Modélisation et inférence Bayésienne pour la maturation d'affinité des anticorps

Marco Molari

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Ecole Normale Supérieure

Modeling and Bayesian inference for antibody affinity maturation

Soutenue par
Marco Molari
Le 29 Octobre 2020

École doctorale n°564
Physique en Île-de-France

Spécialité
Physique

Composition du jury :

Frédéric van Wijland Université de Paris	*Président du jury*
Arup K. Chakraborty Massachusetts Institute of Technology	*Rapporteur*
Armita Nourmohammad University of Washington	*Rapporteur*
Claude Loverdo Sorbonne Université	*Examinatrice*
Marta Łuksza Icahn School of Medicine at Mount Sinai	*Examinatrice*
Simona Cocco École Normale Supérieure	*Directrice de thèse*

1794
ENS | PSL

ACKNOWLEDGEMENTS

It is my belief that, as the African proverb goes, "it takes a village to raise a child". This thesis, which I sometimes jokingly call "my child", is no different, and I truly feel the need to thank my village for raising it with me.

First and foremost, my gratitude goes to my PhD supervisors Simona Cocco and Rémi Monasson. I still remember clearly the day I first came to Paris, as a freshly-graduated master student, ready to start my PhD. With care and patience you guided me through the transition from student to young scientist, teaching me a great deal of subjects while at the same time showing me how honest and high-quality research is made. I will always be grateful for this.

I also want to thank my collaborators Jean Baudry and Klaus Eyer. I cannot overstate how much I learnt from our many interactions. And my sincere gratitude to the jury members Claude Loverdo, Marta Łuksza, Frédéric van Wijland and especially to Arup Chakraborty and Armita Nourmohammad, for their deep review of this work and the interesting discussions we had.

Moving abroad is always a bit of a challenge, but I had the good fortune of meeting many people who made it all more than worthwhile. So many fond moments come to mind and I know that I cannot possibly recall them all here. Just know that these memories are very precious to me, and not at all taken for granted. A huge thank-you to Lorenzo, who was with me even before the start of this adventure. Thank you for the time we spent together, the scientific discussions and the life advices. Your passion in work and life is contagious. The same huge thank-you to Diego, who always provided the most interesting discussion topics and infused me with his love for bikes. Thanks to Jérôme, for his rare combination of kindness and intelligence. Thanks to my thesis-companion Aldo, enduring the last few PhD months together made it all the more fun. Thanks to all of my colleagues in the group: Arnaud, Andrea, Clément, Cyril, Barbara, Elisabetta, Francesca, Jorge, Sébastien, Tobias. And to all the other colleagues in the lab: Natanael, Victor, Jacopo, Max, Carlos, Maria, Meriem, Giulio. All of our scientific and non-scientific interactions were a precious gift for me. Thanks to all of the beautiful people who in these years shared the office space with me: Louis, Quentin, Simone, Elie, Victor, Anna, Ling, Ludwig, Eduarda, Fabio. You made room E227a the best working place. A big thank-you to our coffee-provider Kevin for organizing the PhD and Postdoc seminar with me during these years, for our amazing boardgame sessions and tennis matches. Thanks to Federica for her kindness and the skying trip together. Thanks to Mary, who was there since my first years in Cité Universitaire. Thanks to Juliane and Tridib, who are for me an example of true care for others and love for each other. And the biggest thank-you to my little family of 8 rue Labat: Raly, Ani, Sam and the honorary coloc Guillaume. You became my home far from home.

I also owe my gratitude to many beautiful people in San Marino and Italy, who were able to support me in spite of the distance, and who gave me more strength than they know. Thanks to my parents and to Martina, my safe harbor every time I was back. Thanks to Matteo and Simone, who are an inspiration for me in so many different ways. Thanks to Samuele for checking on me with his good-morning messages. Thanks to Alice, for our long calls that brightened my working days. Thanks to Matteo, Lisa, Silvia and Luca, whom I hold dear since our years in Bologna. Thanks to Ilaria, who did half of this journey with me, and without whom this journey would not even have started. And thanks to Elia, who is the closest thing I have to a brother (I promise I'll do my best to abide by the contract). And thanks to many more whom I could not possibly list here, but that have always been there, always.

And at last, thanks to Fanny. With your constant love and support you were there for me through every page I wrote. Thank you for putting up with your Trauerkloß, and making each step lighter. You make me all the more grateful to have come to Paris.

ABSTRACT

Affinity Maturation (AM) is a biological process through which our Immune System generates potent Antibodies (Abs) against newly encountered pathogens. This process is also at the base of vaccination, one of the most successful and cost-effective medical procedures ever developed, responsible for saving millions of lives every year. At its core AM works by subjecting B-lymphocites to iterative cycles of mutation and selection for Antigen (Ag) binding, thus generating a Darwinian evolutionary process that leads to a progressive increase of their binding affinity for the Ag. The mechanisms that govern AM are very complex, involving non-linear interactions between different cell-types, and still present many open questions whose answers have the potential of improving the way we vaccinate. Two important tools have proven invaluable in investigating maturation. On one hand theoretical models by making use of concepts from Statistical Physics, Stochastic Processes and Population Dynamics provide the link between qualitative hypotheses and their quantitative effects. On the other hand Bayesian Inference can make use of these models to extract complex information from experimental data. In this work we make use of both tools to study AM.

The first chapter of this manuscript is dedicated to a biological description of AM, with particular focus on some of the open questions and on the role that theoretical models played in our understanding of this process.

In the second chaper we study the effect of Ag dosage in vaccination. Experiments show the existence of an optimal intermediate dosage that maximizes the average affinity of the elicited Ab population. It was suggested that this optimality might be explained by the role that Ag availability plays in controlling the selection pressure in the evolution process. Through the work of our collaborators K. Eyer et J. Baudry (ESPCI) we were able to obtain experimental measurements of Ab affinity distributions elicited by different immunization protocols in at different Ag dosages. We introduce a model for AM and make use of Bayesian Inference to fit the relevant parameters on these experimental data. Upon maximum-likelihood fit the model is able to simultaneously reproduce all of these distributions. A theoretical analysis of the model shows that this optimality is indeed generated by a tradeoff between growth rate and maturation speed, which are both controlled by Ag concentration. Moreover the value of inferred parameters provides information on inner mechanisms of maturation that would otherwise be hardly accessible, such as the level of permissiveness in selection.

Chapter 3 is devoted to a study of some stochastic effects in maturation. We first introduce a slightly modified version of the model which allows for more deep theoretical analysis. With this we study the survival probability of a lineage and of the full population in conditions of strong selection, in which the population undergoes a bottleneck. Then we apply the path-integral formulation and retrieve equations for the most-likely evolutionary trajectories in affinity space for cell lineages. Compared to phylogenies this

provides a description of lineages evolution in phenotypic rather than genetic space.

Chapter 4 and 5 contain preliminary results and perspectives. In chapter 4 we show how the observed effect of Ag concentration can emerge from microscopic cellular interactions mechanisms. In chapter 5 we study the effect of Ag concentration and pathogen mutability on breadth acquisition in single-Ag vaccination. Our model suggests that increasing the selection pressure by reducing Ag concentration can improve breadth acquisition. At the same time a critical mutability threshold exist above which single Ag-maturation is ineffective, indicating that multiple Ag variants must be employed to confer breadth against highly-mutable pathogens.

RÉSUMÉ

La maturation d'affinité (MA) est un processus biologique grâce auquel notre système immunitaire génère de puissants anticorps (Acs) contre les nouveaux pathogènes rencontrés. Ce processus est également à la base de la vaccination, l'une des procédures médicales les plus efficaces jamais mises au point, qui sauves des millions de vies chaque année. La MA fonctionne en soumettant les lymphocytes B à des cycles itératifs de mutation et de sélection pour la liaison de l'Antigène (Ag), générant ainsi un processus évolutif Darwinien qui augment progressivement leur affinité pour l'Ag. Les mécanismes qui régulent la MA sont très complexes et présentent encore de nombreuses questions ouvertes dont les réponses pourraient améliorer la manière dont nous vaccinons. Deux outils importants se sont révélés inestimables pour étudier la maturation. D'un côté, les modèles théoriques, qui utilisent des concepts de la physique statistique, des processus stochastiques et de la dynamique des populations, permettent de faire le lien entre les hypothèses qualitatives et leurs effets quantitatifs. D'un autre côté, l'inférence Bayésienne permet d'utiliser ces modèles pour extraire des informations complexes des données expérimentales. Dans cette thèse nous utilisons ces deux outils pour étudier la MA.

Le chapitre 1 de ce manuscrit est consacré à une description biologique de la MA, de certaines questions ouvertes et du rôle que les modèles théoriques ont joué dans notre compréhension de ce processus.

Dans le chapitre 2, nous étudions l'effet du dosage de l'Ag dans la vaccination. Les expériences montrent l'existence d'un dosage intermédiaire optimal qui maximise l'affinité des Acs obtenue, qui pourrait s'expliquer par le rôle que joue la disponibilité d'Ag dans le contrôle de la pression de sélection. Par le travail de nos collaborateurs K. Eyer et J. Baudry (ESPCI) nous avons pu obtenir des mesures expérimentales des distributions d'affinité des Acs, résultant de différents protocoles d'immunisation à différents dosages d'Ag. Nous introduisons un modèle pour la MA et utilisons l'inférence Bayésienne pour ajuster les paramètres sur ces données expérimentales. Le modèle est ainsi capable de reproduire simultanément toutes ces distributions. Une analyse théorique montre que cette optimalité est effectivement générée par un compromis entre le taux de croissance et la vitesse de maturation, qui sont contrôlés par la concentration d'Ag. En outre, la valeur des paramètres déduits fournit des informations sur les mécanismes internes de maturation qui seraient autrement difficilement accessibles, comme le niveau de permissivité dans la sélection.

Le chapitre 3 est consacré à l'étude de certains effets stochastiques dans la MA. Avec une version modifiée du modèle qui permet une analyse théorique plus approfondie nous étudions la probabilité de survie d'une lignée de cellules B et de la population entière dans des conditions de forte sélection. Ensuite, nous appliquons la méthode des intégrales de chemin et obtenons des équations pour les trajectoires évolutives les plus probables dans l'espace d'affinité pour les lignées de cellules. Par rapport aux phy-

logénies, cela fournit une description de l'évolution dans l'espace phénotypique plutôt que génétique.

Les chapitres 4 et 5 contiennent des résultats préliminaires et des perspectives. Dans le chapitre 4, nous montrons comment l'effet observé de la concentration en Ag peut résulter de mécanismes d'interactions cellulaires. Dans le chapitre 5, nous étudions l'effet de la concentration d'Ag et de la mutabilité du pathogène sur l'acquisition de protection. Notre modèle suggère que l'augmentation de la pression de sélection par la réduction de la concentration d'Ag peut améliorer la protection contre les pathogènes mutables. En même temps, il existe un seuil critique de mutabilité au-dessus duquel la maturation d'un seul Ag est inefficace, et multiples variantes d'Ag sont nécessaires pour conférer protection.

PUBLICATIONS

This manuscript contains the research work that I have conducted during my PhD under the supervision of prof. Simona Cocco and prof. Rémi Monasson at the Laboratoire de Physique de l'École Normale Supérieure.

The results presented in chapter 2, concerning the application of inference techniques on experimental maturation data and the study of the effect of antigen concentration, were the object of a recent publication [96]. Moreover the work discussed in chapter 3, concerning bottleneck survival probability and evolutionary trajectories in maturation, is part of a paper currently at the draft stage [97]. Finally the results of chapters 4 and 5, concerning the integration of a microscopic B-T cell interaction model in maturation and the extension of the results to the case of multiple Ags, will be further investigated and finalized for publication after the PhD defense.

Publications:

[96] M. Molari, K. Eyer, J. Baudry, S. Cocco, and R. Monasson. Quantitative modeling of the effect of antigen dosage on B-cell affinity distributions in maturating germinal centers. *eLife*, 9:e55678, 2020

[97] M. Molari, R. Monasson, and S. Cocco. Survival probability and size of lineages in antibody affinity maturation. *arXiv preprint arXiv:2010.11580*, 2020

LIST OF ABBREVIATIONS

AM	Affinity Maturation
GC	Germinal Center
GCR	Germinal Center Reaction
IS	Immune System
Ab	Antibody
Ig	Immunoglobulin
Ag	Antigen
BCR	B-Cell Receptor
LZ / DZ	(Germinal Center) Light Zone / Dark Zone
SHM	Somatic Hyper-Mutation
Tfh cell	T-follicular helper cell
FDC	Follicular Dendritic Cell
MC	Memory Cell
PC	Plasma Cell
Ab-SCs	Antibody-Secreting Cells
TT	Tetanus Toxoid
BNAb	Broadly-Neutralizing Antibody
TLR	Toll-Like Receptor
MHC	Major Histocompatibility Complex
APC	Antigen Presenting Cell

CONTENTS

1

THE BIOLOGY OF AFFINITY MATURATION, OPEN QUESTIONS AND THE CONTRIBUTION OF MODELS

1.1 INTRODUCTION AND CHAPTER OUTLINE

Affinity Maturation (AM) is a mechanism through which our Adaptive Immune System is able to acquire protection against new pathogens it encounters. This protection is obtained through the development of Antibodies (Ab), molecules which are capable of selectively binding a particular pathogen with high affinity, thus helping the Immune System (IS) to identify and fight it. The mechanisms that make up AM are highly complex, involving the interaction of many different cell types, and many aspects are still not fully understood.
AM is also at the base of the efficacy of vaccination. This medical procedure, responsible for saving millions of lives every year, works by artificially eliciting maturation and the development of protective Abs. In spite of its long and successful history, many aspects of vaccination still have to be perfected. This procedure was initially developed from empirical observations, but more recently experimental and theoretical advancements have led to a deeper understanding of AM, with important consequences on the way we vaccinate. The continuation of these advancements promises to improve vaccination techniques and possibly even lead to the solution of currently open problems, such as developing effective immunization techniques that could confer protection against mutable pathogens.

This first introductory chapter contains a short description of the biology of Affinity Maturation, and the mechanisms that our IS employs to develop potent antibodies. Along with known results we also present open questions and aspects that are currently under investigation, with special focus on the role that models play in aiding our understanding of this process. The chapter is divided in two sections:

THE BIOLOGY OF AFFINITY MATURATION : this section contains a succinct biological description of Affinity Maturation. This description includes the structure and role of antibodies, the formation and workings of Germinal Centers, and how AM can be elicited artificially through vaccination.

OPEN QUESTIONS AND THE ROLE OF MODELS : recent experimental advancements allowed for the investigation of AM mechanisms at unprecedented level of details, providing new insight and understanding on this phenomenon. However many open questions still remain. Together with experiments, theoretical models have proven to be important tools in the study of AM. In this section we will present some of these open questions, along with some examples of how models have been used to deepen our understanding of AM.

1.2 THE BIOLOGY OF AFFINITY MATURATION

Even though we are unaware of it, we are constantly surrounded by many pathogens for whom our organism represents an ideal environment to reproduce and thrive. The Immune System (IS) is our line of protection against these entities. It also protects us from threats coming directly from our organism, for example by eliminating malfunctioning cells that could turn into tumors. To carry out this task the IS must continually scan our organism, and distinguish the self from the non-self. It is crucial for this recognition to be highly specific, since mis-identifying the self as a threat can potentially result in auto-immune diseases, and conversely not being able to efficiently detect a pathogen can lead to chronic diseases. In addition to this, certain pathogens can evolve at a fast rate to escape immune recognition, each new mutant constituting a new challenge for our IS. To face these challenges the IS comes equipped with mechanism to *identify*, *fight* and *remember* new pathogens it encountered. The ensemble of these mechanisms constitutes the Adaptive Immune System (AIS).[1]

Antibodies are one of the main weapons of the AIS. These versatile molecules can develop specific binding affinity for many different Antigens (Ags), and thus can be used to identify and fight pathogens. Several decades ago it was experimentally observed that upon encounter of a pathogen, the affinity of Abs in serum for the Ag increases over time [38]. The mechanism that allows this increase is known as *Affinity Maturation* (AM). This process takes place in specific microanatomical structures known as *Germinal Centers* (GCs). Here B-lymphocytes, cells responsible for Ab production, are subject to cycles of mutation and selection for Ag binding that progressively increase the affinity of their receptors (the membrane-bound form of Abs), in what is essentially a Darwinian evolutionary process. Protection against the pathogen is guaranteed by cells that differentiate and exit this cycle. These are Plasma Cells, that fight the ongoing infection by secreting high-affinity antibodies, and Memory Cells, that grant protection against future encounters. They remain quiescent until the pathogen is encountered again, at which point they activate and produce Abs. AM is also at the base of the efficacy of vaccination, which artificially elicits maturation through the administration of a dose of inactivated Ag, thus conferring protection without the risk associated to pathogenic infection.

The one above is an overly simplified description of Affinity Maturation. When analyzed in details this process appears much more complex, requiring the interactions of many different cell types and molecules, with many aspects that are not yet well-understood. In this section we provide a more detailed overview of the main known phases and mechanisms of AM. This overview aims at summarizing the vast body of knowledge accumulated on the subject, which is discussed in more depth in many excellent reviews [149, 34, 13, 90, 37, 147, 132, 148] and books [101, 136].

1 These tasks and strategies are distinct from the one of the Innate Immune System, that comprises fast and generic defense mechanisms capable of recognizing certain common pathogenic motives. It lacks however the ability to recognize and remember pathogens that do not carry these motives.

1.2.1 *Structure and function of Antibodies*

Recognizing and neutralizing pathogens are extremely hard tasks. Pathogens come in a wide variety, and the organism must be able to identify and neutralize them, with minimal damage to other cells in the body. Antibodies (Ab), also known as *Immunoglobulins* (Ig), represent one of the main weapons of the Adaptive Immune System in this fight. These molecules are capable of developing binding specificity for toxins and pathogens, and help the immune system neutralize them. Abs are an important component of humoral immunity. These molecules are secreted by Ab-generating cells and are freely transported by plasma until they meet their target.
Any molecule capable of stimulating Ab generation is called an *Antigen* (Ag). Usual antigens are proteins, glycoproteins and polysaccharides of pathogens. The specific region of these molecules that is recognized and bound by an Ab is called an *epitope*.
The mechanism in which Abs help to control infections are many and complex [77], but there are three main recognized mechanisms (see fig. 1 B) through which they are able to neutralize pathogens. The most direct way is known as *neutralization*: by binding to their targets antibodies can prevent them from interacting with the organism. This is important for example in protection against viruses or toxins. For larger-size threats such as bacteria Abs can act by *opsonization*: they bind molecules on the surface of the pathogen forming a coating that can then be recognized by phagocytic cells such as macrophages. These cells have receptors that can bind a part of the Ab and, recognizing the coated pathogen as a threat, they proceed to ingest and destroy it. The last mechanism of action of Abs is *complement activation*. The Complement System is a set of proteins (more than 30) than normally circulate in an inactivated form in our plasma. If activated these proteins can help fight pathogens by attacking the pathogen membrane or sending inflammatory signal to other cells of the immune system.

As outlined above, Abs have two functions. On one side they need to bind the Ag with high specificity. On the other, they need to provide signals to the Immune System to recruit cells and clear the pathogen. This two-fold role is reflected in their structure. Simply speaking, in their most abundant class (IgG isotype) antibodies are Y-shaped molecules, symmetric along the vertical axis, of size approximately 150 kDa (see fig. 1 A). They are composed of four subunits, two *light chains* and two *heavy chains*. Each of these sub-units can be divided in a *constant* and a *variable* region. Variable regions are located around the two tips of the Y-shape, and are involved in binding. Their nucleotide sequence is subject to a lot of mutations, especially in some short domains on the tip, known as the *Complementary Determining Regions* or CDRs. The variability of these regions is directly linked to the binding specificity of the Ab. The constant region constitutes instead mainly the stem of the Y-shape, and is responsible for signaling and interaction with the rest of the immune system. Secreted Abs exists in different classes called *isotypes*,[2] that differ by the heavy chain constant region. Differences in this region translate in structure and function, with distinct classes being preferentially expressed

2 There are five main isotypes: IgM, IgD, IgG, IgA, IgE.

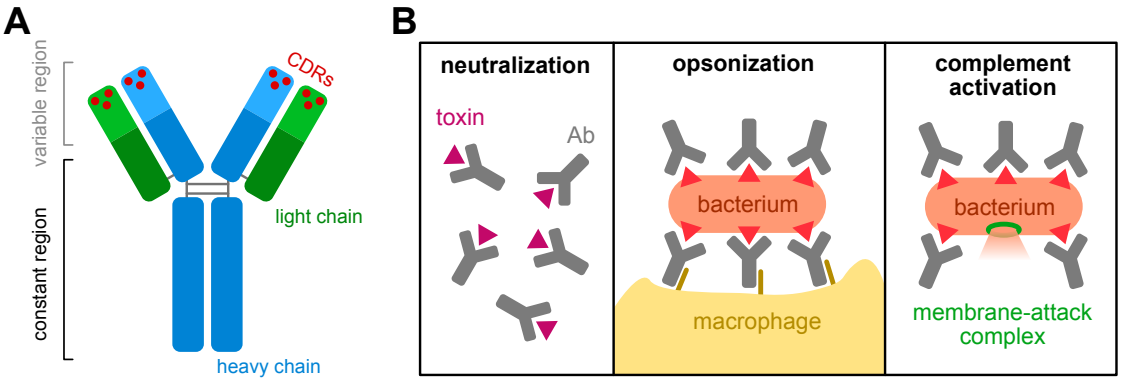

Figure 1: **A**: Schematic representation of the structure of an antibody of IgG isotype. The Ab is composed of two identical pairs of heavy (blue) and light (green) chains. Each chain can be divided in a constant (dark) and a variable (light) region. In the variable part each chain contains three complementary determining regions (CDRs, red). **B**: Three main mechanism of action of Abs. Toxins are neutralized by binding to them and preventing them from reaching their molecular target (left). Abs can recognize Ags on bacteria surface, creating a coating that can be then recognized by macrophages which in turn can ingest and destroy bacteria (middle). Finally, Abs can cause the activation of molecules of the complement system, for example the membrane-attack complex that can create holes in the bacterium membrane (right).

in distinct tissues, and having distinct mechanisms of action.[3]

Immunoglobulins are also present in a membrane-bound form on the surface of B and T lymphocytes. In this case they are known as B or T cell receptors (BCR / TCR). B-lymphocytes (or B-cells) are cells in the immune system whose main effector role, once terminally differentiated, is Ab production. Abs and BCRs expressed by a single B-cell differ only on the constant region, which as discussed above is not involved in binding, and therefore they will have binding specificity for the same target molecule. T-lymphocytes (or T-cells) come in many different sub-classes with very different functions. For example, cytotoxic T-cells for example play an important role in tumor suppression and virus control and act by sending apoptotic signals to cells that are infected or malfunctioning. Regulatory T-cells instead can suppress the activity of other lymphocytes and help control the immune response. In this thesis we will mainly be interested in Helper T-cells, a class of T-cells that helps to orchestrate the immune response, among other ways by providing signals to Ag-stimulated B-cells. Both BCRs and TCRs upon binding their cognate Ag activate signaling pathways inside the cell to communicate the presence of their targets. One major difference between BCRs and TCRs consist in the fact that TCRs bind targets through presentation by other cells. For binding and recognition these targets must be loaded on specific molecules called MHC (from Major Histocompatibility Complex) and exposed on the cell membrane. These molecules come in two classes, with MHC class I molecules being used for the presentation of self-Ag to T-cells that express the CD8 marker (e.g. cytotoxic T-cells), while MHC class II molecules are involved in foreign-Ag presentation to cells with CD_4 marker (e.g. Helper T-cells).

3 For example IgA and IgM are polymers composed of multiple Ig monomers. Since each monomer can bind the target these Abs can act by clumping together pathogens. This is the main action mechanism of IgA against bacteria in the gut [98].

A further difference between BCRs and TCRs is that BCRs target exposed residues on a pathogen, that are close in structure but not necessarily in sequence. Conversely, prior to MHC presentation pathogens are often subject to lysis, and presented residues can also originate from non-exposed parts of the pathogen but are contiguous in sequence.

