine the long-term maintenance of the $CD8^+$ T cell subsets, more data about the underlying dynamics are needed (see chapter 5).

parameter	unit	population parameters	population s.e.	relative s.e. (%)
T_7^E	#cells/ml	9.48×10^4	2.4×10^4	25
Λ	#cells/ml	5.49×10^3	1×10^3	18
δ	day^{-1}	0.62	0.094	15

Table 3.3: Population parameter estimates for the CI model: Population parameters, standard error (s.e.) and relative standard error are given. For model description see Eqs. 3.4.

3.5 The influence of sporadic virus reactivation on CD8$^+$ T cell dynamics

Until now, we always modelled the viral antigen presentation to CD8$^+$ T cells as a continuous process. However, as it is assumed that latent MCMV only sporadically reactivates, the previously described dynamics might only represent the mean viral and cellular turnover. To assess how sporadic viral reactivations would influence the overall CD8$^+$ T cell dynamics, we therefore created a model, in which the viral dynamics are based on impulsive differential equations [Bainov et al. 1993]. This allowed us to test how changes in the reactivation patterns, such as frequency and magnitude of viral reactivation events, would shape the corresponding CD8$^+$ T cell response.

3.5.1 Modelling sporadic viral reactivation

To model viral reactivation we used a system of impulsive differential equations. Here, the increase of the viral load is modelled in pulses at specified time points $T_i,\ i \in \mathbb{N}$ and is included in the dynamics by adding a specified amount of reactivated virus R_i at each time point T_i to the current amount of virus $V(T_i)$. The equations are given as

$$\begin{aligned} \frac{dV}{dt} &= -kVT_{\mathrm{E}}, \quad t \neq T_i \\ \Delta V &= V + R_i, \quad t = T_i \\ \frac{dT_{\mathrm{E}}}{dt} &= \big(\rho_{\mathrm{T}} V - \delta\big) T_{\mathrm{E}} \ , \end{aligned} \tag{3.7}$$

with the nomenclature of parameters being identical to the VD model (see 3.1).

3.5.2 Changes in viral burst patterns can lead to memory inflation

To test if memory inflation can also be caused by sporadic viral reactivation, we parametrised the sporadic reactivation model (see Eqs. 3.7) with the corresponding parameter estimates obtained from fitting the ENR model to the M38 data (see Table 3.2). To describe the viral reactivation events we first considered a simple periodic viral burst pattern, in which the duration between the viral bursts and the amount of released virus stays constant. We found that such a simple pattern always resulted in a "expansion-contraction-maintenance" dynamics (see Fig. 3.3, black line), which is similar to the response observed for non-inflationary T cells.

However, by allowing for changes to occur in the viral reactivation pattern or the CD8$^+$ T cell-mediated clearance of infected cells over time, we could recreate the observed memory inflation dynamics not only in one but in three separate ways: (1) By decreasing the time between two reactivation events, (2) by increasing the amount of virus released per reactivation event and (3) by decreasing the CD8$^+$ T cell dependent virus clearance rate, k (see Fig.3.3). All of these approaches led to the generation and maintenance of a higher long-term level of CD8$^+$ T cells compared to a monophasic reactivation pattern.

Furthermore, we found that changing the viral load per burst led to a proportionally scaling of the mean number of T cells long-term (based on calculations from day 50 to day 100 p.i.), while changes in the reactivation frequency and the T cell-mediated clearance scaled inversely proportional (see Appendix A.4). This means that, for example, a doubled viral load per reactivation event or a halved clearance rate result both in a doubled mean number of T cells long-term (see Fig.3.3 **B & C**). However, even though the scaling of these parameters affected the predicted mean long-term number in the same way, the oscillatory behaviour in terms of amplitude or period was very different (see Fig.3.3 and Appendix A.4).

In summary, our results indicate that memory inflation can also be caused by inherent changes in the MCMV reactivation patterns and/or a change in the CD8$^+$ T cell-mediated clearance of infected cells.

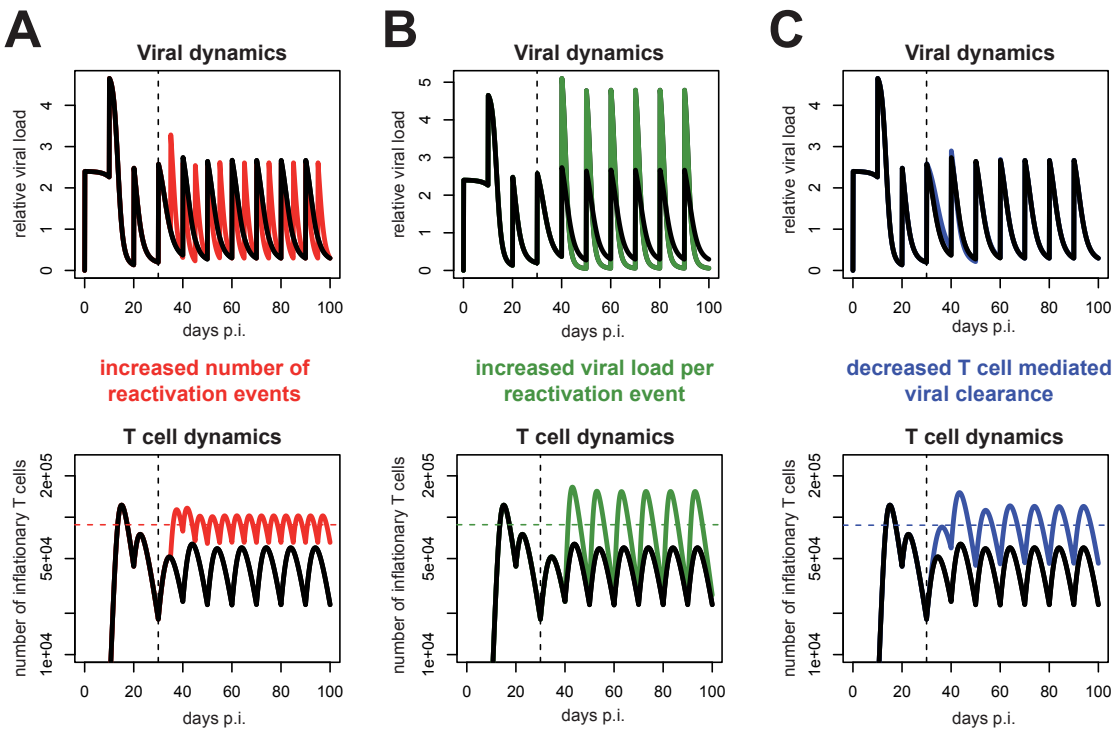

Figure 3.3: Memory inflation-like dynamics can be caused by changes in the sporadic viral reactivation patterns or T cell-mediated clearance: (**A**) Viral reactivation pattern based on a 10 days interval scheme (black) vs a 5 day interval scheme (red). The corresponding dynamics of inflationary CD8$^+$ T cells is shown below. (**B**) The effect of an increased viral burst load. The curves show the viral and cellular dynamics for the default ($R_i = 2.4$, black) versus the doubled viral load ($R_i = 4.8$, green). (**C**) The effect of an decreased T cell mediated viral clearance. The curves show the viral and cellular dynamics for the default ($k = 5 \times 10^{-6}$, black) versus the halfed clearance rate ($k = 2.5 \times 10^{-6}$, blue). Vertical dotted lines indicate the time at day 30 p.i. at which the viral reactivation dynamics changes. Coloured horizontal lines indicate the mean number of CD8$^+$ T cells from day 50 to 100. Parametrisation of the baseline dynamics (black curves) is given by $k = 5 \times 10^{-6}$, $\rho = 0.29$, $\delta = 0.32$, $R_i = 2.4$ and $T_i = 10i,\ i \in \mathbb{N}_0$ based on the sporadic reactivation model given by Eqs. 3.7.

3.6 Discussion

Understanding the interactions between CMV and the adaptive immune system is not only a prerequisite for explaining the phenomenon of memory inflation, but is also an important step towards the precise induction of protective CD8$^+$ T cell responses by CMV-based vaccination approaches.

Because of its promising features, memory inflation is a frequently studied topic in medical research. However, many details about its generation and maintenance have only been partially understood so far. Here, we analysed the responses generated by MCMV infection by mathematical modelling, which allowed us to describe the observed dynamics in a quantitative way, therefore broadening our understanding of the factors determining the dynamics of memory inflation.

One of the major questions regarding the CD8$^+$ T cell dynamics during MCMV infection addresses the maintenance of the inflationary T cell pool. Different biological hypotheses have been proposed to explain this phenomenon [Smith et al. 2014; Torti et al. 2011b; Seckert et al. 2012]. In our study, we could use mathematical modelling to distinguish between the likelihood of different theories. By creating and testing various models describing the possible underlying cellular dynamics and fitting them to the experimental data, we find that our results support the theory of inflationary CD8$^+$ T cell responses being dependent on direct antigen presentation in the blood periphery, as proposed by [Smith et al. 2014].

On top of testing different biological hypotheses, our modelling approach also allowed to analyse the underlying viral dynamics, which are difficult to determine *in vivo*. Our best-fitting model predicts that the inflationary response is shaped by two different viral antigenic stimuli, one of which is dominant during the acute phase of the infection, while the other dictates the long-term behaviour (see Fig. 3.2 **B**). These results correspond to findings in previous studies, which indicated the importance of antigen presentation on non-haematopoietic cells for the maintenance of the inflationary response during persistent infection [Torti et al. 2011b; Smith et al. 2014]. In addition, we find that the non-inflationary pool is maintained differently and seems to be dependent on an external influx, that might potentially be provided by CD8$^+$ T cells residing in secondary lymphoid organs [Torti et al. 2011b].

Since our study is based on a NLMEM approach, we could also analyse the individual variations in the inflationary responses. We find that the general shape and the magnitude of the response is critically dependent on the parameters describing viral replication and T cell-mediated clearance. Here, a reduced clearance and an increased replication rate led more pronounced cellular expansion dynamics. As both rates are likely to be affected by viral evasion mechanisms, one could speculate that the magnitude of the short- and long-term response depends mostly on the ability of the virus to evade the host's immune control.

Because no viral load is detectable experimentally during late stages of infection, it is assumed that CMV reactivates sporadically. To provide more insight into the possible effects of sporadic viral reactivation on the CD8$^+$ T cell dynamics, we set up a mathematical model in which virus reactivation is occurring based on periodic pulses. To our surprise, we found that memory inflation-like responses can also be generated by changes in the periodic reactivation patterns

or the $CD8^+$ T cell efficacy, such as an increase in viral bursts frequency or a decrease in the T cell related clearance efficacy (see Fig. 3.3). As CMV is able to mount different immune-evasive strategies [Reddehase et al. 2002; Hengel et al. 1998], a decrease in the $CD8^+$ T cell dependent clearance rate, possibly due to viral latency [Seckert et al. 2012], might be a plausible cause for the secondary $CD8^+$ T cell expansion observed *in vivo*.

Besides being able to describe the longitudinal dynamics, our modelling approach also allowed us to quantify the viral and cellular turnover during memory inflation. For example, we found that the reservoir of latently infected cells starts building up around 2 weeks p.i. and reaches its maximum capacity around day 25 p.i. While this is later than the dynamics observed by Torti et al., in which a second expansion phase was estimated to occur around day 11-15 p.i. [Torti et al. 2011b], the difference might result from their focus on cellular frequencies instead of cell numbers, which can potentially lead to a bias due to a massive loss in non-inflationary cell numbers occurring around that same time (see Fig. 2.1).

Furthermore, our model predicts a decline in cell numbers which corresponds to a half-life of inflationary cells of approximately 26 days. However, many other studies show that the inflationary pool is maintained in a rather stable manner over longer periods of time [Munks et al. 2006; Torti et al. 2011b; Smith et al. 2014; Walton et al. 2011]. Therefore, our predicted decline might either be an artefact attributed to the experimental data or indicates a long-term oscillatory cellular dynamics. Varying inflationary responses might arise from an increase in viral reactivation events occurring when $CD8^+$ T cell levels falls below a certain threshold [Reddehase et al. 2002]. However, to address the question about the stability of the inflationary cell pool, more long-term data is needed for analysis.

Taken together, our results argue for the following hypothesis regarding the generation of memory inflation: During acute infection, an initial $CD8^+$ T cell response against inflationary epitopes is primed. This response declines as the acute infection terminates, but is triggered again by the successful establishment of CMV latency in non-haematopoietic cells, which goes hand in hand with superior viral immune-evasion and/or diminished $CD8^+$ T cell mediated clearance of infected cells. In the long run, the overall cellular dynamics are shaped by the reactivation cycles of the latent virus.

In our analysis of the non-inflationary response we encountered the problem that the quantification of cellular turnover is difficult to determine because the system quickly reaches a steady state and no further change in the overall response or the subset levels is observed (see Fig. 2.1 and Appendix A.3). Here, additional long-term measurements of $CD8^+$ T cell (subset) numbers would not help to identify the underlying cellular dynamics in terms of proliferation kinetics and cell longevity. To determine these properties, markers indicating cellular turnover, such as Ki67 or BrDU could be used [Torti et al. 2011b]. Additionally, the use of labelled cell population and the quantification of cellular turnover based on our methods developed in chapter 5 would allow to address these questions as well.

Since we are lacking the data of the original expansion phase as well as sufficient long-term data, our current modelling of MCMV-induced T cell behaviour only describes the $CD8^+$ T cell dynamics over the course of 7 to 70 days p.i. Analysing cell numbers over the course of the acute MCMV infection might be used to determine differences in the early expansion dynamics

of inflationary and non-inflationary T cells. Additionally, the measurements of other organs besides the blood would help to determine the local dynamics of CD8$^+$ T cells during MCMV infection. Even though we had additional measurements of CD8$^+$ T cells in lymph nodes and spleen at day 8, 10 and 12 p.i. at our disposal, these data were not sufficient to infer any tissue-dependent dynamics.

In our analysis of sporadic reactivation we only generated responses that were build on virus patterns following a periodic behaviour. However, viral reactivation might be subject to a stochastic expression pattern, with bursts occurring at irregular time intervals. These stochastic bursts might facilitate the viral evasion of the immune system and possibly allow the virus to maintain the state of a persistent infection. However, since all cytomegaloviruses establish life-long infections, a premature death of the host cannot be in the interest of the virus. Therefore, it is an interesting speculation if the development of imperfect immune-evasive strategies is an evolutionary stable strategy for CMVs, meaning that frequent triggering of the adaptive immune response balances viral replication and host survivability, or, put differently, balances short-term and long-term viral infectivity. Since the measured CD8$^+$ T cell levels contain information about viral reactivation patterns, sampling inflationary cells more frequently over the course of the persistent infection would certainly help to characterise the interactions of CMV and the immune system.

In summary, our analysis provides an important step towards the quantification of CD8$^+$ T cell responses induced by MCMV infection and therefore helps to determine potential effects of vector-based vaccination strategies.

CHAPTER 4

$CD8^+$ T cell dynamics induced by variable malaria vaccination approaches

4.1 Introduction

Despite strong efforts in the past, malaria is still one of the most common infectious diseases today and is especially prevalent in tropical and subtropical regions around the world, where it is responsible for millions of infections and hundreds of thousands of deaths per year [WHO 2016]. As malaria can cause periodic fevers, diarrhoea, anaemia or even brain damage, it is estimated that the disease has an enormous social and economical impact in the high-risk regions [Sachs et al. 2002]. Even though a lot of effort, including medical research, is dedicated towards protection from malaria, this goal has proven hard to reach so far.

4.1.1 The infection cycle of malaria

Malaria is caused by *Plasmodium*[1], a parasite with a complex life cycle involving multiple stages in different hosts species (see Fig.4.1 and [Hafalla et al. 2011]). Briefly sketched, *Plasmodium*, in the form of sporozoites, is transmitted into humans (but also to other species) by mosquitoes, where they quickly migrate to the liver tissue to infect hepatocytes. During the liver stage, the parasites replicate intracellularly and mature into merozoites, which are later released from the hosting hepatocytes into the blood stream. In the blood, merozoites start to infect red blood cells (RBCs) in which they replicate and produce new merozoites. The periodic rupturing of harbouring RBCs establishes an ongoing infection cycle in the blood, over the course of which the number of available RBCs continuously declines. The progressive loss of blood cells is one of the reasons for the observable symptoms of malaria, such as characteristic fevers or anaemia. Over the course of the infection, some merozoites will mature into male and female gametocytes, which, once taken up by mosquitoes again, will start to reproduce sexually, thus leading to the formation of new sporozoites.

There are several reasons why the immune system is not able to handle the infection with *Plasmodium* parasites properly. One of them is the short time the parasite spends in its liver stage, which amounts only to a couple of days [Frevert 2004]. As the adaptive immune sys-

[1] There are many different species of *Plasmodium*. We will base our following characterisations on *Plasmodium falciparum*, which is the most common malaria-inducing parasite type.

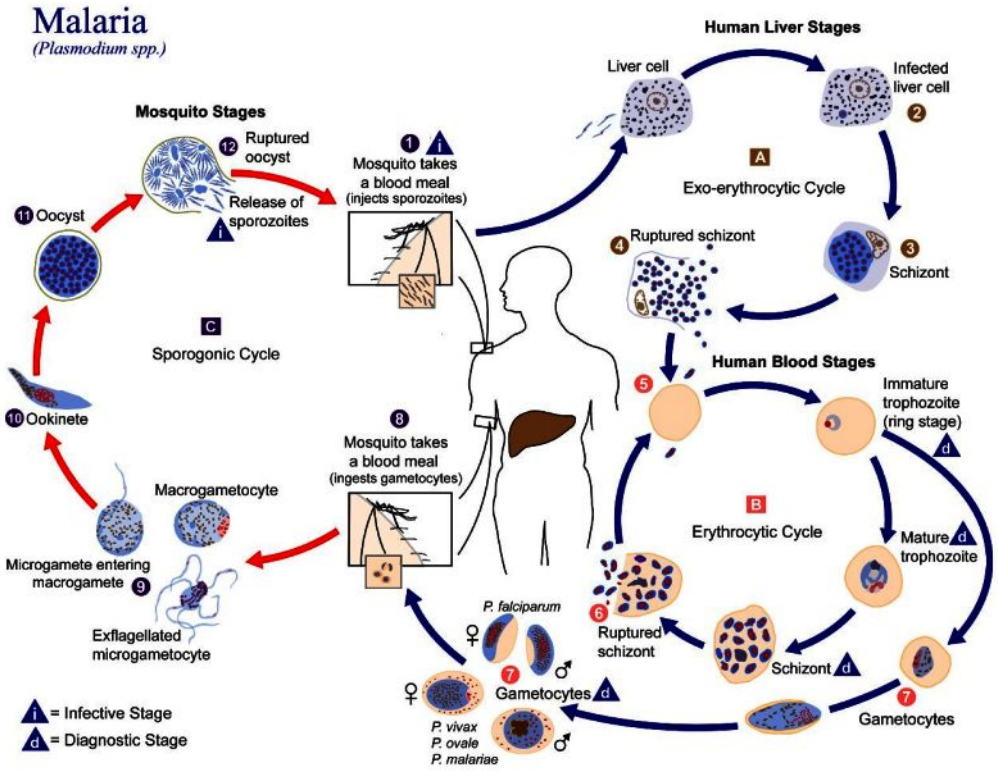

Figure 4.1: The life cycle of *Plasmodium*: After transmission from mosquitoes, sporozoites infect liver cells in which they mature into schizonts. Upon bursting, infected hepatocytes release merozoites (1-4), which start infecting red blood cells where they reproduce intracellularly (5-6). Some merozoites mature into male and female gametocytes, which are taken up by mosquitoes again, where they reproduce sexually and generate new sporozoites (7-12). Figure reproduced and modified with permission from the Centers of Disease Control and Prevention (CDC).

tem needs time to build up, many infected hepatocytes escape the immunological control and release merozoites into the blood stream[2], thereby changing the pattern of the infection. As infected RBCs do not express MHC I molecules, the intracellular parasite reproduction cannot be recognised by CD8$^+$ T cells [Mohandas et al. 2012]. Furthermore, infected RBCs can attach themselves to endothelial cells, which prevents them from entering the spleen, where abnormal blood cells are usually removed [Chotivanich et al. 2002].

4.1.2 Vaccination approaches against malaria

Currently, there exists only one regulatory approved vaccine against malaria ("RTS,S"), which targets the parasite liver stage [Cohen et al. 2010; Morrison 2015]. However, several studies have shown that the vaccine has only intermediate efficacy in adults and low to intermediate efficacy in children and that the granted protection seems to wane quickly [Abdulla et al. 2008; Bojang et al. 2009; Agnandji et al. 2014].

[2]Hypothetically, a single infected hepatocyte that escapes the immune surveillance is enough to establish a blood stage infection.

However, since the liver stage serves as a bottleneck in the parasite life cycle, targeting infected hepatocytes seems to be a promising approach for T cell-based vaccine designs. Using whole sporozoite vaccination strategies, it has been shown that generating high amounts of liver-resident effector cells confers protection against sporozoite challenges in mice and humans [Schmidt et al. 2008; Schmidt et al. 2010; Fernandez-Ruiz et al. 2016]. Therefore, a major goal of T cell-based vaccines for malaria is the build-up of elevated effector-like $CD8^+$ T cell responses surveilling the liver tissues.

Currently, this approach is hampered by a lack of knowledge regarding the influence of the vaccination regimen on the underlying $CD8^+$ T cell dynamics [Sallusto et al. 2010]. For example, it is not known how exactly the number of booster injections or the administered vaccination dose affect the generation of liver-resident $CD8^+$ T cells. The importance of booster injections in mediating protection has been shown in previous studies [Schmidt et al. 2011; Patel et al. 2017], but has not been analysed systematically yet.

4.1.3 Chapter overview

To address this question, we set up an extensive experimental protocol, in which we assessed the influence of different vaccination regimes based on *Plasmodium berghei* radiation attenuated sporozoites (PbRAS) on the dynamics of organ-dependent $CD8^+$ T cell responses (see Materials and Methods 2.3.1). The biological experiments were conducted by the lab of Prof. Dr. A.-K. Mueller, University Hospital Heidelberg. In these experiments, we varied the number of booster injections as well as the dosage and measured the corresponding number of generated $CD8^+$ T cell subsets in liver and spleen at several time points after vaccination. Analysing the experimental data by mathematical modelling allowed us not only to identify the underlying cellular differentiation pathway but also to quantify the influence of the vaccination regimen on the generation of protective $CD8^+$ T cell pools in spleen and liver.

This chapter is structured as follows: In section 4.2, we analyse the organ-dependent $CD8^+$ T cell differentiation and migration pathway and model the cellular responses after vaccination with several normal-dosed booster injections. To this end, we set up a global model of $CD8^+$ T cell differentiation, including all possible $CD8^+$ T cell subset interactions and establish an unbiased model selection algorithm that automatically assesses the appropriateness of different models to explain experimental data. This optimisation routine allows us to determine the most likely cellular differentiation pathway, of which we subsequently quantify the cellular turnover.

In section 4.3, we additionally determine how the vaccination dose influences the generation of protective levels of liver-resident $CD8^+$ T cells. Here, we analyse how the cellular responses are affected by the administration of a high or a subprotective vaccination dose.

In summary, our results allow us to determine the $CD8^+$ T cell dynamics generated after malaria immunisation and can be used to improve future vaccination approaches.

4.2 Modelling the CD8$^+$ T cell response after normal dose PbRAS vaccination

In a first attempt to describe the experimental results by mathematical modelling, we fitted the normal-dose experimental data to a selection of hand-picked models. While we found some models that described the data well, a simultaneously tested random selection of models yielded even better results, meaning that the hand-picked model approach was too biased. Therefore, we implemented a model selection algorithm, which allowed for the unbiased search of the best model describing the experimental data. We applied this algorithm on the global model of cellular turnover, which incorporates all possible CD8$^+$ T cell subset interactions, to find the most suitable submodel for describing the experimental data.

4.2.1 The global model of cellular turnover

Our global model includes five CD8$^+$ T cell subsets localised in different tissues. For the spleen, we distinguish between naïve (T_N), central memory (T_{CM}) and effector/effector memory[3] (T_{EMS}) CD8$^+$ T cells, while in the liver we separate between effector/effector memory (T_{EML}) and tissue-resident (T_{RM}) T cells. The T_{CM} response in the liver was neglected, as previous studies indicated no protection-mediating role for this compartment [Holz et al. 2016].

In our model, we allow each T cell compartment, besides T_N , to proliferate and to differentiate into other compartments (for a sketch of all possible dynamics see Fig. 4.2). Since the experimental data does not allow for the distinction of cellular expansion and death rates, all proliferation rates are net-proliferation rates, i.e. they include cell death.

To model the antigen-dependent stimulation of CD8$^+$ T cells we use an indicator function, which changes the model parameter values if antigen is present. Therefore, each cellular rate consists of two separate parameters and is defined according to the following example:

$$\alpha := \mathbb{1}_{\{t \in D_{AG}\}} \alpha^{ag} + \alpha^{b} \ . \tag{4.1}$$

Here, we use the notation α^{ag} and $\alpha^{b} \in \mathbb{R}$ to describe the antigen-dependent and antigen-independent turnover, respectively (ag stands for antigen, b for baseline). The set D_{AG} contains the time intervals in which antigen is present and is defined as $D_{AG} := \bigcup_{i=1}^{n} [T_i, T_i + L_i]$, where n defines the total number of injections, $T_i \in \mathbb{R}^+,\ i = 1, ..., n$ the time at which injection i is administered and $L_i \in \mathbb{R}^+,\ i = 1, ..., n$ the corresponding duration of antigen presence. The first vaccination shot is always given at $T_1 = 0$ and the respective length is set to $L_1 = 8$, which is a common estimate for the duration of CD8$^+$ T cell expansion during acute infection [De Boer et al. 2001; Kohler 2007].

To describe the interactions of spleen and liver, we used a negative feedback loop from liver to spleen, which we modelled via a Hill equation, γ, taking into account the levels of T_{RM}. It is

[3] The analysed marker expression of CD44, CD62L and CD69 allows no distinction between effector and effector memory T cells.

given as

$$\gamma(\mathrm{T_{RM}}) := 1 - \frac{{\mathrm{T_{RM}}}^2}{{\mathrm{T_{RM}}}^2 + 236,000^2} \ , \tag{4.2}$$

which acts exclusively on the antigen-dependent proliferation of $\mathrm{T_{CM}}$ in the spleen, $\rho^{\mathrm{ag}}_{\mathrm{CM}}$. The values are based on the analysis performed by Dr. F. Graw (see Materials and Methods 2.3.2).

Putting all the parts together, the complete CD8$^+$ T cell turnover dynamics of the global model can be described by the following system of ordinary differential equations:

$$\begin{aligned}
\frac{\mathrm{dT_N}}{\mathrm{dt}} &= -\mathbb{1}_{\{t\in \mathrm{D_{AG}}\}}\Big(\sigma^{\mathrm{ag}}_{\mathrm{N-CM}} + \sigma^{\mathrm{ag}}_{\mathrm{N-EMS}}\Big)\mathrm{T_N} \\
\frac{\mathrm{dT_{CM}}}{\mathrm{dt}} &= \mathbb{1}_{\{t\in \mathrm{D_{AG}}\}}\Big(\sigma^{\mathrm{ag}}_{\mathrm{N-CM}}\mathrm{T_N} + \gamma(\mathrm{T_{RM}})\rho^{\mathrm{ag}}_{\mathrm{CM}}\mathrm{T_{CM}}\Big) + (\rho^{\mathrm{b}}_{\mathrm{CM}} - \sigma_{\mathrm{CM-EMS}})\mathrm{T_{CM}} \\
&\quad + \sigma_{\mathrm{EMS-CM}}\mathrm{T_{EMS}} \\
\frac{\mathrm{dT_{EMS}}}{\mathrm{dt}} &= -\mathbb{1}_{\{t\in \mathrm{D_{AG}}\}}\sigma^{\mathrm{ag}}_{\mathrm{N-EMS}}\mathrm{T_N} + \sigma_{\mathrm{CM-EMS}}\mathrm{T_{CM}} \\
&\quad + \Big(\rho_{\mathrm{EMS}} - \sigma_{\mathrm{EMS-EML}} - \sigma_{\mathrm{EMS-RM}}\Big)\mathrm{T_{EMS}} \\
&\quad + \sigma_{\mathrm{EML-EMS}}\mathrm{T_{EML}} + \mathbb{1}_{\{t\in \mathrm{D_{AG}}\}}\sigma_{\mathrm{RM-EMS}}\mathrm{T_{RM}} \\
\frac{\mathrm{dT_{EML}}}{\mathrm{dt}} &= \sigma_{\mathrm{EMS-EML}}\mathrm{T_{EMS}} + \Big(\rho_{\mathrm{EML}} - \sigma_{\mathrm{EML-EMS}} - \sigma_{\mathrm{EML-RM}}\Big)\mathrm{T_{EML}} \\
&\quad + \mathbb{1}_{\{t\in \mathrm{D_{AG}}\}}\sigma_{\mathrm{RM-EML}}\mathrm{T_{RM}} \\
\frac{\mathrm{dT_{RM}}}{\mathrm{dt}} &= \sigma_{\mathrm{EMS-RM}}\mathrm{T_{EMS}} + \sigma_{\mathrm{EMLRM}}\mathrm{T_{EML}} \\
&\quad + \Big(\rho_{\mathrm{RM}} - \mathbb{1}_{\{t\in \mathrm{D_{AG}}\}}(\sigma_{\mathrm{RM-EMS}} - \sigma_{\mathrm{RM-EML}})\Big)\mathrm{T_{RM}}
\end{aligned} \tag{4.3}$$

Together with the parameters describing the durations of antigen presence after the second and the third vaccination shot, L_2 and L_3, the global model contains a total of 26 different parameters.

4.2.2 Unbiased model selection algorithm

The aim of the unbiased model selection algorithm (UMSA) is to find the most appropriate submodel of the global model (Eqs. 4.3 and Fig. 4.2.1) to describe the experimental data. Based on a specific submodel, the algorithm tests all neighbouring models and updates the current model, if a better one is found. Therefore, the algorithm successively progresses through the space of all possible models. It stops if no better models can be found. The details of the unbiased model selection algorithm are found below and a sketch is provided in Fig. 4.3.

Critical parameter sets and algorithmic memory To avoid the testing of models that are not able to reproduce the cellular dynamics the algorithm takes into account critical parameter sets. Each model has to contain at least one parameter of each critical set in order to be tested. If this condition is not fulfilled the respective model will be discarded.

For our global model given by Eqs. 4.3, we specified the following six critical parameter sets:

- $\mathrm{S}_1 = \{\sigma^{\mathrm{ag}}_{\mathrm{N-CM}}; \sigma^{\mathrm{ag}}_{\mathrm{EMS-CM}}; \sigma^{\mathrm{b}}_{\mathrm{EMS-CM}}\}$ (differentiation into $\mathrm{T_{CM}}$)

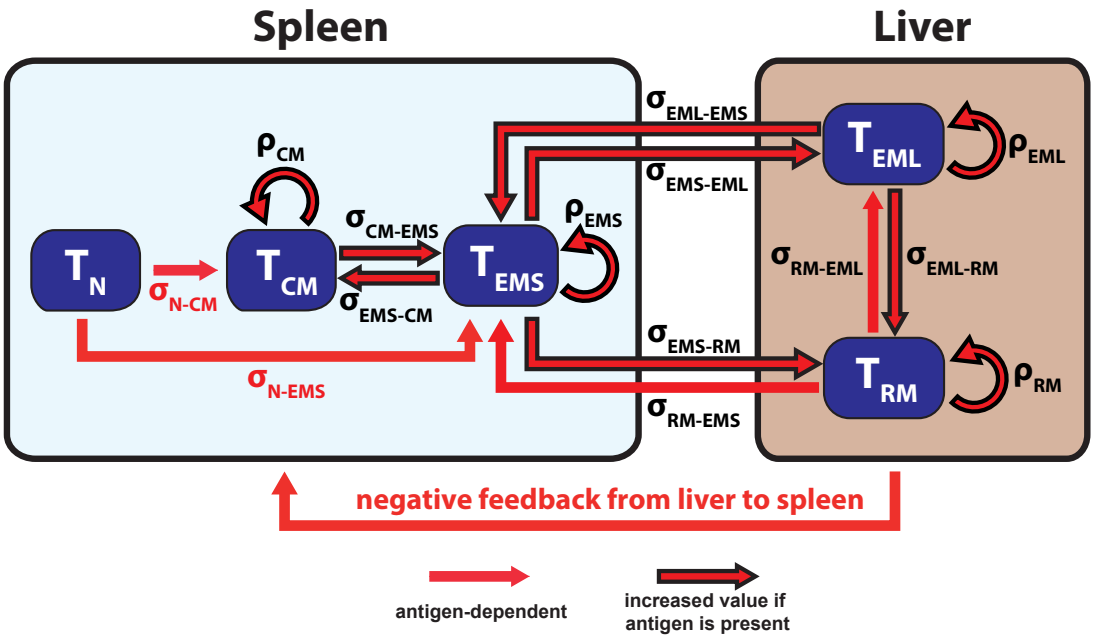

Figure 4.2: Sketch of the global model: This model sketch shows all possible differentiation and net-proliferation dynamics of the CD8$^+$ T cell subsets localised in spleen and liver. Coloured arrows indicate if a rate is active only if antigen is present (red) or also independently of antigen exposure (red arrows with black outline, includes baseline and antigen-dependent turnover).

- $S_2 = \{\sigma^{ag}_{N-EMS}; \sigma^{ag}_{CM-EMS}; \sigma^{b}_{CM-EMS}\}$ (differentiation into T_{EMS})
- $S_3 = \{\sigma^{ag}_{EMS-EML}; \sigma^{b}_{EMS-EML}; \sigma^{ag}_{RM-EML}\}$ (differentiation into T_{EML})
- $S_4 = \{\sigma^{ag}_{EMS-RM}; \sigma^{b}_{EMS-RM}; \sigma^{ag}_{EML-RM}; \sigma^{b}_{EML-RM}\}$ (differentiation into T_{RM})
- $S_5 = \{\rho^{ag}_{EMS}\}$ (antigen-dependent proliferation of T_{EMS})
- $S_6 = \{\rho^{ag}_{EML}\}$ (antigen-dependent proliferation of T_{EML})

Additionally, the algorithm checks if the model was already tested before. If this is the case, the model is skipped and will not be tested again.

The outline of the algorithm Starting from an initial model, the unbiased model selection algorithm is searching for superior models explaining the experimental data based on the combination of three different methods:

- **Forward search:** Adds one parameter to the current model
- **Double forward search:** Adds two parameters to the current model
- **Backward elimination:** Removes one of the current model parameters

At each step, the UMSA generates a new set of models based on the chosen method and tests their performance in comparison to the current model. For example, if the current model contains k

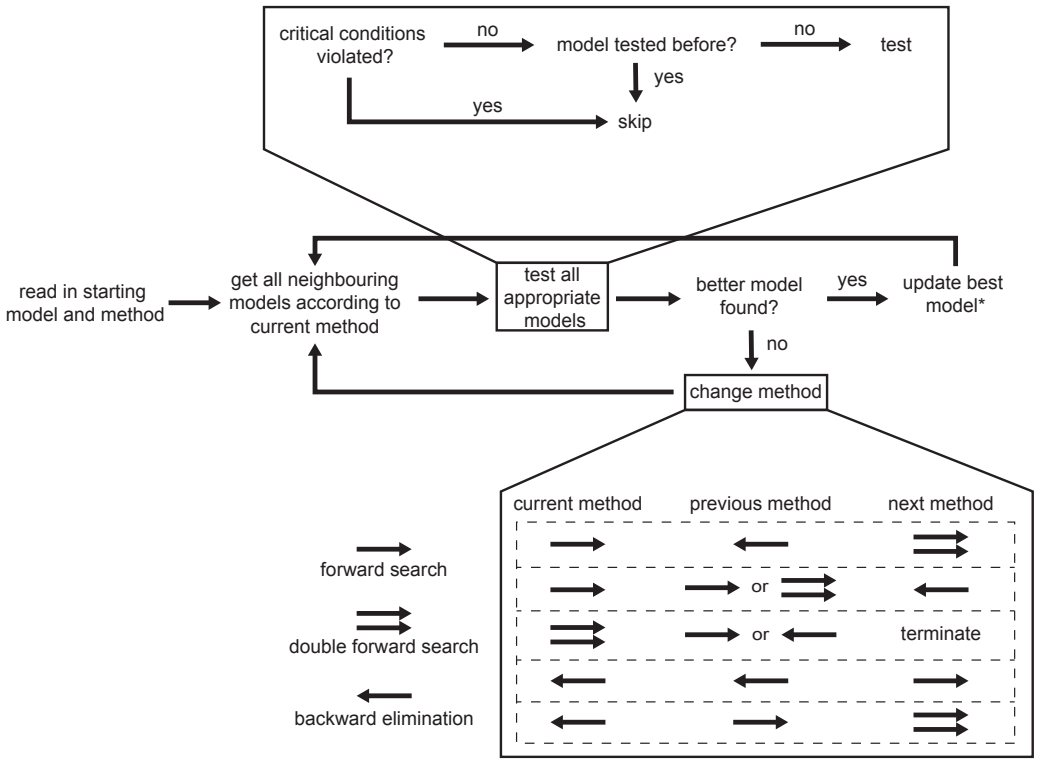

Figure 4.3: The outline of the unbiased model selection algorithm: After reading in the initial model and starting method, the algorithm generates a new set of models at each step based on the current method and tests their performance in comparison to the currently best model. All models that do not violate the critical conditions are used for testing. The estimated parameter values of the best fit and the corresponding AICc value are returned. If a better model is found, the current model is updated and the main routine is repeated with the current method (*exception: If the current method is a double forward search, the algorithm switches to a forward search). If no better model is found, the algorithm switches the method and either continues with the main routine or terminates.

(out of n possible) different parameters, the algorithm will generate k new models for a backwards elimination and $n - k$ models for a forward search. All models that do not violate the critical conditions are used for testing. The testing procedure involves the fitting of the respective model to the experimental data using the `optim`-fitting routine in `R` with 5 different starting conditions for the parameter values (one using the parameter set of the currently best model and four randomly sampled initial conditions). Convergence of the fitting routine was ensured by the structural set-up described in Material and Methods 2.1.1. After testing, the estimated parameter values of the best of the five fits and the corresponding AICc value are returned. The current model is updated if one of the model shows better a performance according to the AICc. The different methods (forward search, double-forward search and backward elimination) are used in successive steps as described below. If none of the different methods finds a better model, the algorithm is halted and the currently best model is returned.

The exact testing sequence of the algorithm is given as follows (see also Fig. 4.3):

- Read in the initially specified model. Jump to the initially specified testing method.
- Forward search
    1. For each non-model parameter, create a new model by adding the respective parameter to the current model. Test all appropriate models.
    2. If a better model is found, update the current model and continue with forward search.
    3. If no better model is found:
        - Continue with backward elimination if the previous method was a forward or a double forward search.
        - Continue with double forward search if the previous method was a backward elimination.
- Double forward search
    1. For each combination of two non-model parameters, create a new model by adding the two respective parameters to the current model. Test all appropriate models.
    2. If a better model is found, update the current model and continue with forward search.
    3. If no better model is found, halt the simulation and return the currently best model and estimated parameter values.
- Backward elimination
    1. For each model parameter, create a new model by removing the respective parameter from the current model. Test all appropriate models.
    2. If a better model is found, update the current model and continue with backward elimination.
    3. If no better model is found:
        - Continue with forward search if the previous method was a backward elimination.
        - Continue with double forward search if the previous method was a forward search.

4.2.3 Model selection and model ranking

In total, we applied our unbiased model selection algorithm to ten different starting models. These models included (a) the global model, (b) 3 minimal models, in which we selected one parameter from each critical set of parameters randomly, and (c) 6 different models, in with we randomly selected one parameter from each critical set of parameters and randomly added five of the remaining parameters. This means that the models at the start comprised 26 unknown parameters in (a), 8 (6 + 2 for the antigen presence durations after the second and third shot L_2 and L_3) parameters in (b) and 13 (11 + 2) parameters in (c). We used backward elimination as the starting method for the global model and forward search for the other starting models.

Over the ten runs performed, approximately 2000 different models were tested and fitted to the normal-dose data (see Materials and Methods 2.3.1). While optimising the models, we applied

the following restrictions regarding the values of the model parameters: First, all parameters, besides the baseline net-proliferations, were not allowed to take negative values. Second, the baseline net-proliferation rate of the T_{RM} compartment, ρ_{RM}^{b} was restricted to non-positive values. This was done to prevent a continuous expansion of the T_{RM} cell pool over time. Third, the upper limit of all antigen-dependent net-proliferation rates was set to five, thus removing models that assume a massive increase in cell numbers from the fitting procedure. This artificial limit corresponds to a division time of 3 hours, which is close to the biological limit found for CD8$^+$ T cell division during antigenic stimulation [Yoon et al. 2010].

Ordering the results of our model selection algorithm by the ΔAICc (see Materials and Methods 2.1.2), we found that one model clearly outperformed the others (see Fig. 4.4 **A**, only the 200 best models are shown). The second best model already had a ΔAICc equal to 4, indicating considerably less support for this model [Burnham et al. 2003]. Interestingly, we found a clear distinction between the possible naïve T cell differentiation pathways, as models assuming a differentiation from T_N into T_{EMS} performed worse than those that assumed a differentiation from T_N into T_{CM} (see Fig. 4.4 **B**).

The best model is shown in Fig. 4.4 **C**. Here, cell differentiation follows an almost linear pathway: Upon antigen encounter, T_N turn into T_{CM} cells which further differentiate into T_{EMS} cells. These cells migrate to the liver and turn into T_{RM} cells, if antigen is present. Our model also predicts an antigen-dependent turnover from T_{EMS} into T_{CM} cells in the spleen. In case antigen is absent, the compartment of T_{RM} cells is self-sustaining, while both T_{EMS} and T_{EML} compartments ultimately get supplied by slowly proliferating T_{CM} cells in the spleen. All model parameters have high Akaike evidence ratios (see Appendix B.1), corroborating the importance of the respective parameters for describing the cellular dynamics. Interestingly, the lowest evidence ratio is found for the antigen-dependent proliferation of T_{RM}, indicating that different possibilities exist to build up this pool given booster injections (see section 4.2.4).

4.2.4 Parameter identifiability

Identifiability of parameters was assessed by performing a profile likelihood approach [Raue et al. 2009]. The corresponding plots are shown in Appendix B.2. All of the baseline parameters turned out to be identifiable, while most of the antigen-dependent parameters were non-identifiable (see Table 4.1). Due to a lack of acute infection data, this non-identifiability is most likely structural and could be resolved if more data points are sampled [Raue et al. 2009].

Analysing the baseline turnovers, we find that T_{CM} cells are estimated to have a doubling time of 2 to 6 days. However, T_{CM} proliferation is only needed to compensate for the loss of T_{EML} cells in the liver, which have a half-life or residence time of 6 to 15 hours, indicating continuous replacement. On the other hand, the number of T_{RM} in the liver tissues is stably maintained over time without any influx from other compartments.

During antigen presence, T_{EMS} and T_{EML} proliferate rapidly with estimated doubling times of 4 to 10 and 3 to 6 hours[4], respectively. Interestingly, the expansion of T_{RM} cells was mostly self-dependent, as the differentiation rate from T_{EML} into T_{RM} was very low (see Table 4.1),

[4]The lower limit of 3 hours for the doubling time of T_{EML} stems from the fact that all proliferation rates were restricted to values smaller than 5 day^{-1} (see section 4.2.3).

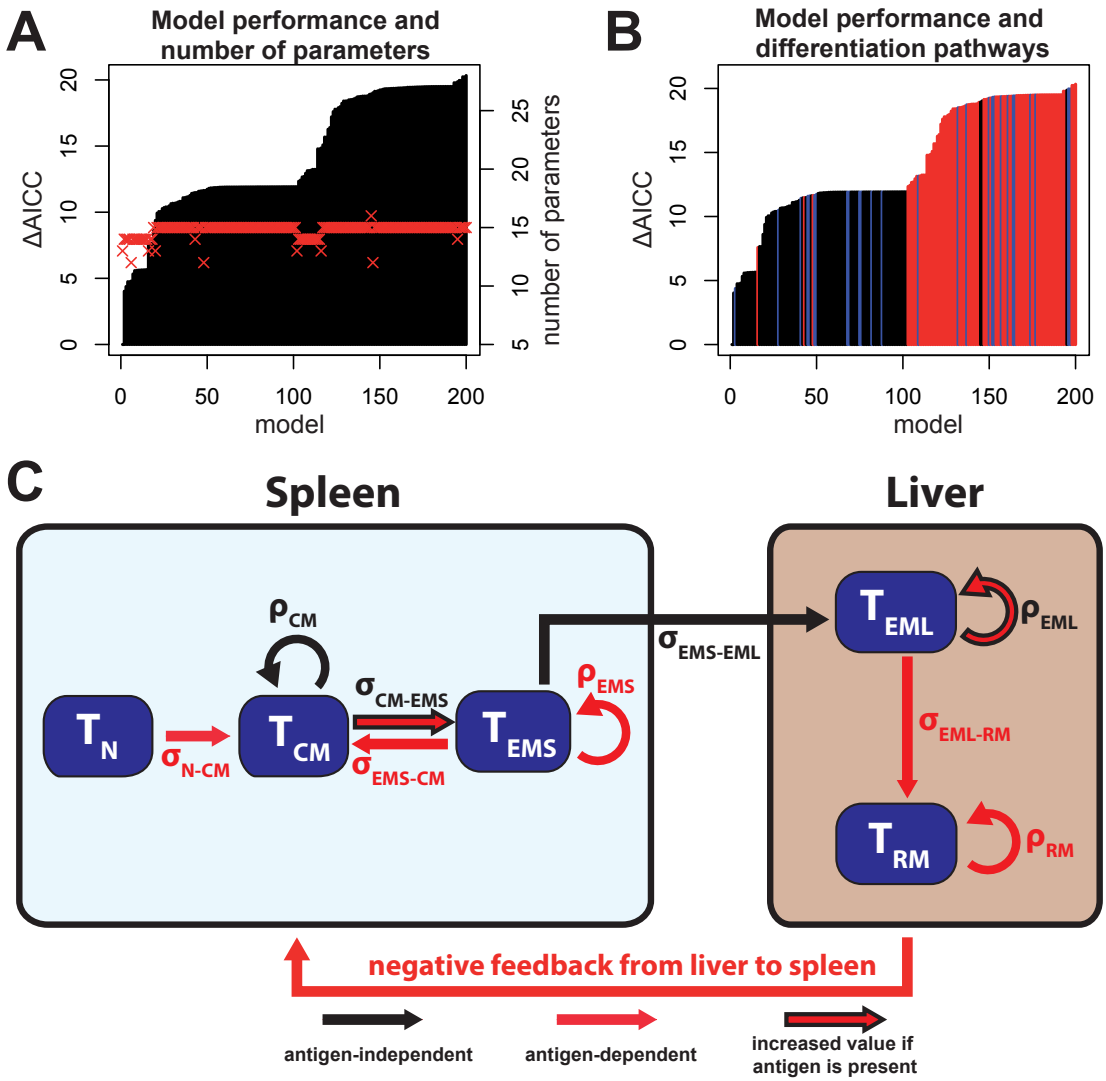

Figure 4.4: Model ranking and the pathway of the best model: (**A**) The ΔAICc (black vertical lines, left axis) and the number of parameters (red crosses, right axis) for the 200 best model fits. (**B**) The ΔAICc for the 200 best model fits, with colour indicating the type of naïve T cell differentiation pathway: $T_N \rightarrow T_{CM}$ (black), $T_N \rightarrow T_{EMS}$ (red) or $T_N \rightarrow T_{CM}$ /T_{EMS} (blue). (**C**) The CD8$^+$ T cell differentiation pathway according to the best model. Coloured arrows indicate if a rate is active only if antigen is present (red) or also independently of antigen exposure (black - no change during antigen presence, red with black outline - increased value if antigen is present)

meaning that T_{EML} were only needed as an initial supply compartment. The antigen-dependent doubling time of T_{RM} was estimated to be between 11 and 42 hours.

Additionally, we found that the duration of antigen presence after the first and second booster injections, L_2 and L_3, were estimated to be only around 0.5 to 3 days, which is much shorter than the initial expansion phase after prime, which lasted for 8 days (initially fixed value, see

parameter	unit	confidence interval
$\sigma^{\mathrm{ag}}_{\mathrm{N-CM}}$	day^{-1}	non-identifiable
$\rho^{\mathrm{b}}_{\mathrm{CM}}$	day^{-1}	[0.105;0.34]
$\sigma^{\mathrm{ag}}_{\mathrm{CM-EMS}}$	day^{-1}	[0.6;∞)
$\sigma^{\mathrm{b}}_{\mathrm{CM-EMS}}$	day^{-1}	[0.11;0.335]
$\sigma^{\mathrm{ag}}_{\mathrm{EMS-CM}}$	day^{-1}	[0.03;∞)
$\rho^{\mathrm{ag}}_{\mathrm{EMS}}$	day^{-1}	[1.6;3.9]
$\sigma^{\mathrm{b}}_{\mathrm{EMS-EML}}$	day^{-1}	[0.16;0.42]
$\rho^{\mathrm{ag}}_{\mathrm{EML}}$	day^{-1}	[2.8;5]
$\rho^{\mathrm{b}}_{\mathrm{EML}}$	day^{-1}	[-2.6;-1.08]
$\sigma^{\mathrm{ag}}_{\mathrm{EML-RM}}$	day^{-1}	[1e-5,0.068]
$\rho^{\mathrm{ag}}_{\mathrm{RM}}$	day^{-1}	[0.4;1.5]
L_2	days	[0.7;3.15]
L_3	days	[0.55;2.7]

Table 4.1: The confidence intervals for the parameters of the best model: Confidence intervals were calculated based on a profile likelihood approach (see Materials and Methods 2.3.3 and Appendix B.2). Here, red colour indicates antigen-dependent and grey colour baseline parameters.

section 4.2.1). These findings would suggest that the immune system is able to handle subsequent infections much more efficiently.

While the model predictions based on the best parameter sets were in good agreement with the experimental data, a closer look revealed that predicted $\mathrm{T_{EML}}$ cell numbers during the first expansion phase after prime were unreasonably high. Here, cell numbers could reach values of 10^9 and more (see Fig. 4.5 **A** and Appendix B.1). Since the mouse liver is estimated to contain only around 2×10^8 hepatocytes [Sohlenius-Sternbeck 2006], these model predictions are beyond any biological plausibility.

Looking for a reason behind this massive cellular expansion, we found a correlation between the $\mathrm{T_{EML}}$ peak cell numbers and the antigen-dependent differentiation from $\mathrm{T_{EML}}$ into $\mathrm{T_{RM}}$, $\sigma^{\mathrm{ag}}_{\mathrm{EML-RM}}$ (see Fig. 4.5 **B**). A low rate corresponds to an accumulation of $\mathrm{T_{EML}}$ in the liver as they differentiate only slowly into $\mathrm{T_{RM}}$. Since no data from the acute phase of infection was available to resolve this problem, we based our following analyses about the dose-dependency on ten different parameter sets that predicted $\mathrm{T_{EML}}$ cell numbers lower than 5×10^7 (see Appendix B.2).

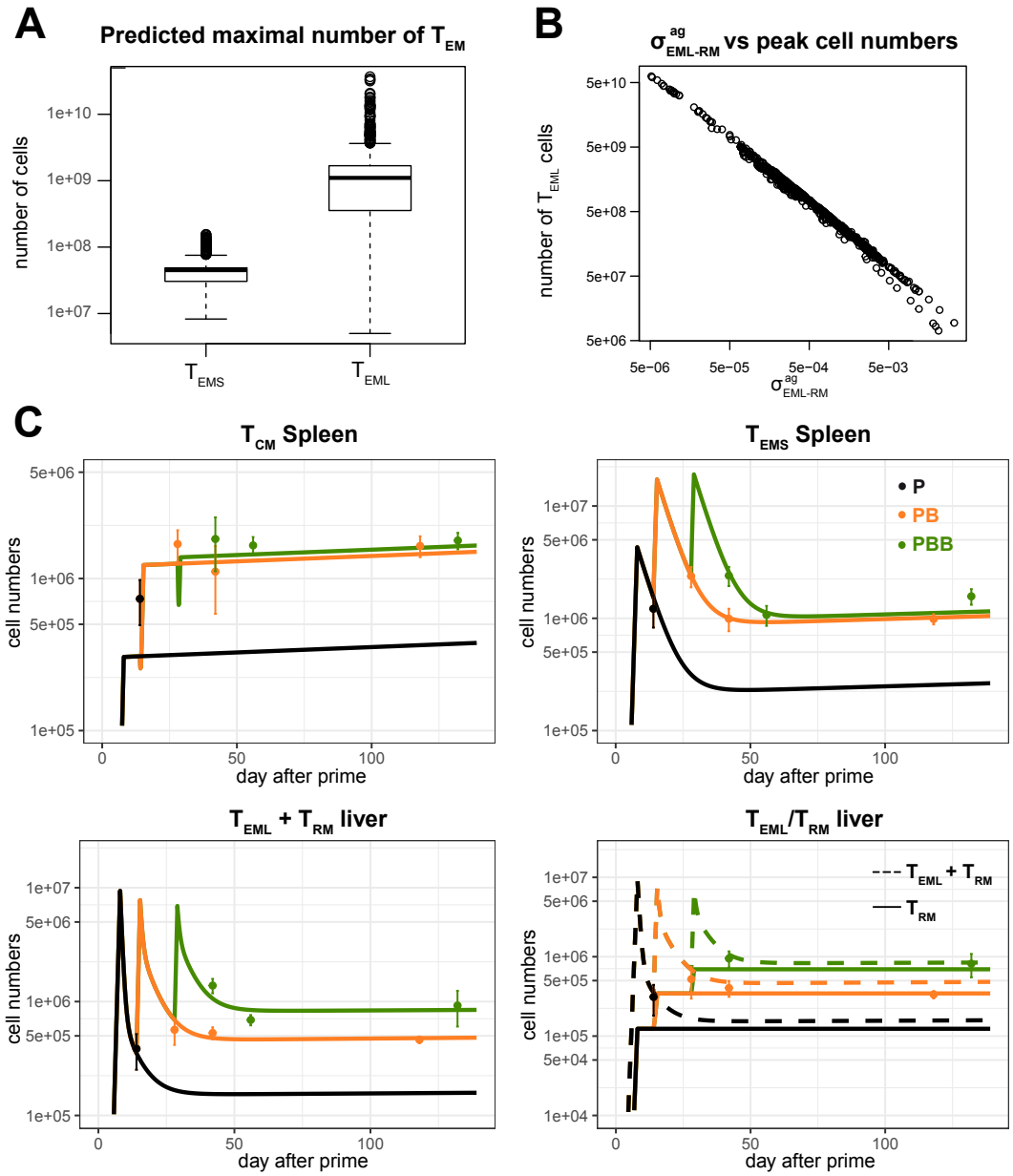

Figure 4.5: Performance of model predictions: (**A**) Boxplots show the distribution of the maximal number of T_{EMS} and T_{EML} cells derived from simulations based on the 800 best parameter sets obtained by the profile likelihood approach. (**B**) Correlation between the antigen-dependent differentiation rate from T_{EML} into T_{RM}, σ^{ag}_{EML-RM}, and the T_{EML} cell numbers at peak. (**C**) Panels show the cellular dynamics of T_{CM} (middle left), T_{EMS} (middle right), T_{EML} + T_{RM} (bottom left) and T_{EML} /T_{RM} compartments (bottom right). Coloured lines represent the simulated time-courses based either on a prime (P, black), prime-boost (PB, orange) or an prime-boost-boost (PBB, green) vaccination strategy. Model parameters are chosen according to set 10 in Appendix B.2. Dots represent the means of the measured T cell numbers based on the respective normal-dose experiments. Error bars indicate standard error of the mean. All experiments were performed by the lab of Prof. Dr. A.-K. Mueller, Centre for Infectious Diseases, Parasitology Unit, University Hospital Heidelberg.

4.3 The influence of the vaccination dose on cellular dynamics

In addition to the normal dose (1×10^4 PbRAS), experiments were conducted with mice receiving a high (1×10^5 PbRAS) or a subprotective dose (1×10^3 PbRAS, see Fig. 2.2). Assuming these experimental data sets could be described based on slight modifications to the previously obtained parameter sets[5] for the normal dose, we generated new models in which a dose dependent factor, D was affecting different parameter combinations. In these models, a specified subset of the normal-dose parameter set was modified by multiplying the respective parameter values with the factor D, thereby increasing or decreasing the corresponding rates. Here, we assumed that the dose factor can only influence antigen-dependent rates. Our best model for the normal dose contains nine antigen-dependent rates, therefore $2^9-1=511$ different parameter combinations were tested. For all of these possibilities, we determined the corresponding dose-dependent factor and AICc by fitting them separately to the high or the subprotective dose[6].

4.3.1 The effect of a high dose on cellular dynamics

The biological experiments showed that a prime-boost high dose vaccination approach resulted in 100% short-term protection (see Table 2.1). Additionally, in a corresponding challenge 104 days after the last booster injection, only one out of eight mice developed symptoms of malaria, indicating that this vaccination approach can also confer long-term protection. According to a recent study, this superior protection can be attributed to high levels of T_{RM} cells in the liver [Fernandez-Ruiz et al. 2016], an observation we also found in our experimental data.

Our modelling approach suggest that these elevated T_{RM} responses are a result of a stronger differentiation stimulus as well as an increased antigen-dependent proliferation of T_{RM} cells (see Table 4.2). We estimate that the administration of a high dose increases the values of the affected rates by approximately $15-20\%$ compared to the rates describing the dynamics for the normal dose. Additionally to the increased differentiation and proliferation, our model suggests that antigen-dependent dynamics are prolonged after the second shot, which results in even higher levels of T_{RM} cells.

The best fit to the high dose data is shown in Appendix B.3.

4.3.2 The effect of a subprotective dose on the cellular dynamics

In our experiments, vaccination approaches based on subprotective dosing were not sufficient to generate protection and all mice challenged with infectious parasites developed symptoms of malaria (see Table 2.1).

Here, mathematical modelling revealed that the most likely explanation for the weak response stems from a drastically reduced differentiation from T_{CM} into T_{EMS} cells in the spleen (see Table

[5] Since no best parameter set could be determined for the normal dose, we applied the dose-dependent fitting procedure to ten parameter sets that had T_{EML} cell numbers lower than 5×10^7 - see section 4.2.4

[6] Since we tested all of these models for the ten different parameter combinations given in Appendix B.2, over 5000 models were fitted in total to each data set.

-2LL	dose factor	σ^{ag}_{N-CM}	σ^{ag}_{CM-EMS}	σ^{ag}_{EML-RM}	ρ^{ag}_{RM}	L2	set
50.0	1.19	X	X	-	X	X	2
50.3	1.19	X	X	-	X	X	3
50.7	1.23	-	X	-	X	X	9
51.8	1.17	X	X	X	X	X	7
52.2	1.16	X	X	X	X	X	4
52.2	1.16	X	X	X	X	X	10
52.9	1.17	X	X	X	X	X	8
53.8	1.16	X	X	X	X	X	5
54.3	1.15	X	X	X	X	X	6
55.0	1.18	X	X	-	X	X	1

Table 4.2: Each of the ten parameter sets in given in Appendix B.2 was used as a baseline dynamics to fit the high dose data. The table shows which dose factor combination resulted in the best fit for the respective parameter sets ('X' means that the parameter is affected, '-' means not affected, only parameters that were affected at least once are shown) and the corresponding dose factor. Results are ordered by negative log-likelihood.

4.3). Our model estimates a decrease of up to $60 - 80\%$ compared to the turnover found for the normal dose. Biologically, this would correspond to an arrest of CD8$^+$ T cell differentiation in the T_{CM} stage, a feature that has also been observed in other experiments [Redeker et al. 2014; Zehn et al. 2009]. Additionally, our model predicts that the accumulation of T_{RM} cells is also massively hampered, as all of the ten model fits seen in Table 4.3 show either a decreased differentiation from T_{EML} into T_{RM} cells or an impaired proliferation of the T_{RM} compartment.

The best fit to the subprotective dose data is shown in Appendix B.4.

4.3.3 Correlating model predictions with protection

Comparing the cellular dynamics in the liver for the different dosages based on a prime-boost-boost approach, reveals that the T_{RM} pool generation is remarkably inefficient for the subprotective dose, as the predicted levels after three subprotective shots did not even reach the level observed after administering one normal dose (see Fig. 4.3 **A** and Appendix B.5). To reach the presumed 100% protection level (which is found after three shots based on a normal dose or two shots based on a high dose), our model predicts that 10-11 shots based on a subprotective dosage would be necessary (see Fig. 4.3 **B**). However, here we assumed that the duration of antigen presence for the following booster injections is the same as after the third shot. As this assumption leads to questionable dynamics of the other cell compartments (see Appendix B.5), a further reduction in the duration of antigen presence (meaning $L_1 > L_2 > L_3 > L_4 > ...$) or a change in the antigen-dependent parameters is likely to affect the cellular dynamics as well. Given the present experimental data, it is not possible to determine these factors. Thus, it re-

-2LL	dose factor	σ^{ag}_{CM-EMS}	σ^{ag}_{EML-RM}	ρ^{ag}_{RM}	set
53.5	0.17	X	-	X	2
53.8	0.18	X	-	X	3
55.3	0.19	X	-	X	7
56.4	0.22	X	X	-	8
56.6	0.22	X	-	X	10
57.6	0.22	X	X	-	5
57.9	0.22	X	X	-	1
58.9	0.30	X	X	-	4
62.3	0.31	X	X	-	6
66.9	0.4	X	X	-	9

Table 4.3: Each of the ten parameter sets given in Appendix B.2 was used as a baseline dynamics to fit the subprotective dose data. The table shows which dose factor combination resulted in the best fit for the respective parameter sets ('X' means that the parameter is affected, '-' means not affected, only parameters that were affected at least once are shown) and the corresponding dose factor. Results are ordered by negative log-likelihood.

mains an open question if a sufficient number of subprotective shots can generate any immunity at all.