The wide variety of pathogens that we encounter, each one presenting a different assortment of targets for recognition, pushes our immune system to develop binding affinity for a very disparate class of molecules. Since the binding is specific the same level of diversity is required in the set of Abs and receptors produced. This diversity is achieved thanks to two main processes. The first is known as *V(D)J recombination*, the mechanism through which different gene segments are combined to generate the immunoglobulin heavy and light chain. To assemble these chains multiple choices must be made between different genes at different positions, which combinatorially increase the variety of possible immunoglobulins that can be created directly from the genome. This diversity is usually sufficient to generate a basic binding affinity for a pathogen. However to reach the high affinity required for an efficient Ab response, a second mechanism is at play for B-cells, namely *Affinity Maturation*. It acts by selecting among all the random mutations the ones that increase affinity for the Ag, as described more in detail in the next sections.

1.2.2 *The Germinal Center Reaction*

Affinity Maturation begins with the formation of Germinal Centers (GCs), transitory microanatomical structures that develop inside of secondary lymphoid organs (such as lymph nodes or the spleen). In the case of immunization these structures are visible few days after Ag administration, reaching their mature state after roughly one week [34]. They have widely varying lifetimes, ranging from 1-2 weeks for soluble protein boosting to months or longer for certain chronic infections [147]. Their size can vary considerably [158] and they comprise many different kinds of cells, but the majority (up to a few thousands [144]) are B-cells. Traditionally GCs were believed to be seeded by a very small number of founder clones, but recent experiments [144] estimated this number to be higher, between 50 and 200. Histological imaging of GCs shows a clear subdivision in two compartments (see fig. 2), termed the Light Zone (LZ) and the Dark Zone (DZ). The name is related to their appearance, with the DZ being more densely packed with cells, whose nuclei are stained in blue in histological images. This subdivision is reflected, albeit in a less marked way [132], also at the level of processes, with the DZ being the main site of cell division and the LZ being associated to Ag binding and selection. B-cells iteratively migrate back and forth between these two zones, as can be observed from imaging experiments [7, 150]. In this way they are subject to repeated cycles of mutation and selection for Ag binding, effectively undergoing a Darwinian evolutionary process that is at the base of the increase in affinity.

To understand the way the GC Reaction (GCR) works we can follow the fate of B-cells during one cycle, starting from cell division in the DZ. GCs B-cells express high level of *Activation-Induced cytidine Deaminase* (AID), an enzyme that increases the natural rate of DNA mutations up to 10^{-3} per base-pair per generation [66, 89, 18], a rate 10^6 times higher than the physiological mutation rate of germline genes. This phenomenon,

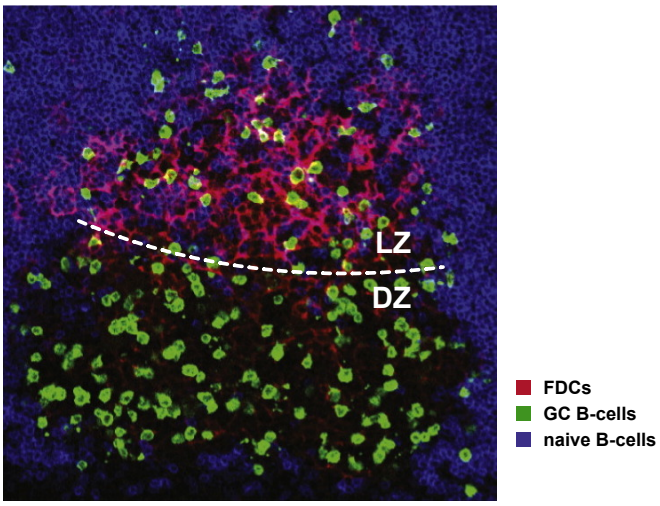

Figure 2: intravital multiphoton image of a GC in a mouse. GC B-cells are marked in green, Follicular Dendritic Cells (FDCs) in red, and naive B-cells not recruited in the GC reaction in blue. T-cells are not marked. The boundary between the Light and the Dark Zone (LZ/DZ) is marked with a white dashed line. Image adapted from [150].

termed *Somatic Hyper-Mutation* (SHM) makes so that with each division each cell accumulates on average about 1 new mutation (mainly point mutations) in the Ig-coding gene [29]. In some cases these mutations can render the receptor unproductive, which leads to cell death [86]. In other cases they might produce functional cells but with modified BCR, introducing a positive or negative change in affinity for the Ag. Importantly, even cells that are functional if isolated from the rest of the GC will undergo apoptosis in few hours [76]. This goes to show that in order to survive cells must receive some form of survival signals (as described below) that will be provided more likely to cells with higher affinity.

After a variable number of divisions (estimated to two on average [45]) B-cells migrate to the GC LZ. In this zone a network of *Follicular Dendritic Cells* (FDCs) [53] is present. In addition to delivering different chemical signals, these cells are also thought to provide a reservoir of Ag for B-cells to bind.[4] FDCs are in fact able to capture and retain Ag on their surface for extended periods of time, even of the order of many months [83, 84]. This unusually long retention time has been explained by recent experimental observation [52] that once captured on the membrane Ag can be internalized in the cell in endosomal compartments, in which it is protected from degradation, and can cyclically resurface becoming again available for B-cells to bind. In the LZ GC B-cells form contacts with FDCs [111], binding the Ag on their surface and trying to extract it with mechanical forces [137]. This provides a first mechanism of affinity discrimination, since

4 Experimental evidence discussed in [132] and [5] shows that presentation through FDCs is not the only way for B-cells to access Ag. While large Ags (more than 70 kDa) cannot freely penetrate into GCs and are mainly available through presentation on FDCs, small soluble Ags like egg white lysozyme (14.4 kDa) could be available also by direct penetration in the GC. Contrarily to the way B-cells interact to membrane bound Ag, which has been largely investigated, the response and fate of GC B-cells that interact with soluble Ag is less well-understood.

the amount of Ag extracted depends on the affinity of the BCR for the Ag.[5]
After binding the Ag is internalized, processed and loaded on MHC-II molecules. This combination of antigenic peptide and MHC molecule is termed a pMHC complex. The difference in affinity of cells influences the amounts of Ag captured, and therefore also the density of pMHC complexes displayed on their surface. This density is probed through the interaction with another class of cells that are present in GCs, namely *T-follicular helper* cells (Tfh cells)[6], that are mainly localized in the LZ. Even though these T-cells only represent a minor fraction of the GC cells, they are essential for AM and in their absence GCs do not form [61]. Tfh cells are highly motile, forming transient and repeated contacts with B-cells [7, 134]. In these contacts they sense the pMHC density on B-cells surface, and preferentially deliver survival signals[7] to cells with higher pMHC density [126, 35], with longer contacts being associated to positive selection [134]. Cells that receive these survival signals can either differentiate, as described in the next section, or migrate back to the GC DZ, where they start a new cycle with divisions. Experiments estimated the fraction of LZ B-cells that recirculate to the DZ to around 30% [150]. Experiments also showed that the amount of Ag captured and presented to T-cells influences the amount of SHM and the number of divisions of B-cells [45, 44]. A schematic depiction of the GCR is provided in fig. 3.

1.2.3 *Differentiation into Memory and Plasma Cells*

GC B-cells do not possess direct effector functions, to offer protection against pathogens they must first differentiate. Indeed, along the course of the maturation process GC B-cells can differentiate into one of two different fates: either to *Plasma Cells* (PCs) or *Memory Cells* (MCs).[8] PCs can be further classified as *Short-Lived* (SLPCs) or *Long-Lived* Plasma Cells (LLPCs). The former are Ab factories, responsible for mounting a fast response against the pathogen by secreting large amounts of Abs. As the name suggests they are active only for a limited time, with a lifetime of several days [141]. Conversely, LLPCs have been shown to live for a long time, being detectable even one year after immunization [135], and possibly much longer. They take residency primarily in the bone marrow and secrete sufficient antibody quantities to confer some level of protection upon pathogen re-encounter [128]. MCs are instead in a quiescent state, but in case of a future pathogen encounter they can re-activate and start antibody production, mounting a faster response than when the pathogen was first encountered [156]. Importantly,

5 The complex interlocking of receptors that forms the bond between a lymphocyte and and an Ag-presenting cell is called an *immune synapse*. The mechanism of Ag extraction involves mechanical pulling forces, which can enhance affinity discrimination, and will be discussed further in the next section.

6 These cells, belonging to the class of T-helper cells that express the CD4 marker, are identifiable by their expression of a transcription factor named BCL-6.

7 The binding of B-cells membrane receptor CD40 by its ligand CD40L expressed on T-cells might play a major role in the delivery of this signal [152].

8 We point out that Affinity Maturation is not the only source of MCs or PCs. These cells can be generated even prior to the GC reaction [131]. For example MCs can form also in genetically modified animals that lack GCs [146], or as a response to T-cell independent Ag [113]. However these cells do not display signature of maturation, and do not harbor many affinity-enhancing mutations. Ig isotype is correlated with the origin of these cells, with MCs that are generated in a GC-independent way displaying IgM isotype, while class-switched cells, i.e. cells that do not display IgM isotype, often originate from GCs.

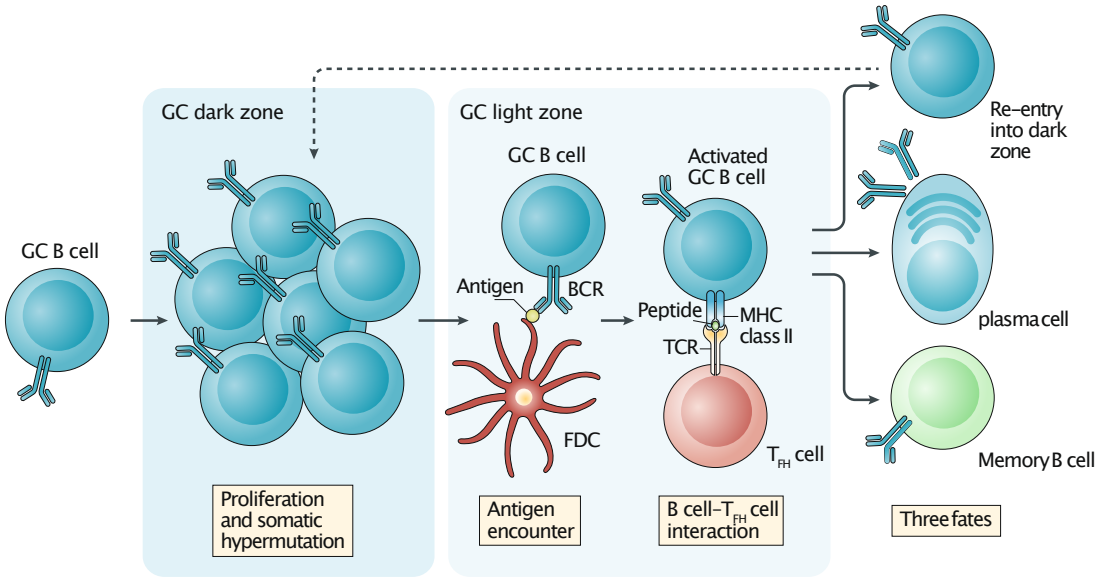

Figure 3: Schematic representation of the main steps of the GCR. In the GC dark zone B-cells undergo proliferation and somatic hypermutation (SHM). They then exit to the light zone, where they encounter antigen on FDCs, present the antigen to T follicular helper cells (Tfh cells) and if successfully selected can undergo three main fates: namely, differentiation into memory B cells or plasma cells, or re-entry into the GC dark zone to undergo further cycles of mutation and selection. Figure and caption adapted from [5]

they can also seed new GC reactions and undergo further maturation [87].[9] The current view [5] is that in a response against T-cell dependent Ag, most SLPCs are generated as a quick response, shortly after Ag administration, directly from naive B-cells that are able to bind Ag with some basic affinity. MCs and LLPCs are instead, for the most part, gradually generated through affinity maturation from B-cells that have entered GCs.
Moreover, MCs and LLPCs have been shown to differ in their average affinities. Experiments in [130] demonstrated that lower affinity GC cells tend to enter the MC pool. In addition to this, in [157] it was shown that GCs undergo a temporal switch, with MCs being produced mainly at the beginning of maturation, and LLPCs mainly at later times. Whether cell's fate depends on affinity or time since the beginning of the GC reaction, the result is that in the MC pool we find cells with lower affinity but also more broadly reactive, while cells that belong to the LLPC pool show more maturation and higher affinity [5].
Together MCs and LLPCs constitute our *Immunological Memory*: they protect us against future pathogen encounter. In [5] they are describes at two "walls" of protection (see fig. 4). LLPCs build up the first protection barrier. They secrete highly selected high-affinity Abs that are protective against reinfection by pathogens that are homologous to the one encountered. Some pathogens however are able to circumvent this first barrier, by developing mutations that decrease their binding affinity for these Abs. This is where the second MC barrier comes into play. By undergoing less selection these cells are very diverse and more broadly reactive, and can be reactivated by many virus variants, and

9 Recent experiments aimed at quantifying the recall of MCs in secondary GCs [91] revealed that only a minority of MCs are successfully recruited in GCs following boost Ag injection. This highlights the importance of designing immunization strategies that would enhance memory recall.

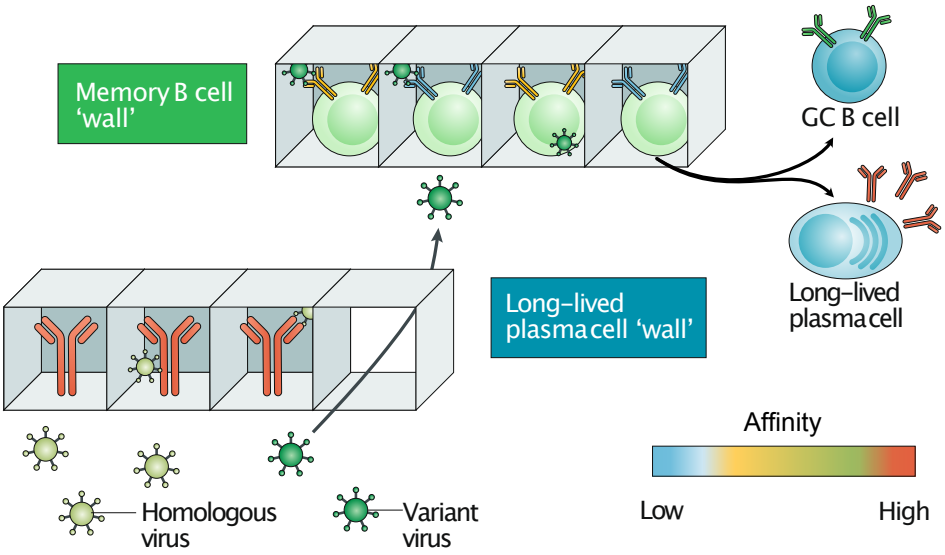

Figure 4: Long-lived plasma cells secrete high-affinity Abs (depicted in red) that form a first 'wall' of defense against reinfection by homologous pathogens. Variant pathogens can find holes in this wall; however, having escaped the antibodies from the long-lived plasma cells, the variant pathogens encounter a second wall formed by memory B cells that were less highly selected and therefore maintain a broader range of antigen affinities and specificities. The memory B cells are activated by the variant pathogen to differentiate into long-lived plasma cells or to re-enter GCs. Figure and caption adapted from [5].

then start Ab production and further affinity maturation.
In this perspective the GC reaction is a process that appears more oriented towards optimization of future immunity rather than immediate pathogen clearance [132]. This is indeed consistent with its central role in vaccination.

1.2.4 *Vaccination*

The two walls of protection that we described above can be raised without the risks associated to direct pathogen encounter through vaccination. The modern version of this procedure is traditionally attributed to Edward Jenner, who in the late 18th century created the first vaccine on the observation that infection with cowpox, a relatively mild disease, would confer protection against smallpox, which instead was often fatal. In fact the word "vaccine" comes from the latin name for cowpox, *variolae vaccinae*, litterly "smallpox of the cow". The importance of this procedure cannot be overstated [48, 103]. Vaccination has been used to fight, and in some cases even fully eradicate diseases, as in the case of smallpox. The World Health Organization estimates that thanks to this procedure an average of 2-3 million deaths are avoided each year.[10]. In addition to this, vaccination is also highly cost-effective, with the benefits deriving from disease protection greatly outweighing the costs [116].
Vaccination works by artificially eliciting AM through the administration of a solution

10 Source: https://www.who.int/en/news-room/fact-sheets/detail/immunization-coverage

containing pathogenic Ag and adjuvants. This solution activates an immune response with consequent formation of GCs and maturation of B-lymphocytes. Adjuvants [12, 30] are an important vaccine ingredient with two main functions, namely stimulation of the IS and the retention of Ag for a longer time. Administration of pure Ag without the presence of an active pathogen is in fact often insufficient to generate an immune response of enough magnitude to elicit GC formation. Vaccines contain molecules that can stimulate immune cells by activating receptors, linked to innate immune system pathways, that recognize common pathogenic features.[11] The second important role of adjuvants is Ag retention and release. When injected directly in their soluble form most Ags degrade quickly, but if instead injected within an adjuvant matrix they are protected and more gradually released, granting a longer Ag availability to IS cells [82].
In spite of its long history, vaccination is far from being completely mastered and many challenges remain to be faced. For example, in spite of many recent advancement we still lack a vaccine for highly-mutable pathogens such as HIV. Since vaccination is based on AM, our best hope of overcoming these challenges lays in a better understanding of this process, to be achieved through a combination of theoretical and experimental advancements.

1.3 OPEN QUESTIONS AND THE ROLE OF MODELS

In this section we discuss some of the open questions and less-understood aspects of AM, such as the influence of Ag dosage on the outcome of maturation, the way affinity discrimination is operated in GC selection, the difference between maturation against simple or complex Ags, and immunization strategies against highly-mutable pathogens. Theoretical and computational models have proven to be potent tools to guide our understanding of AM, and for each of these topics we will present some examples of models and the insight they provided. Moreover at the end of each subsection we will briefly indicate some of its relations to our work. To facilitate reading and browsing in long subsections we added margin notes to indicate the subject discussed.

1.3.1 *The role of models in understanding AM*

Affinity Maturation is a complex process. If we break it down to its minimal components we find that it proceeds through many stochastic and often non-linear interactions between different cellular agents. In such context mathematical models are essential to provide the link between our qualitative understanding of the sub-processes that constitute maturation, as well as to aid the quantitative interpretation of data [26]. They allow us to bridge the gap between the hypothesis one wants to test and their expected effect on experimental observables, a gap that given the complexity of the process can often lead to counter-intuitive results. This is even more important when one considers that with modern experimental high-throughput techniques collected data on a given system can amount to a very large number of measurements, which might be hard to interpret.

11 As a side-effect, the stimulation of these pathways could push B-cells to quickly respond to this danger signal by differentiating into SLPCs. For example in [4] it was shown that stimulations of these pathways resulted in a higher Ab response, but also in a lower Ab affinity and maturation level. This suggests that for vaccines that require a high level of maturation the stimulation of these pathways should be limited [5].

In this context models can help identify and extract relevant patterns from these large datasets. These models often employ tools from Statistical Physics which, by being concerned with the study of large system composed of microscopically interacting particles, offers a natural language to describe these processes.

In the study of AM models have proven their importance since the early days. One notable example is the seminal work of Perelson et al. [64, 114] in which the authors introduced a maturation model in which the mutation rate could be controlled, and by solving the associated control problem showed that optimal maturation required repeated alternated intervals of high mutation rate, in which the population would diversify, and low mutation rates, in which it would mainly evolve by selection, which lead them to advance the hypothesis of recycling of B-lymphocytes through GC LZ and DZ. Direct experimental confirmation of this model prediction came only much later, with visualization of cells movements through intravital microscopy [150].
AM models come in many different kinds and include different ingredients. They have two main purposes: on one side understanding the mechanism of AM, such as the way selection acts or the interaction between cells in GCs, and on the other side optimizing immunization procedures, for example with respect to Ag dosages, or when highly-mutable pathogens are involved. A nice review of GC models, their ingredients and purposes can be found in [22]. In the following section we will describe some of these models alongside the questions they try to answer. Given the wide diversity of existing models we make no claim of completeness, limiting the discussion to some relevant examples from both of the above-mentioned categories.

1.3.2 *The effect of Antigen dosage*

Since B-cells are selected based on their ability to bind the Ag, Ag dosage plays an important role in controlling the outcome of vaccination. Its effect is however non-trivial and has yet to be fully understood. Since Ag dosage is a relatively easy variable to control, this understanding could be directly turned into guidelines on how to optimize vaccination.

optimal Ag dosage

Amongst other effects Ag dosage has an impact on Ab affinity, a fact that was already observed in early experimental investigations. In [38] for example the authors carried out experiments of hapten[12] immunization on rabbits. In these experiments rabbits were immunized with different Ag dosages, and average Ab affinity was measured at different times from immunization. The authors observed that the greatest and fastest affinity increase was observed at the lowest tested Ag dosage. Conversely, at the highest dosage affinities of Abs remained low. Similar experiments in [46] where two Ag dosages were tested provided similar results, with the highest average increase in affinity observed for the low Ag dosage. Here the authors advanced the hypothesis that Ag might act as a selection factor, and when present in limited amounts would preferentially activate cells

12 Haptens (from the Greek *haptein* - to fasten) are small organic molecules that do not elicit Ab production when injected in their unbound form. However when bound to a protein carrier they are able to stimulate the IS enough for Ab production. The produced Abs in this case can be hapten-specific, carrier-specific or hapten-carrier specific, depending on their binding specificity [101].

with higher affinity. They also suggested that a mechanism of tolerance might be at play when Ag is abundant, preventing cells that bind too strongly from activating. Moreover experiments described in [47] associated a higher Ag dose with a higher number of mutations and more diversity in the Ab population, and a lower dose with a more focused selection of high-affinity mutants.
These results can be interpreted with considerations performed in [63], where the authors analyzed the data collected in [38] to quantify the evolution of the mean and variance of the affinity distribution. They observed that at low Ag dosage the average affinity tended to increase and variance remained low, and conversely at high dosages the affinity did not show any strong improvements while the variance remained high. They suggested that these results can be interpreted by role that Ag availability must play in controlling the strength of selection for GC B-cells. When Ag is scarce (low dosage) evolution is in the strong selection regime, with only few high-affinity cells being successfully selected. Therefore in this case one observes a high affinity increase along with a reduction in diversity. On the other hand in the weak selection regime (high dosage) one expects more cells to be able to bind and internalize Ag, which generates a higher diversity and a smaller affinity increase.
Different early experiments also showed that the injection of a high dose of soluble Ag when the maturation is ongoing results in the apoptosis of GC B-cells and the termination of the GCR [133, 118, 51]. In [132] it is suggested that this might be caused by a form of negative selection, a mechanism of protection against the development of self-reactivity. According to this hypothesis strong cross-linking of BCRs in GC B-cell would generate an apoptotic signal, that would indeed negatively select cells that develop affinity for self-Ags that might be encountered in the GC environment.