4.4 Discussion

Over the last decades, there has been a lot of research on the topic of T cell-based vaccines for malaria and the obtained results seem very promising so far. While there are still some major milestones along the way, studies have shown proof of concept and the goals needed to achieve for efficient vaccine creation are more or less clear. What is currently missing is a better understanding of the effects of different vaccination approaches on the both short-term and long-term cellular dynamics, which would allow to optimise existing vaccination protocols to induce specifically tailored immune responses.

Our present study helps to shed light on this topic in several different ways. First, our modelling approach revealed the underlying cellular differentiation and migration pathway, connecting the spleen as the site of the initial priming with the liver serving as the main site of infection. As protection against malaria is most likely mediated by the local accumulation of high levels of hepatic T_{RM} cells [Fernandez-Ruiz et al. 2016], the generation of this compartment is of high interest. Our model suggests that the original founder population of T_{RM} cells in the liver stem from hepatic T_{EM} cells, which are supplied by splenic emigrants. However, after establishment of an initial T_{RM} population, the increase in numbers following booster injections is mostly due to T_{RM} proliferation, meaning this compartment is more or less self-sustaining,

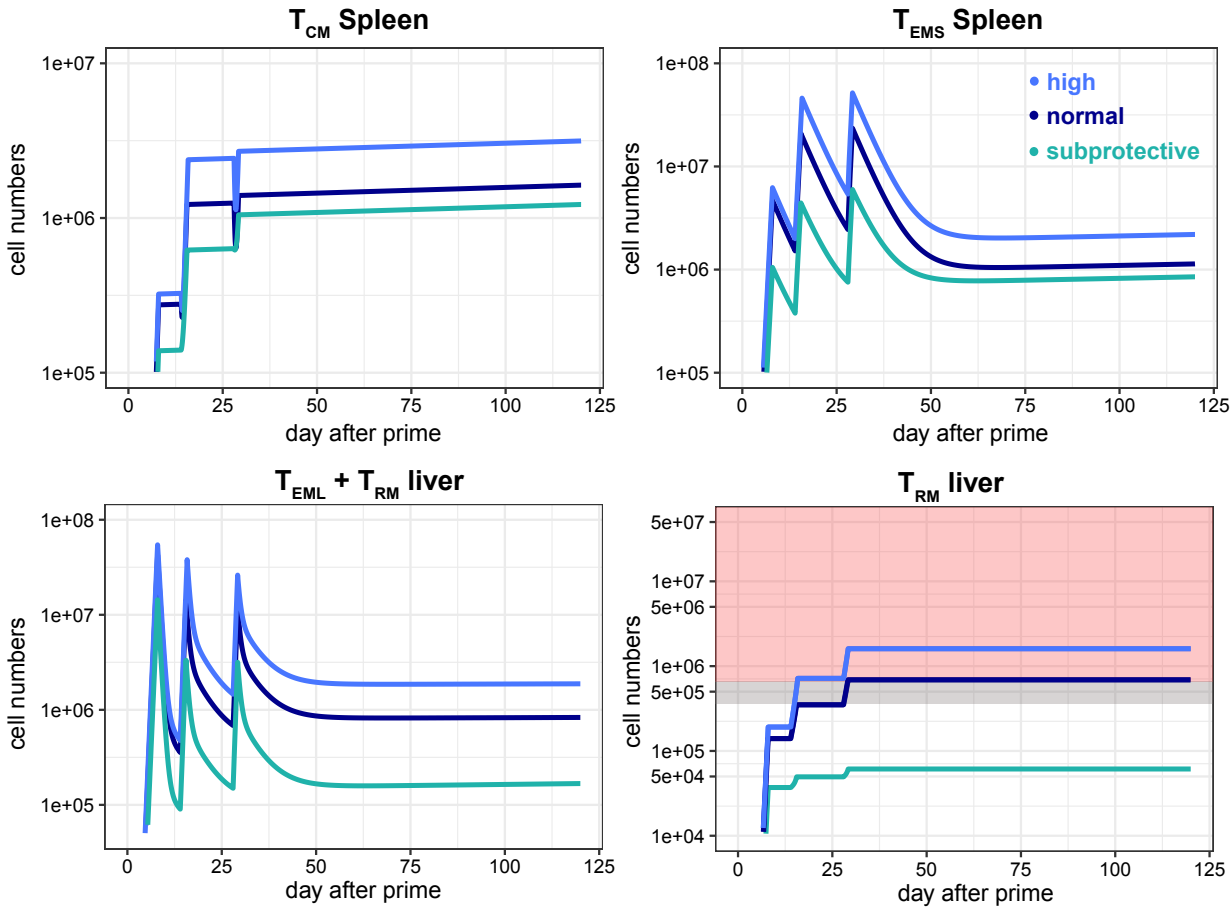

Figure 4.6: Comparing the cellular dynamics based on different vaccination doses: Panels show the cellular dynamics of T_{CM} (top left), T_{EMS} (top right), T_{EML} + T_{RM} (bottom left) and T_{RM} compartments (bottom right) based on a normal dose (dark blue), high dose (light blue) or subprotective dose prime-boost-boost vaccination approach (turquoise). Shaded areas in the T_{RM} plot (bottom right) indicate the regions of 50% (grey) and 100% protection (red) as found in the experimental challenge data (see Materials and Methods 2.1).

therefore diminishing the contribution of the spleen to hepatic T_{RM} levels in following booster injections.

The mathematical modelling of the experimental data revealed an almost linear differentiation pathway, following roughly a $T_N \rightarrow T_{CM} \rightarrow T_{EMS} \rightarrow T_{EML} \rightarrow T_{RM}$ pathway. A similar differentiation pathway has been proposed in other studies [Kohler 2007; Buchholz et al. 2013; Farber et al. 2014]. However, we also find conversion from T_{EMS} to T_{CM} in the spleen, a feature that has been suggested in many publications analysing cellular pathways [Wherry et al. 2003; Akondy et al. 2017; Crauste et al. 2017], however without the preceding T_{CM} stage. Since CD8$^+$ T cell subset characterisation in our study is mostly based on the expression of CD62L, the predicted conversion might simply characterise an upregulation of this surface marker. Also, while our modelling is strongly supporting a $T_N \rightarrow T_{CM}$ pathway (see Fig. 4.4 **B**), we don't want to exclude the possibility of additional $T_N \rightarrow T_{EMS}$ conversion.

In addition to determining the cellular differentiation pathway, our modelling approach also allowed for the quantification of the parameters describing the cellular turnover. Most parameters turned out to be identifiable, especially the baseline parameters determining the long-term dynamics and the antigen-dependent proliferation rate of T_{RM} cells (see Table 4.1 and Appendix B.2). However, we encountered several problems regarding the estimation of the antigen-dependent parameters. These shortcomings are due to a lack of expansion phase data, which allows our model to predict different (including unreasonably high) numbers of T_{EM} cells over the course of acute infections. While we identified the critical parameters which predict the peak amount of T_{EM} cells, new data is necessary to correctly determine the antigen-dependent cellular turnover in the days after prime and booster injections.

Furthermore, we analysed how the generation of the different T cell pools is affected by the vaccination dose. We found that both the high and the subprotective dose data can be mostly explained by a change in the differentiation strength from T_{CM} to T_{EMS} in the spleen. For the high dose, our model suggests that more T_{CM} cells are stimulated to differentiate, therefore increasing the amount of cells that eventually migrate to the liver. On the other hand T_{CM} seem to arrest in their state if a subprotective dose is administered, most likely due to the absence of immunological stimuli, such as antigen presentation or inflammation markers. This arrest of T_{CM} due to weak stimuli is in agreement with previous findings [Redeker et al. 2014; Zehn et al. 2009]. The importance of local antigenic stimulation for the generation of the T_{RM} pool is further supported by the prediction of an increased turnover of T_{EML} into T_{RM} in the liver for the high dose and a decreased turnover for the subprotective dose.

On a more technical note, we implemented an algorithm designed for the unbiased selection of the most appropriate model to describe the experimental data out of a multitude of possible models. Although similar algorithms do exist (e.g. found in `R` packages *glmulti*, *MuMIn* or *leaps*), they are either based on performing regression analysis or test all models of a given (sub-)set. Since our modelling approach was based on ODEs and the number of possible models ranged in the order of several millions, none of the available approaches suited our needs. Therefore, we created a model selection algorithm, which searches the model space in a more efficient way. Evaluated in the present context of modelling the immune response to malaria vaccination, our unbiased model selection algorithm assessed only about 2000 (out of several million possible) models, before returning appropriate results. Additionally, due to the flexibility our own model selection approach provided and the computational resources available to us, our algorithm could test up to 200 models simultaneously, meaning results could be obtained much faster than with any of the other available model selection algorithms.

While our current model of $CD8^+$ T cell differentiation is only a first step towards the set-up of a complete model of the $CD8^+$ T cell responses induced by malaria vaccination, it is very useful in determining potential ambiguities that need to be addressed experimentally. For example, we find that the antigen-dependent parameters cannot be properly quantified unless data from the acute infection phases are gathered. Also, more data is needed to estimate the stability of the T_{RM} pool, which is crucial to determine the generated protection over longer periods of time. Another important aspect is the general role of the T_{CM} compartment in the spleen: While there is no significant difference with respect to the tested vaccination approaches (including

mock, see Fig. 2.3), the measurements of the cell numbers show a high inter-individual variation and are therefore possibly masking any changes that might occur in individuals over time. For future experiments, the use of cell labelling (see chapter 5) might help to quantify the turnover observed in this cell pool, identify its maintenance over time as well as correlate the observed T_{CM} numbers to other compartments, especially the levels of T_{RM} observed in the liver.

Another important point that needs to be addressed experimentally is the malaria specificity of the vaccination-induced CD8$^+$ T cell pool. In our current approach, the number of malaria-specific cells was determined by normalising the measured cell numbers with numbers obtained from naïve mice. An additional experiment, testing the IFN-γ production after stimulation with an immunodominant antigen [Hafalla et al. 2013; Doll et al. 2016], revealed that around $5-15\%$ of the overall response was specific for this epitope, but showed no saturation in either spleen or liver. As saturation was found for the overall CD8$^+$ T cell response, this could either indicate that the pathogen-specific response is masked by CD8$^+$ T cells of other specificities or that the analysed epitope becomes more immunodominant.

In addition to the data that is needed to clarify model and experimental ambiguities, new experiments are needed to test the predictions generated by our current modelling approach, which is of crucial importance to confirm or reject certain model assumptions. Of special interest would be biological experiments with more booster injections given the normal, high and subprotective doses, as they might help to elicit saturation levels in spleen and liver, which are not incorporated in our current modelling approach (see Fig. B.5). Also, experiments with intermediate doses between subprotective/normal or normal/high are needed to find a continuous mathematical function describing the dose-dependent influence on the cellular dynamics, which was not possible with the current data set.

In summary, our results allow us to determine the CD8$^+$ T cell dynamics generated after malaria immunisation and can be used to improve future vaccination approaches.

CHAPTER 5

On the influence of labelling strategies to infer cellular dynamics

The results presented in this chapter have been partially published in [Gabel et al. 2017].

5.1 Introduction

The ability to distinguish cells by certain markers and labels has been an indispensable asset in many biological experiments addressing cellular dynamics and development, including the analysis of $CD8^+$ T cell dynamics (see section 1.3.2). Tracking differently labelled cells not only allows for the identification of lineage pathways [Perie et al. 2014], but also the observation of dynamical changes in cell populations over time [Gerlach et al. 2013]. In addition, the information obtained by labelling can be used to quantify cellular turnover, such as cell activation, proliferation, migration and differentiation dynamics [De Boer et al. 2013; Kaiser et al. 2013; Ganusov et al. 2014].

5.1.1 The application and use of different labelling approaches

There exists a large variety of experimental techniques to label and track individual cell populations. For example, staining cells by a single marker, such as BrDU [Tough et al. 1994; Mohri et al. 1998], the fluorescent dye CFSE [Lyons 2000; Yates et al. 2007], or deuterated glucose and heavy water [Hellerstein et al. 1999; Ribeiro et al. 2002; Mohri et al. 2001], can be used to infer cellular turnover and proliferation dynamics. A more fine-grained approach involving several different markers - e.g. by transferring cell populations bearing congenic markers [Shen et al. 1985; Kearney et al. 1994; Buchholz et al. 2013] or using naturally occurring diverse markers, such as T cell receptor diversity [Maryanski et al. 1996; Lin et al. 1998; Turner et al. 2003; Blattman et al. 2000] - allows to distinguish the dynamics of individual subpopulations. Furthermore, labelling cells by unique, inheritable genetic barcodes makes it possible to follow cellular dynamics on a single cell level and therefore allows to address questions regarding cell heterogeneity and individual cell differentiation pathways [Gerlach et al. 2013; Schumacher et al. 2010; Schepers et al. 2008; Naik et al. 2014].

Labelling cell populations is particularly useful, if the experimental set-up forbids the sampling of longitudinal data, e.g. due to harvesting of organs or cell cultures. In these cases the intra-

individual variability derived from the labelled cell populations can provide enough information to estimate cellular turnover, even if the data was only sampled at a single time point [Kaiser et al. 2013; Gerlach et al. 2013]. Interestingly, it is possible to quantify interacting dynamics, such as linked migration and proliferation, even if not all of the involved cell compartments are observable. Therefore, using multiple labels can compensate for both the lack of time-resolved and missing cellular data; shortcomings with are frequently encountered when analysing immune responses.

5.1.2 Appropriateness of labelling strategies for parameter estimation

Data from labelling experiments have been successfully used to quantify cell population dynamics, including those of $CD8^+$ T cells. The spectrum of the applied labelling approaches in these experiments is astonishingly broad: Strategies using hundreds of uniquely labelled cells [Gerlach et al. 2013] are found alongside approaches having only a handful of markers, which are shared between a large number of cells [Kaiser et al. 2013; Buchholz et al. 2013]. This huge variety raises the interesting question if all labelling strategies are equally well suited to capture the underlying cellular dynamics and hence generate appropriate data for analysis. So far, it has not been determined how these different labelling approaches affect the quantification of cellular dynamics: To what extent would parameter estimation improve if more labels or more cells per label are used? Are all labelling strategies able to capture different kinetics, such as cell proliferation and differentiation, equally well? How robust are labelling approaches with respect to experimental shortcomings such as incomplete transfer or sampling? Answers to these questions are necessary to establish a guideline for the generation of reliable and robust labelling strategies with respect to parameter estimation.

5.1.3 Chapter overview

To address these raised questions, we analysed the appropriateness of different labelling strategies to quantify cellular dynamics by performing a simulation study. Our approaches are based on actual cellular dynamics, and feature a) homoeostatic turnover, as found e.g. in the maintenance of naïve or memory $CD8^+$ T cells, and b) cellular expansion dynamics, mimicking the activation and proliferation of $CD8^+$ T cells during acute infections (see Fig. 5.1 **A** and section 5.2.1). Here, we simulated adoptive transfer experiments based on differently labelled cell populations (see Fig. 5.1 **B**) and used the data derived from stochastically generated cell populations to quantify the model parameters (see section 5.2.1). This approach allowed us to evaluate the impact of different labelling strategies and experimental limitations on parameter estimation.

This chapter is structured as follows: In section 5.2, we establish the mathematical models (section 5.2.1) and the simulation methods (section 5.2.2) used throughout the rest of this chapter.

In section 5.3 we systematically analyse the influence of experimental labelling factors on the estimation of cellular turnover. These factors include the number of labels and the label size (section 5.3.1) as well as the time point of sampling (section 5.3.2). In section 5.4, we determine how experimental limitations affect the parameter estimations. Here, we look at the influence

of incomplete transfer (section 5.4.1), incomplete sampling (section 5.4.2) and missing cellular compartments (section 5.4.3).

In general, our analysis can be used as a guideline for generating suitable labelling approaches that will allow to quantify cellular dynamics even more robustly in future experiments.

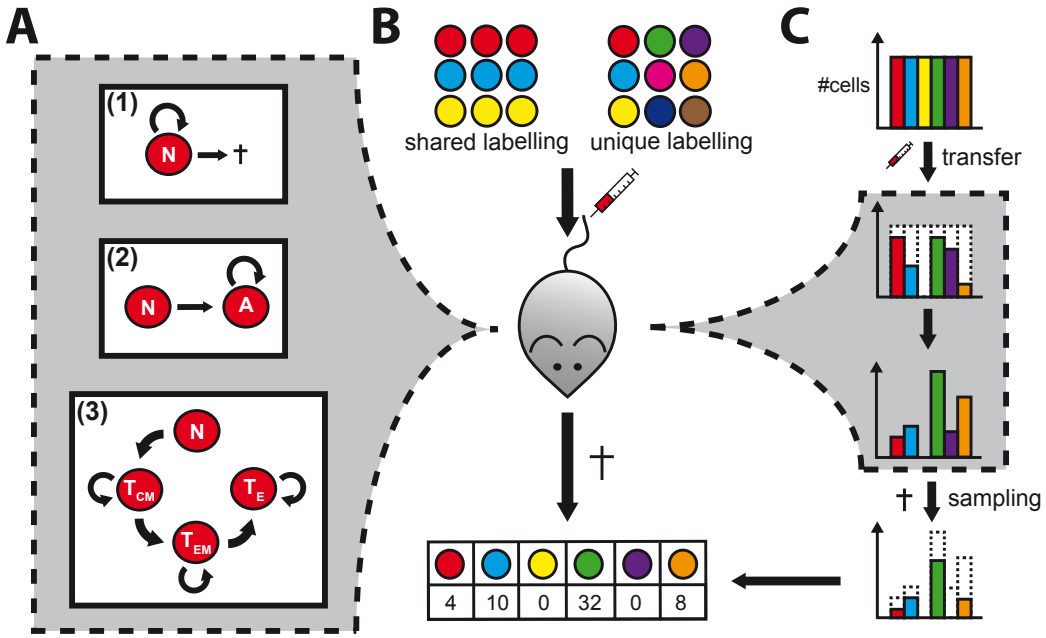

Figure 5.1: Sketch of the cellular dynamics, labelling set-ups and experimental shortcomings: (**A**) The sketches show the three simulated cellular dynamics: (**1**) Homoeostatic turnover: Naïve cells proliferate only to compensate cell death, therefore maintaining a stable number of cells. (**2**) Simple expansion dynamics: By encountering their respective antigen naïve cells are activated and start to proliferate extensively. (**3**) Complex expansion dynamics: Upon activation, naïve cells turn first into central memory precursor cells (T_{CM}) and subsequently differentiate further into effector memory precursor (T_{EM}) and finally effector cells (T_E). All compartments, besides the naïve cells, are proliferating. (**B**) A labelling strategy is defined by the number of different labels and the label size i.e. the number of cells per label. The depicted labelling strategies show a shared and a unique labelling approach. After transfer into a host, these cells are thought to follow one of the three cellular dynamics: homoeostatic turnover, simple or complex expansion dynamics. The label of a cell is passed onto all offspring. At a specific time, cells are sampled and used for evaluation. Data are gathered in the form of *count* data measuring the number of cells of a specific label within the sampled population. (**C**) Possible experimental shortcomings: Cells of the initial labelling strategy can be partially lost during adoptive transfer (incomplete transfer). The transferred fraction of cells subsequently expands according to the underlying dynamical model. Cells can again be lost when sampling (incomplete sampling). Figure adapted from [Gabel et al. 2017].

5.2 Mathematical models and simulation approaches

5.2.1 The mathematical models of cellular dynamics

Here, we distinguish three different scenarios of cellular dynamics, which are based on - but not limited to - $CD8^+$ T cell dynamics. The corresponding models are described as follows.

(1) Homoeostatic turnover: A cell population is considered to be in equilibrium if the total number of cells is more or less constant over time. However, this doesn't imply that there is no cellular turnover, as cells can still be lost due to cell death, differentiation into other types of cells or migration into different compartments or organs. To compensate for this efflux, lost cells are replaced over time, e.g. by remaining proliferating cells or by an influx from elsewhere. Examples of homoeostatic turnover are the dynamics of naïve $CD8^+$ T cells before antigen encounter, or the pool of memory T cells that is maintained after an infection [De Boer et al. 2003]. In the homoeostatic model, we assume that a cell population, here termed naïve cells, N, proliferates with rate ρ and dies with rate δ. The cellular dynamics are described by the following differential equation:

$$\frac{dN}{dt} = (\rho - \delta)N \ . \tag{5.1}$$

To ensure homoeostatic turnover, we always set $\rho = \delta$ in our stochastic simulations. Unless specified otherwise, we assume that $\rho = \delta = 0.5$ day^{-1}. While the actual homoeostatic turnover of naïve $CD8^+$ T cells is much lower [De Boer et al. 2013], this value was chosen to allow for reasonable simulation times. However, as the parameter estimation quality is not affected by moderate changes in the sampling time (see Appendix C.1), all results can be derived without loss of generality.

(2) Simple expansion dynamics: The second analysed dynamics describes the activation and subsequent proliferation of $CD8^+$ T cells due to antigen exposure (see Fig. 5.1 **A**) . After encountering their cognate antigen, naïve T cells are activated and start to fight effectively against the invading pathogen by massively expanding in numbers and simultaneously differentiating into effective subpopulations [Buchholz et al. 2013]. Here, to model a simple expansion dynamics, we distinguish only between naïve, N, and activated T cells, A (based on the notation in [De Boer et al. 2001]). Naïve cells are activated with rate μ, and activated cells start to proliferate with rate ρ. The dynamics of this model can be described by the following system of ordinary differential equations:

$$\begin{aligned} \frac{dN}{dt} &= -\mu N \\ \frac{dA}{dt} &= \mu N + \rho A \end{aligned} \tag{5.2}$$

For simplicity, cell death is neglected in this model. Unless specified otherwise, the rates are always set to $\mu = \rho = 0.3$ day^{-1}.

(3) Complex expansion dynamics: As the compartment of activated $CD8^+$ T cells in the simple expansion model is actually heterogeneous and comprised of functionally diverse subsets (see section 1.3), we set up a third model featuring different types of activated cells. Based on the analysis performed in [Buchholz et al. 2013], we distinguish between central memory precursor (T_{CM}), effector memory precursor (T_{EM}) and effector cells (T_E). The relationship between these compartments is given by the following linear differentiation pathway (see also Fig. 5.1 **A**):

$$\begin{aligned} \frac{dN}{dt} &= -\mu_N N \\ \frac{dT_{CM}}{dt} &= \mu_N N + (\rho_{CM} - \mu_{CM})T_{CM} \\ \frac{dT_{EM}}{dt} &= \mu_{CM} T_{CM} + (\rho_{EM} - \mu_{EM})T_{EM} \\ \frac{dT_E}{dt} &= \mu_{EM} T_{EM} + \rho_E T_E \ , \end{aligned} \tag{5.3}$$

where μ_x and ρ_x describe the differentiation and proliferation rates of the corresponding compartments. We parametrised the model by using the estimates derived from [Buchholz et al. 2013]; the respective values are given as $\mu_N = 2.2$ day^{-1}, $\mu_{CM} = 0.2$ day^{-1}, $\mu_{EM} = 0.04$ day^{-1}, $\rho_{CM} = 0.85$ day^{-1}, $\rho_{EM} = 1.42$ day^{-1} and $\rho_E = 1.6$ day^{-1}.

5.2.2 Simulating labelling experiments

To generate simulated data for our analysis, we performed stochastic simulations of the systems defined by Eqs (5.1)-(5.3) based on the Gillespie algorithm [Gillespie et al. 2006]. Simulations were carried out in `R` using the package *adaptivetau* [R Core Team 2016]. In each simulation, we start with a specified number of differntly labelled naïve cells at time $t = 0$. Over the course of each run, these cells proliferate, differentiate or die in a stochastic manner, according to the underlying model. The label of each individual cell is retained when activating or differentiating and is passed onto the daughter cells when proliferating. At sampling time $T > 0$ the simulation is halted and the number of cells for each label is assessed.

In addition to the model parameters describing the cellular dynamics, each simulation depends on the sampling time, T, at which cells are sampled, and the labelling strategy, which is characterised by the number of different labels, L, and the label size, M. Unless stated otherwise, each simulated labelling strategy is uniformly distributed, i.e. every label has the same number of cells initially.

In case of incomplete transfer we account for the loss of transferred cells by first randomly sampling a fraction of the original labelling strategy and then using this sample as an initial condition in our simulation. The sampling is performed by the built-in `R` function *sample*. Similarly, by sampling a predefined fraction of cells from the stochastically generated simulation output, we can reproduce the experimental limitation known as incomplete sampling. The sampled cell populations[1] are then used to estimate the parameters of the underlying system as described in Materials and Methods 2.4.1.

[1] Given the simple or complex expansion model, the compartment of naïve cells was not used for fitting.

5.3 The influence of experimental factors on parameter identification

In the following section, we look in detail how different experimental factors affect the estimation of model parameters. The statistical quantities used for assessing the appropriateness of labelling strategies are defined in Material and Methods 2.4.2. While the results presented in this section are based on the simple expansion model (see Eqs. 5.2), identical observations can be found for the homoeostatic and complex expansion model (see Appendix C).

5.3.1 The influence of the labelling strategy

Assuming labelling strategies are uniformly distributed, i.e. every label has the same number of cells initially, the applied labelling approaches can be characterised by the following two parameters: The number of labels, L, and the label size, M, i.e the number of cells per label. To quantify the influence of the labelling strategy on the parameter estimation quality, we generated a total of 2450 different labelling approaches by varying the number of labels, L, from 2 to 50 and the number of cells per label, M, from 1 to 50, and subsequently analysed their performance to correctly estimate the underlying cellular dynamics.

We find that increasing the number of labels, L, continuously improves the estimation quality for both the activation, μ and the proliferation rate, ρ of the simple expansion model. Here, an increase reduces the absolute bias, the mean confidence interval length (MCIL) as well as the false coverage rate (FCR, see Fig. 5.2 **C**). On the other hand, increasing the number of cells per label, M initially results in a pronounced decline in these quantities for strategies using a small number of cell per label, but this effect saturates rather quickly (see Fig. 5.2 **D**). In our example, this saturation occurs around $M = 10$ cell per label, indicating that medium label sizes might already be sufficient to robustly estimate cellular turnover. However, using uniquely labelled cells can still help to identify the influence of experimental limitations such as incomplete transfer (see section 5.4).

In summary, our results argue for the use of a large number of different labels with medium to large numbers of cells per label to reliably infer cellular turnover.

Influence of the distribution of labels Assuming there is only a limited number of cells, which can be labelled, we wondered how the initial distribution of label sizes would affect the robustness of the parameter estimation.

To this end, we repeated our analyses by using a fixed total number of cells that were distributed amongst $L = 50$ different markers using either a uniformly, linearly or exponentially distributed label size (see Fig. 5.3 **A**). The evaluation of data from non-uniformly distributed label sizes required the adaptation of the calculations of the corresponding summary statistics (see Appendix C.4). To compare the performance of the different labelling strategies, we then calculated the difference between the statistical quantities (e.g. $\Delta_{MCIL} = \text{MCIL}_{\textbf{Uniform}} - \text{MCIL}_{\textbf{Linear}}$).

We find that of all three distributions, a uniformly distributed labelling strategy yields the most reliable results, irrespective of the total number of cells transferred (Fig. 5.3 **B-E**, results

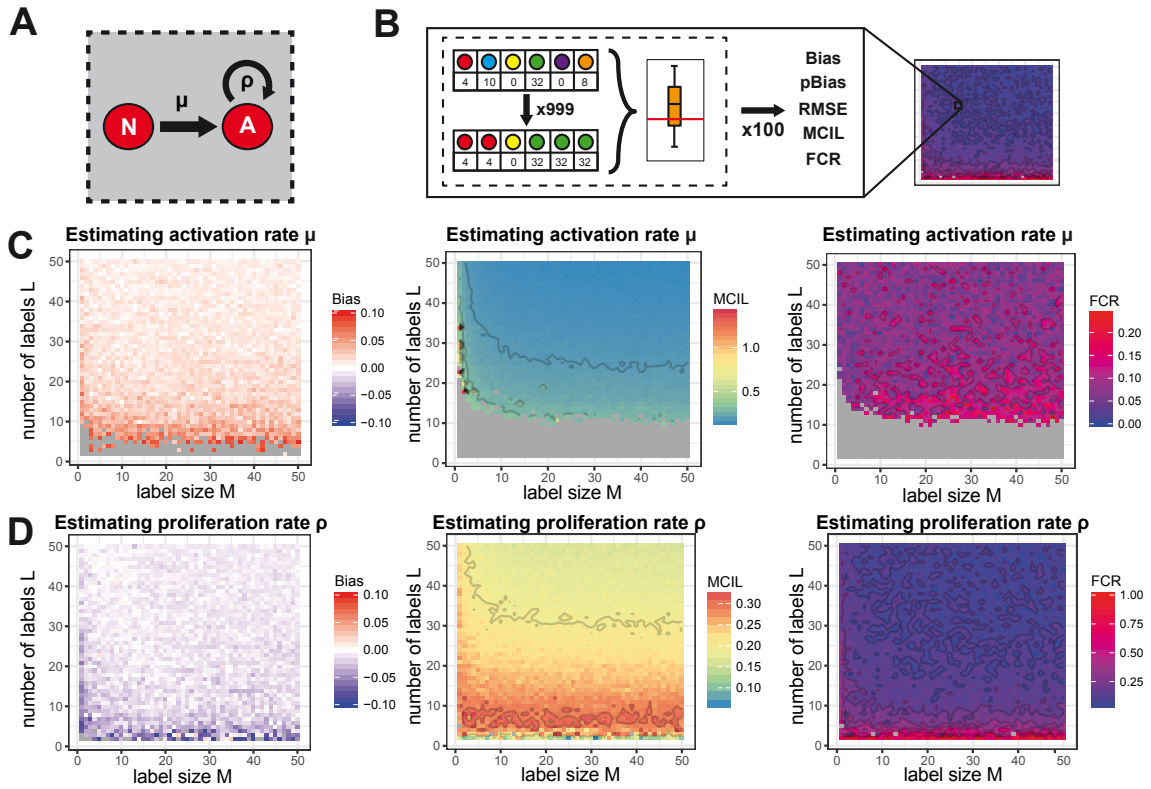

Figure 5.2: The influence of the labelling strategy on parameter estimation: (**A**) Sketch depicting the dynamics of the simple expansion model: Naïve cells are activated with rate μ and activated cells proliferate with rate ρ. (**B**) Calculation of heatmaps: Each labelling strategy is used to generate 100 different stochastically simulated data samples. Each data set is subsequently bootstrapped with 999 repeats to calculate the corresponding statistical quantities (see Materials and Methods 2.4.1 and 2.4.2). (**C-D**) The bias, the mean confidence interval length and the false coverage rate for the estimation of the activation rate, μ (**C**) and the proliferation rate, ρ (**D**) based on simulations given the model depicted in (**A**). Parameters values are given as $\mu = 0.3$, $\rho = 0.3$ and cells were sampled at time $T = 3$. Grey colour indicates values being above or below the depicted range (bias), or that the method is not able to estimate the respective confidence interval for the corresponding parameter combination (MCIL and FCR, see Materials and Methods 2.4.2). Figure adapted from [Gabel et al. 2017].

are shown for Δ_{MCIL}). This observation is found for both the activation, μ, and proliferation rate ρ. As an exponentially distributed labelling strategy performs worst, we conclude that increasing the inequality of label sizes in the initial labelling strategy impairs the parameter estimation quality and therefore the number of cells should always be uniformly distributed amongst the available labels.

5.3.2 The influence of the sampling time

As we assumed that experimental measurements can only be obtained at a single time point, the choice of the sampling time also affects the robustness of parameter estimation. Our results indicate that the sampling time influences model parameters in different ways: Given the simple

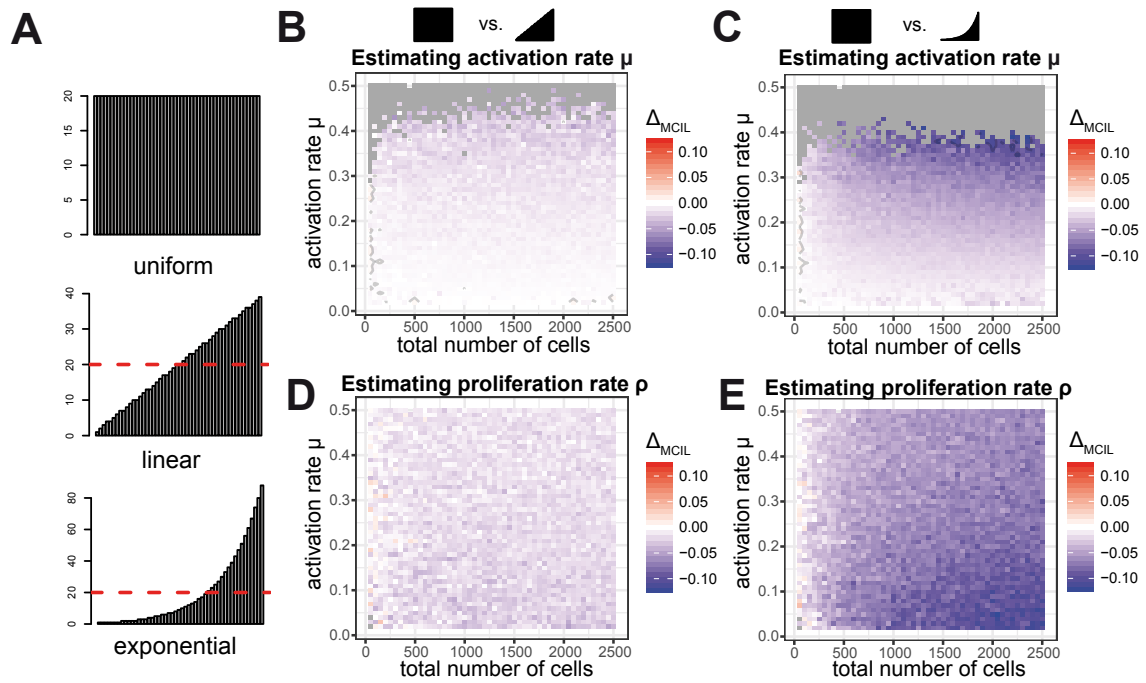

Figure 5.3: Influence of the label size distribution on parameter estimation: (**A**) Examples of the three different label size distributions used. From top to bottom: Uniformly, linearly and exponentially distributed initial labelling strategies. Every example distribution consists of a total of 1000 cells and $L = 50$ labels. The red dotted lines indicate a corresponding uniform distribution. (**B-C**) The difference in the MCIL for the estimation of the activation rate, μ, between the uniformly and linearly (**B**), and the difference between uniformly and exponentially distributed labelling strategies (**C**). (**D,E**) Analogous to (**B,C**) the differences in the MCIL for the estimated proliferation rate, ρ, are depicted. Figure adapted from [Gabel et al. 2017].

expansion model, we find that for the activation rate, μ, the identifiability first increases and then decreases with an increased sampling time, T, regarding the MCIL (see Fig. 5.4 **A**). In contrast to this, we observe that estimation of proliferation dynamics is continuously improved given later sampling times with respect to the MCIL (see Fig. 5.4 **B**) and the percentage bias (see Appendix C.5). Thus, there is a trade-off regarding the time point of sampling leading to either more robust estimates for the activation rate or the proliferation rate (Fig. 5.4 **C**).

This trade-off is also found in simulations based on the complex expansion system (see Appendix C.5). Here, proliferation rates are estimated more reliably as time increases, while the transition from naïve to central memory precursor cells is captured especially well for early sampling time points. In case of the homoeostatic system, no effect of the sampling time on the parameter estimation is observed, and both the proliferation and death rate are estimated reliably irrespective of the sampling time (see Appendix C.5).

In summary, our results show that for most rates the identifiability increases with the sampling time. However, an optimal estimation of all of the involved dynamics might not be possible if some rates are subject to a temporal restriction.

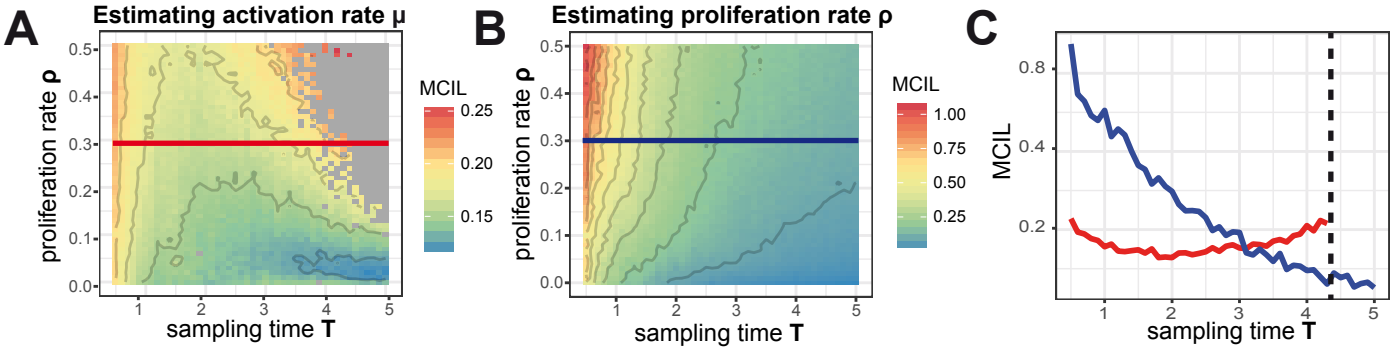

Figure 5.4: Influence of the sampling time on parameter estimation: MCIL of the activation rate, μ (**A**) and proliferation rate, ρ (**B**) based on varying combinations of sampling times, T and proliferation rates, ρ given the model of simple expansion dynamics. (**C**) Cross sections of panels (**A** & **B**) showing the MCIL of the activation rate, μ, (red) and the proliferation rate, ρ, (blue) with respect to the sampling time for a fixed proliferation rate ($\rho = 0.3$). The black dotted line indicates the time after which no estimates for the MCIL of the activation rate could be obtained. The estimation for each parameter combination is based on 100 independent stochastic simulations. Parameters that were kept fixed are $\mu = 0.3$, $L = 50$ and $M = 5$. Grey colour indicate that the method is not able to estimate the respective MCIL for the corresponding parameter combination (see Materials and Methods 2.4.2). Figure adapted from [Gabel et al. 2017].

5.4 The influence of experimental limitations on the identification of parameters

In each experimental set-up using cellular labelling techniques, there are factors beyond the experimentalist's control, which can influence the parameter estimation quality. In the following, we will analyse how three common experimental shortcomings (incomplete transfer, incomplete sampling and missing cellular compartment data) are affecting the identification of cellular dynamics.

5.4.1 The influence of incomplete transfer

During adoptive transfer of cells into a living host it is unlikely that all cells survive the transit. Common experimental obstacles include the failure to reach the target tissue or the host-induced rejection of labelled cells [Moon et al. 2009]. Therefore, it has to assumed that the experimentally observed cellular responses are generated only from a fraction of the originally transferred cell population. To study the influence of incomplete transfer on the robustness of parameter estimates, we adjusted our previously used simulation approach by only transferring a fraction of the initial labelling strategy (see section 5.2.2) and analysed the output given the homoeostatic and the complex expansion model (see Eqs. 5.1 and 5.3). Both labelling strategies are based 800 cells in total, which are distributed based on a unique ($L = 800$ labels with $M = 1$ cell each)

or a shared labelling approach ($L = 8$, $M = 100$)[2]. In the following, it is assumed that cells are only lost during transfer and not during sampling, meaning that all generated cells are recovered from the system for subsequent analysis.

Given the homoeostatic model, we find that incomplete transfer results in an overestimation of both the proliferation and the death rate in either of the two labelling strategies used (see Fig. 5.5 **B**). The ratio of the two rates was also estimated incorrectly: Instead of both rates having roughly the same values, the death rate δ was consistently estimated to be higher than the corresponding proliferation rate ρ. This is due to the fact that the system has to compensate for the apparent loss of cells over time. We also observed that a higher transfer loss increased the variation of the parameter estimates for both labelling approaches.

Given the complex expansion model, the findings look slightly different, but can be explained in the same way. Here, we observe that a loss of cells in transfer mainly affects the initial differentiation from T_N into T_{CM} (see Fig. 5.6). For a unique labelling approach, the subsequent differentiation into T_{EM} is also strongly biased[3]. Here, the low cell numbers could be easily compensated by a reduced activation of naïve cells, as this compartment was not used for fitting.

However, if the transfer fraction is known, it is possible to adjust the estimation procedure, which leads to unbiased parameter estimates (see Fig. 5.5 **C** for the homoeostatic model and Appendix C.6 for the complex expansion model). This works reliably well for both labelling strategies even if large fractions of cells are lost during transfer.

In summary, our results show that neglecting incomplete transfer can lead to a strong bias in some, but not necessarily all parameter estimates.

5.4.2 The influence of incomplete sampling

In many cases, sampling cells from the underlying host system allows to recover only a fraction of the complete cell population, due to the limiting nature of blood or tissue samples and the scattered localisation of cells [Steinert et al. 2015]. To determine the influence of incomplete sampling on the parameter estimation procedure we repeated our analysis by using only fractions of the simulated response for the mathematical evaluation. In the following simulations, we assumed that no cells are lost during transfer.

Given the homoeostatic system, we find that incomplete sampling results in an underestimation of both the proliferation, ρ, and death rate, δ (see Fig. 5.7 **B**). However, the bias in the proliferation rate is more pronounced, indicating that incomplete sampling does not affect both rates equally. This means that incomplete sampling affects the parameter estimation in a different way compared to incomplete transfer, where both rates were equally overestimated (see section 5.4.1).

More diverse effects are observed for the parameter estimation given the complex expansion model (see Fig. 5.7 **C**). Similar to the homoeostatic model, incomplete sampling results in an underestimation of proliferation rates, which is more pronounced given the unique labelling

[2]Similar strategies have already been used in previous labelling experiments [Gerlach et al. 2013; Buchholz et al. 2013]

[3]Note: The bias observed in the parameter estimation of the shared labelling approach is an artefact of the estimation procedure (see section 5.4.3).

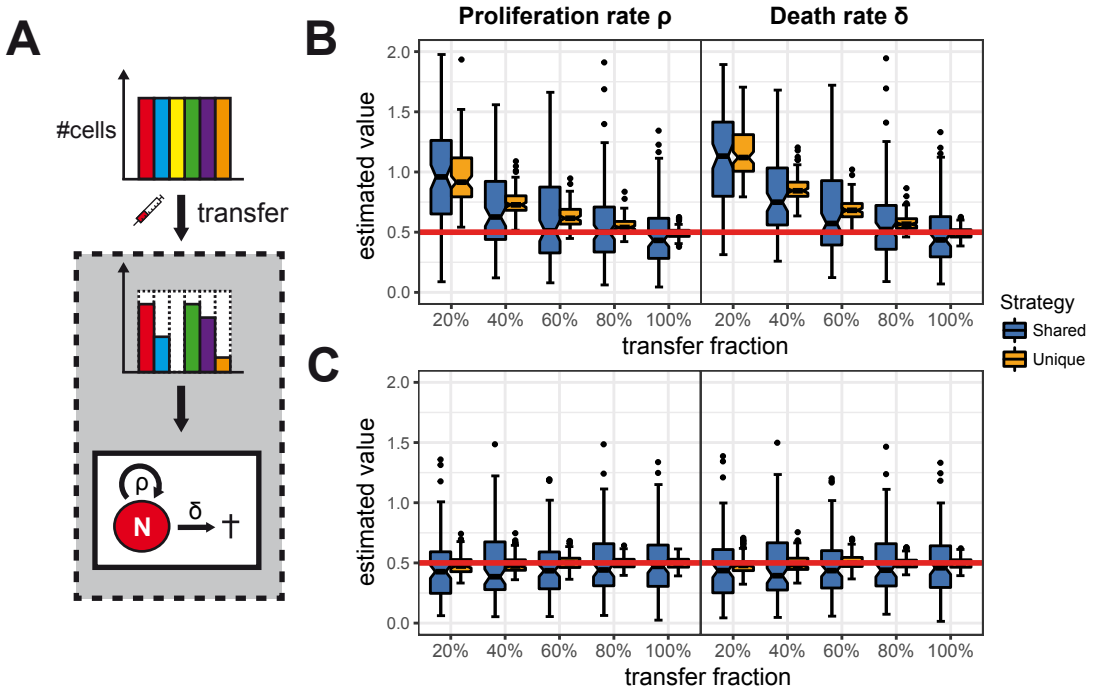

Figure 5.5: The influence of incomplete transfer given the homoeostatic model: (**A**) Sketch depicting the limitation of incomplete transfer: Only a fraction of the initially transferred cell population survives the transit. Subsequently, surviving cells turn over homoeostatically. (**B**) Panels show the distribution of estimates for the proliferation rate, ρ (left), and the death rate, δ (right), based on different transfer fractions. The labelling strategies are based on 800 cells in total, distributed according to either a shared ($L = 8, M = 100$, blue) or a unique labelling approach ($L = 800, M = 1$, orange). Here, the loss in transfer was not accounted for in the parameter estimation procedure. (**C**) Taking the transfer fraction into account resolves the estimation bias. Each boxplot is based on the results of 100 individual stochastic simulations with $\rho = \delta = 0.5$ d^{-1} and cells sampled 8 days after transfer. Red lines indicate the true parameter values. Figure adapted from [Gabel et al. 2017].

approach. Neglecting the bias in the shared labelling approach (see section 5.4.3), the underestimation of the rates increases with every cellular compartment, being lowest for ρ_{CM} and highest for ρ_{E} (see Fig. 5.7 **C**). On the other hand, differentiation rates are estimated reliably, and, given the shared labelling approach, almost no influence of incomplete sampling is observable here.

As for incomplete transfer, the limitation of incomplete sampling can be addressed by rescaling the measured cell numbers according to the sampling fraction (see Appendix C.7). However, it might be technically difficult to retrieve this fraction experimentally.

As parameter estimation based on shared labelling approaches are less affected by incomplete sampling, we would conclude that these strategies are more robust given this experimental shortcoming.

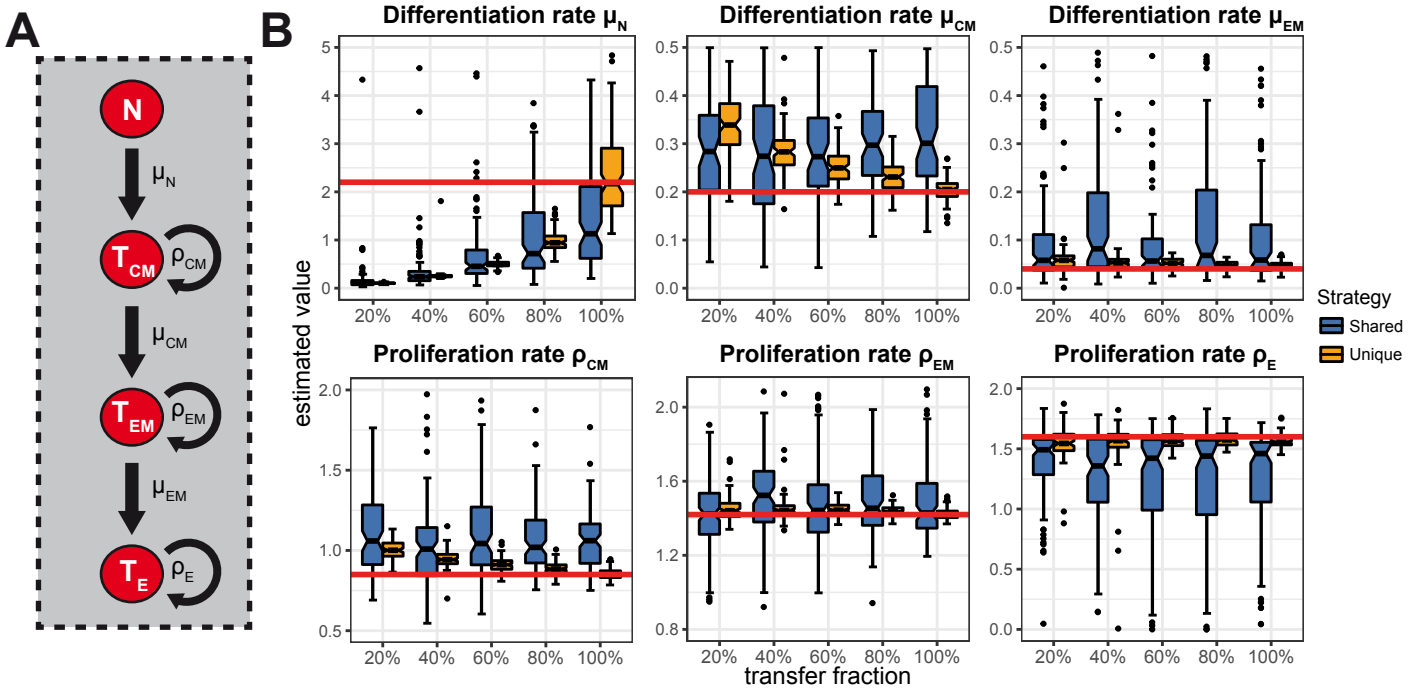

Figure 5.6: The influence of incomplete transfer given the complex expansion model: (**A**) The linear pathway of cell differentiation of the complex expansion model based on [Buchholz et al. 2013]: Naïve cells (N) turn into central memory precursor cells (T_{CM}), which subsequently turn into effector memory precursor (T_{EM}) and effector (T_E) cells. Cells differentiate and proliferate according to the corresponding rates $\mu_.$ and $\rho_.$, respectively. (**B**) Panels show the parameter estimation quality for the differentiation (upper row) and proliferation rates (lower row) given different transfer fractions. The labelling strategies are based on 800 cells in total, distributed according to either a shared ($L = 8, M = 100$, blue) or a unique labelling approach ($L = 800, M = 1$, orange). Here, the loss in transfer was not accounted for in the parameter estimation procedure. Every boxplot is based on the results of 100 individual stochastic simulations. Differentiation and proliferation rates are defined as $\mu_N = 2.2\ d^{-1}$, $\mu_{CM} = 0.2\ d^{-1}$, $\mu_{EM} = 0.04\ d^{-1}$, $\rho_{CM} = 0.85\ d^{-1}$, $\rho_{EM} = 1.42\ d^{-1}$ and $\rho_E = 1.6\ d^{-1}$ [Buchholz et al. 2013]. Red lines indicate the true parameter values. Figure adapted from [Gabel et al. 2017].

5.4.3 The influence of missing compartments

So far we implicitly assumed that experimental observations can be obtained for each cellular compartment. However, due to experimental restrictions, such as spatial or temporal limitations, it might not be possible to measure certain cellular subsets [Kaiser et al. 2013; Steinert et al. 2015].

In the previous sections, we always neglected the number of naïve cells in our analysis when fitting the simple and the complex expansion dynamics to the simulated data (see section 5.2.2). This was done to follow the studies our analysis is based on [Buchholz et al. 2013; Kaiser et al. 2013]. In these studies no data was available for the respective initial model compartment[4], but

[4]In Buchholz et al., the initial compartment consisted of the naïve $CD8^+$ T cell population and in Kaiser et al., the *Salmonella* population located in the gut.

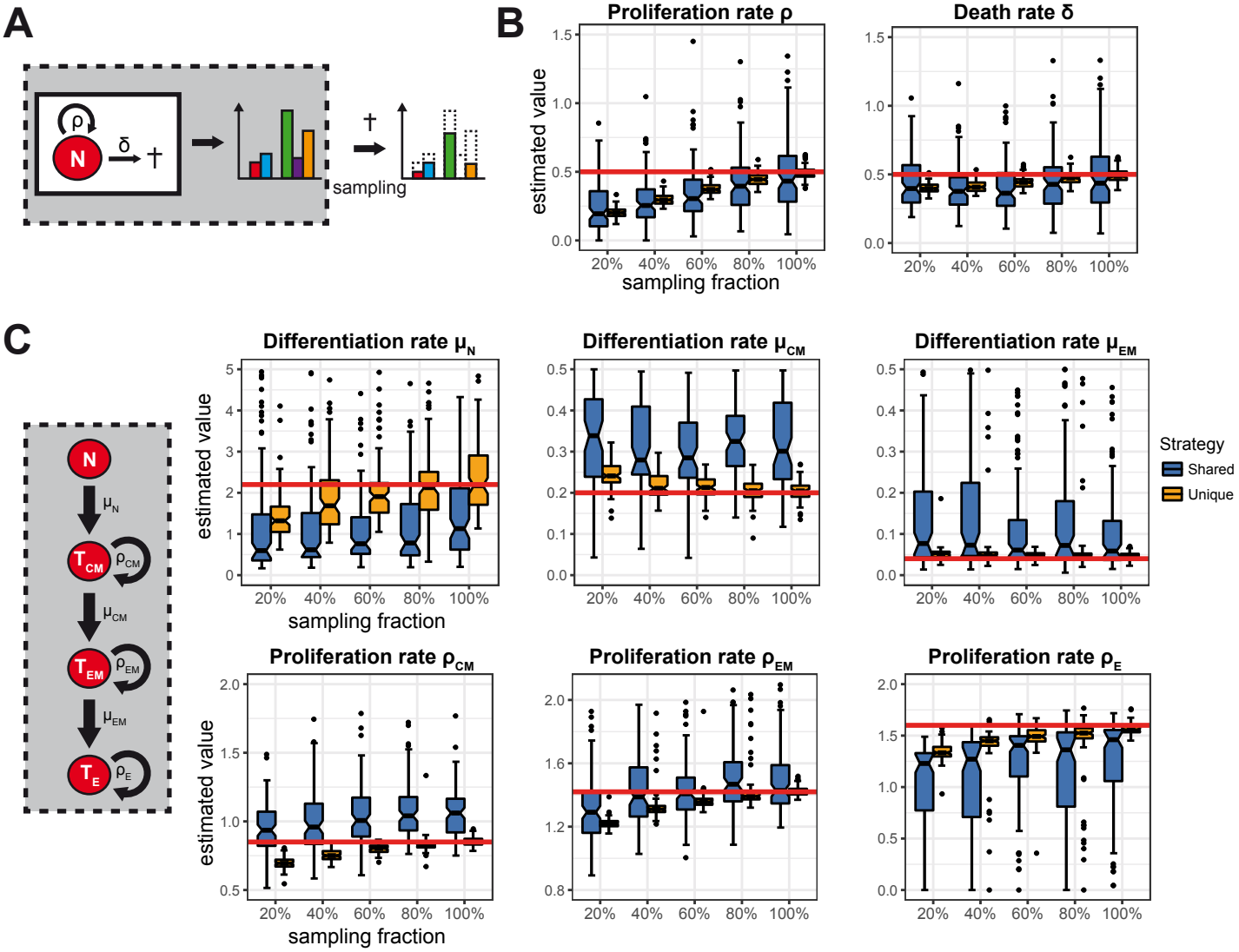

Figure 5.7: The influence of incomplete sampling on parameter estimation: (**A**) Sketch depicting the problem of incomplete sampling: Only a fraction of the labelled cells is found in the sample and will be used for subsequent analysis. (**B**) Panels show the the distribution of estimated proliferation, ρ and death rates, δ given different sampling fractions based on the homoeostatic system shown in **A**. Here, the results from a shared (blue, $L = 8,\ M = 100$) and a unique (orange, $L = 800,\ M = 1$) labelling approach are depicted. (**C**) The corresponding results for the complex expansion model. Every boxplot is based on the results of 100 individual stochastic simulations. For the definition of the model parameters, see captions of Figs. 5.5 and 5.6. Red lines indicate the true parameter values. Figure adapted from [Gabel et al. 2017].

the rates belonging to these compartment could be estimated nevertheless due to the additional information provided by the applied labelling approaches.

We wanted to investigate how much information is lost if the data belonging to a specific model compartment cannot be retrieved from the host system and hence used for mathematical analysis. To this end, we refitted the parameters of the complex expansion model several times

by selectively removing the data of one of the four different T cell compartments. We then compared the estimates based on incomplete data to the case in which all available data was taken into account when fitting the model.

In general, our results show that the rates belonging to the removed compartment show a larger variation and are estimated more poorly compared to parameter estimates obtained by fitting the complete data set (see Fig. 5.8). However, this was only partially true for the shared labelling approach ($L = 8$, $M = 100$). Here, we find that all parameters were estimated more reliably and robustly if the compartment of central memory precursors (T_{CM}) was removed (see Fig. 5.8, blue boxplots). This surprised us, as we did not expect that less information could result in a better fit. However, taking a closer look at the data at hand, we find that correlation coefficients associated with the T_{CM} compartment are highly variable and especially a negative correlation resulted in a poor fitting quality, as apparently the model was not able to reproduce the observed quantities (see Appendix C.8). This shortcoming also explains the consistent bias observed in the previous sections. Therefore, our results argue for wariness with respect to the use of low-numbered and highly variable data points, but also highlight the benefit of the application of labelling strategies in terms of estimating cellular dynamics not present in the data.

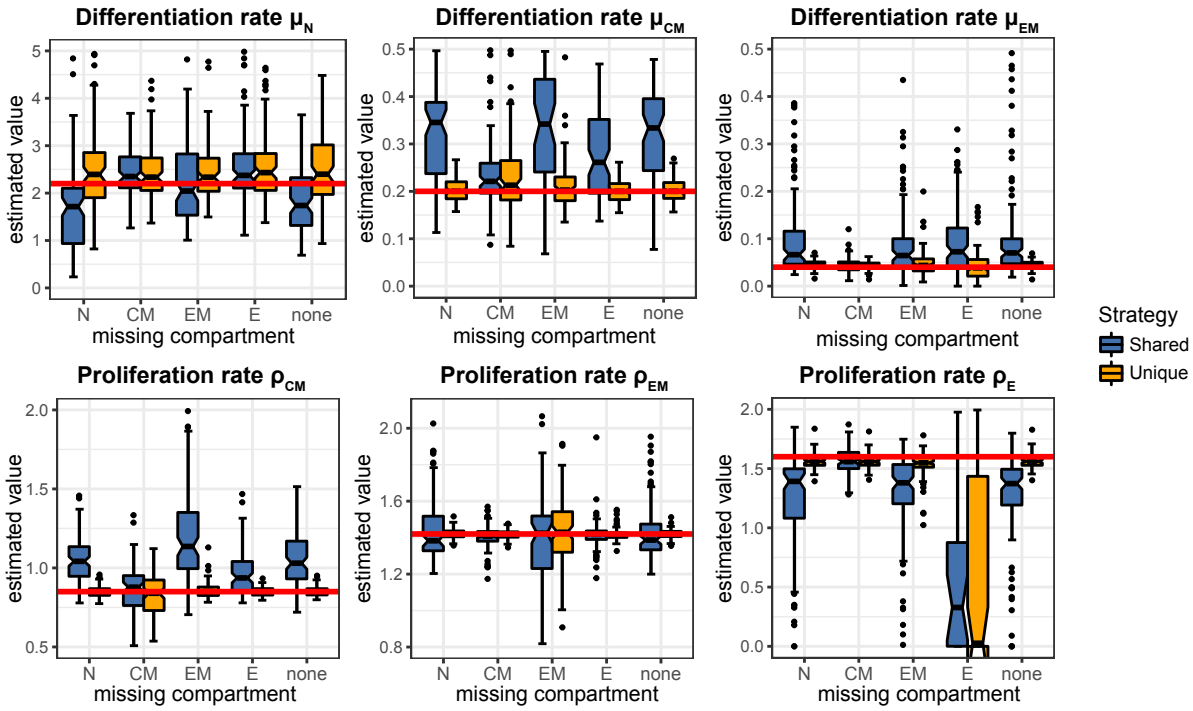

Figure 5.8: The influence of missing data on parameter estimation: The distribution of the estimated parameters for the complex expansion dynamics based on different missing data are shown. Every boxplot is based on the results of 100 individual stochastic simulations. Parameters used to simulate the dynamics and the time point of sampling are defined as before. Red lines indicate the true parameter values.

5.5 Discussion

Over the last decades experimental advances have steadily increased the number of possibilities to label cells in biological experiments. As of today, a huge variety of labelling methods, relying on naturally occurring or artificially induced cellular markers, is available (reviewed in [De Boer et al. 2013; Schumacher et al. 2010]). The possibility to adjust individual experimental factors, such as the number of different labels and the number of cells per label, allows addressing many different biological questions, including the determination of cellular dynamics during immune reponses. In particular, cell labelling strategies can help to quantify cellular processes, such as activation, differentiation and proliferation [Gerlach et al. 2013; De Boer et al. 2013; Buchholz et al. 2013], which is important to understand the behaviour and the possible manipulation of cellular systems. For example, T cell vaccination approaches need to induce elevated effector responses crossing a certain threshold to guarantee protection (see chapter 4). Therefore, the accurate quantification of cellular processes is of utmost importance to reliably induce protective responses. While many different labelling strategies have been used so far, the extent to which they allow the appropriate identification and quantification of cellular dynamics has not been systematically analysed yet.

To address this question, we analysed the performance of different labelling strategies given various cellular systems and experimental shortcomings *in silico*. Here, we simulated three different scenarios of cellular turnover, closely related to dynamics of $CD8^+$ T cell, including homoeostatic cell proliferation and cellular expansion dynamics. As we assume in our analysis that cellular data can only be gathered at a single time point, our scenarios represent the most basic experimental scenario. It is especially in these cases, that cell labelling drastically improves the amount of available information.

Our results indicate that there are some general conclusions about the appropriateness of cellular labelling approaches. First, increasing the number of labels, L is continuously improving the estimation quality in terms of bias and variance (see Fig. 5.2). This makes sense, as each label adds another data point available for the mathematical analysis. Furthermore, the accuracy of parameter estimates also benefits from an increased number of cells per label, M, although this effect saturated rather quickly. Testing two extreme labelling strategies - a shared ($L = 8$ labels with $M = 100$ cells each) and a unique labelling approach ($L = 8$ labels with $M = 1$ cells each), we would conclude that a higher number of labels is more important than a higher number of cells, as the unique labelling strategy performs much better than the shared labelling approach in terms of bias and variance.

Additionally, we found that approaches combining smaller and larger label sizes perform worse than uniformly distributed labels. In general, a strategy using uniformly distributed label sizes seems to generate the most reliable data for analysis. However, using unique labelling approaches can still be useful, as the unique labels allow to estimate the potential fraction of cells that is lost during adoptive transfer [Buchholz et al. 2013].

Furthermore, we found that increasing the sampling time generally allowed to estimate proliferation and differentiation rates more reliably, with the notable exception of time-constrained dynamics, such as the activation dynamics of naïve cells in our examples. Hence, the robust

estimation for all cellular dynamics might not be possible if only one sampling time point is available.