The problem of optimal dosage for vaccine is discussed in a recent paper by Rhodes et al. [121]. Here the authors warn against the fact that current methods for Ag dosage decisions are antiquate and might lead to sub-optimal outcomes. They bring examples of retrospective dose-ranging studies in which the immunogenicity of fractional doses of vaccines were tested [25, 120, 49]. These studies found that lower doses were able to elicit an equivalent or even better response, leading the World Health Organization to recommend fractional vaccine dosages when vaccine availability is limited.[13] Rhodes and colleagues explain that the sub-optimality of dosages come from a long-standing assumption in vaccine development according to which the relationship between vaccine dosage and organism response is a saturating sigmoidal curve, that is to say a too low dosage will elicit no response, while a too large dosage will elicit the maximum response. In reality this relationships seems to have a peak, and experiments suggest that maximal IS stimulation is obtained at intermediate Ag dosages [122, 1]. This wrong assumption would therefore lead to vaccine dosages that are higher than the optimal intermediate dose. The authors advocate for the need of models capable of understanding and predicting the effect of Ag dosage for optimal vaccine development.

controlling Ag administration in time

Another important dimension to consider in terms of Ag availability, other than Ag dosage, is time. The amount of Ag available for B-cells to bind is a dynamic quantity that decreases over time due to different processes such as natural decay and consumption

13 Source: Meeting of the Strategic Advisory Group of Experts on immunization, October 2016 - conclusions and recommendations. https://apps.who.int/iris/handle/10665/251816

by B-cells. A more precise control over this quantity can be achieved through tuning simultaneously the dose of injected Ag and its distribution in time. In [143] for example the authors test different immunization schemes in which the total amount of injected Ag is kept constant, but its time-distribution is varied. In the experiments, in addition to a standard immunization scheme with two bolus[14] injections, three other schemes are tested in which for a week Ag is administered daily at exponentially increasing, constant, or exponentially decreasing dosages. They found that an exponentially increasing administration scheme, mimicking an infection, elicited the highest rate of Ab production. These results were interpreted with the help of a computational model of the GCR, consisting in a collection of differential equations describing different reactions, and with a key ingredient being the dynamics of capture and retention of Ag. The model was able to reproduce the correct hierarchy between the different immunization schemes, and explained the supremacy of the exponential scheme with the fact that in this scheme the Ag was better captured and retained. In addition to this the authors also tested the effect of osmotic pumps, subcutaneous devices that release Ag gradually over a period of weeks. Again they observed that these devices generated a better response than bolus immunization.

The potential advantages of slow Ag delivery methods were discussed in a recent review [28]. Here the authors point out that bolus-injection vaccination, in which the highest amount of available Ag is delivered when the GCR has not yet been initiated, is dissimilar to real infections, in which GCs have at their disposal a constant source of Ag generated by pathogens. They suggest that implementing immunization techniques that provide sustained Ag availability might greatly improve the outcome of immunization through different mechanisms. For example, a continuous Ag delivery would provide a source of non-degraded Ag to GC B-cells. In addition to this, this Ag would also be better retained by the fact that higher affinity cells have the time to develop and secrete Abs, that can bind Ag in immune complexes for a longer time. Continuous Ag delivery also increases the number of Tfh cells, that in turn can provide a higher number of survival signals, decrease selection pressure and generate a more diverse B-cell population. This might be essential when Abs must accumulate many mutations, such as in the case of Abs able to fight highly-mutable pathogens. Indeed, continuous delivery methods have been shown to enhance Ab responses against HIV [27].

Even if bolus vaccination is employed, a good control of Ag dosage is crucial when immunizing against mutable pathogens. In this context the population must be subject to an optimal selection pressure, low enough to grant population survival, but high enough for affinity maturation and breadth acquisition. This tradeoff can translate in optimal intermediate values of Ag concentration, as showed in [155, 154]. However, since immunization against mutable pathogen is a wide subject, we postpone more detailed discussion of these models to section 1.3.5.

To summarize, the amount of available Ag has a strong influence on the outcome of maturation, and its effect can be partially interpreted by considering its role in controlling selection pressure during AM. One consequence of this is the fact that a too high Ag

14 In a *bolus* injection a dose of vaccine is delivered at once. Here the term is used in opposition to delivery methods that distribute the administration over longer periods of time.

dosage generates a sub-optimal immune response. Current immunization techniques do not usually take these effects into account, and predicting the effect of varying the Ag dosage on the outcome of immunization is still an open challenge, and one whose solution could greatly improve our current immunization strategies. Other than controlling dosage, also controlling its dynamic delivery is important for an optimal immunization. In this sense continuous delivery methods constitute a promising research direction.
The part of our work presented in chapter 2 is strongly related to the problem of optimizing immunization through the control of Ag dosage and injection time. Here we combine experiments measuring Ab affinity distribution with quantitative mathematical modeling, to gain insight on the way Ag dosage controls the affinity enhancement.

1.3.3 *GC selection and mechanisms of affinity discrimination*

The affinity increase in the course of AM requires a mechanism of affinity discrimination that favors the proliferation of high-affinity B-cells over ones with lower affinity. Even though many experimental results contributed to exploring these mechanisms, the way selection operates in GCs is still not completely understood. However, it is generally accepted that positive selection in GCs can be traced back to two distinct signals [151, 74, 132]. The first occurs when B-cells bind Ags on FDCs, with the cross-linking of their BCRs delivering an activation signal to the cell. The second is delivered when they present Ag exposed on MHC molecules to cognates Tfh cells, which deliver a rescue signal.

affinity discrimination through BCR-Ag interaction

The binding of Ag on FDCs has the double function of sending an activation signal through BCR cross-linking, but also of capturing and exposing Ag on MHC molecules proportionally to the cell affinity.
While providing an important signal, BCR binding alone does not seem to be sufficient to drive B-cells to differentiate to Ab-producing PCs [68]. This fate decision seems to require an additional signal, that can be provided by either T-cell help or by binding of *Pattern Recognition Receptors* such as TLRs [6]. The fact that BCR binding is not sufficient for differentiation might be a safety mechanism against self-reactivity [71]. In this way Ag binding is not enough to start the production of Abs, but also additional proofs that the Ag is foreign is required, such as T-cell help or the detection of pathogenic features through TLRs. The underlying mechanism might be based on a metabolic clock, set in motion by BCR signaling, that if not deactivated by the second signal eventually leads to cell death. This might also explain the experimental observation that injection of a high dose of soluble Ag, with consequent strong BCR stimulation, causes the death of GC B-cells [133, 118, 51].
The binding of Ag plays also an important role in affinity discrimination, since the quantity of Ag internalized is proportional to the BCR affinity. This discrimination is enhanced by the fact that GC B-cells have an higher affinity threshold for BCR signaling and Ag gathering when compared to naive B-cells [72, 71] which makes so that difference in Ag affinity result in greater difference in the amount of Ag internalized. This is made possible by the characteristic *immune synapse* structure that GC B-cells form when in contact with an *Antigen Presenting Cell* (APC), in which the Ag is localized in cluster and bonds are tested by exerting mechanical pulling forces [111]. The way in which these forces can enhance affinity discrimination has been elucidated in an interesting

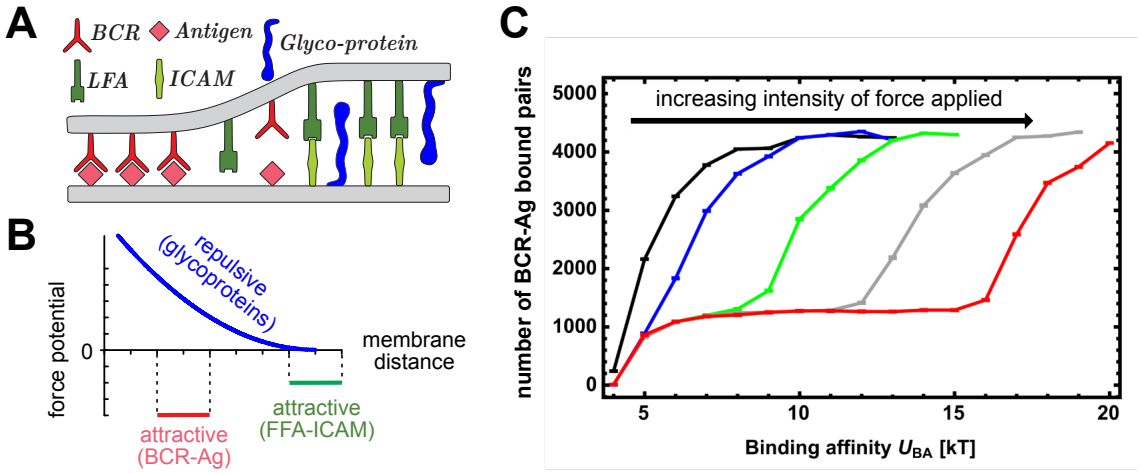

Figure 5: model by Knežević et al. studying how mechanical forces can enhance affinity discrimination. **A**: membrane-membrane interaction between an antigen-presenting cell (APC, bottom) and a B-cell (top). The two membranes are coated with proteins that are free to diffuse, and at specific membrane distances they can form bonds with their cognate pair. **B**: force potential for membrane protein binding. The pairing of BCR and Ag or LFA and ICAM generates an attractive force, albeit at two different membrane distances. The presence of glycoprotein instead generates repulsive forces. **C**: number of BCR-Ag bound pairs as a function of BCR affinity. Different colors correspond to different intensities of mechanical pulling forces applied on Ag-BCR pairs, according to the direction of the arrow. A stronger force widens the range of affinity discrimination. Figure adapted from [67].

model by Knežević et al. [67]. The model reproduces the interaction surface between a GC B-cell and an APC. On the surface of the two cells different kinds of membrane-bound molecules can diffuse, which generate repulsive or attractive forces depending on the local distance between the two membranes (see fig. 5 A and B). Glycoproteins on the two cell membranes generate a repulsive force, while Ag-BCR pairs or ICAM-LFA pairs (two conjugate membrane molecules) generate attractive forces. A total energy for the system can be written, taking into account the elasticity of the B-cell membrane and the forces generated by these molecules. The authors also add a third contribution, that account for pulling forces that act on Ag-BCR pairs. By letting the system evolve using Monte Carlo simulations on the Hamiltonian function thus defined, they observe that the amount of force exerted strongly influences the dependence of the number of Ag-BCR pairs that form in the steady state as a function of their affinity. Indeed, the application of mechanical forces increases the range of affinity discrimination (fig. 5 C).[15]

affinity discrimination through B-T cell interactions

The second important affinity discrimination mechanism occurs through the interaction of B and Tfh cells. Tfh cells can provide a rescue signal to B cells, delivered preferentially to high affinity B-cells displaying on their surface the highest density of peptide-MHC complexes. Indeed, as already mentioned, it has been shown that contact-dependent interactions with Tfh cells are essential for B-cells to be able to recirculate from DZ to the LZ [150], and that these contacts have variable duration, with long-

15 As a side note, the authors point out that affinity discrimination is also influenced by Ag concentration. Their model predict optimal discrimination at intermediate Ag dosage, highlighting another possible role of Ag concentration in GC selection.

lasting contacts being associated to the delivery of survival signals [7, 134]. However the question remains of how, in the chaotic and dynamic GC environment T-cells are able to sense and discriminate between high and low affinity cells. Rather than on a single interaction, the rescue signal might be progressively accumulated in repeated B-T cell interactions. In [152] the authors discuss a possible mechanism proposed in [75] that is based on a positive feedback loop involving the binding of two receptors and their ligands. These are the pair CD40/CD40L and ICOS/ICOSL. ICOS and CD40L are found on the T-cell surface, while their counterpart is expressed on B-cells. In conjunction with BCR signaling, CD40 stimulation[16] has been shown to stimulate cell division [80], thus being an important component of the survival signal. This stimulation could be delivered as follows. The amount of ICOSL and pMHC present on B-cells controls the surface size of B-T cell contacts. An extended contact surface coupled with cognate Ag presentation stimulates T-cells to externalize more CD40 to increase the rescue signal to the B-cell, and reduces the likelihood of T-cells to share contacts with ICOSL-lacking B-cells. In turn, CD40 signaling upregulates the expression of ICOSL in B-cells [75] (a feature unique to GC B-cells), making the B-cell more likely to entangle in further T-cell contacts. This creates a positive feedback loop that translates difference in B-cell affinities in nonlinearly amplified CD40 signaling differences over multiple B-T cell contacts. We depicted this mechanism in fig. 6.
An additional effect of T-cell help that could contribute to affinity discrimination is differential proliferation. In particular cells presenting the highest pMHC density to T-cell help have been shown to undergo a higher number of divisions and higher SHM [45] with a mechanism that depends on the amount of T-cell help received [44]. In the model discussed in [9] it was shown that a selection mechanism based on differential proliferation would result in quicker diversity loss than if selection was operated mainly through removal of low-affinity cells. While both selection mechanisms might be at play, in [132] the authors state that it is unlikely for differential proliferation to be the unique source of positive selection, because too many divisions without further selection would likely lead to the accumulation of too many deleterious mutations.

The interaction mechanisms illustrated above shows how important T-cells are for the rescue and maintenance of the B-cell pool, a fact supported by the observation that the number of Tfh cells correlates with the number of GC B-cells [124, 149]. However some experiments suggest that these interactions might be mutual rather than unidirectional[132], with B-cells being important in Tfh cells recruitment. Results of [160] show that ICOS stimulation is important for recruitment and maintenance of Tfh cells, and this stimulation is mainly delivered by ICOSL on bystander B-cells, and in a non-cognate antigen-independent manner. In this view B-cells form a field in which T-cells are immersed, which continually supports Tfh cells through ICOSL presentation.

cell-intrinsic and cell-extrinsic components in selection

In [132] GC selection is categorized as *cell-intrinsic* and *cell-extrinsic*. The former describes selection mechanisms whose outcome depend uniquely on the B-cell features (mainly its affinity), while the latter also includes external factors such as competitions with other B-cells. These categories loosely match the two kinds of affinity-discrimination

16 Other evidence of the importance of CD40 in the delivery of the survival is the fact that blocking CD40 signaling leads to the dissolution of GCs, and that CD40 stimulation was shown to rescue human GC B-cells [117, 50, 76].

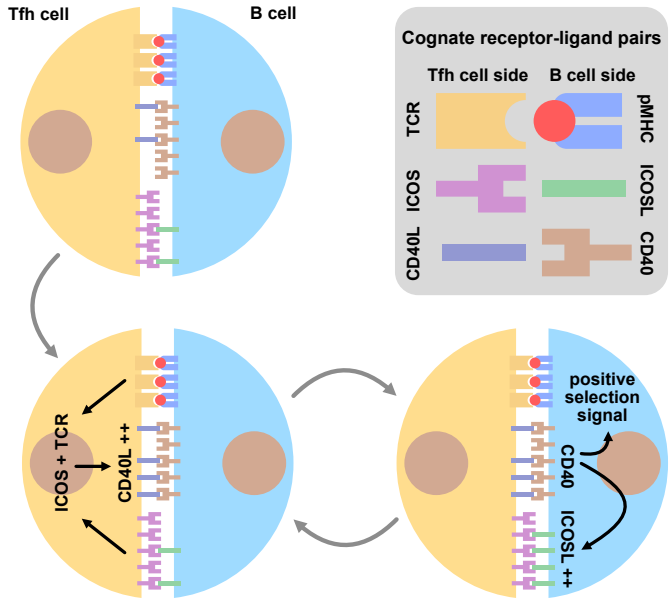

Figure 6: schematic depiction of the positive feedback loop for selection proposed in [75]. The figure represents a contact interaction between a B-cell presenting Ag on MHC-II molecules and a cognate Tfh cell, with membrane receptors and ligands represented according to the legend on the top right. If a sufficient amount of pMHC are presented then the combined engagement of TCR and ICOS stimulates the Tfh cell to externalize a greater amount of CDL40L. This in turns provide a greater survival signal to the B cell and upregulates the production of ICOSL, completing the positive loop.

mechanisms we discussed, with Ag-binding selection being by nature more cell-intrinsic, while selection for T-cell help is more competitive. When comparing these two processes, selection seems to be dominated by competition rather than by cell-intrinsic factors. One important experimental observation in this sense is the fact that in absence of high-affinity competitors, even B-cells with low Ag binding affinity can colonize GCs, but they fail to do so if higher affinity competitors are present [126]. In [33, 126] for examples studies on transgenic mice showed that in absence of competition even affinities as low as 8 μM are sufficient for B-cells to colonize GCs.

Even though the threshold for B-cell activation is low [14], nonetheless in some instances this threshold might be relevant. For example in cases when Ag availability is very limited, or when in a subsequent immunization an Ag mutant is administered for which most of the GC B-cells have low binding affinity. In such cases cell-intrinsic selection might exert the highest level of selection pressure, causing a fast population decrease and affinity maturation. Typically, if the maturation is fast enough for some cells to meet the cell-intrinsic selection threshold then the population size starts to increase again, and maturation continues under the action of competition. If on the other hand maturation is not fast enough, then the population might go extinct before the cell-intrinsic threshold is met. This low-population-size state in which extinction is possible was termed *population bottleneck* and was included in different maturation models [155, 154, 161]. For example, in [161] the authors argue that the mutation rate and selection pressure in GCs might have been set by evolution so as to be optimal for maturation. In particular they suggest that the optimal mutation rate is a result of a tradeoff between avoiding

too many lethal mutations and accumulating enough beneficial ones, while the optimal selection strength is chosen so as to maximize maturation speed while at the same time avoiding population extinction in the bottleneck. They introduce a model for AM in which B-cell growth rate, which includes contributions from lethal mutations, duplication and selection, is approximated to a linear function of the cell binding energy for the Ag, and define the critical affinity as the one at which growth rate is zero. The authors then write the differential equation that governs the evolution of the average number of cells of a given affinity under the approximation of infinite population size. This equation other than the growth rate also includes the contribution of affinity-affecting mutations. Its solution shows that if the initial affinity of the population is below the critical threshold then the total population size will initially be subject to an exponential decrease, followed by a super-exponential increase[17] once the population affinity reaches the critical threshold. The model is then extended to account for the finite size of the population. In this case population survival in bottleneck conditions depends both on the population initial size and on the difference between its initial binding energy and the threshold one, this difference being controlled by Ag availability. The authors numerically solve this extension for different values of mutation rate and initial affinity, and then compare the affinity increment obtained. In doing so they observe the above mentioned optimality in mutation rate and initial affinity.

negative selection and self-reactivity

Additional insight on GC selection can be gained by observing negative selection and apoptosis in the GC. Both the DZ and the LZ have been shown to contain apoptotic B-cells, but with different characteristics [86, 138]. Most of DZ apoptotic cells have acquired mutations that impairs BCR expression, indicating SHM as the major source of apoptosis. LZ apoptotic cells instead possess functional BCRs, and interestingly these include both high and low affinity cells. This suggests that cells can be selected out not only if they fail to bind the Ag (low affinity) but also for other affinity-independent reasons, such as failure to encounter Ag-presenting cells [74]. Indeed, the possibility that stochastic effects might play a role in GC selection was already advanced in the literature [90, 144].

Negative selection mechanisms must also be in action to protect against the development of self-reactive Abs that could arise as a result of SHM [21, 71]. The processes that take care of eliminating self-reactive clones are not completely understood. One possible mechanism might be the already mentioned fact that a too strong BCR stimulation, such as the one that might be caused by the development of self-affinity, activates apoptosis. Another factor contributing might be that T-cells are screened for self-reactivity. However recent experiments showed that GC selection is tolerant to low levels of self reactivity. Slightly self-reactive clones can be tolerated in GCs for long enough times for mutations to remove self-reactivity in favor of foreign affinity, in a process that has been termed *clonal redemption* [23, 24].

In this context an interesting model to discuss is the one presented in [69]. Here the authors show how affinity discrimination can be enhanced if selection is based on repeated productive interactions between B and T cells, and that this mechanism can also protect against the development of self-reactive B-cells. In the model the authors suppose that

17 This super-exponential increase, which is not biologically plausible, arises as a consequence of the authors approximation of the growth rate being a linear function of binding energy. In reality for high enough affinities the growth rate saturates, limiting the population size increase to an exponential increase.

in the LZ a B-cell will undergo multiple encounters with T-cells, and depending on the amount of Ag displayed it will have a certain probability for the encounter to be productive. The authors show that if selection is operated on the basis of the number of productive encounters, then this mechanism can exponentially amplify the difference in affinities into difference of survival probabilities[18]. The number of Tfh cells (or rather their encounter probability) also influences this affinity discrimination. When the encounter probability is small then affinity difference are further amplified, but the total survival probability is smaller and selection more stochastic, generating a tradeoff phenomenon by virtue of which optimal selection is operated at an intermediate encounter probability. The authors also extend the model to account for self-reactivity. In this case the fact that T-cells are screened against self-reactivity makes so that on average cells possessing BCRs with some self-reactivity will be less likely to have productive encounters. This disadvantage will be less marked for cells with a weak self-reactivity but a good foreign-Ag affinity, and this mechanism allows the presence of weakly self-reactive cells.

Another process that could contribute to selection is *Ab-feedback* [162]. This term identifies the process by which Abs secreted earlier in the response can influence subsequent B-cell selection, for example by binding Ag on FDCs and preventing B-cells from accessing it (*Ag masking*). This process is thought to accelerate maturation by providing an additional source of competitive pressure but also to shorten GCR duration and GC volume [11]. Moreover, it is also a possible mechanism through which different GCs could interact, by being permeable to Abs from PCs generated elsewhere. Ab-feedback has also been included in different computational models [155, 162, 11]. By blocking access to a targeted epitopes, Ab feedback might also constitute an incentive for B-cells to target new epitopes. As will be discussed in the next section this principle might be exploited for the generation of BNAbs [93].

Ab-feedback

Taken together all of these considerations show that selection in the GC is a complex process and, albeit much experimental progress has been made in identifying some of the mechanisms at play, it is far from being completely understood. The considerations discussed above also shaped our modeling approach, and are relevant with respect of some of our results. For instance, in the AM model that we present in chapter 2 we include a parameter[19] that encodes the effects of stochasticity in selection, and regulates the probability for a high-affinity cell to fail selection. Inference of this model on data allows us to probe the relative importance of T-cell help competitive selection versus BCR Ag-binding selection, our analysis indicating that in our experiments the former might play the major role. Moreover, in section 3.4 we study how the bottleneck survival probability is controlled by Ag availability and initial population size, and in section 3.5 we consider how the bottleneck state shapes evolutionary trajectories in affinity space. Finally, in chapter 4 we discuss preliminary results indicating that the effect of Ag concentration in tuning selection strength can be reproduced by a simple interaction model between B and T cells, in which the role of stimulatory signals from bystander B-cells in recruiting T-cells is encoded in the dependence of the number of T-cells on the number of B-cells.

18 The way this amplification is achieved shows similarities with the mechanism of *kinetic proofreading* [54, 108].
19 Namely b in eq. (7)

1.3.4 *Maturation in the presence of complex Ag, selection permissiveness and homogenization*

complex Ag makes GC selection permissive

Most of the early AM experiments were performed by immunizing animals with *haptens* (see section 1.3.2 for the definition). Immune response against these simple molecules is often characterized by a big affinity increase [38] and limited genetic diversity of the responding population [59]. However, while remaining easy to study, the response against haptens might fail to capture relevant aspect of maturation against more complex Ags, such as the ones present in an infection or in immunization. In a recent review [41] Finney et al. try to characterize some of these differences. They discuss experiments of vaccination against complex Ags (*Bacillus anthracis* and *influenza hemaglutinin*) performed in [70]. In these experiments it was observed that, as expected, the avidity[20] of GC B-cells gradually increased over time. More surprisingly, when measuring affinity of B-cells at day 16 after immunization, it was observed that around 75% of cells displaying detectable binding affinity had similar and very high avidity, but the remaining 25% covered a very wide avidity range (> 100 fold) with much lower avidities (see fig. 7A). Even amongst the same clone, cells spanned a 40-fold interval of avidity values. In addition to this, a relevant portion of cells displayed no detectable binding affinity. These observations suggested a more *permissive* nature of GC selection, and led Finney and colleagues to formulate a selection model in which the probability of being selected for a GC B-cell is an increasing function of affinity, while at the same time being non-zero even for low-affinity cells (see fig. 7 B).

This permissiveness is also related to the level of clonal homogenization, i.e. the reduction of diversity in the clonal composition of GCs due to selection forces. As stated above, such reduction was indeed observed in hapten responses [60]. However the diversity loss might be less marked in responses against complex Ags, where more heterogeneity is retained. In one famous experiment [144] Tas and colleagues were able to follow the study the clonality of GCs with a clever genetic color-tagging of different clones. Their experiment revealed that indeed diversity loss was occurring at different rates, with some GCs being quickly dominated by a single clone already 9-11 days post injection, while others remained heterogeneous throughout the study period, and that maturation can occur also in the absence of homogenization.
In another interesting experiments of immunization [2] Abbot and colleagues study the effect of precursor frequency and affinity on the competitive potential of GC B-cells. They created a system in which both the affinity and the frequency of a specific precursor could be controlled. Amongst other results, they measured the dominance of this precursor in GCs at different times. At physiological frequencies, homogenization was observed only if the precursor had high affinity. Even in this case clonal dominance was not complete, with some GCs showing a low occupancy of these high-affinity cells.