We also looked at the impact of experimental shortcomings on parameter estimation, namely incomplete transfer, incomplete sampling and the inability to measure certain cellular compartments. All of these shortcomings can bias the parameter estimates if they are not accounted for. Interestingly, we found that the effects on parameter estimation quality were different for incomplete transfer and incomplete sampling. Incomplete transfer seemed to affect dynamics depending on the associated cell compartment. For example, all rates belonging to the effector cell compartment were reliably estimated given our complex expansion model irrespective of the number of cells lost during transfer, while the rates describing the dynamics of the central memory precursor compartment all exhibited a bias, the strength of which being dependent on the transfer fraction (see Fig. 5.6). In contrast to this, incomplete sampling selectively affected types of cellular dynamics irrespective of the associated compartment. Given the complex expansion model, we found a strong influence on the estimation of the proliferation rates, but barely any on the estimation of differentiation dynamics (see Fig. 5.7). As the shared labelling approach was less affected by both incomplete sampling and incomplete transfer, using labelling approaches with medium or large label sizes seems to be a good strategy to obtain robust results given these experimental shortcomings.

If the respective transfer or sampling fraction is known, it is possible to correct the parameter estimation procedure and obtain accurate results (see Fig. 5.5 and Appendix C.7). However, obtaining these fractions might prove experimentally difficult. While transfer loss can be experimentally determined by using unique labelling approaches [Buchholz et al. 2013], assessing the sampling fraction remains difficult, especially if cells are localising in different tissues.

In our analyses, we always assumed that labels are stable and do not interfere with the underlying cellular dynamics. While this is true for certain artificial markers, such as genetic barcodes or congenic markers [Gerlach et al. 2013; Buchholz et al. 2013; Kaiser et al. 2013], this assumption will mostly likely not hold in case of naturally occurring markers. For example, the receptors of $CD8^+$ T cells consist of distinct $\alpha-$ and β-chains [Blattman et al. 2000; Wong et al. 2007; Zarnitsyna et al. 2013], which could potentially be exploited as cellular labels. However, β-chains are associated with binding affinity and can therefore influence the activation of $CD8^+$ T cells [Stone et al. 2009; Chervin et al. 2013]. As α-chains are less involved in antigen binding, it remains an interesting question to address their applicability as cellular markers.

In summary, our results show that the set-up of labelling strategies influences the accuracy with which cellular rates can be determined. To optimise the experimental output for mathematical analysis, uniformly distributed labelling strategies with at least intermediate numbers of cells per label should be used. Furthermore, the parameter estimation quality benefits from every additional label added. Using such labelling approaches should allow for the generation of robust and reliable data, even in the case of experimental shortcomings. We hope that the application of the labelling strategies proposed here will contribute towards improving the quantification of cellular dynamics in the future.

CHAPTER 6

General discussion

6.1 Implications for the design of T cell-based vaccines

The successful establishment of protection-mediating T cell-based vaccines would be a major step towards the control of some of the most dangerous infectious diseases today. However, to reach this goal, more research is necessary, as many properties regarding the efficient induction of protective $CD8^+$ T cell levels, as well as their short- and long-term maintenance are currently only poorly understood. We would like to highlight in this last section how the knowledge gained by our studies can be used in the context of T cell-based vaccination approaches.

6.1.1 On the induction of protective $CD8^+$ T cell responses

Given diseases like malaria or HIV, the generation of immunological memory is not sufficient for protection. Rather, it was found that high levels of effector-like $CD8^+$ T cells are necessary to prevent these infections [Hansen et al. 2011; Fernandez-Ruiz et al. 2016]. Therefore, a major goal in medical research regarding T cell-based vaccines lies in the elicitation of elevated $CD8^+$ T cell responses which are able to provide immediate effector function and are efficiently surveilling the cell tissues at risk.

In order to guarantee protection, T cell-based vaccination approaches have to make sure that the induced immune response crosses a certain threshold. Due to this, setting up successful vaccination strategies requires a quantitative understanding of cause and effect of the $CD8^+$ T cell dynamics elicited by vaccination approaches. Here, mathematical modelling is an almost indispensable tool to accurately determine the underlying cellular turnover.

In our study analysing the $CD8^+$ T cell dynamics in MCMV infection, we find that the shaping of the inflationary response is depending on different viral stimuli, one corresponding to the acute and the other one to the latent phase of viral reproduction. The magnitude in cell numbers reached over the course of the infection seems to be mostly dependent on the efficiency of viral reproduction, which in our model was determined by the rates describing viral replication speed and T cell-mediated clearance. Therefore, our results would suggest that the success of vaccines using CMV vectors to generate elevated effector-like responses critically depends on prolonged acute viral replication, which presumably leads to an increase in the size of the latent viral reservoir later on. If this is in fact true, the magnitude of the induced long-term $CD8^+$ T

cell response by CMV could potentially be boosted by the administration of immunosuppressing drugs shortly before or after vaccination.

As we found in our study about malaria vaccination approaches, other factors such as the dosage and the frequency of the vaccination regimen can also determine the level of the generated $CD8^+$ T cell response. Here, our results indicate that an efficient induction of liver-resident memory T cells strongly depends on both the administration of booster injections and high PbRAS dosages. Our modelling suggests that the failure to generate protection after only one vaccination shot stems from the organ-dependent cellular kinetics: As naïve $CD8^+$ T cells are mostly activated in the secondary lymphoid organs, the antigens provided by the first vaccination dose are only sufficient to establish a founder population of T_{RM} cells in the liver. These cells, however, are expanding rapidly upon antigen re-encounter, which is provided by the administration of booster injections. Therefore we would conclude that T cell-based vaccination approaches targeting malaria necessarily need to include at least one booster injection in order to mediate protection.

Here, quantifying the proliferation dynamics of T_{RM} cells more accurately would allow to robustly determine the appropriate number of injections. As we showed, improving the mathematical estimation quality can be accomplished by the use of labelled cell populations. In general, every approach studying the efficiency and efficacy of T cell-based vaccination approaches could benefit from the application of suitable labelling strategies. Sticking with the example of Malaria vaccination regimens, the use of cellular labelling strategies would help to confirm our proposed differentiation pathway, to estimate the antigen-dependent parameters more robustly and to determine the contribution of the spleen to the T_{RM} response in the liver in subsequent booster injections. Each of these findings would help in the establishment of a protective T cell-based vaccine.

6.1.2 On the long-term efficacy of T cell-based vaccines

The induction of long-term protection is an essential goal for all vaccination approaches. For many currently available vaccines, it is known that protection is waning over time and its loss needs to be compensated by the administration of well-timed booster injections. Therefore, analysing the maintenance of pathogen-specific $CD8^+$ T cell responses is crucial to determine the efficacy of T cell-based vaccination approaches over time.

In both of our studies dealing with MCMV and *Plasmodium* infection, long-term data of the respective $CD8^+$ T cell responses was available, which allowed us to predict the stability of the protection-mediating T cell levels. For the pool of T_{RM} induced after malaria vaccination, our experimental data and the mathematical modelling suggested that the cell numbers of this compartment stays more or less stable over the course of months. Additionally, experimental challenges more than 100 days after the administration of the last booster injection showed that protective levels reached early after vaccination are also effective long-term, indicating a long half-life of the induced T_{RM} population.

For CMV, previous studies also indicated a constant maintenance of the elicited inflationary responses [Munks et al. 2006; Torti et al. 2011b; Hansen et al. 2011]. However, our results

suggest that the inflationary pool is declining over time (see chapter 3). Not discarding the possibility of an experimental artefact, a likely explanation would be a varying, maybe oscillating, maintenance of the long-term $CD8^+$ T cell response, possibly due to viral reactivation events. This finding would have implications for the mediated protection based on CMV-induced T cell based-vaccines, as variable levels of effector-lie responses will most likely result in variable levels of protection over time as well.

Analysing the inherent turnover of a cell population in steady state is difficult, as the overall cell numbers do not indicate the underlying dynamics. To infer such information, measurements of proliferation and survival markers, such as Ki67 or BrDU [Torti et al. 2011b], or deuterium labelling [Borghans et al. 2017] can be used for quantifying $CD8^+$ T cells dynamics. Our study in chapter 5 showed that cellular turnover in homoeostatic cell populations can also successfully obtained by the application of appropriate labelling strategies. In our study, this even allowed us disentangle corresponding proliferation and death rates based on the variation in the labelled subsets. As our results show that the quantification of cellular dynamics strongly benefits from the use of specific labelling strategies, its application should be considered in experiments determining the maintenance of cellular subsets generated by T cell-based responses, including the pool of T_{RM} after malaria vaccination or inflationary cells in MCMV infection.

6.2 Future work

A scientist's work is never done.

This adapted version of an old American proverb[1] is supposed to capture an inherent property of science, where the answer to a particular problem almost inevitably seems to raise a new set of questions. The analyses performed in this thesis provide no exception thereof. Thus, we would like to highlight the most pressing questions that warrant further investigations and discuss potential experimental and mathematical approaches that could address these topics in future studies.

One shortcoming of our analysis of $CD8^+$ T cell dynamics in MCMV infection was that the model predictions were solely based on measurements obtained from the blood of mice. While we had a few additional measurements of cell numbers in different organs, these data were neither enough to predict organ-dependent dynamics nor was it possible to relate the cell numbers observed in these organs to the numbers measured in the blood. Should future work indicate a critical involvement of a non-vasculatory organ in the generation and maintenance of memory inflation, it would be necessary to obtain frequent organ-specific measurements to reliably quantify the cellular dynamics. Assuming that sampling of cellular data involves the harvest of host animals, the use of appropriate labelling strategies would strongly increase the amount of data available for analysis. Furthermore, the stability of inflationary responses over time should be addressed experimentally. To this end, frequent longitudinal measurements need to be obtained. As we show in our analysis, the observable $CD8^+$ T cell response holds information about possible viral reactivation events and could therefore help to determine how viral latency triggers memory inflation.

[1] The original saying is *Man may work from sun to sun, but woman's work is never done.*

Gathering more longitudinal data is also important for our analysis of malaria vaccination approaches. Here, the numbers of tissue-resident memory cells are of special importance, as they would allow to address questions about the genesis and maintenance of this pool. New data are also needed to test the mathematical predictions with respect to pathway and turnover identification. Here, the use of cellular labelling strategies would be a tremendous help. In addition to determining the optimal dosage and frequency of immunisations approaches, the influence of the interval between two booster injections should also be analysed. Our approaches were mostly based on a 14 day interval scheme, but should be consistently extended to shorter and longer, e.g. 7 or 28 days, intervals. The data of these experiments would allow our modelling approach to incorporate the influence of the timing of injections as well.

While our theoretical analysis of the influence of labelling strategies on the estimation of cellular parameters takes many possible dynamics into account, a systematic expansion of additional cellular interactions, e.g. by including reversible differentiation processes or modelling proliferation by logistic growth, is needed to determine the applicability of the previously obtained results to more complex systems. Also, quantifying the impact of labelling strategies on parameter estimation given multiple measurement time points will help to assess the benefit of labelling approaches in situations in which data is less scarce. As both the analyses about $CD8^+$ T cell dynamics in malaria and MCMV are lacking data from the initial expansion phase, combining these experiments with our proposed labelling strategies would allow to address many of our raised points simultaneously.

In summary, the analyses provided in this thesis are an important contribution to the understanding of $CD8^+$ T cell dynamics and differentiation in the context of immunity and vaccination. We therefore hope that our results will aid in the successful development of effective T cell-based vaccines against various kinds of infectious diseases.

APPENDIX A

A.1 Model fits to the non-inflationary data

The dynamics of the M45-specific $CD8^+$ T cells as predicted by the constant influx model (see Eqs. 3.4 and Fig. A.1).

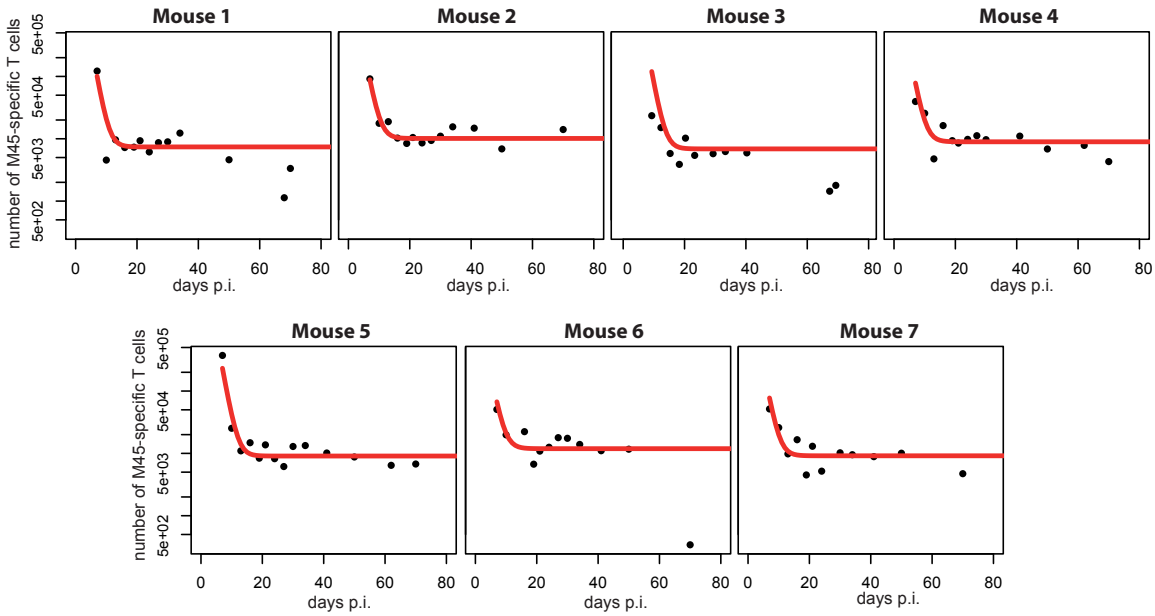

Figure A.1: Individual fits to the non-inflationary data: Panels show the model prediction (red lines) compared to the individual experimental data (black dots). All experiments were performed by the lab of Prof. Dr. A. Oxenius, Institute of Microbiology, ETH Zürich. Model fits were estimated based on the software Monolix. For the respective parameter sets see Table A.2.

A.2 Individual parameter estimates

The individual parameters sets are given in Table A.1 (inflationary data) and Table A.2 (non-inflationary data).

mouse	V_7	T_7	ρ_V	k	β	ρ_R	ρ_T	δ
1	1.73	61,611	0.15	2.95e-06	5.39e-04	0.414	0.29	0.323
2	1.7	39,356	0.13	3.01e-06	4.79e-04	0.413	0.31	0.318
3	1.79	35,789	0.09	3.12e-06	5.16e-04	0.414	0.29	0.326
4	1.69	33,498	0.1	3.06e-06	4.68e-04	0.413	0.3	0.32
5	1.73	55,775	0.11	3.04e-06	5.15e-04	0.413	0.29	0.324
6	1.7	24,215	0.09	3.14e-06	5.16e-04	0.415	0.29	0.325
7	1.71	30,817	0.18	2.86e-06	5.1e-04	0.414	0.3	0.322

Table A.1: Individual parameter estimates for the ENR model: Parameters obtained from fitting the ENR model to the inflationary data. For model description see Eqs. 3.3.

mouse	T_7^E	Λ	δ
1	100,640	4781	0.64
2	88,659	6176	0.61
3	119,900	4327	0.63
4	78,411	5508	0.62
5	232,120	5751	0.63
6	67,265	7110	0.6
7	77,534	5611	0.61

Table A.2: Individual parameter estimates for the CI model: Parameters obtained from fitting the CI model to the non-inflationary data. For model description see Eqs. 3.4.

A.3 Subset dynamics of MCMV-specific T cells

M45- and M38-specific CD8$^+$ T cells were gated for their expression of surface markers CD62L and KLRG1 and the distribution of the overall pool was assessed at every measurement time. Fig. A.2 shows the mean distribution over time. Cell subset are classified by belonging either to central memory (CD62L+/KLRG1-), effector memory (CD62L-/KLRG1-) or effector cells (CD62L-/KLRG1+).

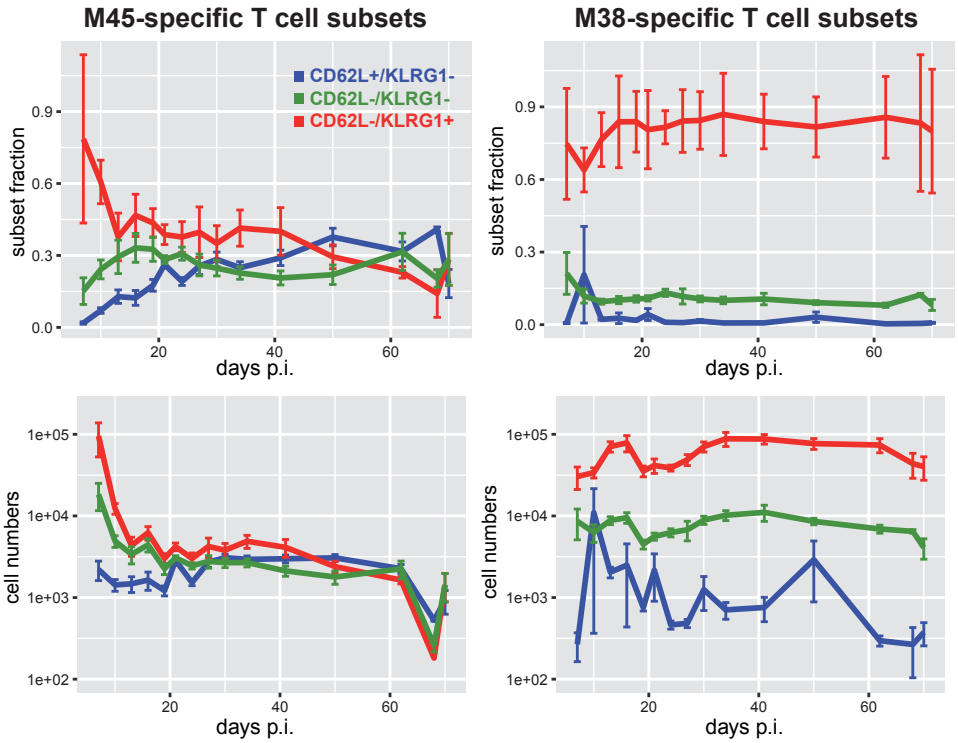

Figure A.2: Subset dynamics of MCMV-specific T cells: The dynamics of M45- and M38-specific CD8$^+$ T cells is shown over time. The top row shows the percentage of the respective subsets over time and the bottom row the corresponding number of cells. Subsets are analysed with respect to the expression of surface markers CD62L and KLRG1. Mean ± SEM is shown. All experiments were performed by the lab of Prof. Dr. A. Oxenius, Institute of Microbiology, ETH Zürich.

A.4 The influence of periodic reactivation events on the corresponding CD8$^+$ T cell dynamics

Based on the sporadic reactivation model given by Eq. 3.7 with parametrisation given as $k = 5 \times 10^{-6}$, $\rho = 0.29$, $\delta = 0.32$, $R_i = 2.4$ and $T_i = 10i$, $i \in \mathbb{N}_0$, we analysed how the scaling of parameters determining the viral response would affect the oscillating CD8$^+$ T cell dynamics. Here, we looked at changes in the viral burst frequency, T_i, the released viral load per burst, R_i and the CD8$^+$ T cell mediated clearance, k. The effects on the long-term mean cell number, the amplitude and the minimal cell number are shown in Fig. A.3 and are based on the model predictions from day 50 to day 100 p.i. Scaling is performed directly for the viral load per burst, but inversely for the reactivation frequency and the clearance rate. (meaning a scaling factor of m actually corresponds to a scaling factor of $\frac{1}{m}$)).

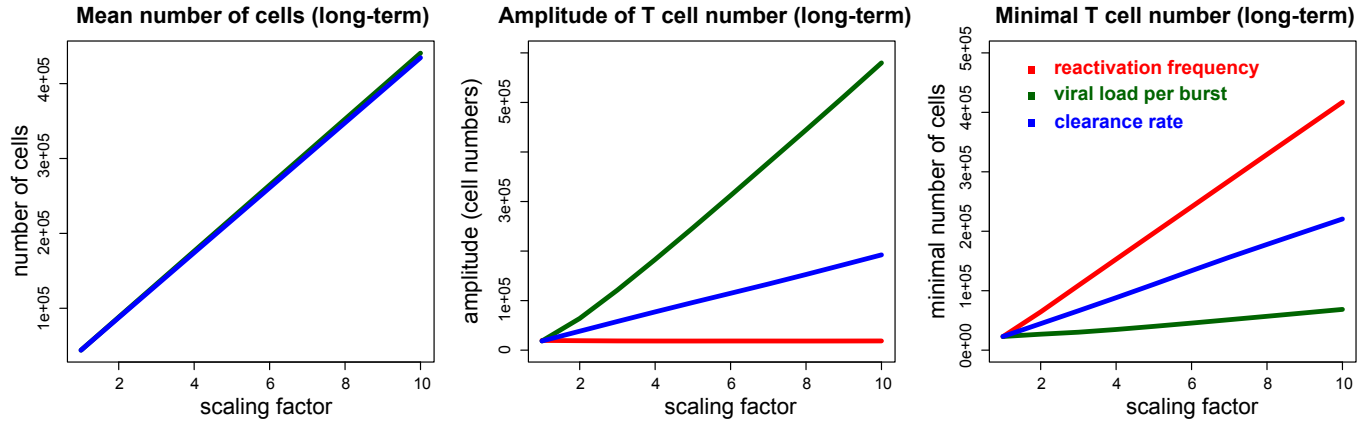

Figure A.3: The effects of changes in the sporadic viral reactivation patterns on the CD8$^+$ T cell dynamics: The long-term mean number of cells (**A**), the amplitude (**B**) and the minimal cell numbers (**C**) are shown. All properties are calculated based on the CD8$^+$ T cell numbers from day 50 to day 100 p.i. Scaled parameters include the viral reactivation frequency (red), the viral load per reactivation event (green) and the T cell mediated clearance rate (blue). Scaling is performed directly for the viral load, but inversely for the reactivation frequency and the clearance rate.

APPENDIX B

B.1 Cell numbers during expansion phases

The fit in Fig. B.1 is based on the parameter combination that gave the lowest negative log-likelihood value. The number of effector cell in the liver is predicted to be around 4×10^9.

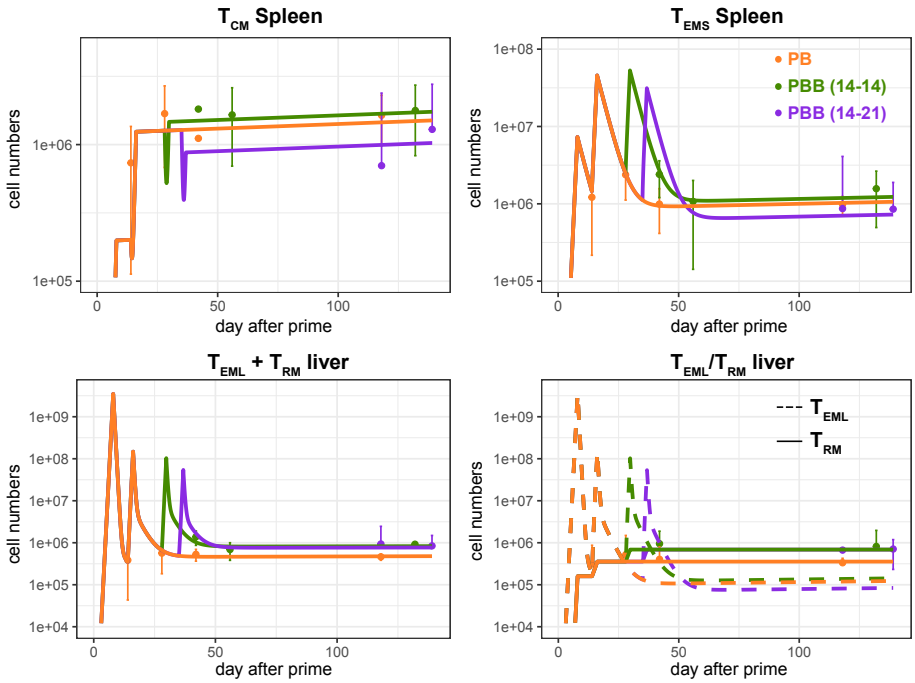

Figure B.1: The best fit obtained by the profile likelihood approach: Panels show the cellular dynamics of T_{CM} (top left), T_{EMS} (top right), T_{EML} + T_{RM} (bottom left) and T_{EML} /T_{RM} compartments (bottom right). Lines are based on a prime-boost (PB, orange) or a prime-boost-boost (PBB, green and purple) vaccination strategy using a normal dosage (1×10^4 PbRAS). The dots represent the mean measured cell numbers. Error bars indicate the 95% confidence interval based on the t-distribution. All experiments were performed by the lab of Prof. Dr. A.-K. Mueller, Centre for Infectious Diseases, Parasitology Unit, University Hospital Heidelberg

B.2 Profile likelihoods

Fig. B.2 shows the profiles for the parameters of the best fitting model, based on the calculations given in Materials and Methods 2.3.3. The profiles were used to calculate the confidence intervals given in Table 4.1.

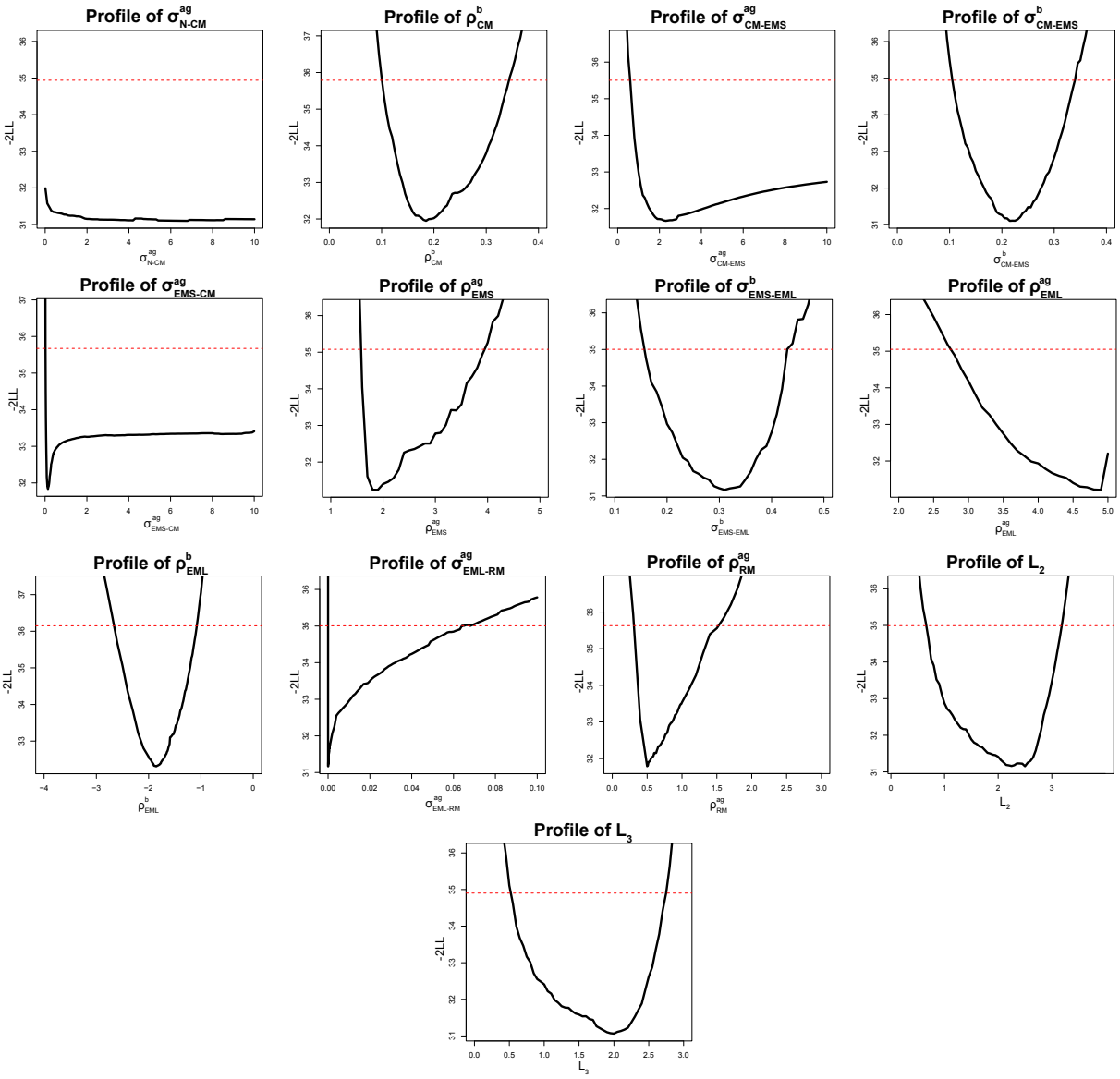

Figure B.2: The profiles of the model parameters showing the negative log-likelihood arising from keeping the respective parameter fixed to specified values. Here, confidence intervals contain all values falling below the red dotted line (see Materials and Methods 2.3.3 and [Raue et al. 2009]).

B.3 Akaike weights and evidence ratio

For the calculation of the evidence ratios and normalised probabilities see Materials and Methods 2.1.2.

parameter	evidence ratio	normalised probability
σ^{ag}_{N-CM}	44.9	0.978
σ^{ag}_{N-EMS}	0.09	0.083
ρ^{ag}_{CM}	0.068	0.064
ρ^{b}_{CM}	31000	0.999
σ^{ag}_{CM-EMS}	17000	0.999
σ^{b}_{CM-EMS}	70000	0.999
σ^{ag}_{EMS-CM}	39000	0.999
σ^{b}_{EMS-CM}	0.039	0.038
ρ^{b}_{EMS}	0.11	0.102
$\sigma^{ag}_{EMS-EML}$	0.047	0.046
$\sigma^{b}_{EMS-EML}$	41000	0.999
$\sigma^{ag}_{EML-EMS}$	42000	0.04
$\sigma^{b}_{EML-EMS}$	0.041	0.039
ρ^{b}_{EML}	21000	0.999
σ^{ag}_{EML-RM}	44.9	0.978
σ^{b}_{EML-RM}	0.04	0.039
σ^{ag}_{RM-EML}	0.042	0.041
ρ^{ag}_{RM}	15.2	0.938
ρ^{b}_{RM}	0.088	0.082
σ^{ag}_{EMS-RM}	0.066	0.062
σ^{b}_{EMS-RM}	0.042	0.041
σ^{ag}_{RM-EMS}	0.046	0.044
ρ^{ag}_{EMS} / ρ^{ag}_{EML} / L_2 / L_3	always included	1

Table B.1: The evidence ratios based on the Akaike weights showing how much more likely a model is that includes the respective parameter (see Materials and Methods 2.1.2). Akaike weights were calculated based on the 500 best models. The normalised probability shows how large the support is for the group of models containing the respective parameter given all other models. Blue colour indicates the parameters of the best fitting model.