One of the possible mechanisms that could generate this permissiveness is *bystander activation* [58], i.e. the delivery of activation signals from T-cells to neighboring B-cells in

20 While *affinity* characterizes the interaction between the Ab and a single Ag molecule, *avidity* is related to the cumulative binding strength of an Ab with multiple Ag molecule. Abs with similar affinities can have different avidities if one binds a single Ag molecule, while another can simultaneously bind multiple such molecules. One factor that influences avidity is the Ab isotype, since some isotypes such as IgA and IgM are polymers composed of multiple Ig monomers.

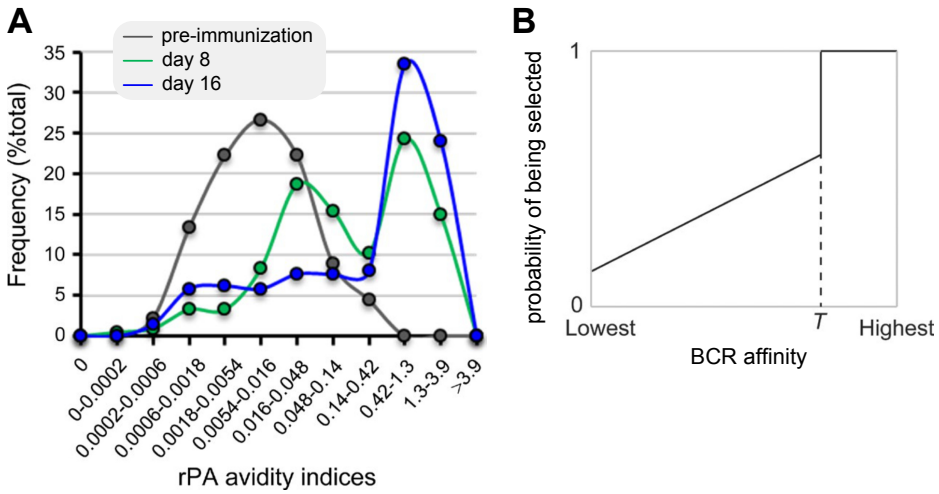

Figure 7: **A**: experimental distribution of avidities of IgGs from mature B-cells after immunization against *bacillus anthracis* protective antigen (rPA), from experiments performed in [70]. Distributions corresponds to cells extracted before immunization (gray) or at day 8 (green) or 16 (blue) after immunization. As expected from maturation, the average avidity increases over time. However even at day 16 a consistent part of the population (around 25%) is composed of low-avidity B-cells remain. **B**: GC selection model proposed in [41]. The authors suggest that even low-affinity B-cells might have a non-zero probability of being positively selected. Figure adapted from [70] and [41].

a non-Ag-specific manner. This could occur because of membrane Ag transfer between B-cells [119], or simply for physical proximity of specific and non-specific B-cells when a high amount of T-cell help is available [16]. This mechanism was invoked as a possible explanation for different experimental observations, in which responder clones following vaccinations showed no detectable affinity for the administered Ag. For example in [56] the authors analyzed data from human influenza vaccination. Using samples collected at multiple time-points they could identify a subset of responder clones. Surprisingly, a large fraction of these responders had no binding affinity for the Ag. Similarly, after vaccinating mice against Tetanus Toxoid in [39] the authors analyzed IgG-secreting cells elicited by a recall pure Ag injection. It is reasonable to believe that most of these fast responding and class-switched cells must have been generated during previous immunization. Again, only a minor fraction (around 33%) of cells showed relevant Ag affinity ($K_d < 500$ nM). It is however not yet clear the role that bystander activation might play in GC selection, and more experimental investigation is needed.

The permissive nature of GC selection might be especially important in the fight against mutable pathogens. The development of broadly-reactive Abs often requires the accumulation of multiple mutations, during which Abs remain in a low-affinity state. Not selecting these cells out might be crucial for the development of broadly-binding Abs [13]. Moreover when facing mutable pathogens, rather than uniquely maximizing affinity it might be advantageous also to maintain diversity in the responding population [132, 58]. Especially given the fact that, as previously discussed, low-affinity cells preferentially differentiate to join the MC compartment, building up the second "wall of

defense" [5].

maturation by precursor selection

Another interesting experiments that highlight a difference between simple and complex Ag immunization is the one performed by Murugan et al. [102]. While usually in responses against haptens maturation relies mainly on the accumulation of few key mutations that confer a high affinity increase, in complex Ags maturation might occur mainly by selection and expansion of high-affinity precursors that were already present at the beginning of maturation. Using single-cell Ig sequencing Murugan and colleagues studied the response to immunization against *Plasmodium Falciparum*, the malaria parasite, through the administration of the sporozoites chemoprophylaxis vaccine.[21] They found that while the average affinity was increasing over time, responders did not show signs of efficient maturation. Cells were harboring many mutations but in most cases these did not seem to increase the affinity for the Ag.[22] Instead, most of the high-affinity memory cells were originating from unmutated precursors. The affinity increase in this case was due, rather than to the appearance of advantageous mutations, to the efficient selection of these rare high-affinity germline B-cells, that even outcompete mutants with improved affinity, but that developed from lower-affinity precursor.
To interpret these results the authors introduced a very interesting model for AM, in which different "Ag complexities" could be considered. Ag complexity was encoded the number of beneficial mutations that were to be acquired for a given total affinity improvement, with the underlying assumptions that for simple Ags like haptens even a single mutation can confer a high affinity increase [8, 17], while for complex Ags the accumulation of more mutations is needed. In the model for each Ag injection 50 GCs were simulated, with MCs produced from a GCR having the possibility to seed future GCs (see fig. 8 A). By visualizing the final affinity distribution of the produced MCs the authors observed that for complex Ags most of the high-affinity individuals consisted of unmutated high-affinity precursor (fig. 8 B), that were gradually selected in the course of maturation, colonizing a higher number of GCs in subsequent injections. To confirm this they re-executed the simulation removing the possibility for cells to accumulate beneficial or deleterious mutations. This did not create a relevant difference in the final MC affinity distribution. The situation was different when simpler Ags where considered. In this case high-affinity cells were mainly generated through accumulation of beneficial mutations (see fig. 8 C).

Finally, let us discuss the relevance of these considerations with respect to our modeling approach. In our maturation model, later presented in chapter 2, we incorporated a parameter that encoded selection permissiveness.[23] By applying the model on experimental data we unveiled many of the signatures of maturation against complex Ag. As will be discussed later, together with a high level of permissiveness, we also observed a varying degree of homogenization, that correlated with the presence of a high-affinity precursor.

21 We point out that response to this parasite is peculiar, in that vaccination of pathogen-experienced individuals seems to be less effective that for pathogen-naive individual. It is speculated that this might be related to the atypical MCs produced during infection [159].

22 The absence of correlation between the amount of mutations accumulated and the Ab affinity was also observed in experiments of TT immunization [43].

23 Namely parameter α in eq. (7)

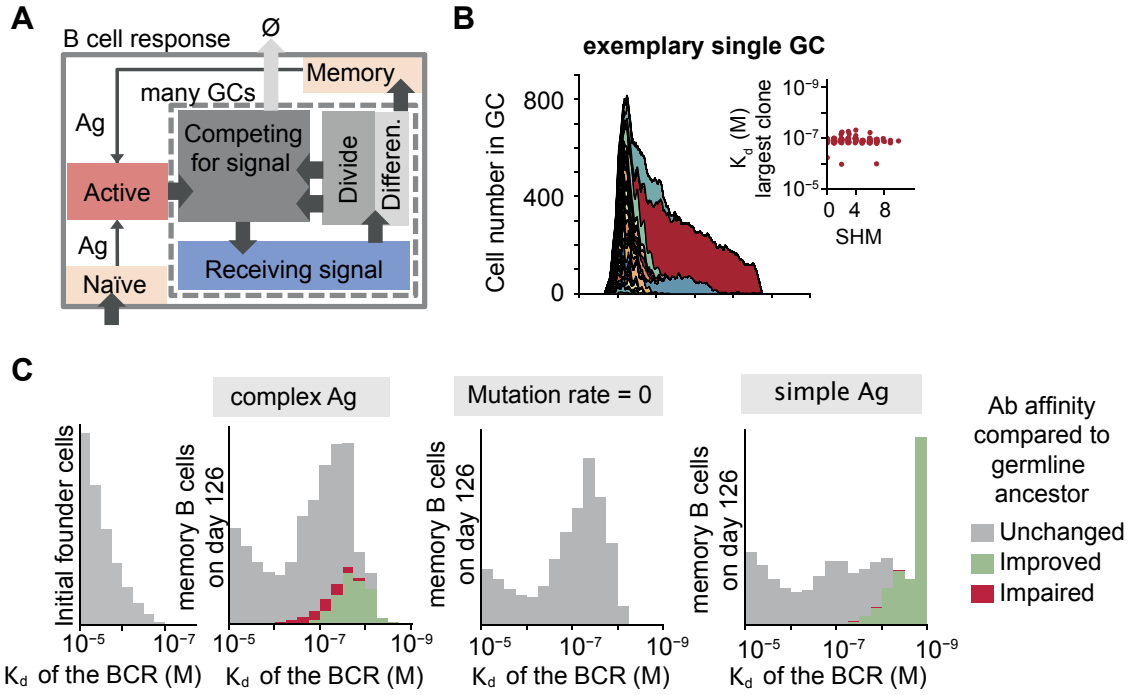

Figure 8: **A**: schematic representation of the AM simulation by Murugan and colleagues [102]. Many GCs are simulated in parallel, with B-cells undergoing cycles of mutation and selection. When a new Ag injection is administrated cells that differentiated into MCs can be reactivated and colonize new simulated GCs. **B**: example of clonal dynamic in a simulated GC. Notice that for the most expanded clone (red) the number of mutations does not correlate strongly with affinity, indicating that it was not selected as a result of accumulated mutations. **C**: affinity of the MC pool at the end of the immunization scheme for three different simulated scenario. When a complex Ag is considered (left), for which affinity increase requires the accumulation of many small mutations, then maturation occurs mostly by selection of pre-existing high-affinity clones. In fact even if mutations are turned off (middle) a similar maturation is observed. If a simple Ag is considered instead (right), for which few key mutations can confer a big affinity increase, then maturation occurs by accumulation of beneficial mutations. Figure adapted from [102].

1.3.5 *Immunizing against mutable pathogens*

One of the most important research frontiers in immunization protocols concerns the development of vaccination strategies against highly-mutable pathogens. These pathogens can escape immune recognition by developing mutations on targeted epitopes, and as such are extremely hard for our IS to fight and eliminate. Nonetheless, in some rare cases the IS manages to develop antibodies capable of binding conserved pathogenic regions which are not affected by mutations, and therefore are able to recognize a broad variety of virus mutants. These potent Abs, known as *Broadly Neutralizing Abs* (BNAbs), are however naturally generated late in the response, and given the diversity of the viral population at late infection stage are insufficient for total pathogen removal. However, they can be protective against infection if they can be elicited via an appropriate immunization strategy. The development of such strategy is therefore one of the main therapeutical research goals, pursued through experimental efforts often guided by theoretical models.

HIV infection and BNAbs

The most notorious example of this kinds of pathogens is the *Human Immunodeficiency Virus* HIV-1, which causes acquired immunodeficiency syndrome (AIDS). To be able to neutralize this pathogen Ab must bind functional sites on its envelope spike gp160, encoded by the ENV gene. Binding of these regions is made hard by different protection mechanisms [100], the most important of which is the fact that initial responses are usually directed to spike regions which are highly mutable, a fact that earned the virus the title of "nature's master of disguise" [85]. Other possible mechanisms include the shielding of conserved residues, which complicates access and binding.
In a typical immune response against the virus [100], the first effective Abs appear some months after infection. They are specific for the viral population present in the host, and exert a sufficient selective pressure to elicit the development of escape mutants. At this point the IS and viral populations start a process of coevolution, each population exerting selection pressure on the other in an evolutionary "arms race", whose signature can be detected in the repertoire evolution [109]. Some years after infection, BNAbs start to appear in a small percentage of infected individuals (around 1%). These Abs are effective against a wide variety of viral sub-species. At this point the viral population diversity is too great for them to clear the infection, however they could contribute to keep viral load in check. The study of these Abs revealed that they are typically characterized by a high level of SHM in framework and CDR regions, and that these mutations are essential for the Ab affinity [65].
The time needed for appearance of BNAbs and their relatively high mutational load indicates that their development requires extensive maturation [148] that may transit through low-affinity states. This suggests that the permissive nature of GC selection (discussed in section 1.3.4), which allows also low-affinity clones to remain in the GC for extended periods of time, might be essential for BNAbs development [13].
Moreover the fact that they appear only after that the virus population has developed sufficient diversity indicates that if immunization strategies exist that can elicit the appearance of BNAbs, these must make use of multiple virus variants.

This raises many questions, such as how many variants should be used, how should they be chosen, in what concentration should they be present and in what temporal

order should be administered. The answer to these questions requires understanding of how AM occurs in the presence of multiple Ag variants, and how it can be directed to develop maximum breadth. Many theoretical models have been developed to tackle these questions. A fundamental ingredient of these models is the way Abs and different strains of Ag interacts, and how mutations in these two populations influence the binding affinities. Different modeling choices in this sense are reviewed in [123]. Here we present some examples of relevant models that tackle the problem of optimizing vaccination.

Ag variants administered sequentially or in a cocktail vaccine

As a first example, in [155] the authors consider the problem of HIV vaccination, introducing a model of AM in the presence of multiple Ag variants, and study the optimal dosage and time administration of these variants. They represent Ags as string of binary values $\{+1, -1\}$, where -1 corresponds to a mutation. The string is divided in three regions. One region represents conserved residues, that do not mutate between variants. A second region encodes for variable residues, that instead can freely mutate. A third region corresponds to shielding residues, and an Ab mutations that increases affinity for these residues will as a consequence decrease affinity for the binding of constant regions. Abs are represented as real-valued vectors, with each component indicating the binding energy for each viral residue. The total binding energy between any Ab and virus is the sum of these contributions, with an inversion of sign for viral residues that are mutated. AM proceeds by rounds of duplication, mutation and selection. When an Ab develops an affinity-affecting mutation the binding energy for one randomly picked viral residue is varied with an additive contribution, with deleterious mutations being more likely than beneficial ones. Selection is instead performed in two steps: Ag-binding and competition for T-cell help. The probability of passing these steps depends on Ag availability and on the cell binding affinity, and for competitive selection also on the affinity of the rest of the population. They also consider Ab-feedback in their model, in which case competition includes also Abs produced during the course of the GCR. Moreover, when multiple Ag variants are administered together two possibilities are considered: cells either encounter and bind only one Ag variant, or all of the variants together.
With these ingredients the authors compare the outcomes of simulations of three different immunization schemes. In one the wild-type (WT) Ag, containing no mutations, is administered in a cocktail injection together with two other Ag variants. In a second scheme a first WT injection is followed by a second injection with a cocktail of two mutants. In a third scheme the WT and the two Ag mutants are administered separately in three sequential injections. For these three schemes the quality of the resulting immunization is measured both in terms of Ab breadth and of probability of surviving the population bottleneck. They find that Abs with maximal breadth can be more consistently elicited if Ag variants are sequentially administered, and at an intermediate Ag dosage. This outcome is interpreted in terms of *optimal frustration*: when multiple Ag variants are administered they apply conflicting selection forces to the population, each one pushing towards maximizing the binding affinity for a particular variant. When this frustration is too intense, meaning a strong pressure is applied for the simultaneous binding of all the variants, then the most common outcome is extinction. Conversely if the frustration is too low, and selection forces do not push the population towards simultaneous variant binding, then strain-specific Abs are developed. In the model this optimal frustration is obtained at intermediate Ag concentration, and by temporally sep-

arating conflicting selection forces via sequential immunization.
Recently, some experimental evidence has been collected in support of this view. For example in [95] experiments of rabbit immunization with different variants of the HIV-1 ENV spike proteins showed that sequential immunization elicited a more broad serum neutralizing activity when compared to cocktail immunization.

optimal frustration and number of viral strains in a vaccine

The principle of intermediate optimal frustration discussed above was also invoked in [127]. In this paper the authors study the effect of immunization with a cocktail of different Ag variants, focusing on optimizing the number of variants, their mutational distance and Ag concentration. In this case affinity of Ags and Abs is modeled using a shape-space representation (see fig. 9 A). Both Abs and Ags are defined by a vectors in an 8-dimensional space which encodes for their biochemical properties, in such a manner that the binding energy of an Ab-Ag pair is proportional to the square norm of their distance. The dimensions of this space are divided in four *conserved* dimensions and four *variable* dimensions. The position of Ags is constrained to lay on the coordinate zero in the conserved dimension, while coordinates in the variable ones are randomly extracted from a gaussian with zero mean and same variance. Abs initial vectors are instead initially randomly picked so as to lay on an hypersphere centered on the origin. As in the previous model Abs evolve in iterative rounds of duplication, mutation and selection. Affinity-affecting mutations are performed by shifting the Ab position in the shape-space adding a random vector, whose components are randomly extracted from a gaussian. As a consequence of the high dimensionality of the shape-space mutations are on average deleterious. Selection operates in two steps, similarly to the previous model, with one step encoding Ag binding and the other competition for T-cell help. As in the previous work, survival depends on both Ag concentration and affinity for one or more Ags. Given that the Ag population is on average centered around the origin of the space, it follows that breadth acquisition for Abs corresponds to accumulating mutations that shift them closer to the origin. By simulating maturation with a different number of Ag variants in the administered cocktail the authors observe the existence of three regimes. When a few variants are administered the Ab population quickly develops into few strain-specific Abs with sub-optimal breadth. If on the other hand a big number of variants are administered conflicting selection forces quickly lead the population to extinction. Optimal breadth is obtained at an intermediate number of variants, corresponding to an intermediate level of frustration in the selection forces (see fig. 9 B and C). For any fixed number of administered variants a similar optimality is observed also in concentration. This is a consequence of the role that this parameter plays in controlling the selection strength. The authors observe that the optimal concentration depends strongly on the number of variants used, which potentially complicates vaccination. The optimal concentration can be made more robust if the phenotypic distance between the Ag variants is reduced when their number is increased.
In the article the authors provide an interesting interpretation of the observed optimality, describing maturation in this context as a dynamical system with two strong attractors. The first, corresponding to a high level of frustration, leads to population extinction, while the second, at low frustration level, leads to strain-specific Abs. If the model correctly captures maturation in the presence of multiple Ag mutants, then the complexity of eliciting BNAbs through vaccination can be understood in terms of the complexity of initializing the system on a trajectories that remains as long as possible in an unstable

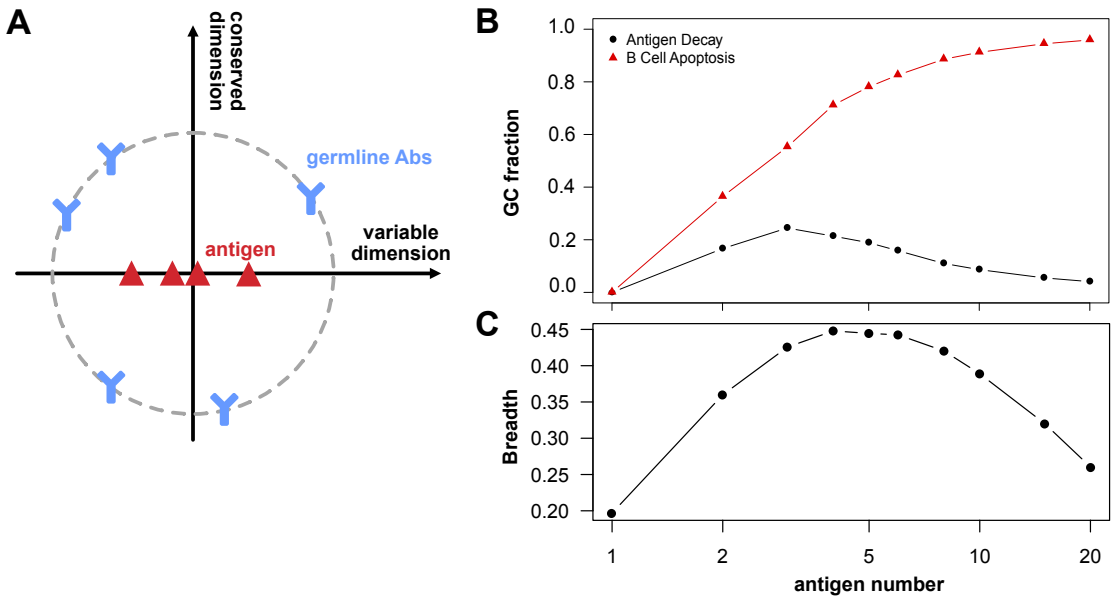

Figure 9: **A**: simplified depiction of the shape-space representation for Ab-Ag affinity used in [127]. Affinities are a decreasing function of the distance between an Ag and an Ab in the space. Administered Ags are constrained to have coordinate zero on the conserved dimension, while Abs have initially all the same distance from the origin of the space, corresponding to the position of maximal average breadth. Affinity-affecting mutations change the position of an Ab by a random shift. **B**: fraction of GC simulation that terminate by Ag decay or by B-cell apoptosis. Simulations can terminate in three ways, either by Ag decay if the simulation times exceeds a maximum threshold, or by Ag consumption (not shown) if the B-cell population exceeds a maximum threshold, or by B-cell apoptosis if the population goes extinguished. At small numbers of Ag variants the frustration is small and most of the simulations terminate by Ag consumption. At high Ag number instead frustration is high and most of the simulations terminate in population extinction. Termination by Ag decay, which correspond to long-lasting maturations, are only obtained at intermediate Ag number and frustration levels. **C**: maximal immunization breadth is obtained at intermediate number of Ag variants, i.e. intermediate levels of frustration. Figure adapted from [127].

equilibrium between these two attractors.

exploiting Ab-feedback

Another possible strategy to push the B-cell population to develop affinity for a conserved but sub-dominant epitope was proposed in [94]. Here the author studies the effect of Ab-feedback in GC reactions and suggests that exploiting epitope-masking of the immunodominant epitope, achieved by injecting epitope-specific Abs, could direct maturation towards the sub-dominant epitope. The model employed is part of a family of agent-based spatial models [94, 92, 20] in which different cell types are embedded and move in a simulated GC environment. In this environment B-cells can bind Ag deposited on FDCs in an affinity-dependent manner, interact with T-cells to receive help, divide and mutate. A space-shape representation, similar to the one employed in the previous model, is used to encode Ag-Ab interaction. Only two epitopes are considered in the model, which are differentiated by their accessibility, the hardly-accessible epitope being harder for B-cells to encounter. In each Ag encounter B-cells bind preferentially

the Ag for which they have the highest affinity, which is progressively consumed. If Abs are present in the GC, either because artificially administered or because secreted from previously produced PCs, they can bind and mask Ag of the corresponding epitope and prevent B-cells from binding to it.
Simulations of the model without the presence of Ab feedback or injected Abs shows that the response focuses mainly on binding the easily accessible epitope. When Ab feedback is included Abs produced early in the response contribute to the masking of this epitope, producing a weak but detectable response against the sub-dominant epitope.[24] This response remains weak because the masking starts to operate only late in the course of the GCR, when high-affinity Abs for the dominant epitopes start to be available. If instead Abs against the dominant epitope are administered then the response becomes skewed towards the sub-dominant epitope, in a manner dependent on the concentration of these Abs, with as a side effect the total magnitude of the response being smaller, since less total Ag is available for binding. This result leads the author to speculate that complementing cocktail vaccine with Abs against dominant epitopes might help to focus the response against less dominant but more conserved epitopes.

coevolution models and diversity in vaccine viral strains

Even though not directly conceived for studying vaccination, also models of coevolution between the HIV population and the IS could provide suggestions on how to better prompt the IS to produce BNAbs. For example in [79] the authors simulate the coevolution of virus strains and Abs. In simple terms, viruses and Abs are represented as strings of characters. A virus presents two epitopes, a conserved and a variable one. Abs can bind either of the two, with the affinity depending on how well the two strings match. The fitness of an Ab increases with its average affinity for the viral population, while conversely viral fitness decreases with the average affinity with which the virus is recognized by the Ab population. With this setup the authors find that the development of BNAbs is favored when the viral population is highly diverse. In these conditions Abs are better pushed to developing affinity for the conserved viral epitope, which leads the authors to advocate for the use of a cocktail vaccine with many diverse mutants.[25]
The fact that diversity in the viral population favors the appearance of BNAbs was also suggested by another coevolution model [110]. Here viruses and Abs are represented as binary vectors of $\{+1, -1\}$ values, with viruses possessing a conserved region that is the same for all viral variants. The binding affinity between a virus and an Ab is expressed as a dot-product between the vectors, with weights for each component representing residues accessibilities. Similarly to the previous case the two populations coevolve, with the fitness of an Ab (virus) depending on how well it recognizes (is recognized by) the opposite population. The authors write stochastic differential equations that describe the evolution of the the average binding affinity for the variable and conserved regions, and show that under some approximation this out-of-equilibrium evolution admits a

24 As stated by the author, the positive effect of slow Ag administration schemes on the development of BNAbs [27] could be partially explained by the fact that in this case the longer duration of the response makes so that Ab feedback has the time to shift the response towards sub-dominant more conserved epitopes.

25 The difference between this results and what found in [155] is to be traced back to the different hypotheses that these models employ. One major difference is the fact that in this model Abs target either the conserved or the variable viral epitope, while in [155] given the spatial proximity of conserved and variable residues in HIV the authors assumed that Abs interact with both a conserved and variable epitopes at the same time.

stationary solution in which each population adapts to the other at the same rate.[26] The authors then extend the model to account for different Ab lineages, and consider the probability that the introduction of a new Ab lineage results in fixation of this lineage in the population. They find that this probability depends both on its relative fitness with respect to the other lineages, but also on their rate of adaptation to the viral population and on the viral population rate of "counter-adaptation". When considering BNAbs these do not possess a particularly higher fitness than other lineages, but their advantage resides in the fact that the viral rate of "counter-adaptation" is reduced by the fact that virus cannot escape binding by developing mutations on the conserved epitope. This advantage is more marked, and the fixation probability higher, if the viral population is more diverse, which leads the authors to suggest that the efficacy of an immunization strategy in conferring breadth might increase if it makes use of many different viral strains.

We point out that in some cases the results obtained through these models depend on the assumptions they employ on how maturation plays out when multiple Ags are present, and in some cases varying these assumptions leads to different results. This highlights the importance of experimental investigation of maturation mechanisms in the presence of Ag variants.

We conclude this section by mentioning that chapter 5 of our work contains preliminary results that are related to mutable pathogen immunization. There we investigate the effect of single-Ag immunization against pathogens with different mutability. Our model predicts the existence of a critical mutability level above which single-Ag immunization is not beneficial for breadth acquisition.

26 Interestingly, the authors show how this rate of adaptation can be experimentally measured by studying its effect on antibody neutralizing affinity against time-shifted viral populations. They in fact show that the rate of adaptation of the two populations, which they describe in terms of *fitness flux* and *transfer flux*, can be measured by comparing the neutralizing power of Abs at any given time against viral populations from past and future time-points (and symmetrically of viral populations against time-shifted Abs populations).

2

EFFECTS OF AG DOSAGE: MODELING AND DATA ANALYSIS

2.1 INTRODUCTION AND CHAPTER OUTLINE

In this chapter we present an application of inference techniques to the study of Affinity Maturation, with special focus on the effect of Ag dosage.
In the previous chapter we discussed experimental evidence demonstrating that Ag dosage plays a non-trivial role in AM (see section 1.3.2). In particular experiments show that optimal maturation is obtained at intermediate values of the dosage [63], and suggest that this phenomenon could be a consequence of the selection pressure exerted by the availability of Ag. Since B-cells have to bind Ag to survive selection, this pressure must be stronger when Ag is scarce [46, 112, 143]. The importance of reaching a quantitative understanding of this non-trivial effect of Ag concentration, through modeling and inference, lies in the hope that this understanding could eventually be used to improve human vaccination [121].
A second open issue concerning AM is to characterize in a quantitative way the selection acting in the GC, in particular how *permissive* it is [13, 90, 148, 58] (see section 1.3.4). Through mechanisms such as bystander activation [19, 39, 40] GC selection can indeed allow intermediate and low affinity clones to survive [144]. These phenomena generate a wider diversity than previously appreciated, especially when considering complex Ags displaying different epitopes [70]. Permissiveness could for example be useful against mutable pathogens, where maintaining a pool of general cross-reactive cells might be a better strategy than only selecting for the best strain-specific binders [41].

Building from these premises and taking inspiration from other models in the literature [155, 154] here we introduce a stochastic model for Affinity Maturation in the context of vaccination. The model encodes explicitly the role of Ag concentration in tuning the selection strength. With the use of a new technique they developed [40] our collaborators were able to obtain experimental measurements of affinity distributions of Antibody-Secreting Cells (Ab-SCs) generated following different immunization protocols, in which both Ag dosage and time between injection was varied. This rich information is then exploited in our inference procedure. By performing the deterministic limit of the model we obtain an expression for the probability of a single experimental observation as a function of model parameters. This expression can be inverted to numerically evaluate the Maximum-Likelihood Estimate of the unknown parameters. The inferred model can quantitatively reproduce the data, including the existence of an optimal Ag dosage. Importantly, this optimality can then be interpreted with a theoretical analysis of the deterministic limit of the model. We show that at constant levels of Ag concentration the affinity of the B-cell population in our model tends to evolve as a traveling wave. By formulating this limit as an eigenvalue problem we can derive how this limit depends on Ag concentration. This analysis reveals that Ag availability controls the GC population growth and maturation rates, generating a tradeoff phenomenon of which the observed optimality is a consequence. In addition to this, inference allows

us to investigate hardly-accessible features of maturation, such as the permissiveness of GC selection. In accordance with recent advancement in the literature we find that also in our case GC selection seems to be permissive.

The chapter is organized in the following sections:

Foreword: Bayesian inference : a very short introduction on Bayesian Inference, illustrating the relation between Bayes Theorem and Maximum Likelihood Estimate.

The experimental dataset : contains a description of the technique used to obtain the experimental measurements on which our inference is based, and qualitative observations on the resulting dataset.

Stochastic model of AM to predict affinity distributions : here we introduce a stochastic model for AM, in which the effect of Ag concentration is explicitly encoded. The model offers predictions for affinity distributions for comparison with experiments.

Model deterministic limit : in the limit of infinite size we obtain a deterministic version of the model, on which our inference technique is based.

Inferring model parameters : using the deterministic model we can define a likelihood function for the model parameters. With an implementation of the *Parallel Tempering* technique we numerically maximize this likelihood as a function of unknown model parameters. With the inferred value of model parameters the model is able to quantitatively reproduce experimental affinity distributions.

Analysis of deterministic model offers insight on the effect of Ag dosage : we study the asymptotic model behavior at constant Ag concentration, expressing it in the form of an eigenvalue equation. The solution of this equation shows how Ag concentration controls both population growth and maturation speed, generating a tradeoff that is at the base of the observed optimality of intermediate Ag dosages.

Inference as a tool to investigate AM mechanisms : after fitting the model to data, the value of inferred parameters provides insight on the inner workings of GC selection and AM. In particular we discuss the permissiveness of GC selection, the degree of homogenizing selection and the relative contributions of Ag-binding and competitive selection.

Conclusion : it contains a summary of the main results of the paper, along with a discussion on the limitations of the model.

The work discussed in this chapter is the object of a publication [96], from which the chapter is partially adapted.

In the preface of his book "Probability theory: the logic of science" [62] E.T. Jaynes states that inference is concerned with the *optimal processing of incomplete information*. In real-life problems often we do not possess enough experimental evidence to answer questions directly, by means of deductive reasoning. In most of the cases, the best strategy is to integrate all the available information on the system to get as close to the answer as possible. *Statistical inference* is indeed concerned with performing this information integration within the mathematical framework of probability theory.

Probabilities can be defined as *degrees of belief* in propositions.[1] If x is a proposition then we indicate with $P(x)$ this degree of belief. Starting from *Cox axioms* [31] (presented in a simplified version in [81]) one can show that degrees of belief can be mapped into probabilities, satisfying the usual probabilities rules. In particular, one of the consequences of these axioms is the *chain rule*, that states that $P(x, y) = P(x|y)\,P(y)$. Here $P(x, y)$ represents the probability that both propositions x and y are true at the same time, and $P(x|y)$ is the probability that proposition x is true provided that we known proposition y to be true. From the chain rule we can easily derive the following:

$$P(y|x) = \frac{P(x|y)\,P(y)}{P(x)} \tag{1}$$

This simple relation, known as *Bayes Theorem*, is at the base of all *Bayesian inference*. It offers a prescription on how to update the degree of belief in y after new evidence x is found. The term $P(y|x)$, known as *posterior probability*, is this updated probability after evidence is observed. This is proportional to the likelihood of observing the evidence if y were true, i.e. $P(x|y)$, times the *prior probability* of y, i.e. $P(y)$. This latter term accounts for our prior belief before any evidence is observed.

Most importantly, this theorem can be used to solve *inverse probability problems*. These consist in inferring the value of hidden variables using the combination of a generative model for the process in question and of the observation of accessible variables.
To make things more explicit let us suppose we are interested in knowing the value of some hidden parameter h that controls a process that generates a set of independent observations $D = (d_1, \ldots, d_N)$.[2] If we know the way this hidden parameter influences each observation, i.e. we know the direct probability $P(d_n|h)$, then Bayes theorem states that:

$$\log P(h|D) = \sum_{n=1}^{N} \log P(d_n|h) + \log P(h) + \text{const} \tag{2}$$

The most likely value of the hidden parameter given the evidence is then given by the value that maximizes the posterior probability:

$$h^* = \arg\max_h \left[\sum_{n=1}^{N} \log P(d_n|h) + \log P(h) \right] \tag{3}$$

1 This is known as the *Bayesian* or *subjectivist* approach. Alternatively one can define probabilities as empirical frequencies of repeated random trials, which is known as the *frequentist* approach. There is currently an open debate between these two positions.

2 Note that this can be easily generalized to the case in which h and d are vectors.

If no prior expectation on the value of h is present then a uniform prior $P(h) = \text{const}$ is conventionally assumed, which reduces the problem to maximizing the likelihood term containing the sum. In this case the procedure is known as *Maximum Likelihood Estimation* (MLE). If instead a non-uniform prior is assumed then the procedure is termed *Maximum a-posteriori* (MAP).[3]

In this chapter MLE will be used to infer the value of parameters of our maturation model from the information contained in experimental affinity measurements.

2.3 THE EXPERIMENTAL DATASET

In the previous section we discussed how the informative power of inference depends crucially on two aspects: the accuracy of the assumptions made when defining the underlying model on one side, and the number and quality of the experimental measurements performed on the other. In our case, the work of our collaborators provided us with very rich information on the affinity distribution of Ab-SCs elicited by immunization. This section is dedicated to a short description of the dataset and the measurement technique they employed. For a more detailed description one can refer to [40].

2.3.1 *Experimental technique*

The aim of the experiments is to probe the effect of Ag dosage and immunization schedule on the resulting immunization quality, defined as the affinity of the Abs produced in response to Ag exposure. To do so our collaborators performed the procedure, sketched in fig. 10. They tested three different immunization schemes (described below) on mice to immunize them against Tetanus Toxoid (TT). At the end of the procedure each mouse was sacrificed and cells from the spleen were harvested. IgG Secreting Cells (IgG-SCs) were isolated, enriched and compartmentalized in picoliter-size droplets using a microfluidic device. Together with the cell, the droplet also contained paramagnetic beads coated with an apposite nano-body to bind the IgG light-chain and capture the secreted antibodies. By applying an external magnetic field the beads can be aligned to form an elongated chain, termed beadline. In addition to this the droplet also contained TT and a binder for the IgG constant region, both tagged with a fluorescent label (green for TT and red for the Fc binder). Secreted Abs, whose concentration increased over time, captured Ag according to their affinity, and were in turn captured on the beadline. By measuring the fluorescent signal one can quantify the magnitude of the Ag and Ab relocation. With an appropriate calibration curve one can relate the ratio of this relocation as the concentration of secreted Abs increases to the affinity of the secreted Abs. In our analysis affinities are quantified either in terms of the dissociation constant K_d between the Ab and the Ag, or of the binding energy ϵ. These two quantities are related through the standard conversion $\epsilon = \log K_d$, where the dissociation constant is expressed in Molar units,[4] and energies are in units of $k_B T$, where k_B is Boltzmann constant and T is the

3 Notice that, since the term in the sum is extensive in the number of observations N, in the limit of a large number of data the prior becomes irrelevant.

4 Notice that other choices of units would shift energies by a constant amount, and would not affect energy differences.

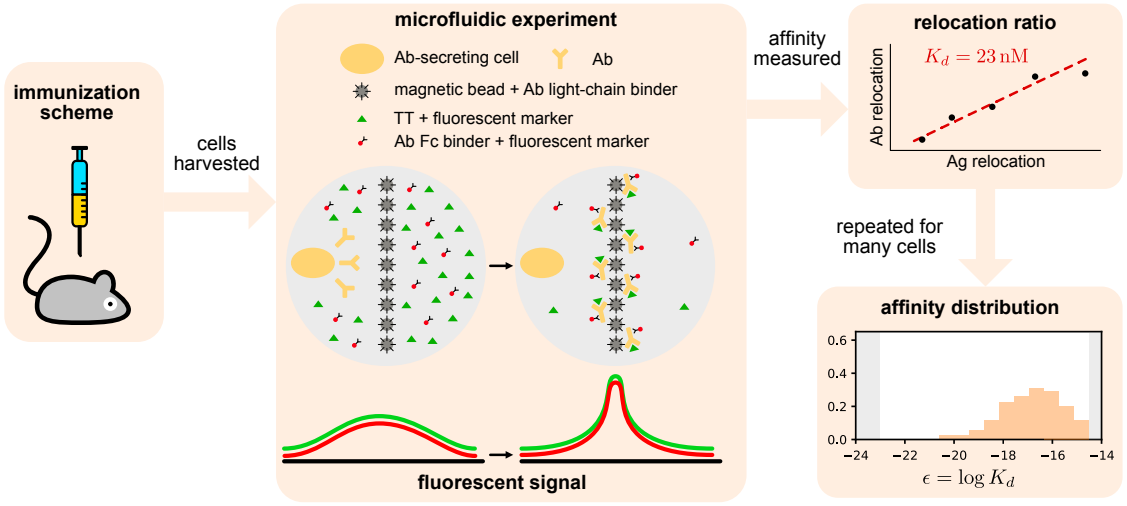

Figure 10: schematic depiction of the experimental technique employed by our collaborators to perform single-cell affinity measurements. A mouse is subject to an immunization scheme. At the end the mouse is sacrificed, IgG Secreting Cells are harvested, enriched and inserted in droplets with fluorescent markers that are used to measure the colocalization of the secreted antibodies and antigen. Through this colocalization the affinity of the antibodies can be determined. This measure is repeated for all the harvested cells, and the full affinity distribution is obtained.

organism temperature.[5] The affinity can be determined in an instrumental sensitivity window that ranges from 0.1 nM to 500 nM (or equivalently binding energies between around -23 and -14.5 units of $k_B T$). Affinities lower than the detection threshold are not detected, while affinities higher than this threshold can be detected but not precisely quantified, therefore in our analysis they were set equal to the threshold 0.1 nM. The experimental apparatus is capable of performing the measurement in parallel on tens of thousands of droplets, which gives as outcome the full affinity distribution of the measured IgG-SCs.

Next-generation sequencing techniques allow one to extract the genetic sequence of a large number of Abs [42], but they do not provide phenotypic information such as binding affinity. Conversely, conventional experiments to quantify affinities either measure an average quantity (e.g. average serum affinity), or have limited throughput. The technique introduced by our collaborators is able to perform affinity measurements of many hundreds of cells at a time, providing an accurate characterization of the affinity distribution of the Ab-SCs elicited by immunization. These distribution are then fully exploited by our inference procedure, described in the following sections.

The experimental apparatus does not reveal the origin of the Ab-SCs analyzed, i.e. whether they derive from reactivated MCs or LLPCs (see section 1.2.3). However the fact that they are class-switched (i.e. they are of IgG isotype) suggests that for the most part they have undergone some prior maturation and do not originate directly from activate naive cells. The fractions of recalled MC and LLPCs in the Ab-SCs population will likely depend on the time and conditions at which the population is probed.

5 This choice is standard in biophysics, and it simplifies the expression of Boltzmann factor into $e^{-\epsilon}$. In practice $1\,k_B T \sim 10^{-24}$ kcal.

2.3.2 *Immunization schemes*

To investigate the effect of Ag dosage and injection schedule on the outcome of immunization we test three different schemes, depicted on the left column of fig. 11.
The first two schemes are used to study the effect of Ag dosage. Scheme 1 consists of two injections. The first contains a dose D of Ag (TT) in Complete Freund's Adjuvant (CFA). After a 4 weeks interval a second injection is administered at the same Ag dose D, in Incomplete Freund's Adjuvant (IFA). Four days after this second injection the mouse is sacrificed, IgG-SCs from the spleen are harvested and their affinity measured using the technique described above. This scheme was tested at 5 different values of Ag dosage $D = 0.01, 0.1, 0.5, 1$ and $10\,\mu g$ TT. Scheme 2 is identical to scheme 1 up to the second injection. At this point after an additional 4 weeks delay a boost injection of $1\,\mu g$ pure TT is administered. The next day the mouse is sacrificed and cells harvested for analysis. Tested dosages for this scheme are $D = 0, 0.01, 0.1, 0.5, 1, 3,$ and $10\,\mu g$ TT. In the case of null dosage injections are performed according to the scheme, but they contain only the adjuvant and no antigen, except for the boost injection which also in this case contains $1\,\mu g$ TT.
Finally, in scheme 3 we study the effect of injection delay. In this scheme mice are subject to a first injection at a constant Ag dosage of $10\,\mu g$ TT in CFA. At this point, after a variable time interval ΔT, a second injection of $10\,\mu g$ TT in IFA is administered. Four weeks later mice are given a boost injection of pure $1\,\mu g$ TT, and the next day cells are harvested and analyzed. Tested time intervals are $\Delta T = 1, 2, 4$ and 8 weeks.

2.3.3 *Qualitative observations: the influence of Ag dosage and injection delay*

For each tested mouse our experimental apparatus measures the full affinity distributions of Ab-SCs (full distributions of pooled data are displayed in later sections in fig. 21). In fig. 11 we report the average binding energy of these distributions for single mice (orange crosses) and for pooled data from mice tested under the same condition (orange empty dots). These are reported as a function of the varied quantity, either Ag dosage D or injection time delay ΔT, depending on the scheme considered. Notice that the number of mice employed per condition can vary (the number for each condition is reported in fig. 21).
As a first qualitative observation one notices that in scheme 1 the average affinity seems to be maximal (i.e. minimum binding energy) at intermediate values of the Ag dosage D. The same behavior can be observed for scheme 2, even though in this case data are more noisy. From the results on scheme 3 instead it appears that delaying the second injection of some weeks is beneficial compared to administering shortly after the first. To interpret these behaviors we introduce in the next section a model for affinity maturation.

2.4 STOCHASTIC MODEL OF AM TO PREDICT AFFINITY DISTRIBUTIONS

Here we introduce a stochastic model for Affinity Maturation, which takes inspiration from [155, 154], with the aim of quantitatively explaining the experimental data. The model describes the evolution of a population of B-cells inside a Germinal Center. These cells are subject to cycles of duplication, mutation, selection and differentiation, repro-

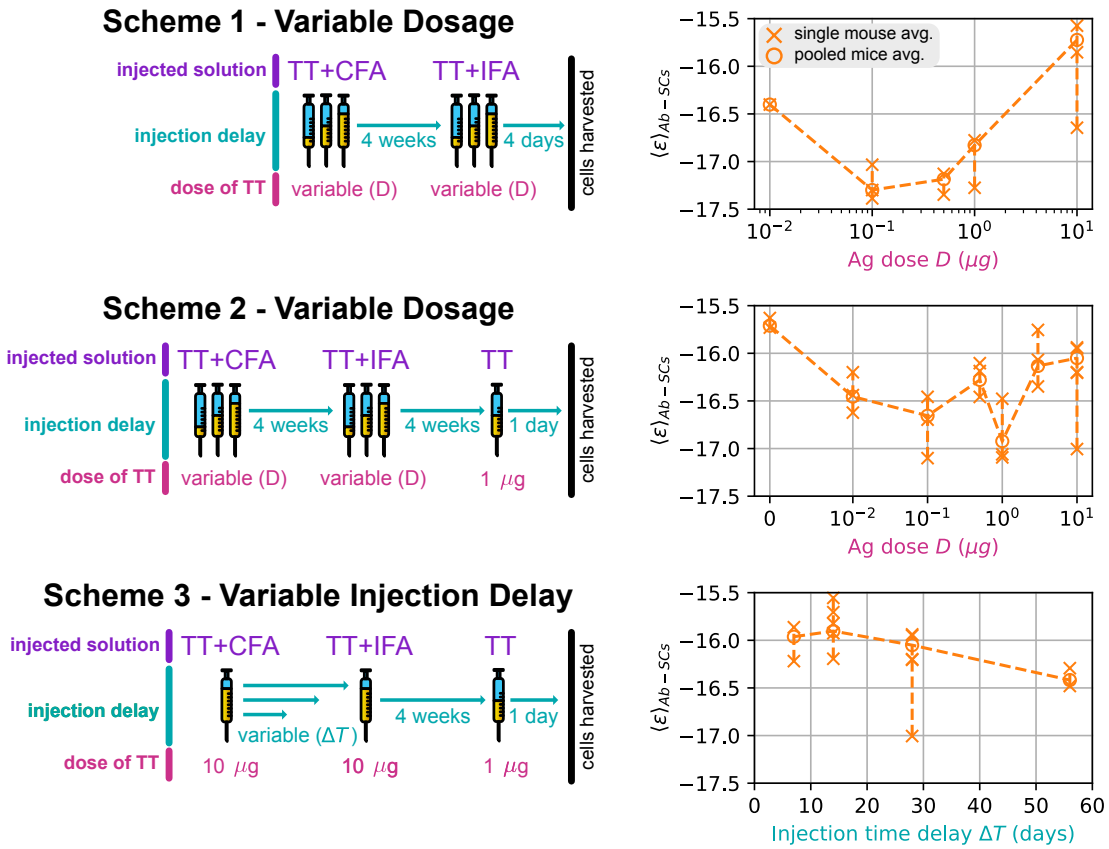

Figure 11: schematic depiction of the three immunization schemes considered (left column) and the effect of the varied quantity (either Ag dosage D or injection delay ΔT) on the average binding energy $\langle \epsilon \rangle$ of the elicited Ab-SCs population (right column). In each scheme depiction we indicate the composition of each injection, the delay between injections and the dosage of Ag. In the corresponding plot on the right, the average binding energy of measurements from a single mouse is indicated with an orange cross (number of mice per condition tested can vary, see fig. 21), while the average over pooled measurements from the same condition are represented as empty dots. Crosses are linked by a vertical dashed line to convey a measure of individual variability.

ducing the biological process described in section 1.2. In our model each cell is characterized solely by the binding affinity of its B-cell Receptor (BCR) for the antigen. The aim of the model is to follow the evolution of the distribution of affinities of cells in the population during the maturation process. This is important for comparison with experimental data, which themselves consists in affinity distributions.
For the sake of readability we limit here the description to the model definition, and move the discussion about the detailed numerical implementation and choice of parameters value from the literature in appendix A.1. All parameters values are also reported in table 1.