B.4 Parameter sets used for dose fitting

All of the parameters sets below generated T_{EML} cell numbers below the threshold of 5×10^7 cells.

parameter	unit	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10
σ^{ag}_{N-CM}	day^{-1}	2.9e-05	2.2e-01	2.2e-01	5.6e-06	1.5e-05	2.8e-06	1.4e-01	1.2e-01	1.4e-06	1.1e-01
ρ^{b}_{CM}	day^{-1}	0.16	0.14	0.14	0.14	0.17	0.14	0.13	0.13	0.12	0.13
σ^{ag}_{CM-EMS}	day^{-1}	3.74	3.02	3.11	5.99	4.1	5.21	3.32	3.59	6.47	3.52
σ^{b}_{CM-EMS}	day^{-1}	0.16	0.14	0.14	0.14	0.17	0.14	0.13	0.13	0.12	0.13
σ^{ag}_{EMS-CM}	day^{-1}	0.25	0.27	0.28	0.45	0.24	0.44	0.32	0.37	0.7	0.37
ρ^{ag}_{EMS}	day^{-1}	3.2	1.95	1.95	3.4	3.3	3.5	2	2	3.6	2.03
$\sigma^{b}_{EMS-EML}$	day^{-1}	0.23	0.2	0.2	0.2	0.24	0.2	0.19	0.18	0.17	0.18
ρ^{ag}_{EML}	day^{-1}	4.99	3.7	3.6	4.99	5	4.99	3.5	3.4	4.99	3.3
ρ^{b}_{EML}	day^{-1}	-1.8	-1.62	-1.57	-1.61	-1.79	-1.53	-1.5	-1.4	-1.35	-1.43
σ^{ag}_{EML-RM}	day^{-1}	8.5e-3	5.1e-3	6.3e-3	1e-2	1e-2	1.2e-2	9.4e-3	1.2e-2	7.8e-3	1.8e-2
ρ^{ag}_{RM}	day^{-1}	0.74	0.53	0.53	0.92	0.74	0.96	0.56	0.59	1.19	0.59
L_2	days	1.06	1.52	1.5	0.89	1.04	0.84	1.42	1.36	0.72	1.34
L_3	days	0.86	1.2	1.18	0.7	0.86	0.67	1.12	1.06	0.55	1.04
T_{EML} at peak	#	4e7	4.2e7	3.2e7	3.3e7	3e7	2.8e7	2e7	1.5e7	4.3e7	9.2e6

Table B.2: The ten best parameter sets that generated T_{EML} cell numbers below the threshold of 5×10^7 cells. The corresponding peak T_{EML} cell numbers are given in the last row.

B.5 Best fit to the high dose data

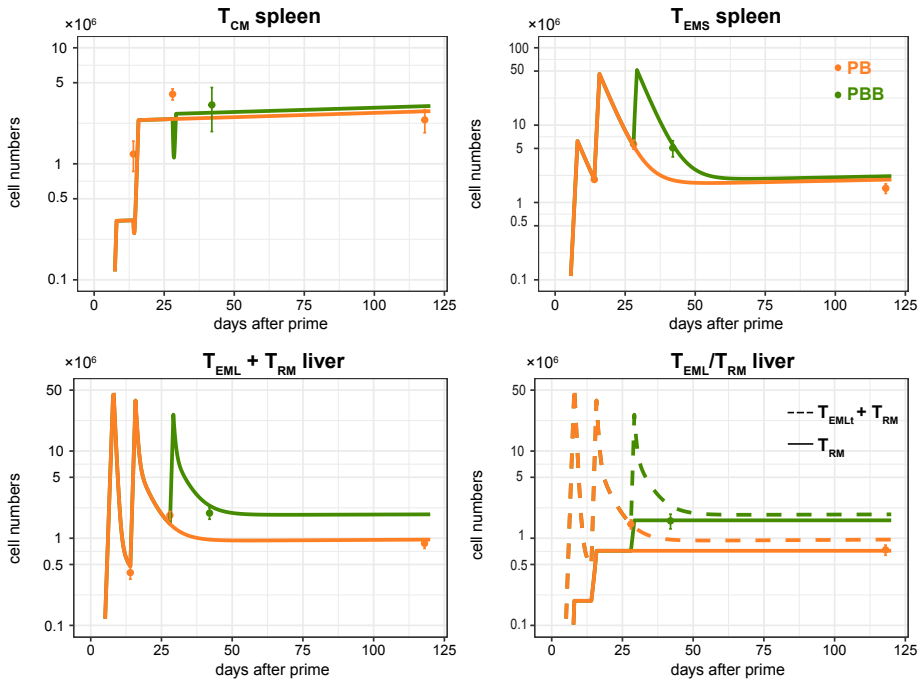

Figure B.3: Model fit to the high dose data: Panels show the cellular dynamics of T_{CM} (top left), T_{EMS} (top right), T_{EML} + T_{RM} (bottom left) and T_{EML} /T_{RM} compartments (bottom right). Lines are based on a prime-boost (PB, orange) or a prime-boost-boost (PBB, green) vaccination strategy using a high dosage (1×10^5 PbRAS). The dots represent the mean measured cell numbers. Error bars indicate the 95% confidence interval based on the t-distribution. All experiments were performed by the lab of Prof. Dr. A.-K. Mueller, Centre for Infectious Diseases, Parasitology Unit, University Hospital Heidelberg

B.6 Best fit to the sub dose data

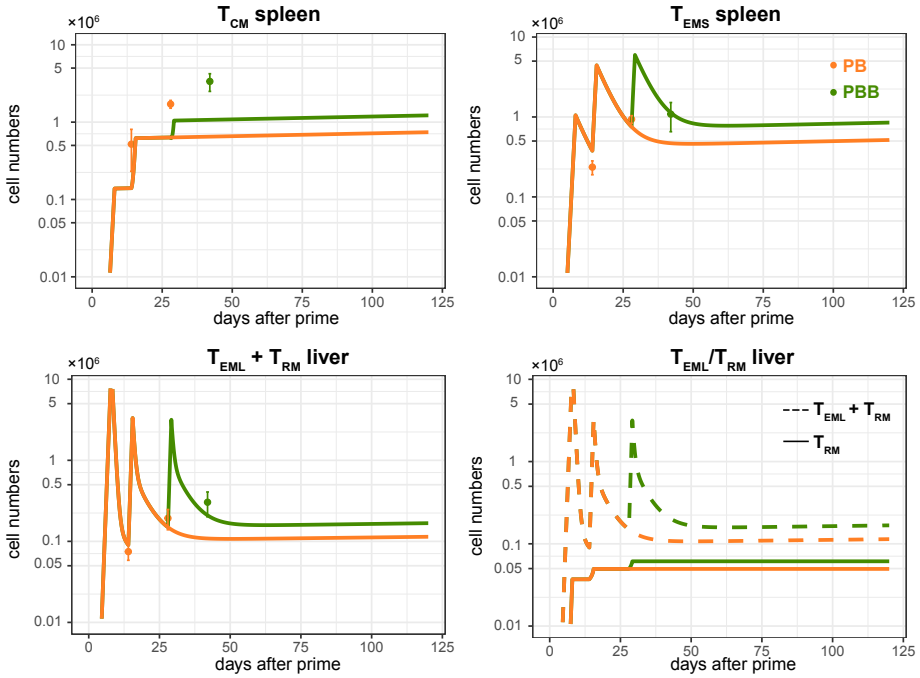

Figure B.4: Model fit to the subprotective dose data: Panels show the cellular dynamics of T_{CM} (top left), T_{EMS} (top right), $T_{EML} + T_{RM}$ (bottom left) and T_{EML} /T_{RM} compartments (bottom right). Lines are based on a prime-boost (PB, orange) or a prime-boost-boost (PBB, green) vaccination strategy using a subprotective dosage (1×10^3 PbRAS). The dots represent the mean measured cell numbers. Error bars indicate the 95% confidence interval based on the t-distribution. All experiments were performed by the lab of Prof. Dr. A.-K. Mueller, Centre for Infectious Diseases, Parasitology Unit, University Hospital Heidelberg

B.7 Comparing different dosage dynamics

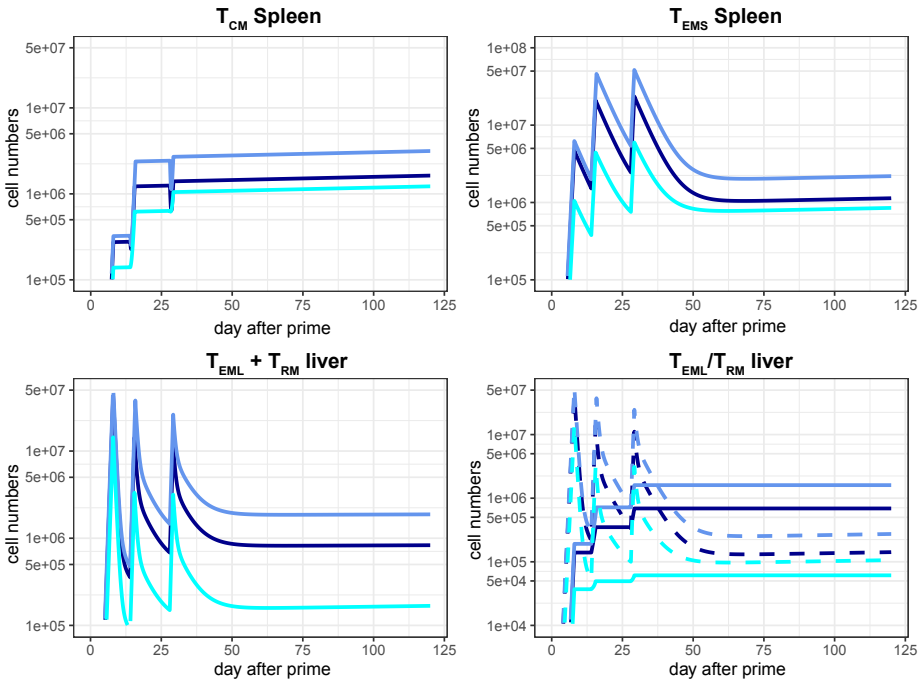

Figure B.5: Comparing the cellular dynamics of different dosages: Panels show the cellular dynamics of T_{CM} (top left), T_{EMS} (top right), T_{EML} + T_{RM} (bottom left) and T_{EML} /T_{RM} compartments (bottom right). Lines are based on a prime-boost-boost vaccination strategy using either a normal (1×10^4 PbRAS - dark blue) or high dosage (1×10^5 PbRAS - light blue). The subprotective (1×10^3 PbRAS - turquoise) was simulated to receive 11 shots in total following a 14 days vaccination scheme.

APPENDIX C

C.1 The influence of the sampling time given the homoeostatic model

Moderate changes in the sampling time do not affect the parameter estimation quality given the homoeostatic model (see Fig. C.1).

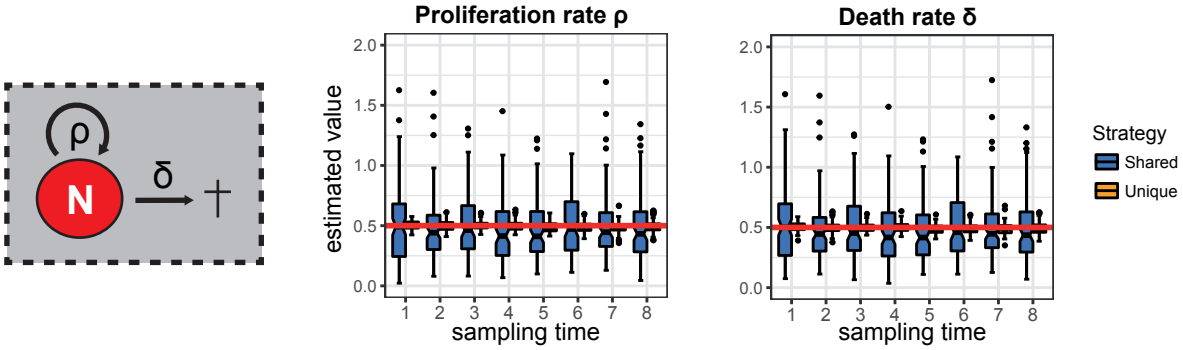

Figure C.1: The influence of the sampling time on parameter estimation given the homoeostatic model: Panels show the distribution of estimates for the proliferation rate, ρ (middle), and the death rate, δ (right), for different sampling times. Parameter estimates for two labelling strategies with $N = 800$ cells, using either a shared ($L = 8, M = 100$, blue) or an unique labelling approach ($L = 800, M = 1$, orange) are shown. The parametrisation of the model is given as $\rho = \delta = 0.5$ day^{-1} as indicated by the red horizontal lines. Figure adapted from [Gabel et al. 2017].

C.2 Solving the master equations

We will show the derivation of the master equation and the fitting procedure based on the complex expansion model (see Eqs. 5.3). Here, we also considered the death rates within the individual compartments. The other two scenarios, homoeostatic turnover (Eqs. 5.1) and simple expansion dynamics (Eqs. 5.2) can easily be derived from the following calculations by setting certain parameters equal to zero.

In the complex expansion model, we distinguish between naïve (N), central memory precursor ($\mathrm{T_{CM}}$), effector memory precursor ($\mathrm{T_{EM}}$) and effector cells ($\mathrm{T_E}$). The relation between these compartments is given by the following linear differentiation pathway:

$$\begin{aligned}
\frac{dN}{dt} &= -\mu_{\mathrm{N}}\mathrm{N} \\
\frac{d\mathrm{T_{CM}}}{dt} &= \mu_{\mathrm{N}}\mathrm{N} + (\rho_{\mathrm{CM}} - \mu_{\mathrm{CM}})\mathrm{T_{CM}} \\
\frac{d\mathrm{T_{EM}}}{dt} &= \mu_{\mathrm{CM}}\mathrm{T_{CM}} + (\rho_{\mathrm{EM}} - \mu_{\mathrm{EM}})\mathrm{T_{EM}} \\
\frac{d\mathrm{T_E}}{dt} &= \mu_{\mathrm{EM}}\mathrm{T_{EM}} + \rho_{\mathrm{E}}\mathrm{T_E}\ .
\end{aligned} \tag{C.1}$$

The system given in Eq. (C.1) can also be formulated stochastically by a continuous-time Markov process which is described by the master equations for the state probabilities $p_{k,l,m,n}(t)$ derived from the corresponding transition probabilities [Haag 2017]. The master equation for the state probability $p_{k,l,m,n}(t)$, which describes the probability that we observe $\mathrm{N} = k$, $\mathrm{T_{CM}} = l$, $\mathrm{T_{EM}} = m$ and $\mathrm{T_E} = n$ cells in the different compartments at time t, is then determined by

$$\begin{aligned}
\frac{dp_{k,l,m,n}}{dt} = {} & \mu_N(k+1)p_{k+1,l-1,m,n} + \mu_{\mathrm{CM}}(l+1)p_{k,l+1,m-1,n} + \mu_{\mathrm{EM}}(m+1)p_{k,l,m+1,n-1} \\
& + \rho_{\mathrm{CM}}(l-1)p_{k,l-1,m,n} + \rho_{\mathrm{EM}}(m-1)p_{k,l,m-1,n} + \rho_{\mathrm{E}}(n-1)p_{k,l,m,n-1} \\
& + \delta_{\mathrm{CM}}(l+1)p_{k,l+1,m,n} + \delta_{\mathrm{EM}}(m+1)p_{k,l,m+1,n} + \delta_{\mathrm{E}}(n+1)p_{k,l,m,n+1} \\
& - (\mu_{\mathrm{N}}k + (\rho_{\mathrm{CM}} + \delta_{\mathrm{CM}} + \mu_{\mathrm{CM}})l + (\rho_{\mathrm{EM}} + \delta_{\mathrm{EM}} + \mu_{\mathrm{EM}})m + (\rho_{\mathrm{E}} + \delta_{\mathrm{E}})n)\, p_{k,l}\ .
\end{aligned} \tag{C.2}$$

We can derive the probability generating function that is defined as

$$F(z_0, z_1, z_2, z_3) = \sum_{(n_0,n_1,n_2,n_3)} z_0^{n_0} z_1^{n_1} z_2^{n_2} z_3^{n_3} P(n_0, n_1, n_2, n_3; t)\ , \tag{C.3}$$

with the state vector (n_0, n_1, n_2, n_3) characterising the populations of naïve (n_0), central memory precursor (n_1), effector memory precursor (n_2) and effector cell (n_3) at time point t. Inserting Eq. (C.3) in the master equation Eq. (C.2) leads to the following partial differential equation:

$$\partial_t F(z_0, z_1, z_2, z_3) = \sum_{i=1}^{3} \mu_i(z_i - z_{i-1})\partial_{z_{i-1}}F + \delta_i(1 - z_i)\partial_{z_i}F + \rho_i(z_i^2 - z_i)\partial_{z_i}F \tag{C.4}$$

On the other hand, the probability generating function can be derived with regard to z_i and evaluated at $z_i = 1,\ i = 1, ..., 4$ which gives the first factorial moments

$$\frac{dF}{dz_i}|_{z_i=1} = \sum_{(n_0,n_1,n_2,n_3)} n_i P(n_0, n_1, n_2, n_2; t) \tag{C.5}$$

The second derivative yields the second moments, and so on. By defining $x_N := \partial z_0 F|_{z_i=1}$, $y_N := \partial z_0^2 F|_{z_i=1}$ and $c_{N,CM} := \partial z_0 z_1 F|_{z_i=1}$, ..., we can split the partial differential equation in Eq. (C.4) into a system of ordinary differential equations:

$$\begin{aligned}
\frac{dx_{\mathrm{N}}}{dt} &= -\mu_{\mathrm{CM}} x_{\mathrm{N}} \\
\frac{dx_{\mathrm{CM}}}{dt} &= \mu_{\mathrm{CM}} x_{\mathrm{N}} + (-\mu_{\mathrm{EM}} - \delta_{\mathrm{CM}} + \rho_{\mathrm{CM}}) x_{\mathrm{CM}} \\
\frac{dx_{\mathrm{EM}}}{dt} &= \mu_{\mathrm{EM}} x_{\mathrm{CM}} + (-\mu_{\mathrm{E}} - \delta_{\mathrm{EM}} + \rho_{\mathrm{EM}}) x_{\mathrm{EM}} \\
\frac{dx_{\mathrm{E}}}{dt} &= \mu_{\mathrm{E}} x_{\mathrm{EM}} + (-\delta_{\mathrm{E}} + \rho_{\mathrm{E}}) x_{\mathrm{E}}
\end{aligned}$$

$$\begin{aligned}
\frac{dy_{\mathrm{N}}}{dt} &= -\mu_{\mathrm{CM}} y_{\mathrm{N}} \\
\frac{dy_{\mathrm{CM}}}{dt} &= \mu_{\mathrm{CM}} c_{\mathrm{N,CM}} + \rho_{\mathrm{CM}} x_{\mathrm{CM}} + (-\mu_{\mathrm{EM}} - \delta_{\mathrm{CM}} + \rho_{\mathrm{CM}}) y_{\mathrm{CM}} \\
\frac{dy_{\mathrm{EM}}}{dt} &= \mu_{\mathrm{EM}} c_{\mathrm{CM,EM}} + \rho_{\mathrm{EM}} x_{\mathrm{EM}} + (-\mu_{\mathrm{E}} - \delta_{\mathrm{EM}} + \rho_{\mathrm{EM}}) y_{\mathrm{EM}} \\
\frac{dy_{\mathrm{E}}}{dt} &= \mu_{\mathrm{E}} c_{\mathrm{EM,E}} + \rho_{\mathrm{E}} x_{\mathrm{E}} + (-\delta_{\mathrm{E}} + \rho_{\mathrm{E}}) y_{\mathrm{E}}
\end{aligned} \tag{C.6}$$

$$\begin{aligned}
\frac{dc_{\mathrm{N,CM}}}{dt} &= \mu_{\mathrm{CM}} y_{\mathrm{N}} + (-\mu_{\mathrm{CM}} - \mu_{\mathrm{EM}} - \delta_{\mathrm{CM}} + \rho_{\mathrm{CM}}) c_{\mathrm{N,CM}} \\
\frac{dc_{\mathrm{N,EM}}}{dt} &= \mu_{\mathrm{EM}} c_{\mathrm{N,CM}} + (-\mu_{\mathrm{CM}} - \mu_{\mathrm{E}} - \delta_{\mathrm{EM}} + \rho_{\mathrm{EM}}) c_{\mathrm{N,EM}} \\
\frac{dc_{\mathrm{N,E}}}{dt} &= \mu_{\mathrm{E}} c_{\mathrm{N,EM}} + (-\mu_{\mathrm{CM}} - \delta_{\mathrm{CM}} + \rho_{\mathrm{CM}}) c_{\mathrm{N,E}} \\
\frac{dc_{\mathrm{CM,EM}}}{dt} &= \mu_{\mathrm{CM}} c_{\mathrm{N,EM}} + \mu_{\mathrm{EM}} y_{\mathrm{CM}} + (-\mu_{\mathrm{EM}} - \mu_{\mathrm{E}} - \delta_{\mathrm{CM}} - \delta_{\mathrm{EM}} + \rho_{\mathrm{CM}} + \rho_{\mathrm{EM}}) c_{\mathrm{CM,EM}} \\
\frac{dc_{\mathrm{CM,E}}}{dt} &= \mu_{\mathrm{CM}} c_{\mathrm{N,E}} + \mu_{\mathrm{E}} c_{\mathrm{CM,EM}} + (-\mu_{\mathrm{EM}} - \delta_{\mathrm{CM}} - \delta_{\mathrm{E}} + \rho_{\mathrm{CM}} + \rho_{\mathrm{E}}) c_{\mathrm{CM,E}} \\
\frac{dc_{\mathrm{EM,E}}}{dt} &= \mu_{\mathrm{EM}} c_{\mathrm{CM,E}} + \mu_{\mathrm{E}} y_{\mathrm{EM}} + (-\mu_{\mathrm{E}} - \delta_{\mathrm{EM}} - \delta_{\mathrm{E}} + \rho_{EM} + \rho_{\mathrm{E}}) c_{EM,\mathrm{E}} \ .
\end{aligned}$$

The solutions to these equations can be used to calculate the expected mean, the variance and the covariance of the different cell populations. For the naïve cell population, these values are obtained by putting

$$\begin{aligned}
\mathbb{E}[\mathrm{N}] &= x_{\mathrm{N}}(t) \\
\mathbf{Var}[\mathrm{N}] &= y_{\mathrm{N}}(t) + x_{\mathrm{N}}(t) - x_{\mathrm{N}}(t)^2 \\
\mathbf{Cov}[\mathrm{N}, \mathrm{CM}] &= c_{\mathrm{N,CM}}(t) - x_{\mathrm{N}}(t) x_{\mathrm{CM}}(t) \ .
\end{aligned} \tag{C.7}$$

The summary statistics used for fitting are the expected mean, the coefficients of variation (CV) and the correlation coefficients (CC). The latter two are calculated by

$$\begin{aligned} \mathbf{CV}[\mathrm{N}] &= \frac{\sqrt{\mathbf{Var}[\mathrm{N}]}}{\mathbb{E}[\mathrm{N}]} \\ \mathbf{CC}[\mathrm{N},\mathrm{CM}] &= \frac{\mathbf{Cov}[\mathrm{N},\mathrm{CM}]}{\sqrt{\mathbf{Var}[\mathrm{N}]\mathbf{Var}[\mathrm{CM}]}} \ . \end{aligned} \tag{C.8}$$

The summary statistics for the remaining cell populations are calculated accordingly.

C.3 The influence of the labelling strategy on parameter estimation

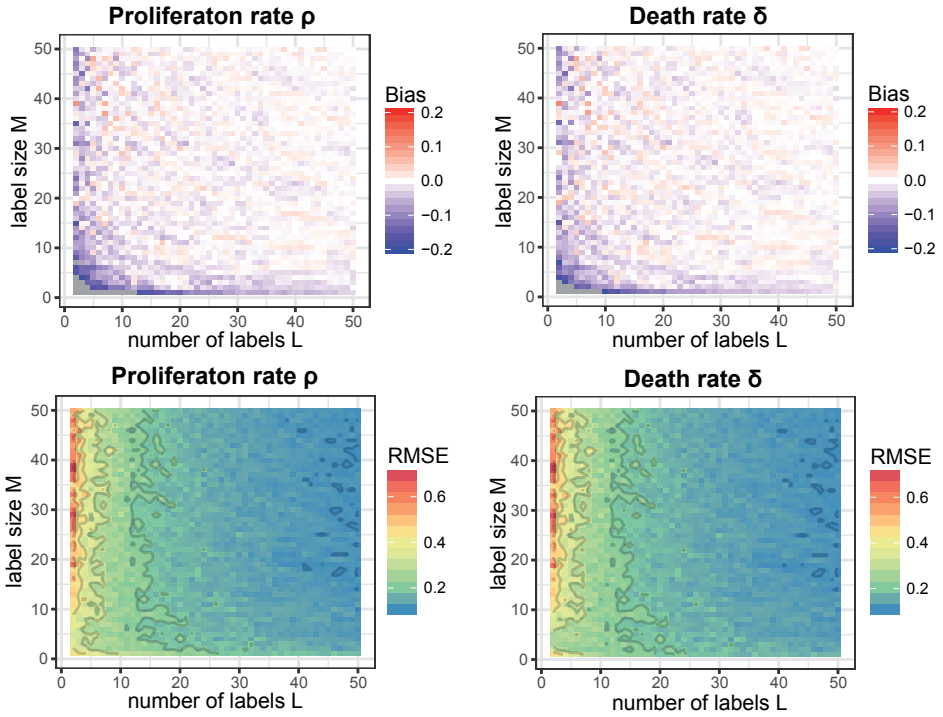

Figure C.2: The influence of the labelling strategy on parameter estimation given the homoeostatic model: Panels show the estimation quality of the proliferation rate, ρ, (left column), and the death rate, δ, (right column) with respect to the bias (top row) and RMSE (bottom row). The parametrisation of the model is given as $\rho = \delta = 0.5\ \mathrm{day}^{-1}$ and the simulated data is obtained at sampling time $T = 8$. Grey colour indicates values outside the shown scale.

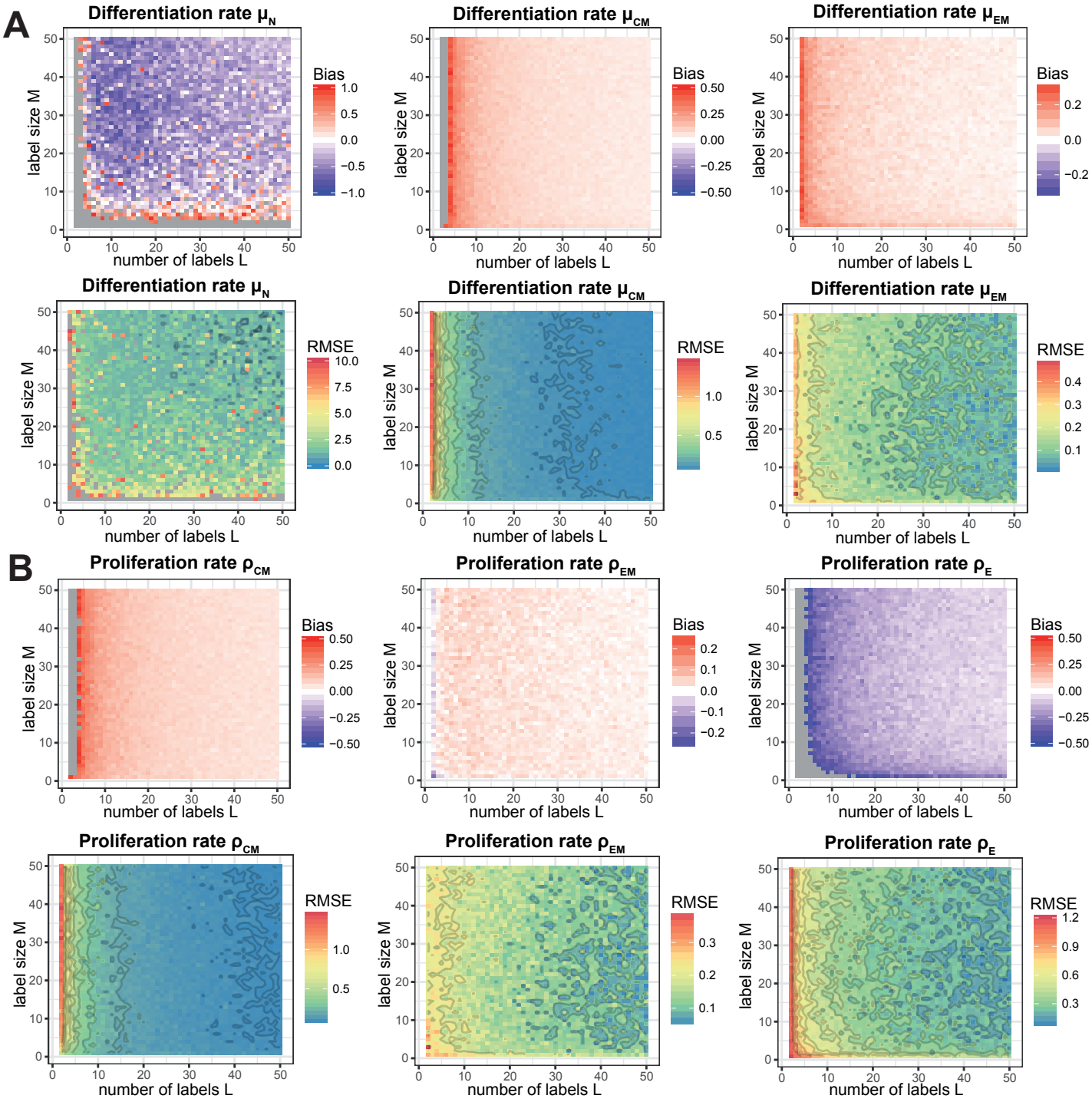

Figure C.3: The influence of the labelling strategy on parameter estimation given the complex expansion model: (A) Panels show the estimation quality of the differentiation rate, μ_{N}, μ_{CM} and μ_{EM} with respect to the bias (top row) and RMSE (bottom row). **(B)** Similar as in **(A)**, but for the proliferation rates ρ_{CM}, ρ_{EM} and ρ_{E}. The parametrisation of the model is given as $\mu_{\mathrm{N}} = 2.2$ day^{-1}, $\mu_{\mathrm{CM}} = 0.2$ day^{-1}, $\mu_{\mathrm{EM}} = 0.04$ day^{-1}, $\rho_{\mathrm{CM}} = 0.85$ day^{-1}, $\rho_{\mathrm{EM}} = 1.42$ day^{-1} and $\rho_{\mathrm{E}} = 1.6$ day^{-1} and the simulated data is obtained at sampling time $T = 8$. Grey colour indicates values outside the shown scale.

C.4 Deriving estimates from non-uniformly distributed labelling strategies

If the initial labelling strategy is not uniformly distributed, the obtained output data need to be adjusted to derive the summary statistics. The idea is to rescale all values to achieve comparability. This is possible because the calculated mean and variance of Eq. C.7 scale linearly with the initial number of transferred cells, if and only if all other cell compartments are empty at time $t = 0$. For our models this condition is fulfilled since we assumed no activated/differentiated cells exist in the beginning.

To be more exact, if X_M describes the number of further differentiated cells at sampling time T, derived from an initial population of M equally labelled cells, it holds that:

$$\mathbb{E}(X_M) = M\mathbb{E}(X_1) \tag{C.9}$$

$$\mathbf{Var}(X_M) = M\mathbf{Var}(X_1)\ . \tag{C.10}$$

This means the distribution of shared labels at time T can be derived from the calculated quantities based on a unique label.