2.4.1 *Antigen dynamics*

The first model ingredient we describe is the dynamics of *Ag concentration* C inside of the simulated GC. In the model the amount of Ag available controls the strength of selection, and evolves during the GCR under the action of different mechanisms, which are schematized in fig. 12A.
At the time of injection an amount C_{inj} of Ag is added to the Ag reservoir. This amount is related to the injected dosage D by a proportionality constant $D = \alpha C_{inj}$. We express α as a mass, which makes concentration dimensionless. The reservoir is constituted by Ag trapped in the adjuvant matrix, from which it is quickly released at a rate k^+ and becomes available for cells to bind. Due to recycling of Ag from surface of Follicular Dendritic Cells (FDCs) to endosomal compartments [52, 90] available Ag decays at a slow rate $k_{\varnothing}^-$, and is consumed by B-cells at a faster rate, $k_B^- N_B$, proportional to the number N_B of B-cells in the GC. As the amount of Ag is depleted, selection of B-cell is more and more stringent, and the GC eventually dies out.
The evolution of the reservoir Ag concentration C_{res} and the available Ag concentration C_{av} are regulated by the following pair of equations:

$$\frac{d}{dt} C_{res}(t) = -k^+ C_{res}(t) \tag{4}$$

$$\frac{d}{dt} C_{av}(t) = k^+ C_{res}(t) - (k_{\varnothing}^- + k_B^- N_t^B)\, C_{av}(t) \tag{5}$$

Complemented with the initial condition at injection time $C_{res}(t = t_{inj}) = C_{inj}$ and $C_{av}(t = t_{inj}) = 0$. An example of this dynamics for an injection of $C_{inj} = 10$ is reported in fig. 12B. Additional details on the numerical implementation of these equations and the choice of parameters are given in appendix A.1.

2.4.2 *GC affinity maturation*

Our simulated GCs are initially colonized by a number N_{found} of founder clones from the population of naive responders [144]. Their binding energies ϵ are independently drawn from a Gaussian distribution, with mean μ_{naive} and standard deviation σ_{naive} (Histogram 1 in fig. 13). During the initial phase of colonization and expansion these founder clones duplicate uniformly without mutation to produce a population of N_i B-cells. We do not model this initial phase other than for Ag consumption (see appendix A.1 for details), and start our simulation T_{GC} days after Ag injection, when the GC is mature [149, 34].

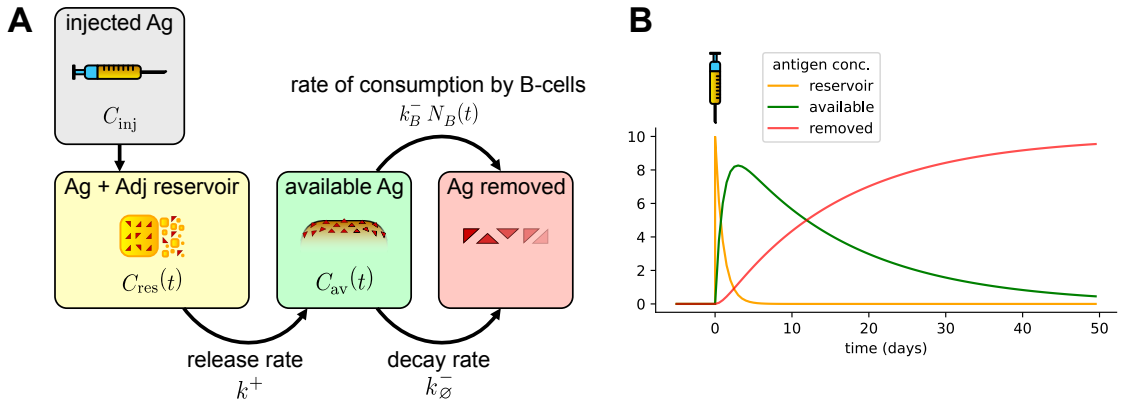

Figure 12: **A**: schematic representation of the Ag dynamics. After injection Ag is stored the adjuvant matrix, from which is released at a rate k^+ and becomes available for B-cells to bind. Available Ag is either consumed by B-cells at a rate $k_B^- N_B$ proportional to the population size, or it decays naturally at a rate $k_\varnothing^-$. **B**: example of Ag dynamics. An injection at $C_{\text{inj}} = 10$ is administered at $t = 0$. In this example the population size is kept constant $N_B = 2500$.

During each evolution round (of duration T_{turn} hours), all cells are assumed to divide twice, independently of their affinity[6]. At division B-cells have probability p_{mut} of developing mutations through a process known as *Somatic Hyper-Mutation*. Mutations can be lethal, neutral, or affinity-affecting with probabilities equal to, respectively, p_l, p_s, or p_a [161]. In the latter case the binding energy of the cell is added a random contribution, $\epsilon \to \epsilon + \Delta\epsilon$, drawn from a log-normal distribution $K_{aa}(\Delta\epsilon)$ (cf. eq. (132)) displayed in fig. 14A. Most affinity-affecting mutations are deleterious, *i.e.* correspond to $\Delta\epsilon > 0$. This makes so that mutations lower the average affinity of the population, but might also increase the affinity of few individuals (Histogram 2 in fig. 13).
After duplication B-cells are first selected according to their capability to bind Ag exposed on Follicular Dendritic Cells (FDCs) (fig. 13 top right). The probability for a cell to survive this selection step is a decreasing function of its binding energy ϵ and increases with the concentration C of Ag on FDCs. It is given by:

$$P_{\text{Ag}}(\epsilon) = \frac{Ce^{-\epsilon}}{Ce^{-\epsilon} + e^{-\epsilon_{\text{Ag}}}} \tag{6}$$

where ϵ_{Ag} is a threshold binding energy. As a consequence, cells with high binding energy (larger than $\epsilon_{\text{Ag}} + \log C$) are likely to be removed from the population (compare histograms 2 and 3 in fig. 13). This survival probability is displayed for three different values of Ag concentration in fig. 14C.
Following internalization, B-cells load the Ag on MHC molecules on their surface [111, 104, 15]. By probing these molecules T follicular helper cells provide survival signals to the B-cells with high Ag affinity (fig. 13, "Competition for T-cell help") [7, 134, 149, 35]. In our model the probability that a B-cell with binding energy ϵ survives this second step of selection is given by:

$$P_T(\epsilon, \bar{\epsilon}) = a + (1 - a - b)\,\frac{Ce^{-\epsilon}}{Ce^{-\epsilon} + e^{-\bar{\epsilon}}}\,, \qquad \text{with} \qquad e^{-\bar{\epsilon}} = \langle e^{-\epsilon} \rangle_{GC} \tag{7}$$

6 This is a simplification since experiments suggests that the number of divisions might correlate with cell affinity [44]. However, as also discussed in section 2.9.2, this amounts to introducing an effective fitness difference which can in practice be qualitatively accounted for by the selection terms in the model.

Where $\langle\cdot\rangle_{GC}$ indicates the average over the current GC population. The threshold energy $\bar{\epsilon}$ depends on the current state of the B-cell population in the GC, as a result of the competition amongst these cells for getting the survival signal from T-helper cells, see Histogram 4 in fig. 13. Parameter a represents the probability for any B-cells to be selected due to stochastic effects (e.g. bystander activation [55]) even with very low affinity; it is introduced to reproduce the observation that selection in GCs is permissive in the presence of complex Ags such as the ones found in vaccines [41]. Parameter b instead represents the probability for a B-cell to fail selection at high affinity. The introduction of b comes from the experimental observation that part of the population of apoptotic cells in GCs has high affinity for the antigen [86]; the removal of these cells could result from stochastic effects [74]. The survival probability is displayed in fig. 14D. After this step, if the number of cells in the population N_B exceeds the threshold value N_{max} each cell is removed with uniform probability $1 - N_{max}/N_B$, so that on average only N_{max} cells are left. Imposing a finite carrying capacity to the GC takes into account limitations on its growth, due to the availability of metabolic resources or the finite amount of T-cell help.
Clones that successfully survive selection differentiate with probability p_{diff} in either Ab-producing Plasma Cells (PCs) or long-lived Memory Cells (MCs), or start a new evolution cycle with probability $1 - p_{diff}$. The probabilities of differentiation into MC and PC, respectively, $\mu_{MC}(t)$ and $\mu_{PC}(t) = 1 - \mu_{MC}(t)$, depend on the time following Ag injection t (early vs. late response) [157]. The MC cell fate is more likely at the beginning of evolution and the PC is more likely towards the end, effectively resulting in a temporal switch occurring around day 11 after injection [157]. This temporal switch is implemented by defining the differentiation probability as:

$$\mu_{MC}(t) = p_{diff} \frac{1}{1 + \exp\{\frac{t - \tau_{diff}}{\Delta\tau_{diff}}\}} \tag{8}$$

$$\mu_{PC}(t) = p_{diff} \frac{1}{1 + \exp\{-\frac{t - \tau_{diff}}{\Delta\tau_{diff}}\}} \tag{9}$$

These functions are displayed in fig. 14B. The MC and PC populations (Histograms 5 in fig. 13) grow at each evolution step, as more and more clones differentiate.

2.4.3 *GC reinitialization*

In the model Ag depletion will eventually lead the GC reaction to a stop. If a second Ag injection is performed a new GC is initiated. The population of N_{found} founder clones for the new GC is composed of both new GC B-cells with naive precursors having sufficient affinity to bind the Ag, and reactivated MCs accumulated during the past evolution [87, 36, 58]. The probability for a founder cell to be extracted from the MC pool is $p_{mem} = N_{mem}/(N_{mem} + N_i)$, where N_{mem} is the number of MCs accumulated up to the time of the second injection. This hypothesis reflects the fact that we expect more reactivated MCs to colonize the newly-formed GC if more MCs were produced in the previous maturation. However one could also consider this ratio to be constant. This second case is discussed in appendix A.2. This initial exchange of MCs is the only interaction between the two GCs, which evolve independently at later times.

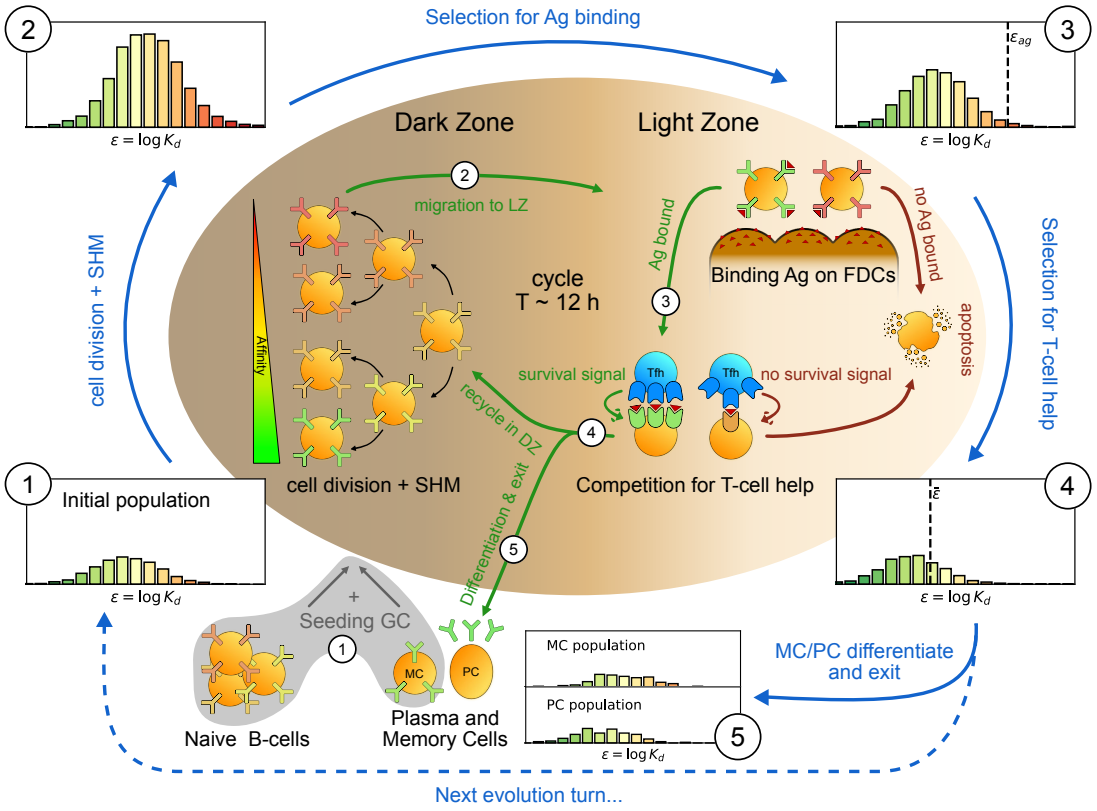

Figure 13: sketch of the Germinal Center Reaction (inner part) and effects of the main reaction steps on the distribution of the binding energies (ϵ, equivalent to the logarithm of the dissociation constant $\log K_d$) of the B-cell population (histograms on the outer part). A red-to-green color-scale is used to depict the affinity of both B-cell receptors in the inner part of the scheme and in the outer binding-energy histograms. Upon Ag administration GCs start to form, seeded by cells from the naive pool having enough affinity to bind the Ag. If the Ag has already been encountered also reactivated Memory Cells (MC) created during previous GC reactions can take part in the seeding. At the beginning of the evolution round cells duplicate twice in the GC Dark Zone and, due to Somatic Hypermutation, have a high probability of developing a mutation affecting their affinity. Most of the mutations have deleterious effects but, rarely, a mutation can improve affinity. As a result the initial population (1) grows in size and decreases its average affinity (2). After duplication cells migrate to the Light Zone, where they try to bind Ag displayed on the surface of Follicular Dendritic Cells. Failure to bind Ag eventually triggers apoptosis. The probability for a cell to successfully bind the Ag depends both on its affinity for the Ag and on the amount of Ag available. Cells with binding energy higher than a threshold value ϵ_{Ag} are stochastically removed (3). The Ag concentration shifts this threshold by a quantity $\log C$. B-cells able to bind the Ag will then internalize it and display it on MHC-II complexes for T-cells to recognize, and then compete to receive T-cell help. We model this competition by stochastic removal of cells with binding energy above a threshold $\bar{\epsilon}$ that depends on the affinity of the rest of the population (4). As before Ag concentration shifts this threshold. Moreover to account for the finite total amount of T-cell help available we also enforce a finite carrying capacity at this step. Surviving cells may then differentiate into either MC that could seed future GCs or Ab-producing Plasma Cells (PC). MCs and PCs are collected in the MC/PC populations (5), while the rest of non-differentiated cells will re-enter the Dark Zone and undergo further cycles of evolution. Eventually Ag depletion will drive the population to extinction.

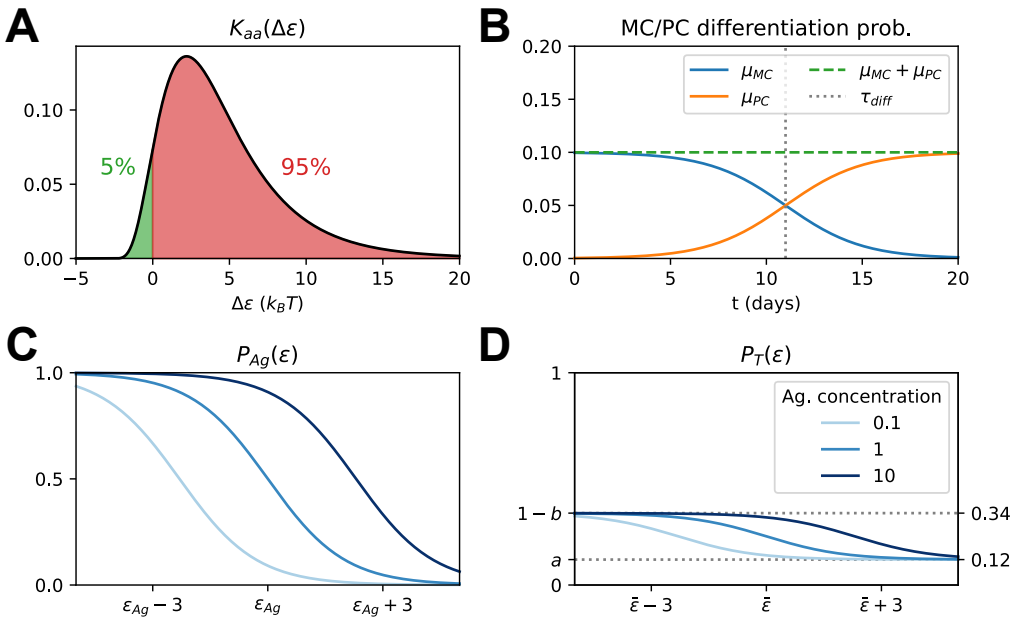

Figure 14: **A**: log-normal probability distribution for the effect $\Delta\epsilon$ of affinity-affecting mutations. Only 5% of mutations decrease the binding energy, and thus have a positive effect on affinity. **B**: probability of MC and PC differentiation (μ_{MC} and μ_{PC} respectively) as a function of time from Ag administration. The total differentiation probability is constant, but a time-switch from MC to PC production occurs around time τ_{diff}. **C**: probability for a cell with binding energy ϵ to survive Ag-binding selection for three different values of Ag concentration (see legend on the right). Notice how concentration controls the strength of selection by shifting the threshold binding energy by a value $\log C$. **D**: probability for a cell with binding energy ϵ to survive T-cell help selection for three different values of Ag concentration (see legend). The threshold value $\bar{\epsilon}$ depends at each time on the binding energy distribution of the population, and thus encodes competition.

2.4.4 *Elicited Ab-SCs*

Administering a recall Ag injection some time after vaccination generates responders Ab-Secreting Cells (Ab-SCs), which are the object of affinity measurements in our experiments. These cells comprise both MCs, that can be stimulated to differentiate and produce Abs upon new Ag encounter [87, 36, 90, 58], and residual PCs formed during previous maturations; PCs belonging to the long-lived pool are capable of surviving up to a human lifetime in the absence of division [159, 32]. In our model the affinity distribution of Ab-SCs is assumed to be a weighted mixture of the MC and PC populations, with fractions equal respectively to g and $1-g$, where the value of g is expected to depend on the conditions under which the system is probed.
Moreover in our experiments cells are harvested from the spleen, and can in principle originate from multiple GCs. To account for this phenomenon when comparing results to experiments we carry out several parallel stochastic simulations of GCs ($N_{GCs} = 20$). These GCs are initialized with different populations of founders, and produce different Ab-SC populations. Their distribution of affinities, averaged over the GCs, can be compared to experimental results. We choose not to introduce interactions between the evolving GCs, due to the lack of experimental quantification of possible GC-crosstalk.

2.4.5 *Three model variants*

In our analysis we will compare three different variants of the model, which differ only by the way selection is performed.

- In variant (**A**) B-cells are selected according to the two-step process consisting of Ag-binding selection and competitive selection for T-cell help, i.e. eqs. (6) and (7).
- In variant (**B**) we consider the same two-step selection but without permissiveness, i.e. we set $a = b = 0$.
- In variant (**C**) we neglect Ag-binding selection and consider only competition for T-cell help, i.e. eq. (7), but allowing for permissiveness.

Notice that these variants contain a different number of model parameters, which must be taken into account when comparing their respective likelihood. We chose to compare these variants to investigate how important permissiveness and Ag-binding selection (excluded respectively in model variants B and C) are to quantitatively fit the data.

2.4.6 *Values of model parameters*

The values of all but nine model parameters were extracted from existing literature, see description in appendix A.1 and table 1. The remaining nine parameters, which were either not precisely known or strongly dependent on our experimental protocol, were fitted from the experimental data through a Maximum-Likelihood inference procedure described in section 2.6. This procedure was executed on each of the three model variants. The nine inferred parameters are:

- the conversion factor α, which allows for conversion between experimental administered Ag dosage D, measured in micrograms, and the dimensionless administered Ag concentration of our model, $C = D/\alpha$.

- the Ag consumption rate per B-cell k_B^-, which controls the GC lifetime and also the extent of the affinity maturation.
- the mean μ_{naive} and variance σ^2_{naive} of the Gaussian binding energy distribution for the GC seeder clones, elicited directly from the naive population.
- the binding energy threshold ϵ_{Ag} for a B-cell to be able to bind Ag with sufficient affinity to internalize it (cfr eq. (6)). This parameter does not appear in model variant C, where only competitive selection for T cell help is implemented.
- The T-cell selection characteristic coefficients, a and b, encoding respectively the baseline probabilities to survive or not survive selection, see eq. (7) and fig. 14D.
- The weight parameters g_{recall}, g_{imm}, representing the MC fraction in the measured population of IgG-SCs for the two protocols, respectively for schemes 2 and 3 with measurement one day after boost, and scheme 1 with measurement 4 days after second injection.

2.4.7 *Example of model evolution at high and low Ag dosage*

To illustrate the model evolution in fig. 15 we report two examples of simulations, obtained with model variant C on a protocol consisting of a single Ag injection. The simulations differ only by the administered Ag dosages $D = 1$ (blue) or 10 (orange) µg of TT. The founder clones population is the same in the two cases in order to eliminate differences coming from variations in the affinities of the initial population.

For both simulations, the main phases in the evolution of the GC can be summarized as follows. At the beginning of GC simulation (day 6 after injection) most of the Ag has been released from the adjuvant matrix and is available for cells to bind. The GC is at maximum capacity and the driving contribution to Ag depletion is consumption by B-cells, which occurs at a rate $k_B^- N_{max}$. This depletion gradually increases selection pressure. This can be understood by inspecting eqs. (6) and (7), in which Ag concentration C acts by shifting the threshold selection energy of a quantity $\log C$. Eventually this concentration reaches a critical value at which selection pressure becomes strong enough to reduce the population size (despite the duplication step) and eventually drive GCs to extinction.

By observing the differences between the two simulations one notices that a lower Ag dosage causes a faster affinity maturation (compare the evolution of average binding energy in fig. 15B), but at the same time generates a shorter-lived GC reaction (compare the population size and time of extinctions in fig. 15B). This is evident when observing binding energy histograms in fig. 15A, in which area reflects population size. The low-dosage case (blue) features a population that evolves faster when compared to the high-dosage case (orange), while at the same time it goes extinct sooner. These effects concur to shape the final MCs and PCs binding energy distribution (fig. 15C). Protection against future pathogen encounters will be granted by these cells, and as such their affinity distribution can be used as an indicator to estimate the success and quality of

Values of model parameters

symbol	value	meaning	source
T_{turn}	12 h	duration of an evolution turn	[155]
T_{GC}	6 d	time for GC formation after injection	[34, 60, 88]
N_{max}	2500	GC max population size	[37, 144]
N_i	2500	initial GC population size	[37, 144]
N_{found}	100	number of GC founder clones	[144, 90]
p_{diff}	10%	probability of differentiation	[155, 94, 114]
τ_{diff}	11 d	switch time in MC/PC differentiation	[157]
$\Delta\tau_{diff}$	2 d	switching timescale in MC/PC differentiation	[157]
p_{mut}	14%	prob. of mutation per division	[155, 89, 66]
p_s, p_l, p_{aa}	50%, 30%, 20%	probability of a mutation to be silent/lethal/affinity-affecting	[161, 155, 154]
$K_{aa}(\Delta\epsilon)$	cf. eq. (132)	distribution of affinity-affecting mutations	[115]
k^+	0.98 d^{-1}	Ag release rate	[82]
$k^-_\varnothing$	1.22×10^{-2} d^{-1}	Ag decay rate	[145]
a	0.12	baseline selection success probability	max-likelihood fit
b	0.66	baseline selection failure probability	max-likelihood fit
μ_{naive}	-14.60	mean binding energy of seeder clones generated by naive precursors	max-likelihood fit
σ_{naive}	1.66	standard deviation of the seeder clones binding energy distribution	max-likelihood fit
k^-_B	2.07×10^{-5} d^{-1}	Ag consumption rate per B-cell	max-likelihood fit
α	2.3×10^{-2} μg	concentration to dosage conversion factor	max-likelihood fit
g_{recall}	0.56	MC fraction in Ab-SC population for measurement 1 day after boost	max-likelihood fit
g_{imm}	0	MC fraction in Ab-SC population for measurement 4 days after second injection	max-likelihood fit
ϵ_{Ag}	-13.59	threshold Ag binding energy (variant A)	max-likelihood fit

Table 1: List of parameters in the model and of their values. Binding energies are expressed in units of $k_B T$, and times in days (d) or hours (h). The last nine parameters were inferred within model variant C, except ϵ_{Ag}, whose reported value refers to variant A, which includes Ag-binding selection.

the immunization procedure.

The fact that Ag concentration controls both maturation speed and GC lifetime generates a tradeoff phenomenon, which makes so that in our model the optimal average binding energy of the MC and PC population is achieved at intermediate Ag dosages. Intuitively this can be understood by considering the two extreme cases of very low and very high dosages. In the former case selection pressure will be very high and only few high-affinity cells will be selected during the maturation round. However the GC will quickly go extinct and these cells will not have time to expand, maturate and differentiate. In the case of too high dosage instead selection pressure will be very low and a lot of mediocre or intermediate affinity clones will accumulate. The high-affinity clones obtained at the end of the evolution process, when Ag gets depleted and selection pressure increases, will be in minority. Only at intermediate dosages a one obtains a good combination of good maturation speed and sufficient GC lifetime.

In order to turn this intuition in a quantitative explanation in the next section we introduce a deterministic version of the model.

2.5 MODEL DETERMINISTIC LIMIT

In this section we introduce the big population size limit of the model, in which the model evolution becomes deterministic. This limit has a double role in our analysis. Firstly, it can be used to approximate the average evolution of the system, without the need to perform several stochastic simulations. This is exploited in the inference technique described in section 2.6. Secondly, the asymptotic behavior offers insight on the role of Ag concentration in controlling the population evolution, as shown in section 2.7.

2.5.1 *Limit of big population size*

When the population size is big enough one can approximate the binding energy distribution of the population with a continuous distribution. In this case the state of the population is completely determined by the *density function* $\rho(\epsilon)$ which is the product between the population binding energy distribution and the population size. It represents the density of cells in energy space. This density function evolves under the action of different operators, each one encoding the effect of one of the stochastic processes that make up the evolution round.

- Cell duplication is represented by the amplification operator, consisting in a simple multiplication:

$$\mathbf{A}\,[\rho](\epsilon) = 2\,\rho(\epsilon) \tag{10}$$

- Mutations are encoded by convolution of the distribution of energies with a mutation kernel:

$$\mathbf{M}\,[\rho](\epsilon) = \int d\Delta\epsilon\; K_{\text{eff}}(\Delta\epsilon)\;\rho(\epsilon - \Delta\epsilon) \tag{11}$$

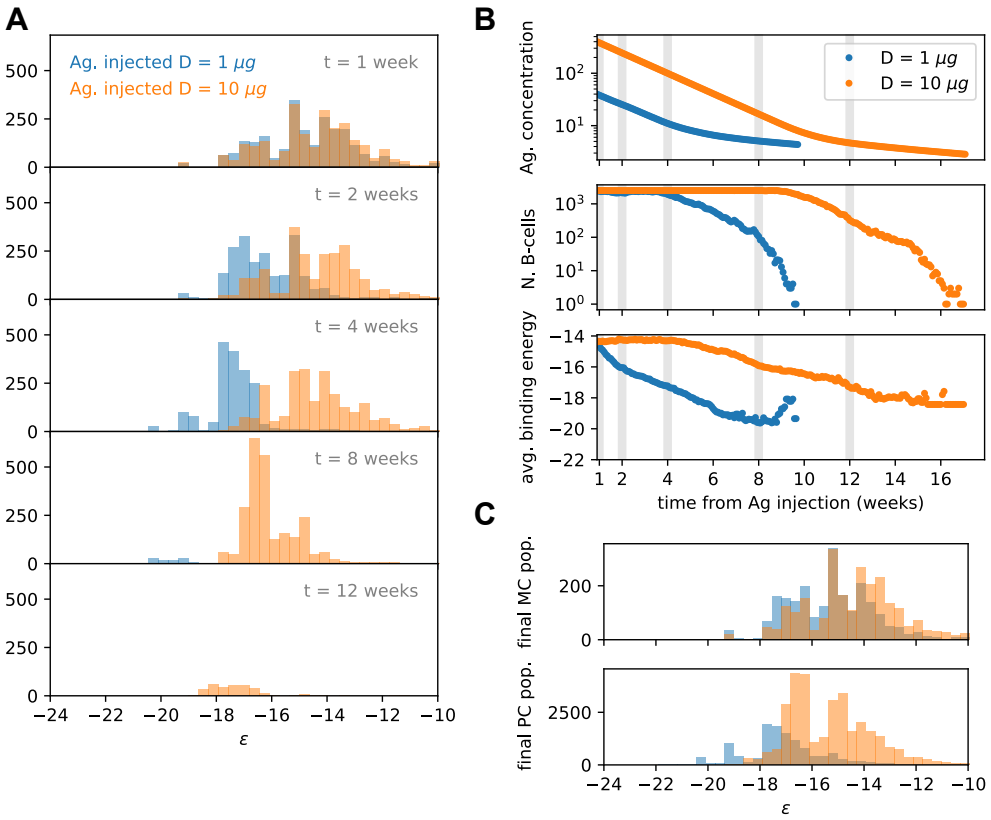

Figure 15: **A**: histogram of the B-cell populations binding energy at different times (1,2,4,8,12 weeks after Ag administration) for two simulations of the model at two different values of administered Ag dosage (1 µg - blue, 10 µg - orange). Ag Dosage D is converted to Ag concentration C through the inferred proportionality constant $\alpha = D/C = 23\,\mathrm{ng}$. Notice that low dosage entails a faster maturation, albeit having a shorter total duration. **B**: Evolution of Ag concentration (top), number of B-cells in germinal center (middle) and average binding energy of the population (bottom) for the same two simulations as a function of time from Ag administration. Vertical grey lines corresponds to time points for which the full affinity distribution is displayed in panel A. **C**: cumulative final populations of differentiated cells at the end of evolution (memory cells - top, plasma cells - bottom) for the two simulations. Colors encode Ag dosage as in panel B and C. Simulations were performed with model variant C and parameters given in table 1.

This kernel $K_{\text{eff}}(\Delta\epsilon)$ includes the effect of silent, affinity affecting and lethal mutations, and is defined as:

$$K_{\text{eff}}(\Delta\epsilon) = p_{\text{mut}}\, p_{\text{aa}}\; K_{\text{aa}}(\Delta\epsilon) + (1 - p_{\text{mut}} + p_{\text{mut}}\, p_{\text{s}})\; \delta(\Delta\epsilon) \tag{12}$$

The first term in the sum accounts for affinity-affecting mutations, that occurs with probability $p_{\text{mut}}\, p_{\text{aa}}$. The second term encodes both silent mutations (probability $p_{\text{mut}}\, p_{\text{s}}$) and the case of no mutation (probability $1 - p_{\text{mut}}$). The contribution of lethal mutations is encoded by the fact that the kernel is not normalized, but rather $\int d\Delta\epsilon\, K_{\text{eff}}(\Delta\epsilon) = 1 - p_{\text{mut}}\, p_{\text{l}}$.

- Selection for Ag binding and T-cell help are encoded simply by a product with the respective probabilities (cf. eqs. (6) and (7)), where for the latter the probability depends on $\bar{\epsilon}$ which in turns depends on the distribution of binding energies, making the operator not linear:

$$\mathbf{S}_{\text{Ag}}\,[\rho](\epsilon) = P_{\text{Ag}}(\epsilon) \times \rho(\epsilon) \tag{13}$$

$$\mathbf{S}_{\text{T}}\,[\rho](\epsilon) = P_{\text{T}}(\epsilon, \bar{\epsilon}) \times \rho(\epsilon)\,, \quad \text{with} \quad e^{-\bar{\epsilon}} = \frac{\int d\epsilon\, \rho(\epsilon)\, e^{-\epsilon}}{\int d\epsilon\, \rho(\epsilon)}\,. \tag{14}$$

- Finally, the carrying capacity (eq. (15)) and the differentiation (eq. (16)) processes correspond to multiplications:

$$\mathbf{N}\,[\rho](\epsilon) = \min\{1, N_{\text{B}}^{\text{max}}/N_{\text{B}}\} \times \rho(\epsilon)\,, \quad \text{with} \quad N_{\text{B}} = \int d\epsilon\; \rho(\epsilon)\,, \tag{15}$$

$$\mathbf{D}\,[\rho](\epsilon) = (1 - p_{\text{diff}}) \times \rho(\epsilon) \tag{16}$$

The distribution of binding energies at round t then evolves through $\rho_{t+1} = \mathbf{E}\,[\rho_t]$, where the complete evolution operator is $\mathbf{E} = \mathbf{D}\,\mathbf{N}\,\mathbf{S}_{\text{T}}\,\mathbf{S}_{\text{Ag}}\,\mathbf{R}$, we indicate with $\mathbf{R} = \mathbf{M}\,\mathbf{A}\,\mathbf{M}\,\mathbf{A}$ the operator encoding for two rounds of mutations and amplification. The evolution operator features, in order of application, two rounds of amplification and mutation, Ag-binding selection, T-cell help selection, carrying capacity and differentiation. Notice that for model variant C, there is no Ag-binding selection, and $\mathbf{S}_{\text{Ag}}$ is replaced with the identity operator.

2.5.2 *Deterministic model reproduces average of stochastic simulations*

In fig. 16 we compare the evolution predicted by the model deterministic limit to the average of 1000 independent stochastic simulations, performed under model variant C and with parameter values reported in table 1. The two are in good agreement at big population size, as can be noticed by comparing the evolutions of the population size (fig. 16A) and average binding energy (fig. 16C). As the size becomes smaller finite-size effect become more and more important, and the two predictions diverge. This divergence comes mainly from an overestimation of the selection pressure in our theoretical solution. In fact, the threshold binding energy for T-cell selection $\bar{\epsilon}$ (see eq. (7)) is sensitive to the high-affinity tail of the population affinity distribution. As the population size diminishes this tail gets progressively less populated than predicted by the infinite-size limit, and the value of $\bar{\epsilon}$ deviates from its theoretical prediction. This slight decrease of

selection pressure at the end of the GC lifetime increases slightly the survival time (see fig. 49) and generates a slight slow-down in maturation.
However, as stated in the previous section, the focus of our interest lays in the affinity distributions of memory and plasma cells, which quantifies the quality of protection conferred by immunization. The prediction for the average binding energy of the two populations remains very accurate during the course of evolution (see fig. 16B and D). This good match is maintained also at the level of full distributions, as discussed later in section 2.6.5 when comparing stochastic simulations (green histograms) to the prediction of the deterministic model (blue distributions) in fig. 21.

This accuracy is not trivial, especially in the context of fitness waves [107]. In many systems, stochastic fluctuations may play a major role, e.g. when the evolving population passes through a bottleneck, and transiently has very low size, before increasing again. Fluctuations may also acquire crucial importance when the evolution lasts so long that the leading edge of the fitness wave has time to exponentially amplify and govern the bulk of the population. This accuracy in our case is due to two main reasons. The first one is that memory and plasma cells are gradually collected during maturation, with most of them originating at times of big population size, in which the model is accurate. Moreover our experimental data are the outcome of many different GC reactions, which partially averages out fluctuations. This partially justifies the use of the deterministic limit on the data. The second reason is linked to the limited maturation observed in our experiments, when compared for example to experiments of immunization against haptens in which affinity can increase of different orders of magnitude [38]. This limited maturation has been associated to complex Ags [70] and is linked in our model to the high level of selection permissiveness, encoded by parameters a and b. This permissiveness slows down maturation (see appendix A.4 and section 2.8.1), and with it also the expansion of stochastic fluctuations from the leading edge to the bulk of the fitness wave. These fluctuations, which are not captured by the deterministic limit of the model, would generate a greater discrepancy if permissiveness was lower.

2.6 INFERRING MODEL PARAMETERS

In this section we make use Bayesian inference to extract from experimental data the most likely value of model parameters, and to compare different model variants through their respective likelihood. This is made possible by the deterministic model limit, which offers a computationally inexpensive way of estimating the likelihood function. This estimation is used in our stochastic likelihood maximization algorithm.

2.6.1 *The likelihood function*

The deterministic model described in the previous section offers a prediction for the average evolution of the density function ρ, from which in turn the final distributions of MCs φ_{MC} and PCs φ_{PC} can be extracted. Experimentally measured Ab-SCs are expected to originate from both population, since they could be both long-lived PCs or reactivated MCs. In our model the parameter g, introduced in section 2.4.4, controls the

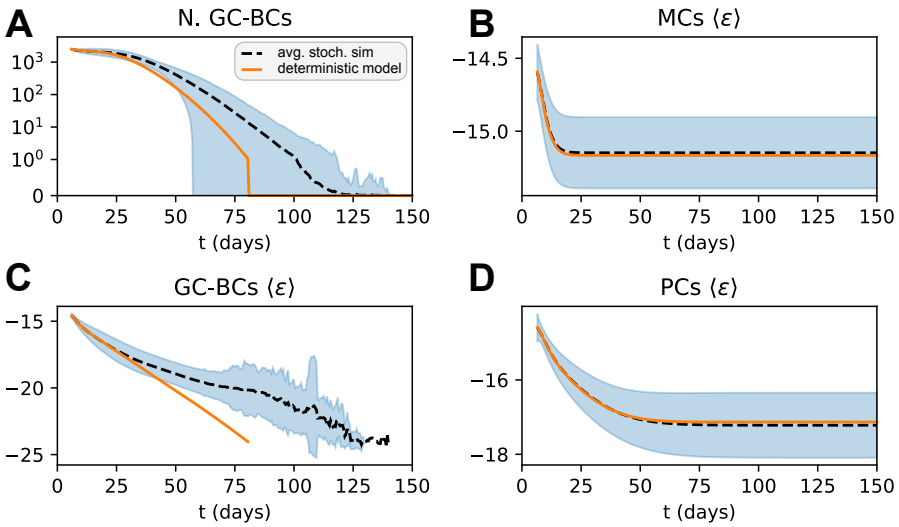

Figure 16: Comparison between stochastic simulations and deterministic limit of the model (variant C) for a single injection at dosage D = 1 µg TT. All quantities are reported as a function of days from Ag injection. The black dashed line and blue shaded area represent the average and standard deviation of the quantities over 1000 independent stochastic simulations, while the orange line represents the prediction of the deterministic limit of the model. Panels represent the number (**A**) and the average binding energy (**C**) of the GC B-cell population, and the average binding energy of the MC (**B**) and PC (**D**) population.

relative contribution of the two populations in the Ab-SCs. According to its definition, the expected distribution of Ab-SCs is given by:

$$\varphi_{\text{Ab-SCs}}(\epsilon) = g\,\varphi_{\text{MC}}(\epsilon) + (1-g)\,\varphi_{\text{PC}}(\epsilon) \tag{17}$$

We expect g to depend on the measurement protocol employed, and thus define two parameters g_{imm} and g_{recall}, which refer respectively to the case of measurement 1 day after boost (scheme 2 and 3) or 4 days after second injection (scheme 1). Finally, for a good comparison between model prediction and experimental data one needs to take into account the experimental sensitivity range of our instrument (cf. section 2.3.1). This is done by restricting and renormalizing the probability distribution in the sensitivity range $-23 = \epsilon_{\text{min}} \leqslant \epsilon \leqslant \epsilon_{\text{max}} = -14.5$. Notice that a measurement equal to ϵ_{min} could in truth originate from any lower value of the energy, since our apparatus cannot distinguish energies lower than this threshold (cf. section 2.3.1). This is not properly taken into account by this renormalization. In our dataset, however, only 4 such measurements are present, and have a very weak influence on the results.
With this procedure one obtains a prediction for the probability of the binding energy of Ab-SCs $\varphi_{\text{Ab-SCs}}(\epsilon|\sigma,\mathbf{p})$ as a function of the immunization scheme details σ (i.e. injected dosage, injection delay, measurement protocol...) and of the unknown model parameters $\mathbf{p}$. The likelihood function is then defined as:

$$\log \mathcal{L}(\mathbf{p}) = \sum_{\sigma\in\mathcal{S}} \sum_{\epsilon\to\sigma} \log \varphi_{\text{Ab-SCs}}(\epsilon|\sigma,\mathbf{p}) \tag{18}$$

Where $\mathcal{S}$ is the set of all immunization schemes tested in our experiments, which consists of 5 different dosages in scheme 1, plus 7 different dosages in scheme 2, plus 4

different injection delays in scheme 3. One must also consider that one of the conditions, at injected dosage $D = 10\mu g$ TT and injection delay $\Delta T = 4$ weeks is repeated, being present in both schemes 2 and 3. This leaves us with a total of 15 different conditions. With $\epsilon \to \sigma$ we indicate the sum over all affinity measurements obtained under condition σ. Notice that with this definition the total log-likelihood is sensitive to the number of measurements, which can vary considerably between different conditions. As such when performing the maximization we favor accuracy over the distributions with the higher number of measurements.
Using the deterministic model to approximate the average of the stochastic model greatly reduces the burden of numerically evaluating the likelihood, which would otherwise require averaging over many stochastic simulations. However at the end of the procedure it is necessary to check that for the inferred parameters values the stochastic and deterministic versions of the model correctly match. This can be done by comparing predictions of the two models for the affinity distributions, which is later displayed in fig. 21.
Decoupling the likelihood of single measurements also constitutes an approximation, since the fate of cells is coupled by competition. This effect is however attenuated by the fact that data are pooled from different mice, and even in a single mouse we expect cells to be generated by different GCRs.

2.6.2 *Numerical likelihood maximization through Parallel Tempering*

The ML procedure requires us to maximize the likelihood as a function of the nine unknown model parameters $\mathbf{p} = (\mu_{naive}, \sigma_{naive}, \epsilon_{Ag}, k_B^-, \alpha, a, b, g_{recall}, g_{imm})$, whose meaning is described in section 2.4.6. We do this with a procedure which makes use of the full experimental affinity distribution, and is based on an implementation of the *Parallel Tempering* technique. This technique, used in the context of molecular dynamics simulations [140], consists in instantiating different copies of a system and letting them evolve at different temperatures. The copies are allowed to exchange their states with adequate probabilities so that low-energy states are correctly sampled at low temperatures. This is particularly advantageous when the energy landscape is rugged, and low-temperature simulations tend to get stuck in local energy minima, while high-temperature simulation explore the whole space without being able to locate the minima precisely. Allowing for state-exchange between different temperature simulations makes so that high-temperature simulations can help low-temperature ones exit local minima in the energy space and converge to the global optimum, as schematized in fig. 17A. In our case, we consider the space of all possible values of the parameters as our configuration space, and use the log-likelihood as a proxy for minus the energy.

Our algorithm can be summarized as follows. A number $N = 10$ of copies of the parameter set is initialized, with values set to $\mathbf{p}_{init} = (\mu_{naive} = -14.6, \sigma_{naive} = 1.6, \epsilon_{Ag} = -13.6, k_B^- = 2 \times 10^{-5}\, d^{-1}, \alpha = 0.025\,\mu g, a = b = 0.2, g_{recall} = g_{imm} = 0.5)$. At each copy is assigned a simulation temperature T logarithmically evenly spaced between 10^3 and 10^{-3}. The maximization procedure consists in 10 000 rounds of iterative parameter changes and temperature exchanges. In the parameter change phase of the round a random additive change $\Delta\mathbf{p}$ is proposed to the parameter vector $\mathbf{p}$, evaluated according to the rule specified below. The likelihood difference $\Delta \log \mathcal{L}$ between the modified and

the original set of parameters is evaluated, and the change is accepted with probability $\min\{\exp\{\Delta \log \mathcal{L}/T\}, 1\}$, where T is the temperature associated to the parameter set. After this, in the temperature exchange phase the difference in log-likelihood $\Delta \log \mathcal{L}$ and inverse temperature $\Delta\beta$ is evaluated for any two parameters sets with consecutive temperatures. An exchange of the two parameters sets is then operated with probability $\min\{\exp\{-\Delta\beta\ \Delta \log \mathcal{L}\}, 1\}$. This is done so that on average high-likelihood parameter values will drift to low temperature simulations. At the end of the last round the value of the parameters $\mathbf{p}_{\text{best}}$ that maximized the likelihood is returned. The pseudocode that describes this procedure is reported in algorithm 1.
Proposed parameter changes $\Delta\mathbf{p}$ are generated as a function of noise level η:

- For μ_{naive}, ϵ_{Ag} variation is performed by adding a random number extracted with uniform probability in the interval $[-10\,\eta, +10\,\eta]$.
- For positive parameters σ_{naive}, α, k_B^- the variation is performed multiplicatively by introducing a percentage change of the parameter uniformly extracted in the interval $[-\eta, \eta]$.
- For the fractions g_{recall}, g_{imm}, a and b variations are performed by adding a random number uniformly extracted in the interval $[-\eta, +\eta]$, after which two additional constraints are enforced. First, since the variables are fractions, the result is constrained in the interval $[0, 1]$. Secondly, for parameters a and b is imposed $a + b \leqslant 1$, which descends from the definition of the survival probability eq. (7).

We make the value of the noise level η depend on the temperature of the system considered, so that higher-temperature simulations also have a higher level of noise, allowing them a faster exploration of the parameter space. Conversely, low temperature simulation will perform only a fine-tuning of the parameters, which permits a more precise convergence to the maximum-likelihood value. In particular we set η to be logarithmically evenly spaced between 10^{-2} and 10^{-1}. Moreover, the proposed parameter change vector $\Delta\mathbf{p}$ contains a joint change of all parameters for the four higher-temperature simulations (i.e. $\Delta\mathbf{p}$ has no null components), while changes affect only one randomly chosen parameter at a time for the rest of the simulations (i.e. $\Delta\mathbf{p}$ has only one randomly chosen non-null component). This is again done to allow for a fast exploration of the space on high-temperature simulations, and a precise convergence to high-likelihood values on low-temperature ones.

In fig. 17B we report the trajectories in energy space of all the parameter sets for the inference performed using model variant C, in which 8 parameters are inferred (all but the threshold binding energy ϵ_{Ag}, which is removed in this variant). The energy for each parameter set is evaluated from the difference in log-likelihood with the best value, $\Delta E = \log \mathcal{L}(\mathbf{p}_{\text{best}}) - \log \mathcal{L}(\mathbf{p})$. Grey dots represent points in which trajectories exchange their parameters set. This mostly occurs when a high-temperature trajectory detains a lower energy parameter set than its low-temperature neighboring trajectory. Notice how trajectories at high temperature explore the space by being able to visit low-likelihood (high-energy) zones of the parameter space. Conversely, low-temperature trajectories gradually converge to the value of the parameters that maximizes the likelihood.
In fig. 18 we display the evolution of the parameter set that reaches the highest log-likelihood during the maximization procedure. Panel A features its log-likelihood and

Algorithm 1: stochastic likelihood maximization

Given the initial parameters choice $\mathbf{p}_{\text{init}}$;
for $i = 1$ **to** 10 **do**
  Initialize copy i with parameter set $\mathbf{p}_i^0 = \mathbf{p}_{\text{init}}$;
  Set its temperature to $T_i = 10^{(11-2i)/3}$;
for $t = 1$ **to** $10\,000$ **do**
  for $i = 1$ **to** 10 **do**
    Generate a new randomly mutated parameter set $\mathbf{p}'_i = \mathbf{p}_i^{t-1} + \Delta\mathbf{p}_i$, where the mutation $\Delta\mathbf{p}_i$ is produced according to the rule specified in the text;
    Evaluate the log-likelihood difference induced by the mutation $\Delta\log\mathcal{L} = \log\mathcal{L}(\mathbf{p}'_i) - \log\mathcal{L}(\mathbf{p}_i^{t-1})$;
    With probability $P = \min\{\exp\{\Delta\log\mathcal{L}/T_i\}, 1\}$ accept the change $\mathbf{p}_i^t \leftarrow \mathbf{p}'_i$ or else keep $\mathbf{p}_i^t \leftarrow \mathbf{p}_i^{t-1}$;
  for $i = 1$ **to** 9 **do**
    Evaluate the likelihood difference $\Delta\log\mathcal{L} = \log\mathcal{L}(\mathbf{p}_{i+1}^t) - \log\mathcal{L}(\mathbf{p}_i^t)$ between two adjacent copies of the model;
    Evaluate the inverse temperature difference $\Delta\beta = 1/T_{i+1} - 1/T_i$;
    With probability $P = \min\{\exp\{-\Delta\beta\ \Delta\log\mathcal{L}\}, 1\}$ perform the state exchange $\mathbf{p}_i^t \leftrightarrow \mathbf{p}_{i+1}^t$;
Find $(i^*, t^*) = \arg\max_{(i,t)}\{\log\mathcal{L}(\mathbf{p}_i^t)\}$;
return $\mathbf{p}_{\text{best}} = \mathbf{p}_{i^*}^{t^*}$

temperature as a function of the round of maximization. This parameter set is initially assigned to high-temperature simulations, which explore the configuration space causing big likelihood fluctuations. Through consecutive modifications of the parameters a high-likelihood configuration is reached, and the parameter set is exchanged through simulations and assigned to a low temperature one. At this point the algorithm behaves almost as a gradient descent, accepting almost only parameters changes that increase the likelihood. The peak likelihood value is reached at round 9970, when the trajectory was at the lowest temperature. In panel B the evolution of parameter values is displayed. Notice how big fluctuations correlate with high-temperature state. Orange shaded areas cover one standard deviation of the posterior distribution around the maximum-likelihood estimate (MLE) of the parameters and convey a relative measure of confidence intervals. More precisely, this is obtained by evaluating the effect of small ($\pm 5\%$) single parameters variations around the MLE on the posterior distribution, which is displayed in fig. 19. The fact that this variation is always negative suggests that the algorithm successfully converged to a likelihood maximum. We perform then a quadratic fit on this variation (orange dashed lines in the figure) which corresponds to approximating the posterior with a Gaussian distribution around the ML value. The standard deviation of this Gaussian can be used to estimate how sensitive the likelihood is to the variation of different parameters.