Now considering a non-uniformly distributed labelling strategy made up of L different labels, we denote M_i as the initial number of cells of label i and X_i as the corresponding sampled cell population at time T. We define the quantities

$$E_1 := \frac{\sum_{i=1}^{L} X_i}{\sum_{i=1}^{L} M_i} \tag{C.11}$$

$$V_1 := \sum_{i=1}^{L} (X_i - M_i E_1)^2\ . \tag{C.12}$$

It is immediately clear that the quantity E_1 is an unbiased estimator of the expected value of a cell population starting with $M = 1$ cell. On the other hand, V_1 is a biased estimator as we see

by calculating

$$
\begin{aligned}
\mathbb{E}[V_1] &= \sum_{i=1}^{L} \mathbb{E}[X_i - M_i E_1[^2 = \sum_{i=1}^{L} \mathbb{E}[X_i - \mathbb{E}[X_i] + \mathbb{E}[X_i] - M_i E_1]^2 \\
&= \sum_{i=1}^{L} \mathbb{E}[(X_i - \mathbb{E}[X_i])^2 + 2(X_i - \mathbb{E}[X_i])(\mathbb{E}[X_i] - M_i E_1) + (\mathbb{E}[X_i] - M_i E_1)]^2 \\
&= \sum_{i=1}^{L} \mathrm{Var}[X_i] + \sum_{i=1}^{L} \mathbb{E}[2(X_i - \mathbb{E}[X_i])(\mathbb{E}[X_i] - M_i E_1)] + \sum_{i=1}^{L} \mathrm{Var}[M_i E_1] \\
&= M\mathrm{Var}[X_1] - \frac{2}{M} \sum_{i=1}^{L} \mathbb{E}\left[(X_i - \mathbb{E}[X_i])(M_i \sum_{j=1}^{L} X_j - M\mathbb{E}[X_i]) \right] \\
&\quad + \frac{(\sum_{i=1}^{L} M_i^2)}{M} \mathrm{Var}[X_1] \\
&= M\mathrm{Var}[X_1] - \frac{2}{M} \sum_{i=1}^{L} \mathbb{E}\left[(X_i - \mathbb{E}[X_i])(\sum_{j=1}^{L} M_i (X_j - \mathbb{E}[X_i])) \right] \\
&\quad + \frac{(\sum_{i=1}^{L} M_i^2)}{M} \mathrm{Var}[X_1] \\
&= M\mathrm{Var}[X_1] - \frac{2}{M} \sum_{i=1}^{L} M_i \mathbb{E}[(X_i - \mathbb{E}[X_i])^2] + \frac{(\sum_{i=1}^{L} M_i^2)}{M} \mathrm{Var}[X_1] \\
&= M\mathrm{Var}[X_1] - \frac{2\sum_{i=1}^{L} M_i^2}{M} \mathrm{Var}[X_1] + \frac{(\sum_{i=1}^{L} M_i^2)}{M} \mathrm{Var}[X_1] \\
&= \frac{M^2 - \sum_{i=1}^{L} M_i^2}{M} \mathrm{Var}(X_1) \ .
\end{aligned}
$$

Hence, correcting $V1$ by $\frac{M}{M^2 - \sum_{i=1}^{L} M_i^2}$ leads to an unbiased estimator of the variance of X_1, which we then used in the fitting procedure.

C.5 The influence of the sampling time on parameter estimation

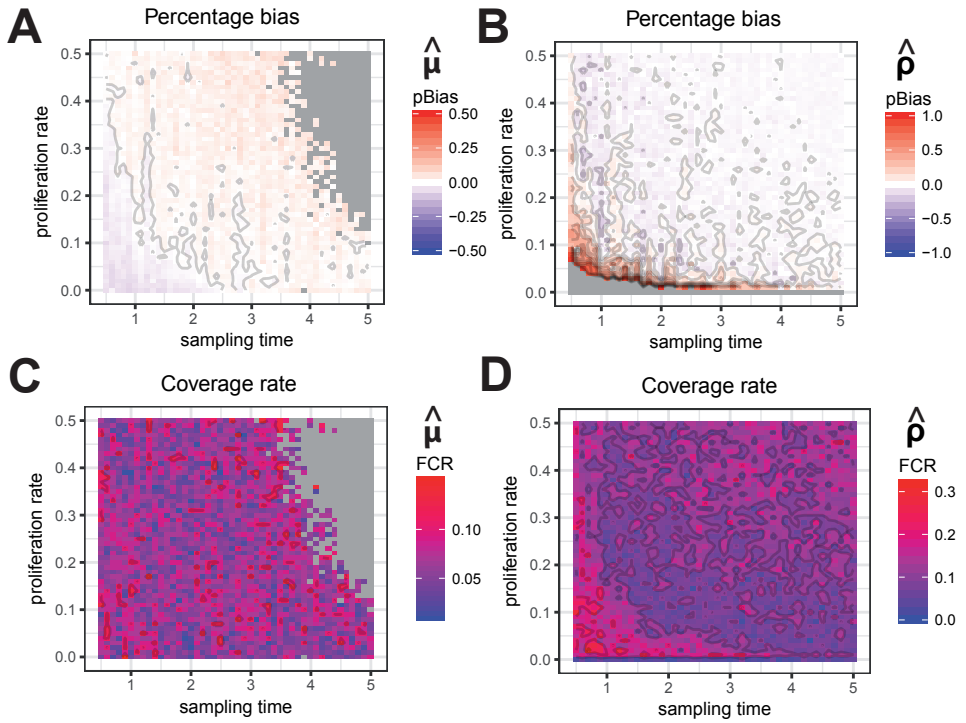

Figure C.4: Influence of the sampling time on parameter estimates given the simple expansion model: (**A** & **B**) The percentage bias for the estimation of the activation rate, $\hat{\mu}$ (**A**) and the proliferation rate, $\hat{\rho}$ (**B**) based on varying combinations of sampling times, T and proliferation rates, ρ. Grey areas in panel **B** indicate parameter combinations resulting in a percentage bias larger than one. The corresponding false coverage rates are shown in panel (**C**) and (**D**), respectively. The estimated parameter is shown above the colour scale. Figure adapted from [Gabel et al. 2017].

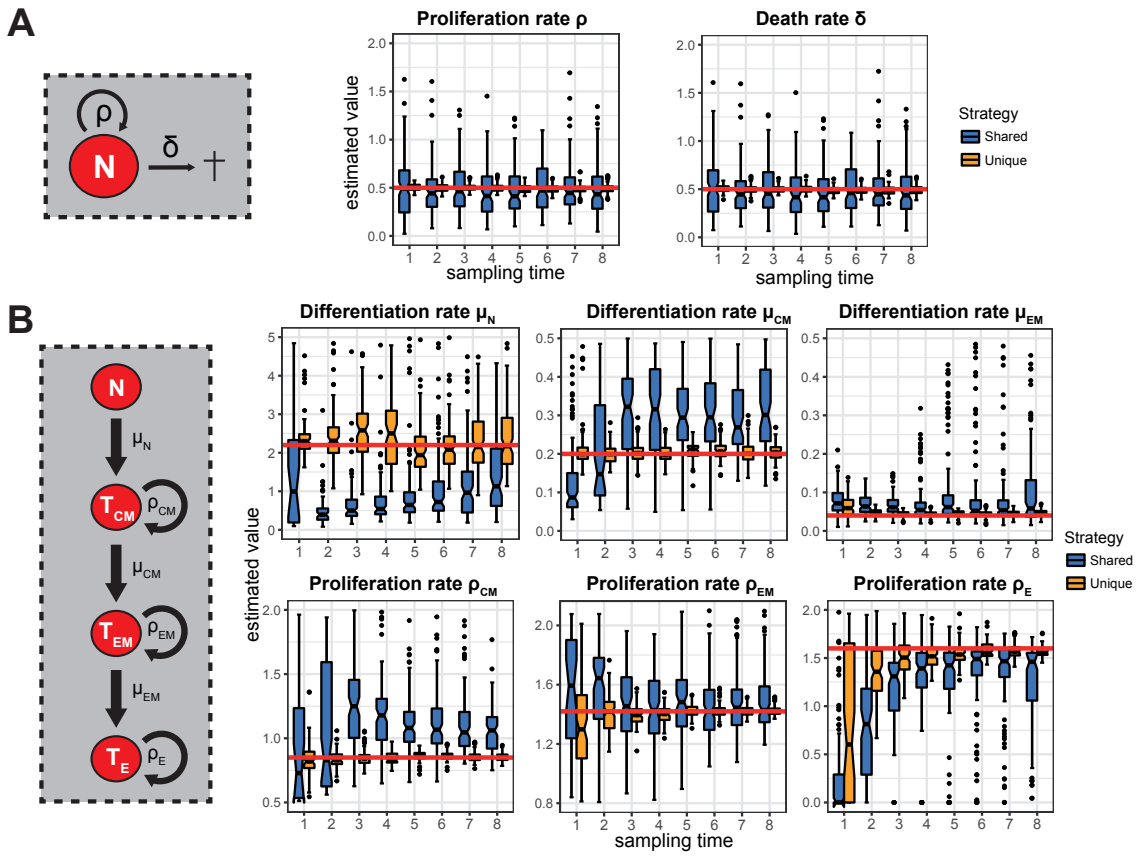

Figure C.5: Parameter estimates for the homoeostatic and complex expansion system given different sampling times: (**A**) Estimation of proliferation and death rates given the depicted homoeostatic system. (**B**) The complex expansion system with corresponding differentiation and proliferation rates. Panels show the estimated rates given different sampling times using a shared ($L = 8,\ M = 100$, blue) and a unique ($L = 800,\ M = 1$, orange) labelling strategy. Here, no loss in transfer or sampling was assumed. Every boxplot is based on the results of 100 independent stochastic simulations. Red lines indicate the true parameter values.

C.6 Adjusting for transfer loss

Losing cells during adoptive transfer can be considered as taking a sample from the transferred cell population, i.e. the initial population of labelled cells. The presumable number of cells that each labelled population contributes to the transferred labelling strategy can be calculated by using the hypergeometric distribution, which needs three input variables: the number of cells in the chosen label, M, the overall number of cells, N and the considered fraction of cells, p, that will survive the transfer. While M and N are known from the initial labelled population, the transfer fraction p needs to be experimentally determined. If all values are known, we can calculate the expected number and variance of the transferred cells by

$$\begin{aligned} \mathbb{E}(\#\text{ of transferred cells}) &= pM \\ \mathbf{Var}(\#\text{ of transferred cells}) &= p(1-p)\frac{M(N-M)}{N-1}\ . \end{aligned} \tag{C.13}$$

Using Eqs. C.7, these values can be used as presumable initial conditions for the system C.6. For results based on adjusted transfer fractions see Figs. 5.5 and C.6.

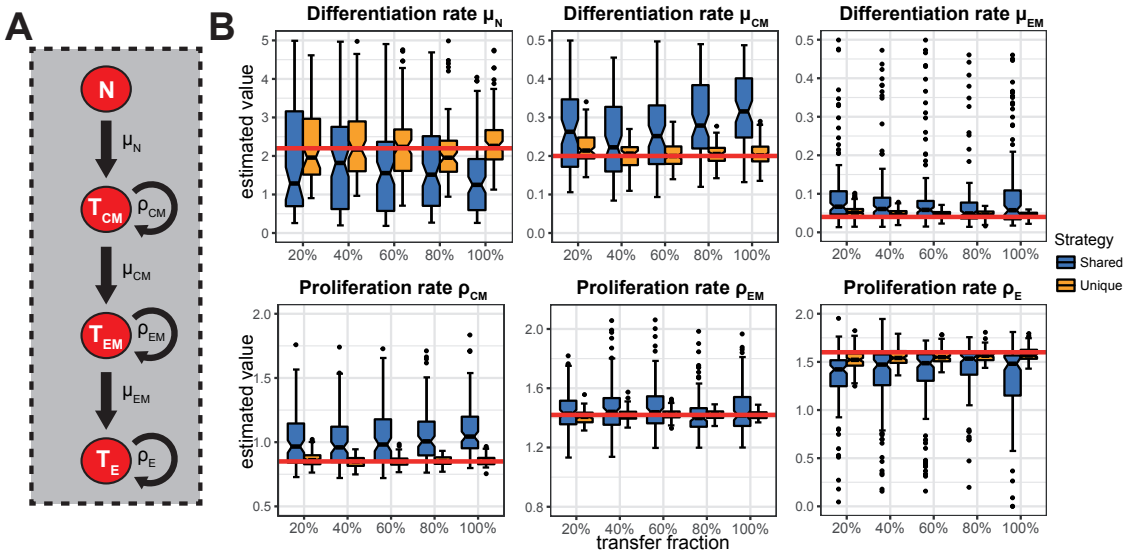

Figure C.6: Parameter estimates for the complex expansion system corrected by the pre-known transfer fraction: (**A**) The complex expansion system with corresponding differentiation and proliferation rates (**B**) Panels show the estimated rates given different transfer fractions using a shared ($L = 8,\ M = 100$, blue) and a unique ($L = 800,\ M = 1$, orange) labelling strategy. Here, we accounted for the loss in transfer. Every boxplot is based on the results of 100 independent stochastic simulations. Red lines indicate the true parameter values. Figure adapted from [Gabel et al. 2017].

C.7 Adjusting for sampling loss

Assuming the sampling fraction, p, is known, then all measured cell numbers can be rescaled by dividing by the sampling fraction.

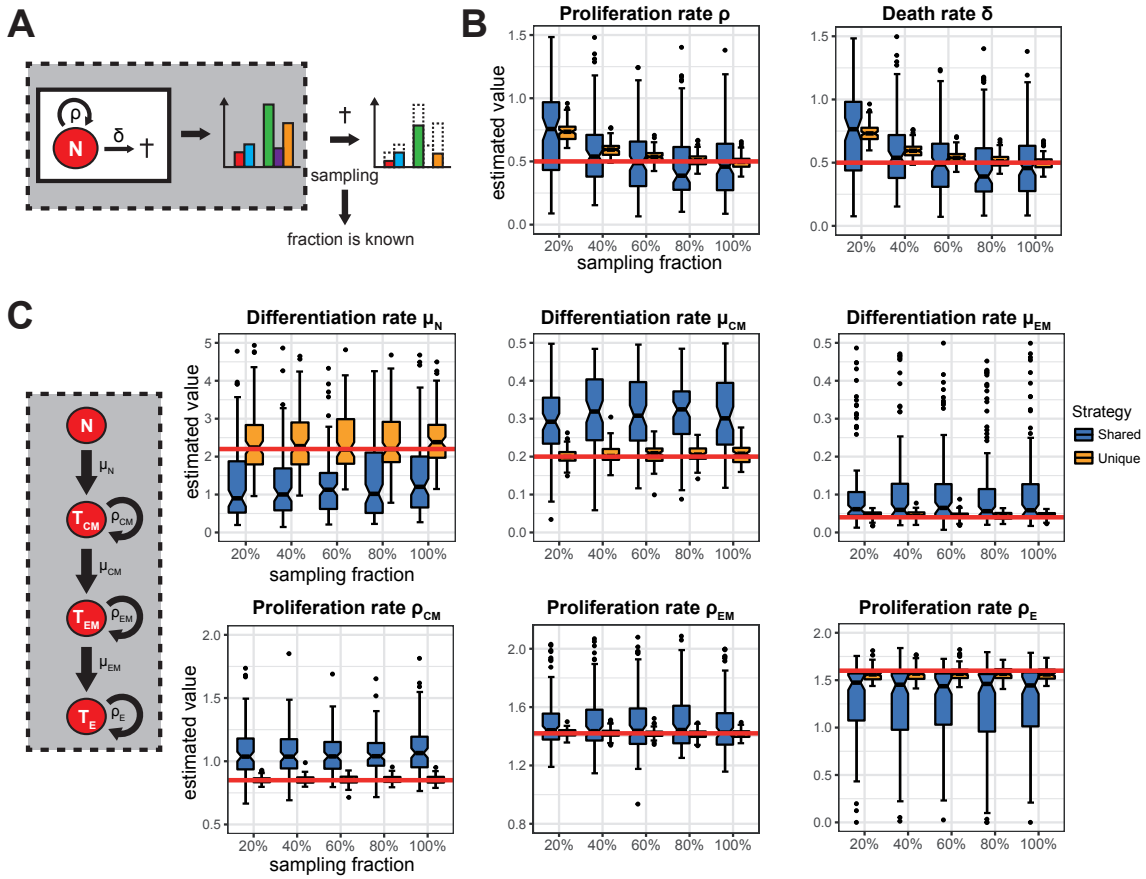

Figure C.7: Parameter estimates for the homoeostatic and the complex expansion system corrected by the pre-known sampling fraction: (**A**) Schematic depicting the problem of incomplete sampling: Only a fraction of the labelled cells is sampled and can be used for analysis. Here, the sampling fraction is known. (**B**) Estimates for proliferation and death rate given the homoeostatic system corrected by the sampling fraction. (**C**) The complex expansion system with corresponding differentiation and proliferation rates. Panels show the estimated rates given different transfer fractions using a shared ($L = 8$, $M = 100$, blue) and a unique ($L = 800$, $M = 1$, orange) labelling strategy. Here, we accounted for the loss in sampling. Every boxplot is based on the results of 100 independent stochastic simulations. Red lines indicate the true parameter values. Figure adapted from [Gabel et al. 2017].

C.8 Failed parameter estimation due to negative correlation

The bias observed in parameter estimates for the shared labelling approach ($L = 8,\ M = 100$) given the complex expansion model is a direct consequence of the inability of the model to reproduce negative correlation coefficients with respect to the central memory precursor cells and the effector compartments. Here, a negative correlation between T_{CM} and T_{EM} /T_E compartments resulted in an impaired parameter estimation (see Fig. C.8, left and middle panel). No other correlation between the summary statistics and the goodness of the fit was found, not even for the correlation of T_{EM} and T_E compartments (see Fig. C.8, right panel).

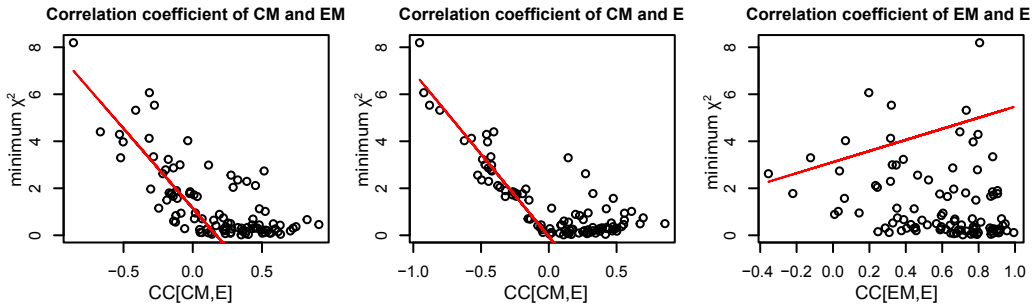

Figure C.8: A negative correlation is responsible for the bias observed in the parameter estimates: Panels show the dependency of the correlation coefficients (**CC**[CM, EM], **CC**[CM, E] and **CC**[EM, E]) and the estimation quality with respect to the minimum χ^2 value. Red lines show a linear fit to all data points having a negative correlation coefficient (i.e. **CC**< 0)

Acknowledgments

Five years ago, the last thing I wanted to do was a PhD.

At that time, I was just finishing my Diploma thesis and was sure I had seen enough of the academic world. However, there was a new group leader coming to lunch with us and he was very persistent about presenting his research projects to me. After stalling him (unsuccessfully) for quite some time, I reluctantly agreed to take a look at the topics he was working on. To be honest, I mostly did it to put an end to his frequent attempts of recruiting me into his group. So I went there, being sure that this was the end of it. I was wrong. As it turned out, it was only the beginning.

Five years later, as I write these lines, I'm about to finish my PhD thesis and I'm again sure I have seen enough of the academic world. However, looking back at the last years I spent working at the university, I have to admit that I had an amazing time. I worked on interesting projects, learned tons of new stuff and met a whole bunch of nice and fun people.

Therefore, my first and most sincerest thanks has to go to Frederik, whose persistence made me change my mind about starting a PhD.

Now, I'm glad I did it.

Besides his remarkable persistence, I have to thank Frederik for being the best supervisor I could have wished for. He was always available to discuss my problems and helped me finding solutions for them. He was one of the most useful sources for all biological and mathematical questions. He made me change my point of view when I was being too negative and critical about my work. But most importantly, he always genuinely cared about my work and I think that's what made all the difference for me and my PhD in the end.

I would also like to thank Prof. Dr. Ursula Kummer for always taking her time guiding me through the pitfalls of university bureaucracy and Prof. Dr. Thomas Höfer for giving me the chance to present my research in our joined immunology meetings.

Furthermore, I want to thank my collaborators on the experimental side, who spend countless hours in the lab gathering the data I needed for my mathematical evaluations. Therefore, my deepest gratitude goes to Nicolas Baumann and Annette Oxenius for providing the MCMV data and to Roland Frank and Ann-Kristin Mueller for the Malaria project. I especially want to thank the latter two for including us mathematical modellers in the process of setting up the biological experiments, which was a great experience.

Additionally, I would like to thank the HGS MathComp graduate school, not only for providing travel funds, but also for hosting many interesting and fun events, at which I met lots of nice people and friends. Also, speaking of university institutions, the Hochschulsport facilities,

the Cafe Botanik and the Botanical garden absolutely deserve a shout-out from my side.

In my experience, it's not important how you spend your time but who you're spending it with. Therefore, I was exceptionally lucky during my times at Bioquant as I had amazing company. Handing out my personal thanks, I would like to start with my fellow group members (in order of their appearance): I would like to thank Pete, for providing many helpful (and some Scotty) ideas, for being the sheriff and superhero the people of Bioquant need, and for having the second longest name I've heard in my whole life; Neha, for teaching me how to make Indian lemonade, how to dance Indian wedding dances and how to politely ask for office supplies - the Indian way ("Michael! Scissors!"); Sam, for being my companion in many botanical endeavours and ant-astic adventures, both of which started of as mindless fun and somehow turned into serial killings along the way.

Continuing with non-group members that earned a special nomination in these acknowledgments for various reasons: Thanks are going to Jan, for being one of the oldest and most dedicated members of Coffee Break Club (Rule number one: You always call for Coffee Break Club!); Arne, for making sure I was prominently included in the Bioquant&Jan WhatsApp group, for turning our cupboard into a creepy shrine and for making me laugh by falling down the stairs; Vineet, for having the most fashionable hair and for saying the most ridiculous things imaginable (there's still a blog about both somewhere in the depths of the internet) and Priyata, for being a distant colleague all along.

However, that's only the tip of the iceberg. Many other people need to be mentioned. Therefore, here's the complete list of my PhD crew (given in alphabetical order): Aline, Alistair, Anika, Anna, Annika, Asha, Astrid, Carina, Diego, Dimitra, Don Juan, Enoch, Fabian, Fereydoon, Gesa, Iva, Jana, Jonas, Juliane, Lea, Lilija, Malin, Martin, Mia, Nadia, Niklas, Nils, Paola, Philipp, Priscila, Rebekka, Sara, Sarah, Sophia, Stefano, Susanne, Susanne, Tobias, Ulli, Verena, Victoria and Yue.

Last but not least, a big thanks has to go to all my pre-existing friends and family, who supported me along the way.

I apologise to everybody I might have forgotten to include in these acknowledgements (I'm pretty sure, there are some). However, if things are continuing as before, I will make up for it by mentioning you in my habilitation ;-)

Bibliography

Abdulla, S. et al. (2008). "Safety and immunogenicity of RTS,S/AS02D malaria vaccine in infants". In: *N. Engl. J. Med.* 359.24, pp. 2533–2544.

Agnandji, S. T. et al. (2014). "Efficacy and safety of the RTS,S/AS01 malaria vaccine during 18 months after vaccination: a phase 3 randomized, controlled trial in children and young infants at 11 African sites". In: *PLoS Med.* 11.7, e1001685.

Ahmed, R., M. J. Bevan, S. L. Reiner, and D. T. Fearon (2009). "The precursors of memory: models and controversies". In: *Nat. Rev. Immunol.* 9.9, pp. 662–668.

Aiaenssens, E.M et al. (2017). "Taxonomy of prokaryotic viruses: 2016 update from the ICTV bacterial and archaeal viruses subcommittee". English. In: *Archives of Virology* 162.4, pp. 1153–1157. ISSN: 0304-8608 and 1432-8798.

Akondy, R. S. et al. (2017). "Origin and differentiation of human memory CD8 T cells after vaccination". In: *Nature*.

Allan, R. S., J. Waithman, S. Bedoui, C. M. Jones, J. A. Villadangos, Y. Zhan, A. M. Lew, K. Shortman, W. R. Heath, and F. R. Carbone (2006). "Migratory dendritic cells transfer antigen to a lymph node-resident dendritic cell population for efficient CTL priming". In: *Immunity* 25.1, pp. 153–162.

Althaus, C. L., V. V. Ganusov, and R. J. De Boer (2007). "Dynamics of CD8+ T cell responses during acute and chronic lymphocytic choriomeningitis virus infection". In: *J. Immunol.* 179.5, pp. 2944–2951.

Antia, R., C. T. Bergstrom, S. S. Pilyugin, S. M. Kaech, and R. Ahmed (2003). "Models of CD8+ responses: 1. What is the antigen-independent proliferation program". In: *J. Theor. Biol.* 221.4, pp. 585–598.

Arias, C. F., M. A. Herrero, F. J. Acosta, and C. Fernandez-Arias (2014). "A mathematical model for a T cell fate decision algorithm during immune response". In: *J. Theor. Biol.* 349, pp. 109–120.

Arsenio, J., B. Kakaradov, P. J. Metz, S. H. Kim, G. W. Yeo, and J. T. Chang (2014). "Early specification of CD8+ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses". In: *Nat. Immunol.* 15.4, pp. 365–372.

Avriel, Mordecai (1976). *Nonlinear programming. analysis and methods.* eng. Prentice-Hall series in automatic computation. Englewood Cliffs, NJ: Prentice-Hall, XV, 512 S. ISBN: 0-13-623603-0 and 978-0-13-623603-0.

Bachmann, M. F., P. Wolint, K. Schwarz, P. Jager, and A. Oxenius (2005a). "Functional properties and lineage relationship of CD8+ T cell subsets identified by expression of IL-7 receptor alpha and CD62L". In: *J. Immunol.* 175.7, pp. 4686–4696.

Bachmann, M. F., P. Wolint, K. Schwarz, and A. Oxenius (2005b). "Recall proliferation potential of memory CD8+ T cells and antiviral protection". In: *J. Immunol.* 175.7, pp. 4677–4685.

Bainov, D. D. and P. S. Simeonov (1993). "Impulsive differential equations: periodic solutions and applications". In: *Longman Scientific and Technical*.

Bassett, J. D., S. L. Swift, H. VanSeggelen, J. A. Hammill, A. J. McGray, C. Evelegh, Y. Wan, and J. L. Bramson (2012). "Combined mTOR inhibition and OX40 agonism enhances CD8(+) T cell memory and protective immunity produced by recombinant adenovirus vaccines". In: *Mol. Ther.* 20.4, pp. 860–869.

Berkson, Joseph (1980). *Minimum Chi-Square, not Maximum Likelihood*. English. Institute of Mathematical Statistics, pp. 457–487.

Blattman, J. N., D. J. Sourdive, K. Murali-Krishna, R. Ahmed, and J. D. Altman (2000). "Evolution of the T cell repertoire during primary, memory, and recall responses to viral infection". In: *J. Immunol.* 165.11, pp. 6081–6090.

Bojang, K., P. Milligan, M. Pinder, T. Doherty, A. Leach, O. Ofori-Anyinam, M. Lievens, K. Kester, K. Schaecher, W. R. Ballou, and J. Cohen (2009). "Five-year safety and immunogenicity of GlaxoSmithKline's candidate malaria vaccine RTS,S/AS02 following administration to semi-immune adult men living in a malaria-endemic region of The Gambia". In: *Hum Vaccin* 5.4, pp. 242–247.

Bolinger, B., S. Sims, G. O'Hara, C. de Lara, E. Tchilian, S. Firner, D. Engeler, B. Ludewig, and P. Klenerman (2013). "A new model for CD8+ T cell memory inflation based upon a recombinant adenoviral vector". In: *J. Immunol.* 190.8, pp. 4162–4174.

Borghans, J. and R. M. Ribeiro (2017). "The maths of memory". In: *Elife* 6.

Bouchnita, A, G Bocharov, A Meyerhans, and V Volpert (2017). "Hybrid approach to model the spatial regulation of T cell responses". English. In: *BMC IMMUNOLOGY* 18.Suppl 1, 12 S. ISSN: 1471-2172 and 1471-2172.

Bouneaud, C., Z. Garcia, P. Kourilsky, and C. Pannetier (2005). "Lineage relationships, homeostasis, and recall capacities of central- and effector-memory CD8 T cells in vivo". In: *J. Exp. Med.* 201.4, pp. 579–590.

Boyman, O., J. H. Cho, and J. Sprent (2010). "The role of interleukin-2 in memory CD8 cell differentiation". In: *Adv. Exp. Med. Biol.* 684, pp. 28–41.

Brune, W. and C. E. Andoniou (2017). "Die Another Day: Inhibition of Cell Death Pathways by Cytomegalovirus". In: *Viruses* 9.9.

Buchholz, V. R., M. Flossdorf, I. Hensel, L. Kretschmer, B. Weissbrich, P. Graf, A. Verschoor, M. Schiemann, T. Hofer, and D. H. Busch (2013). "Disparate individual fates compose robust CD8+ T cell immunity". In: *Science* 340.6132, pp. 630–635.

Burnham, Kenneth P. and David Raymond Anderson (2003). *Model selection and multimodel inference. a practical information theoretic approach*. eng. 2. ed., [corr. print.] Literaturverz. S. 455 - 484. New York ; Berlin ; Heidelberg [u.a.]: Springer, XXVI, 488 S. ISBN: 0-387-95364-7 and 978-0-387-95364-9.

Burton, A., D. G. Altman, P. Royston, and R. L. Holder (2006). "The design of simulation studies in medical statistics". In: *Stat Med* 25.24, pp. 4279–4292.

Campbell, A. E., V. J. Cavanaugh, and J. S. Slater (2008). "The salivary glands as a privileged site of cytomegalovirus immune evasion and persistence". In: *Med. Microbiol. Immunol.* 197.2, pp. 205–213.

Carpenter, J. and J. Bithell (2000). "Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians". In: *Stat Med* 19.9, pp. 1141–1164.

Chervin, A. S., J. D. Stone, C. M. Soto, B. Engels, H. Schreiber, E. J. Roy, and D. M. Kranz (2013). "Design of T-cell receptor libraries with diverse binding properties to examine adoptive T-cell responses". In: *Gene Ther.* 20.6, pp. 634–644.

Chotivanich, K., R. Udomsangpetch, R. McGready, S. Proux, P. Newton, S. Pukrittayakamee, S. Looareesuwan, and N. J. White (2002). "Central role of the spleen in malaria parasite clearance". In: *J. Infect. Dis.* 185.10, pp. 1538–1541.

Clem, A. S. (2011). "Fundamentals of vaccine immunology". In: *J Glob Infect Dis* 3.1, pp. 73–78.

Cohen, J., V. Nussenzweig, R. Nussenzweig, J. Vekemans, and A. Leach (2010). "From the circumsporozoite protein to the RTS, S/AS candidate vaccine". In: *Hum Vaccin* 6.1, pp. 90–96.

Colugnati, F. A., S. A. Staras, S. C. Dollard, and M. J. Cannon (2007). "Incidence of cytomegalovirus infection among the general population and pregnant women in the United States". In: *BMC Infect. Dis.* 7, p. 71.

Crauste, F., J. Mafille, L. Boucinha, S. Djebali, O. Gandrillon, J. Marvel, and C. Arpin (2017). "Identification of Nascent Memory CD8 T Cells and Modeling of Their Ontogeny". In: *Cell Syst* 4.3, pp. 306–317.

Currie, J., M. Castro, G. Lythe, E. Palmer, and C. Molina-Paris (2012). "A stochastic T cell response criterion". In: *J R Soc Interface* 9.76, pp. 2856–2870.

Davison, A. J., R. Eberle, B. Ehlers, G. S. Hayward, D. J. McGeoch, A. C. Minson, P. E. Pellett, B. Roizman, M. J. Studdert, and E. Thiry (2009). "The order Herpesvirales". In: *Arch. Virol.* 154.1, pp. 171–177.

De Boer, R. J., M. Oprea, R. Antia, K. Murali-Krishna, R. Ahmed, and A. S. Perelson (2001). "Recruitment times, proliferation, and apoptosis rates during the CD8(+) T-cell response to lymphocytic choriomeningitis virus". In: *J. Virol.* 75.22, pp. 10663–10669.

De Boer, R. J., D. Homann, and A. S. Perelson (2003). "Different dynamics of CD4+ and CD8+ T cell responses during and after acute lymphocytic choriomeningitis virus infection". In: *J. Immunol.* 171.8, pp. 3928–3935.

De Boer, R. J. and A. S. Perelson (2013). "Quantifying T lymphocyte turnover". In: *J. Theor. Biol.* 327, pp. 45–87.

Dekhtiarenko, I. et al. (2016). "Peptide Processing Is Critical for T-Cell Memory Inflation and May Be Optimized to Improve Immune Protection by CMV-Based Vaccine Vectors". In: *PLoS Pathog.* 12.12, e1006072.

Delyon, Bernard, Marc Lavielle, and Eric Moulines (1999). "Convergence of a Stochastic Approximation Version of the EM Algorithm". English. In: *The Annals of Statistics* 27.1, pp. 94–128. ISSN: 0090-5364 and 2168-8966.

Doll, K. L., L. L. Pewe, S. P. Kurup, and J. T. Harty (2016). "Discriminating Protective from Nonprotective Plasmodium-Specific CD8+ T Cell Responses". In: *J. Immunol.* 196.10, pp. 4253–4262.

Farber, D. L., N. A. Yudanin, and N. P. Restifo (2014). "Human memory T cells: generation, compartmentalization and homeostasis". In: *Nat. Rev. Immunol.* 14.1, pp. 24–35.

Fernandez-Ruiz, D. et al. (2016). "Liver-Resident Memory CD8+ T Cells Form a Front-Line Defense against Malaria Liver-Stage Infection". In: *Immunity* 45.4, pp. 889–902.

Frevert, U. (2004). "Sneaking in through the back entrance: the biology of malaria liver stages". In: *Trends Parasitol.* 20.9, pp. 417–424.

Fruh, K. and L. Picker (2017). "CD8+ T cell programming by cytomegalovirus vectors: applications in prophylactic and therapeutic vaccination". In: *Curr. Opin. Immunol.* 47, pp. 52–56.

Gabel, M., R. R. Regoes, and F. Graw (2017). "More or less-On the influence of labelling strategies to infer cell population dynamics". In: *PLoS ONE* 12.10, e0185523.

Ganusov, V. V. and J. Auerbach (2014). "Mathematical modeling reveals kinetics of lymphocyte recirculation in the whole organism". In: *PLoS Comput. Biol.* 10.5, e1003586.

Gebhardt, T., L. M. Wakim, L. Eidsmo, P. C. Reading, W. R. Heath, and F. R. Carbone (2009). "Memory T cells in nonlymphoid tissue that provide enhanced local immunity during infection with herpes simplex virus". In: *Nat. Immunol.* 10.5, pp. 524–530.

Gerlach, C., J. W. van Heijst, E. Swart, D. Sie, N. Armstrong, R. M. Kerkhoven, D. Zehn, M. J. Bevan, K. Schepers, and T. N. Schumacher (2010). "One naive T cell, multiple fates in CD8+ T cell differentiation". In: *J. Exp. Med.* 207.6, pp. 1235–1246.

Gerlach, C., J. C. Rohr, L. Perie, N. van Rooij, J. W. van Heijst, A. Velds, J. Urbanus, S. H. Naik, H. Jacobs, J. B. Beltman, R. J. de Boer, and T. N. Schumacher (2013). "Heterogeneous differentiation patterns of individual CD8+ T cells". In: *Science* 340.6132, pp. 635–639.

Gilbert, S. C. (2012). "T-cell-inducing vaccines - what's the future". In: *Immunology* 135.1, pp. 19–26.

Gillespie, Daniel, Yang Cao, and Linda Petzold (2006). "Efficient step size selection for the tau-leaping simulation method". English. In: *Journal of Chemical Physics* 124.4, pp. 044109–044109–11. ISSN: 0021-9606 ; 1520-9032. DOI: 10.1063/1.2159468.

Haag, Günter (2017). *Modelling with the Master Equation. Solution Methods and Applications in Social and Natural Sciences.* eng. SpringerLink : Bücher. Cham: Springer, Online–Ressource (XXV, 342 p. 109 illus., 19 illus. in color, online resource). ISBN: 978-3-319-60300-1. DOI: 10.1007/978-3-319-60300-1. URL: http://dx.doi.org/10.1007/978-3-319-60300-1.

Haan, J. M. den, S. M. Lehar, and M. J. Bevan (2000). "CD8(+) but not CD8(-) dendritic cells cross-prime cytotoxic T cells in vivo". In: *J. Exp. Med.* 192.12, pp. 1685–1696.

Hafalla, J. C., O. Silvie, and K. Matuschewski (2011). "Cell biology and immunology of malaria". In: *Immunol. Rev.* 240.1, pp. 297–316.

Hafalla, J. C., K. Bauza, J. Friesen, G. Gonzalez-Aseguinolaza, A. V. Hill, and K. Matuschewski (2013). "Identification of targets of CD8+ T cell responses to malaria liver stages by genome-wide epitope profiling". In: *PLoS Pathog.* 9.5, e1003303.

Hansen, S. G., J. C. Ford, M. S. Lewis, A. B. Ventura, C. M. Hughes, L. Coyne-Johnson, N. Whizin, K. Oswald, R. Shoemaker, T. Swanson, A. W. Legasse, M. J. Chiuchiolo, C. L. Parks, M. K. Axthelm, J. A. Nelson, M. A. Jarvis, M. Piatak, J. D. Lifson, and L. J. Picker (2011). "Profound early control of highly pathogenic SIV by an effector memory T-cell vaccine". In: *Nature* 473.7348, pp. 523–527.

Hansen, S. G. et al. (2013a). "Cytomegalovirus vectors violate CD8+ T cell epitope recognition paradigms". In: *Science* 340.6135, p. 1237874.

Hansen, S. G. et al. (2013b). "Immune clearance of highly pathogenic SIV infection". In: *Nature* 502.7469, pp. 100–104.

Hellerstein, M., M. B. Hanley, D. Cesar, S. Siler, C. Papageorgopoulos, E. Wieder, D. Schmidt, R. Hoh, R. Neese, D. Macallan, S. Deeks, and J. M. McCune (1999). "Directly measured kinetics of circulating T lymphocytes in normal and HIV-1-infected humans". In: *Nat. Med.* 5.1, pp. 83–89.

Hengel, H., W. Brune, and U. H. Koszinowski (1998). "Immune evasion by cytomegalovirus–survival strategies of a highly adapted opportunist". In: *Trends Microbiol.* 6.5, pp. 190–197.

Hertoghs, K. M., P. D. Moerland, A. van Stijn, E. B. Remmerswaal, S. L. Yong, P. J. van de Berg, S. M. van Ham, F. Baas, I. J. ten Berge, and R. A. van Lier (2010). "Molecular profiling of cytomegalovirus-induced human CD8+ T cell differentiation". In: *J. Clin. Invest.* 120.11, pp. 4077–4090.

Hoffman, S. L., L. M. Goh, T. C. Luke, I. Schneider, T. P. Le, D. L. Doolan, J. Sacci, P. de la Vega, M. Dowler, C. Paul, D. M. Gordon, J. A. Stoute, L. W. Church, M. Sedegah, D. G. Heppner, W. R. Ballou, and T. L. Richie (2002). "Protection of humans against malaria by immunization with radiation-attenuated Plasmodium falciparum sporozoites". In: *J. Infect. Dis.* 185.8, pp. 1155–1164.

Holz, L. E., D. Fernandez-Ruiz, and W. R. Heath (2016). "Protective immunity to liver-stage malaria". In: *Clin Transl Immunology* 5.10, e105.

Huster, K. M., V. Busch, M. Schiemann, K. Linkemann, K. M. Kerksiek, H. Wagner, and D. H. Busch (2004). "Selective expression of IL-7 receptor on memory T cells identifies early CD40L-dependent generation of distinct CD8+ memory T cell subsets". In: *Proc. Natl. Acad. Sci. U.S.A.* 101.15, pp. 5610–5615.

Hutchinson, S., S. Sims, G. O'Hara, J. Silk, U. Gileadi, V. Cerundolo, and P. Klenerman (2011). "A dominant role for the immunoproteasome in CD8+ T cell responses to murine cytomegalovirus". In: *PLoS ONE* 6.2, e14646.

Jameson, S. C. and D. Masopust (2009). "Diversity in T cell memory: an embarrassment of riches". In: *Immunity* 31.6, pp. 859–871.

Jung, Y. W., R. L. Rutishauser, N. S. Joshi, A. M. Haberman, and S. M. Kaech (2010). "Differential localization of effector and memory CD8 T cell subsets in lymphoid organs during acute viral infection". In: *J. Immunol.* 185.9, pp. 5315–5325.

Kaech, S. M. and E. J. Wherry (2007). "Heterogeneity and cell-fate decisions in effector and memory CD8+ T cell differentiation during viral infection". In: *Immunity* 27.3, pp. 393–405.

Kaiser, P., E. Slack, A. J. Grant, W. D. Hardt, and R. R. Regoes (2013). "Lymph node colonization dynamics after oral Salmonella Typhimurium infection in mice". In: *PLoS Pathog.* 9.9, e1003532.

Kambayashi, T., E. Assarsson, A. E. Lukacher, H. G. Ljunggren, and P. E. Jensen (2003). "Memory CD8+ T cells provide an early source of IFN-gamma". In: *J. Immunol.* 170.5, pp. 2399–2408.

Karrer, U., S. Sierro, M. Wagner, A. Oxenius, H. Hengel, U. H. Koszinowski, R. E. Phillips, and P. Klenerman (2003). "Memory inflation: continuous accumulation of antiviral CD8+ T cells over time". In: *J. Immunol.* 170.4, pp. 2022–2029.

Kearney, E. R., K. A. Pape, D. Y. Loh, and M. K. Jenkins (1994). "Visualization of peptide-specific T cell immunity and peripheral tolerance induction in vivo". In: *Immunity* 1.4, pp. 327–339.

Kim, J., A. R. Kim, and E. C. Shin (2015). "Cytomegalovirus Infection and Memory T Cell Inflation". In: *Immune Netw* 15.4, pp. 186–190.

Klenerman P. ; Oxenius, A. (2016). "T cell responses to cytomegalovirus". In: *Nature Reviews Immunology.*

Kohler, B. (2007). "Mathematically modeling dynamics of T cell responses: predictions concerning the generation of memory cells". In: *J. Theor. Biol.* 245.4, pp. 669–676.

Korber, B. T., N. L. Letvin, and B. F. Haynes (2009). "T-cell vaccine strategies for human immunodeficiency virus, the virus with a thousand faces". In: *J. Virol.* 83.17, pp. 8300–8314.

Koup, R. A. and D. C. Douek (2011). "Vaccine design for CD8 T lymphocyte responses". In: *Cold Spring Harb Perspect Med* 1.1, a007252.

Krammer, P. H., R. Arnold, and I. N. Lavrik (2007). "Life and death in peripheral T cells". In: *Nat. Rev. Immunol.* 7.7, pp. 532–542.

Krmpotic, A., D. H. Busch, I. Bubi?, F. Gebhardt, H. Hengel, M. Hasan, A. A. Scalzo, U. H. Koszinowski, and S. Jonji? (2002). "MCMV glycoprotein gp40 confers virus resistance to CD8+ T cells and NK cells in vivo". In: *Nat. Immunol.* 3.6, pp. 529–535.

Kuby, Janis (1997). *Immunology.* eng. 3. ed. New York: Freeman, XXIV, 664 S. ISBN: 0-7167-2868-0 and 978-0-7167-2868-9.

Kurts, C., B. W. Robinson, and P. A. Knolle (2010). "Cross-priming in health and disease". In: *Nat. Rev. Immunol.* 10.6, pp. 403–414.

Kurz, S. K. and M. J. Reddehase (1999). "Patchwork pattern of transcriptional reactivation in the lungs indicates sequential checkpoints in the transition from murine cytomegalovirus latency to recurrence". In: *J. Virol.* 73.10, pp. 8612–8622.

Lin, M. Y. and R. M. Welsh (1998). "Stability and diversity of T cell receptor repertoire usage during lymphocytic choriomeningitis virus infection of mice". In: *J. Exp. Med.* 188.11, pp. 1993–2005.

Liu, J., K. L. O'Brien, D. M. Lynch, N. L. Simmons, A. La Porte, A. M. Riggs, P. Abbink, R. T. Coffey, L. E. Grandpre, M. S. Seaman, G. Landucci, D. N. Forthal, D. C. Montefiori, A. Carville, K. G. Mansfield, M. J. Havenga, M. G. Pau, J. Goudsmit, and D. H. Barouch (2009). "Immune control of an SIV challenge by a T-cell-based vaccine in rhesus monkeys". In: *Nature* 457.7225, pp. 87–91.

Loewendorf, A. I., R. Arens, J. F. Purton, C. D. Surh, and C. A. Benedict (2011). "Dissecting the requirements for maintenance of the CMV-specific memory T-cell pool". In: *Viral Immunol.* 24.4, pp. 351–355.

Luciani, F., S. Valensin, R. Vescovini, P. Sansoni, F. Fagnoni, C. Franceschi, M. Bonafe, and G. Turchetti (2001). "A stochastic model for CD8(+)T cell dynamics in human immunosenescence: implications for survival and longevity". In: *J. Theor. Biol.* 213.4, pp. 587–597.

Lyons, A. B. (2000). "Analysing cell division in vivo and in vitro using flow cytometric measurement of CFSE dye dilution". In: *J. Immunol. Methods* 243.1-2, pp. 147–154.

Mahnke, Y. D., T. M. Brodie, F. Sallusto, M. Roederer, and E. Lugli (2013). "The who's who of T-cell differentiation: human memory T-cell subsets". In: *Eur. J. Immunol.* 43.11, pp. 2797–2809.

Maryanski, J. L., C. V. Jongeneel, P. Bucher, J. L. Casanova, and P. R. Walker (1996). "Single-cell PCR analysis of TCR repertoires selected by antigen in vivo: a high magnitude CD8 response is comprised of very few clones". In: *Immunity* 4.1, pp. 47–55.

Masopust, D., V. Vezys, E. J. Wherry, and R. Ahmed (2007). "A brief history of CD8 T cells". In: *Eur. J. Immunol.* 37 Suppl 1, S103–110.

Masopust, D. and J. M. Schenkel (2013). "The integration of T cell migration, differentiation and function". In: *Nat. Rev. Immunol.* 13.5, pp. 309–320.

Mohandas, N. and X. An (2012). "Malaria and human red blood cells". In: *Med. Microbiol. Immunol.* 201.4, pp. 593–598.

Mohri, H., S. Bonhoeffer, S. Monard, A. S. Perelson, and D. D. Ho (1998). "Rapid turnover of T lymphocytes in SIV-infected rhesus macaques". In: *Science* 279.5354, pp. 1223–1227.

Mohri, H., A. S. Perelson, K. Tung, R. M. Ribeiro, B. Ramratnam, M. Markowitz, R. Kost, A. Hurley, L. Weinberger, D. Cesar, M. K. Hellerstein, and D. D. Ho (2001). "Increased turnover of T lymphocytes in HIV-1 infection and its reduction by antiretroviral therapy". In: *J. Exp. Med.* 194.9, pp. 1277–1287.

Moon, J. J., H. H. Chu, J. Hataye, A. J. Pagan, M. Pepper, J. B. McLachlan, T. Zell, and M. K. Jenkins (2009). "Tracking epitope-specific T cells". In: *Nat Protoc* 4.4, pp. 565–581.

Morrison, C. (2015). "Landmark green light for Mosquirix malaria vaccine". In: *Nat. Biotechnol.* 33.10, pp. 1015–1016.

Munks, M. W., K. S. Cho, A. K. Pinto, S. Sierro, P. Klenerman, and A. B. Hill (2006). "Four distinct patterns of memory CD8 T cell responses to chronic murine cytomegalovirus infection". In: *J. Immunol.* 177.1, pp. 450–458.

Naik, S. H., T. N. Schumacher, and L. Perie (2014). "Cellular barcoding: a technical appraisal". In: *Exp. Hematol.* 42.8, pp. 598–608.

Obar, J. J., K. M. Khanna, and L. Lefrancois (2008). "Endogenous naive CD8+ T cell precursor frequency regulates primary and memory responses to infection". In: *Immunity* 28.6, pp. 859–869.

O'Hara, G. A., S. P. Welten, P. Klenerman, and R. Arens (2012). "Memory T cell inflation: understanding cause and effect". In: *Trends Immunol.* 33.2, pp. 84–90.

Park, J. H., Q. Yu, B. Erman, J. S. Appelbaum, D. Montoya-Durango, H. L. Grimes, and A. Singer (2004). "Suppression of IL7Ralpha transcription by IL-7 and other prosurvival

cytokines: a novel mechanism for maximizing IL-7-dependent T cell survival". In: *Immunity* 21.2, pp. 289–302.

Patel, H., N. Yadav, R. Parmar, S. Patel, A. P. Singh, N. Shrivastava, and S. K. Dalai (2017). "Frequent inoculations with radiation attenuated sporozoite is essential for inducing sterile protection that correlates with a threshold level of Plasmodia liver-stage specific CD8+ T cells". In: *Cell. Immunol.* 317, pp. 48–54.

Perie, L., P. D. Hodgkin, S. H. Naik, T. N. Schumacher, R. J. de Boer, and K. R. Duffy (2014). "Determining lineage pathways from cellular barcoding experiments". In: *Cell Rep* 6.4, pp. 617–624.

Pinheiro, José C. and Douglas M. Bates (1995). "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model". English. In: *Journal of Computational and Graphical Statistics* 4.1, pp. 12–35. ISSN: 1061-8600 and 1537-2715.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Ratner, A. and W. R. Clark (1993). "Role of TNF-alpha in CD8+ cytotoxic T lymphocyte-mediated lysis". In: *J. Immunol.* 150.10, pp. 4303–4314.

Raue, A., C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmuller, and J. Timmer (2009). "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood". In: *Bioinformatics* 25.15, pp. 1923–1929.

Raue, H. P., C. Beadling, J. Haun, and M. K. Slifka (2013). "Cytokine-mediated programmed proliferation of virus-specific CD8(+) memory T cells". In: *Immunity* 38.1, pp. 131–139.

Reddehase, M. J. (2002). "Antigens and immunoevasins: opponents in cytomegalovirus immune surveillance". In: *Nat. Rev. Immunol.* 2.11, pp. 831–844.

Reddehase, M. J., J. Podlech, and N. K. Grzimek (2002). "Mouse models of cytomegalovirus latency: overview". In: *J. Clin. Virol.* 25 Suppl 2, pp. 23–36.

Redeker, A., S. P. Welten, and R. Arens (2014). "Viral inoculum dose impacts memory T-cell inflation". In: *Eur. J. Immunol.* 44.4, pp. 1046–1057.

Ribeiro, R. M., H. Mohri, D. D. Ho, and A. S. Perelson (2002). "In vivo dynamics of T cell activation, proliferation, and death in HIV-1 infection: why are CD4+ but not CD8+ T cells depleted?" In: *Proc. Natl. Acad. Sci. U.S.A.* 99.24, pp. 15572–15577.

Roberts, A. D., K. H. Ely, and D. L. Woodland (2005). "Differential contributions of central and effector memory T cells to recall responses". In: *J. Exp. Med.* 202.1, pp. 123–133.

Rocha, B. and C. Tanchot (2006). "The Tower of Babel of CD8+ T-cell memory: known facts, deserted roads, muddy waters, and possible dead ends". In: *Immunol. Rev.* 211, pp. 182–196.

Sachs, J. and P. Malaney (2002). "The economic and social burden of malaria". In: *Nature* 415.6872, pp. 680–685.

Sallusto, F., J. Geginat, and A. Lanzavecchia (2004). "Central memory and effector memory T cell subsets: function, generation, and maintenance". In: *Annu. Rev. Immunol.* 22, pp. 745–763.

Sallusto, F., A. Lanzavecchia, K. Araki, and R. Ahmed (2010). "From vaccines to memory and back". In: *Immunity* 33.4, pp. 451–463.

Schepers, K., E. Swart, J. W. van Heijst, C. Gerlach, M. Castrucci, D. Sie, M. Heimerikx, A. Velds, R. M. Kerkhoven, R. Arens, and T. N. Schumacher (2008). "Dissecting T cell lineage relationships by cellular barcoding". In: *J. Exp. Med.* 205.10, pp. 2309–2318.

Schmidt, N. W., R. L. Podyminogin, N. S. Butler, V. P. Badovinac, B. J. Tucker, K. S. Bahjat, P. Lauer, A. Reyes-Sandoval, C. L. Hutchings, A. C. Moore, S. C. Gilbert, A. V. Hill, L. C. Bartholomay, and J. T. Harty (2008). "Memory CD8 T cell responses exceeding a large but definable threshold provide long-term immunity to malaria". In: *Proc. Natl. Acad. Sci. U.S.A.* 105.37, pp. 14017–14022.

Schmidt, N. W., N. S. Butler, V. P. Badovinac, and J. T. Harty (2010). "Extreme CD8 T cell requirements for anti-malarial liver-stage immunity following immunization with radiation attenuated sporozoites". In: *PLoS Pathog.* 6.7, e1000998.

Schmidt, N. W. and J. T. Harty (2011). "Cutting edge: attrition of Plasmodium-specific memory CD8 T cells results in decreased protection that is rescued by booster immunization". In: *J. Immunol.* 186.7, pp. 3836–3840.

Schumacher, T. N., C. Gerlach, and J. W. van Heijst (2010). "Mapping the life histories of T cells". In: *Nat. Rev. Immunol.* 10.9, pp. 621–631.

Seckert, C. K., A. Renzaho, H. M. Tervo, C. Krause, P. Deegen, B. Kuhnapfel, M. J. Reddehase, and N. K. Grzimek (2009). "Liver sinusoidal endothelial cells are a site of murine cytomegalovirus latency and reactivation". In: *J. Virol.* 83.17, pp. 8869–8884.

Seckert, C. K., S. I. Schader, S. Ebert, D. Thomas, K. Freitag, A. Renzaho, J. Podlech, M. J. Reddehase, and R. Holtappels (2011). "Antigen-presenting cells of haematopoietic origin prime cytomegalovirus-specific CD8 T-cells but are not sufficient for driving memory inflation during viral latency". In: *J. Gen. Virol.* 92.Pt 9, pp. 1994–2005.

Seckert, C. K., M. Griessl, J. K. Buttner, S. Scheller, C. O. Simon, K. A. Kropp, A. Renzaho, B. Kuhnapfel, N. K. Grzimek, and M. J. Reddehase (2012). "Viral latency drives 'memory inflation': a unifying hypothesis linking two hallmarks of cytomegalovirus infection". In: *Med. Microbiol. Immunol.* 201.4, pp. 551–566.

Seder, R. A. et al. (2013). "Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine". In: *Science* 341.6152, pp. 1359–1365.

Shen, F. W., Y. Saga, G. Litman, G. Freeman, J. S. Tung, H. Cantor, and E. A. Boyse (1985). "Cloning of Ly-5 cDNA". In: *Proc. Natl. Acad. Sci. U.S.A.* 82.21, pp. 7360–7363.

Sierro, S., R. Rothkopf, and P. Klenerman (2005). "Evolution of diverse antiviral CD8+ T cell populations after murine cytomegalovirus infection". In: *Eur. J. Immunol.* 35.4, pp. 1113–1123.

Sims, S. and P. Klenerman (2015). "Increasing inflationary T-cell responses following transient depletion of MCMV-specific memory T cells". In: *Eur. J. Immunol.* 45.1, pp. 113–118.

Smith, C. J., H. Turula, and C. M. Snyder (2014). "Systemic hematogenous maintenance of memory inflation by MCMV infection". In: *PLoS Pathog.* 10.7, e1004233.

Smith, C. M., N. S. Wilson, J. Waithman, J. A. Villadangos, F. R. Carbone, W. R. Heath, and G. T. Belz (2004). "Cognate CD4(+) T cell licensing of dendritic cells in CD8(+) T cell immunity". In: *Nat. Immunol.* 5.11, pp. 1143–1148.

Snyder, C. M., K. S. Cho, E. L. Bonnett, S. van Dommelen, G. R. Shellam, and A. B. Hill (2008). "Memory inflation during chronic viral infection is maintained by continuous production of short-lived, functional T cells". In: *Immunity* 29.4, pp. 650–659.

Snyder, C. M., K. S. Cho, E. L. Bonnett, J. E. Allan, and A. B. Hill (2011). "Sustained CD8+ T cell memory inflation after infection with a single-cycle cytomegalovirus". In: *PLoS Pathog.* 7.10, e1002295.

Sohlenius-Sternbeck, A. K. (2006). "Determination of the hepatocellularity number for human, dog, rabbit, rat and mouse livers from protein concentration measurements". In: *Toxicol In Vitro* 20.8, pp. 1582–1586.

Steinert, E. M., J. M. Schenkel, K. A. Fraser, L. K. Beura, L. S. Manlove, B. Z. Igyarto, P. J. Southern, and D. Masopust (2015). "Quantifying Memory CD8 T Cells Reveals Regionalization of Immunosurveillance". In: *Cell* 161.4, pp. 737–749.

Stemberger, C., M. Neuenhahn, V. R. Buchholz, and D. H. Busch (2007). "Origin of CD8+ effector and memory T cell subsets". In: *Cell. Mol. Immunol.* 4.6, pp. 399–405.

Stone, J. D., A. S. Chervin, and D. M. Kranz (2009). "T-cell receptor binding affinities and kinetics: impact on T-cell activity and specificity". In: *Immunology* 126.2, pp. 165–176.

Stromberg, S. P. and R. Antia (2012). "On the role of CD8 T cells in the control of persistent infections". In: *Biophys. J.* 103.8, pp. 1802–1810.

Surh, C. D. and J. Sprent (2008). "Homeostasis of naive and memory T cells". In: *Immunity* 29.6, pp. 848–862.

Thom, J. T., T. C. Weber, S. M. Walton, N. Torti, and A. Oxenius (2015). "The Salivary Gland Acts as a Sink for Tissue-Resident Memory CD8(+) T Cells, Facilitating Protection from Local Cytomegalovirus Infection". In: *Cell Rep* 13.6, pp. 1125–1136.

Thom, J. T. and A. Oxenius (2016). "Tissue-resident memory T cells in cytomegalovirus infection". In: *Curr Opin Virol* 16, pp. 63–69.

Torti, N., S. M. Walton, K. M. Murphy, and A. Oxenius (2011a). "Batf3 transcription factor-dependent DC subsets in murine CMV infection: differential impact on T-cell priming and memory inflation". In: *Eur. J. Immunol.* 41.9, pp. 2612–2618.

Torti, N., S. M. Walton, T. Brocker, T. Rulicke, and A. Oxenius (2011b). "Non-hematopoietic cells in lymph nodes drive memory CD8 T cell inflation during murine cytomegalovirus infection". In: *PLoS Pathog.* 7.10, e1002313.

Tough, D. F. and J. Sprent (1994). "Turnover of naive- and memory-phenotype T cells". In: *J. Exp. Med.* 179.4, pp. 1127–1135.

Turner, S. J., G. Diaz, R. Cross, and P. C. Doherty (2003). "Analysis of clonotype distribution and persistence for an influenza virus-specific CD8+ T cell response". In: *Immunity* 18.4, pp. 549–559.

Vieira Braga, F. A., K. M. Hertoghs, R. A. van Lier, and K. P. van Gisbergen (2015). "Molecular characterization of HCMV-specific immune responses: Parallels between CD8(+) T cells, CD4(+) T cells, and NK cells". In: *Eur. J. Immunol.* 45.9, pp. 2433–2445.

Walton, S. M., N. Torti, S. Mandaric, and A. Oxenius (2011). "T-cell help permits memory CD8(+) T-cell inflation during cytomegalovirus latency". In: *Eur. J. Immunol.* 41.8, pp. 2248–2259.

Wherry, E. J., V. Teichgraber, T. C. Becker, D. Masopust, S. M. Kaech, R. Antia, U. H. von Andrian, and R. Ahmed (2003). "Lineage relationship and protective immunity of memory CD8 T cell subsets". In: *Nat. Immunol.* 4.3, pp. 225–234.

WHO (2016). "World Malaria Report 2016". In: *World Health Organization*.

Wong, J., D. Mathis, and C. Benoist (2007). "TCR-based lineage tracing: no evidence for conversion of conventional into regulatory T cells in response to a natural self-antigen in pancreatic islets". In: *J. Exp. Med.* 204.9, pp. 2039–2045.

Woodland, D. L. and J. E. Kohlmeier (2009). "Migration, maintenance and recall of memory T cells in peripheral tissues". In: *Nat. Rev. Immunol.* 9.3, pp. 153–161.

Yates, A., C. Chan, J. Strid, S. Moon, R. Callard, A. J. George, and J. Stark (2007). "Reconstruction of cell population dynamics using CFSE". In: *BMC Bioinformatics* 8, p. 196.

Yoon, H., T. S. Kim, and T. J. Braciale (2010). "The cell cycle time of CD8+ T cells responding in vivo is controlled by the type of antigenic stimulus". In: *PLoS ONE* 5.11, e15423.

Zarnitsyna, V. I., B. D. Evavold, L. N. Schoettle, J. N. Blattman, and R. Antia (2013). "Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire". In: *Front Immunol* 4, p. 485.

Zehn, D., S. Y. Lee, and M. J. Bevan (2009). "Complete but curtailed T-cell response to very low-affinity antigen". In: *Nature* 458.7235, pp. 211–214.

Zhang, S., J. Xiang, S. Theuns, L. M. Desmarets, I. Trus, and H. J. Nauwynck (2016). "MCMV exploits the spleen as a transfer hub for systemic dissemination upon oronasal inoculation". In: *Virus Res.* 217, pp. 47–54.