This likelihood maximization procedure, whose code is freely available at `https://github.com/mmolari/affinity_maturation`, is very general and can be easily extended to the inference of any set of parameters in our model, or to other experimental datasets.

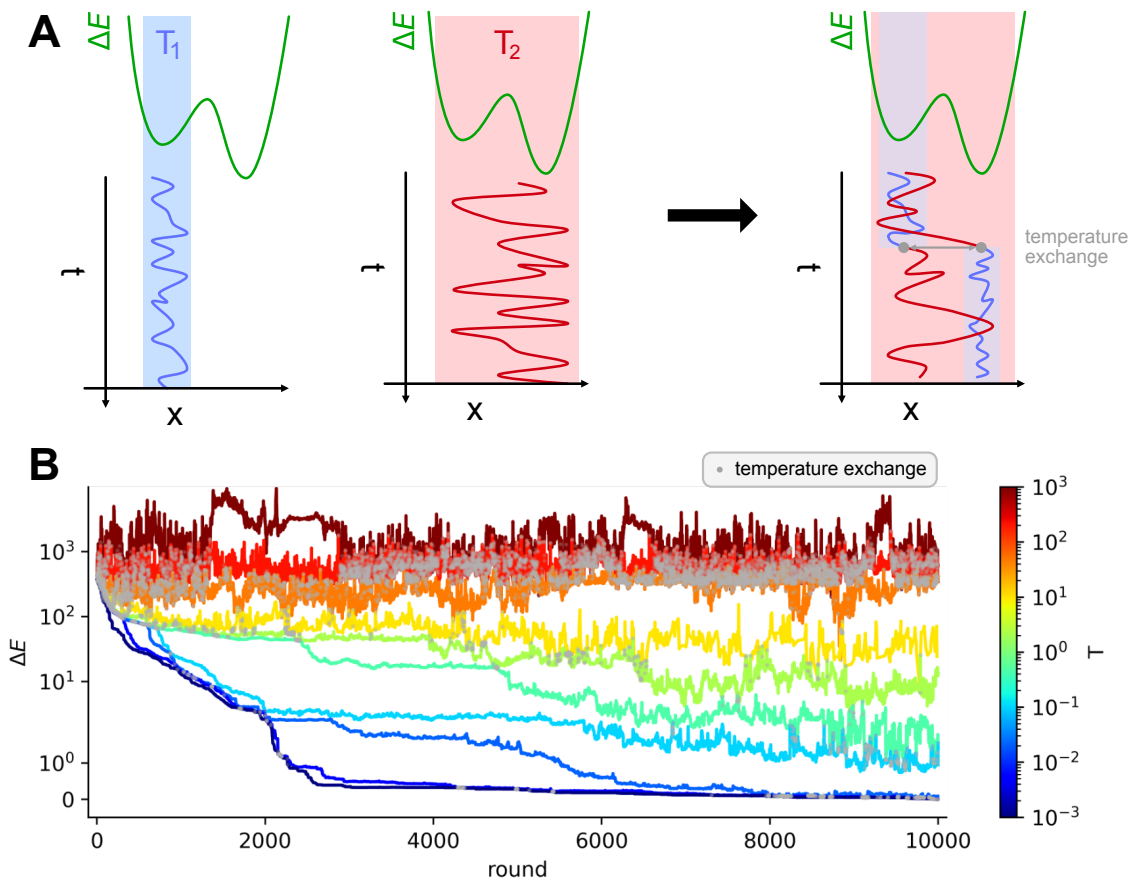

Figure 17: **A**: intuitive representation of the advantage of Parallel Tempering. When a Monte Carlo simulation is run at low temperature (T_1) the system reaches a low-energy state but can get stuck in local energy minima. Conversely at high temperature (T_2) the system is free to explore a larger portion of the space, but is unable to localize the energy minimum. By allowing the states to exchange their temperature when favorable, the high temperature simulation can help the low temperature simulation exit from a local minimum trap and converge to the global energy minimum. **B**: Simulation trajectories in energy space as a function of the simulation round for inference on variant C of the model. The energy difference is defined as $\Delta E = -(\log \mathcal{L} - \log \mathcal{L}_{\max})$, where $\mathcal{L}_{\max}$ is the maximum likelihood recovered by the inference algorithm. Colors encode different temperatures according to the colorbar on the right. Grey dots mark points in which trajectories exchange temperatures. To display both large variations and values equal to zero the energy scale is logarithmic for values of $\Delta E > 1$ and linear for energies $0 \leqslant \Delta E \leqslant 1$.

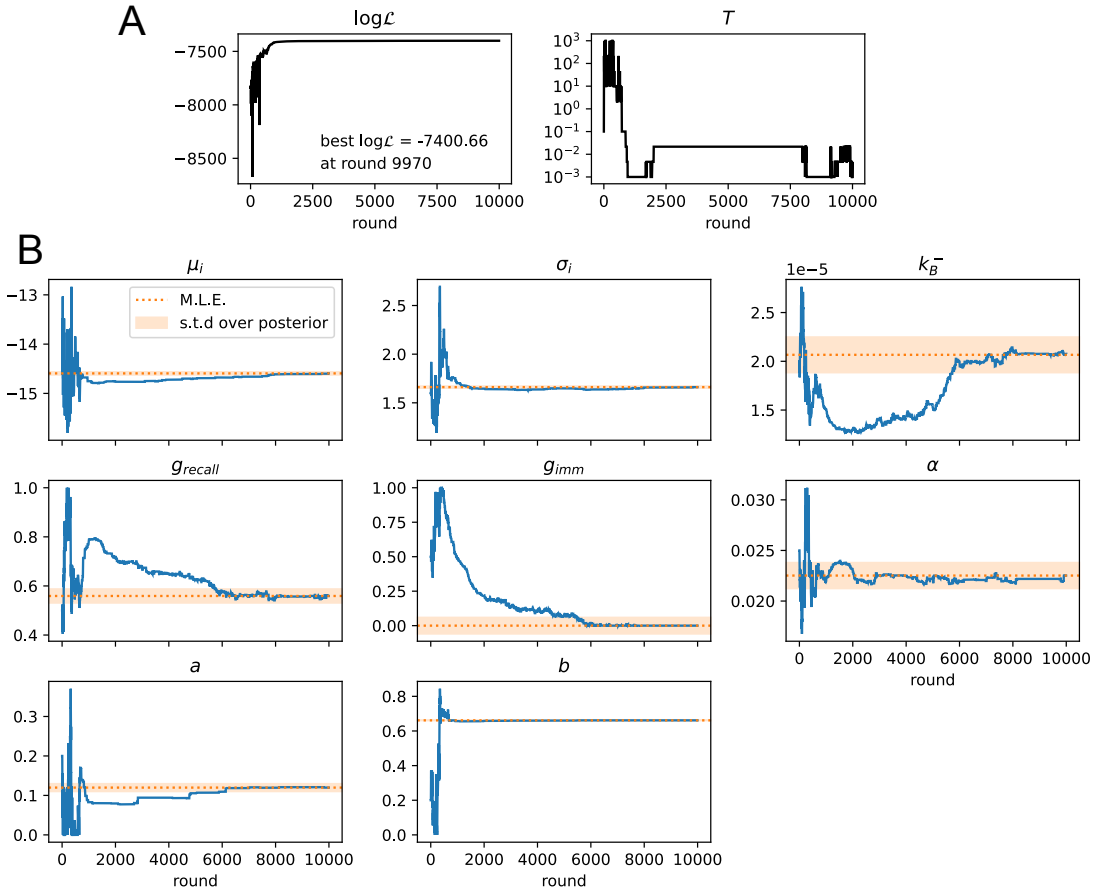

Figure 18: Convergence of the stochastic likelihood maximization procedure for variant C of the model. In this variant 8 of the model parameters are inferred (μ_{naive}, σ_{naive}, k_B^-, α, a, b, g_{recall}, g_{imm}). **A**: values of the log-likelihood $\log \mathcal{L}$ and the temperature T for the parameter set that reached peak likelihood. **B**: evolution of the parameter values during the maximization procedure for the same set (blue lines). Maximum likelihood (ML) estimates of the parameters are marked as orange dashed lines. Orange shaded area cover one standard deviation of the posterior distribution around the ML value, evaluated through a Gaussian approximation of the posterior distribution and the quadratic fit of the log-likelihood displayed in fig. 19.

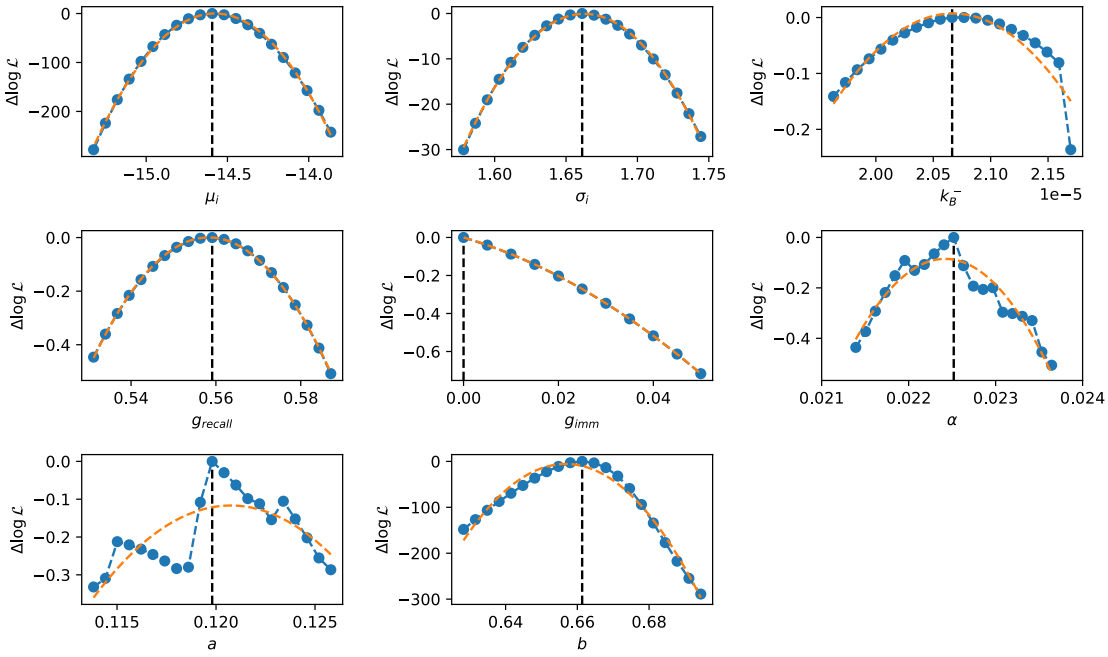

Figure 19: Variations of the posterior distribution around the maximum-likelihood (ML) value found by the stochastic maximization procedure for model variant C. Plots represent the likelihood variation $\Delta \log \mathcal{L}$ of the posterior distribution for a small ($\pm 5\%$) variation of single parameters around their ML values (vertical black dashed lines). The peaked shapes of $\Delta \log \mathcal{L}$ confirm the convergence of the maximization procedure for all parameters. Orange dashed curves represent quadratic fit of $\Delta \log \mathcal{L}$, used to estimate the orange confidence interval in fig. 18B.

2.6.3 *Comparison between model variants*

In table 2 we report the inferred value of model parameters for inference performed under the three model variants described in section 2.4.5. Variant B, in which selection is not stochastic or permissive (i.e. parameters a and b in eq. (7) are set equal to zero) is the one with the lowest likelihood, indicating that these two parameters play an essential role in matching the model to data, as later discussed in section 2.8.1. Variants A and C instead have a comparable likelihood, and comparable values of model parameters. The difference between the two is that in variant C Ag-binding selection (cf. eq. (6)) is not implemented and parameter ϵ_{Ag} is not present. The likelihood of variant A is slightly higher, but this is to be expected since in this variant one more parameter is present. A fair comparison between these variants must indeed take into account the different number of parameters. We therefore resort to the so-called *Bayesian Information Criterion* (BIC) [125] and *Akaike Information Criterion* (AIC) [3]. These criteria take into account the number of parameters by estimating the volume in the parameter space around the peak in likelihood. They are defined as:

$$\mathrm{BIC} = k \ln n - 2 \ln \mathcal{L}, \qquad \mathrm{AIC} = 2k - 2 \ln \mathcal{L} \tag{19}$$

Where k is the number of model parameters (9 for variant A, 7 for B and 8 for C) and n is the total number of experimental measurements ($n = 4950$ for our experimental dataset). Values of these estimators for the three variants are reported in table 2. Based on these values we conclude that the variant to be chosen, with lowest BIC and AIC, is C. Including Ag-binding selection improves slightly the likelihood, but less than expected from the introduction of an extra parameter (ϵ_{Ag}). On the contrary, the large increase in BIC when forbidding permissiveness shows that non-zero values for a, b are definitely needed to fit the data. Variant C will be the one considered for further analysis.

2.6.4 *Consistency check through artificial data generation*

To test the reliability of our inference algorithm we generated 10 artificial datasets using our stochastic model and applied to them the inference procedure. We then compared the inferred values of the parameters with the real value used to generate the data.
To make sure that the amount of experimental measurements for each condition that we have at our disposal are sufficient to infer the model parameters, when generating the data we took into account the size and composition of our experimental dataset. In particular we generate each artificial dataset as follows.
For every experimentally tested conditions (15 conditions in total, consisting in 5 tested dosages for scheme 1, 7 dosages for scheme 2 and 4 delays for scheme 3, with condition $D = 10\,\mu g$ and $\Delta T = 4$ weeks shared between scheme 2 and 3) we run as many stochastic simulations as mice tested for that particular condition. For each simulation we then extract from the responders population a number of cells equal to the one obtained from each mouse. This extraction is done keeping into account the experimental detection range (see section 2.3.1), therefore we exclude cells having affinity lower that $K_d = 500\,nM$, and set any affinity higher than the high-affinity threshold $K_d = 0.1\,nM$ equal to the threshold. Stochastic simulations of each scheme are done in model scenario C, using the standard value of the parameters with one exception: to account for the fact that each mouse contains multiple GCs we raised the number of founder clones

inferred parameters value for the three model variants

	full model (**A**)	no permissivity (**B**)	no Ag-binding selection (**C**)
ϵ_B	-13.59	-13.53	
k_B^- (d^{-1})	1.16×10^{-5}	2.95×10^{-4}	2.07×10^{-5}
μ_i	-14.59	-15.04	-14.59
σ_i	1.64	1.15	1.66
g_{recall}	0.65	0.96	0.56
g_{imm}	0.03	0.61	0
a	0.27		0.12
b	0.66		0.66
α (μg)	0.114	0.018	0.023
$\log \mathcal{L}$	-7400.37	-7459.39	-7400.67
BIC	14 877	14 978	14 869
AIC	14 819	14 933	14 817

Table 2: results of the inference procedure for the three model variants. These variants differ for the way selection is performed: in variant B selection is not permissive (parameters a and b are set equal to zero) and in variant C no Ag-binding selection is performed (parameter ϵ_{Ag} absent). The table reports the inferred values of parameters in the three variants. Empty spaces indicate parameters that are not present in the model variant. For dimensional parameters units are indicated in brackets. In the last three lines we report the corresponding value of the log-likelihood, along with the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) estimators used for comparing the three variants. Both these estimators indicate a slight preference for variant C.

inference on artificially generated datasets

	value in generated data	mean of inferred values	std of inferred values
μ_i	-14.594	-14.764	0.215
σ_i	1.661	1.593	0.112
k_B^- (d^{-1})	2.07×10^{-5}	1.82×10^{-5}	7.06×10^{-6}
g_{recall}	0.559	0.549	0.11
g_{imm}	0	0.009	0.022
α (μg)	0.023	0.032	0.019
a	0.12	0.125	0.094
b	0.661	0.659	0.008
$\log \mathcal{L}$	-7400.665	-7347.233	142.319

Table 3: average results of the inference procedure on 10 artificially generated datasets. On the first column we report the parameters values used to generate the datasets, equal to the inferred parameter values under model variant C. On the second and third column we report the mean and standard deviation of the 10 inferred values. Units are indicated in round brackets for dimensional parameters. These are the dosage conversion factor α, measured in micrograms of Ag, and the Ag consumption rate k_B^-, in inverse of days.

in each GC to 2500 instead of 100. This amounts to introducing a greater diversity in the initial population, which in turns reduces the stochasticity in the evolution outcome while at the same time leaving the average behavior unchanged (see appendix A.3). This is similar in spirit to averaging between multiple GCs, as it is the case for the experimental measurements of cells extracted from the spleen of mice, and at the same time this greatly reduces the computational burden by requiring only one stochastic simulation per condition.

This generation procedure was re-executed 10 times, resulting in 10 different datasets. In fig. 20B we show the binding energy distribution of the 10 generated datasets for condition $D = 0.5\,\mu g$ of Ag in scheme 2 (histograms in gray). The histograms are close to the prediction of the deterministic model (blue curves), but with some deviations due to stochastic sampling. The average binding energy of the population for all considered condition is reported in fig. 20A. Again, the average binding energy of the generated populations (gray crosses) is close to the prediction of the deterministic model (blue dot), and as expected the noise is higher for the conditions where a smaller amount of experimental measurement was performed (compare with number of measurements displayed in fig. 21).

We then run the inference procedure on each artificially generated dataset, using the same setup and initial condition as the ones used to infer model parameters on the real data, under scenario C. The average results of the inference are reported in table 3. For every model parameter the average inferred value is close to the real value used to generate the data, demonstrating that the amount of experimental measurements at our disposal are on average sufficient for a good recovery of the model parameters.

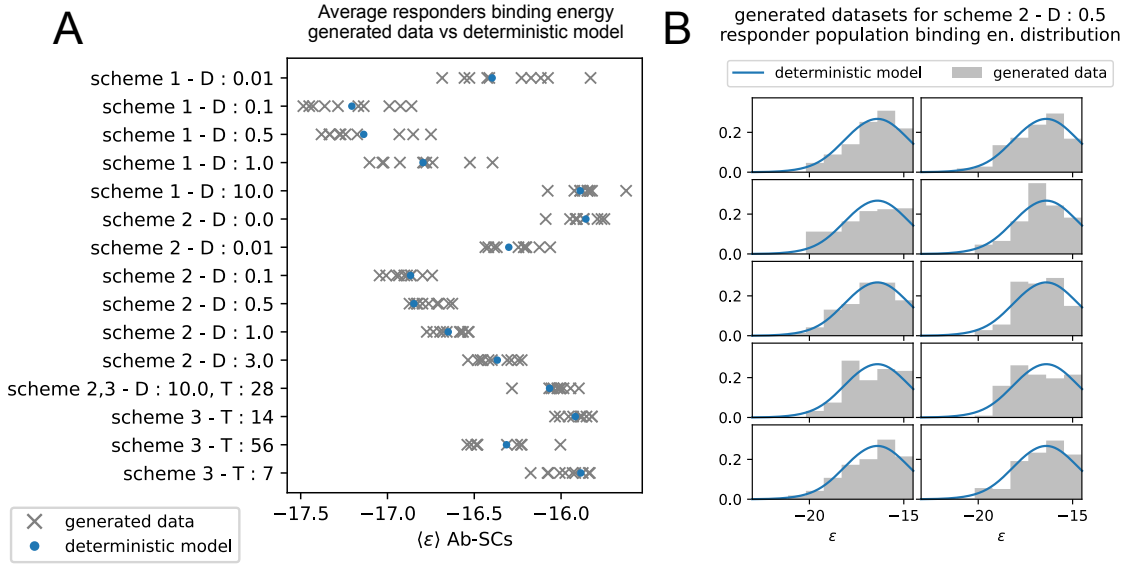

Figure 20: **A**: average binding energy of responders populations in the 10 artificially generated datasets (gray) for the 15 experimentally tested conditions, vs the same quantity as predicted by simulations of the deterministic model. The condition to which the measurement are referred is reported on the y-axis, in the form of the scheme used, the Ag dosage injected (D) and/or the time delay between injection (T). **B**: distribution of generated binding energies in the 10 generated dataset (gray), compared to the distribution of binding energies predicted by the deterministic model (blue), for the condition corresponding to the injection of $D = 0.5\,\mu g$ of Ag in scheme 2. Distributions are limited to the experimental sensitivity range $-23 < \epsilon < -14.5$.

2.6.5 *Inferred model reproduces data*

We conclude this section by showing that upon inference of the unknown model parameters the model is able to quantitatively reproduce the data.

In fig. 21 we compare experimental affinity distributions (orange historams) with the predictions obtained from the model (variant C) using the MLE of the parameters. For each distribution we report the number of measurements included and the number of mice from which measurements were extracted. Distributions are grouped in three columns, each one referring to a different immunization scheme. The value of the scheme variable quantity is reported for each distribution (Ag dosage D in micrograms of Ag in pink for scheme 1 and 2, and injection delay ΔT in days for scheme 3). Distributions are normalized considering only the experimental sensitivity window $0.1\,\mathrm{nM} \leqslant K_d \leqslant 500\,\mathrm{nM}$ (or equivalently $-23.03 \geqslant \epsilon \geqslant -14.51$, see section 2.3.1) delimited in the figure by gray shaded areas. Notice that all of these distributions were used when maximizing the likelihood.
We compare the distributions to the prediction offered by of our model, both from stochastic simulations (green histograms, average over 1000 simulations) and from the deterministic model limit (blue curves). Both represent the binding energy distribution of Ab-SCs, obtained as described in section 2.4.4 (stochastic model) and section 2.4.4 (deterministic limit), and for good comparison with data both are normalized only considering the experimental sensitivity window. Under all tested immunization schemes we observe a very good agreement between theory and experiments, and between stochastic and deterministic versions of the model.

The same agreement is maintained when confronting average quantities. In fig. 22 we compare the average binding energy and high affinity fraction for model and data. The latter is defined as the fraction of cells that has affinity higher than 50 nM, or equivalently binding energy lower than -16.8. As done in fig. 11 experimental measurements are reported both as average for single mice (orange crosses) and average over all pooled mice under the same condition (orange empty dots), where the former is used to convey a measure of individual variability. For good comparison with data averages for the model are evaluated keeping into account the experimental sensitivity range. Results for the deterministic model are represented as a blue line, while green shaded areas indicate results for stochastic simulations. In particular the light shade of green covers one standard deviation around the mean for single GC simulations. Cells extracted from the spleen can however in principle originate from different GCs. The number of GCs in a spleen can vary, but around 20 to 50 GCs per spleen section were reported in [158]. Therefore, to quantify the expected variation when cells ensue from different GCs, in a darker shade we report the standard deviation of the mean of 20 stochastic simulations.

The agreement between the stochastic model and theoretical predictions indicates that, at least for the quantities considered and for our values of model parameters, the infinite size limit is a good approximation of the average stochastic evolution.
Most importantly, the agreement between model and data under all schemes shows that our model can correctly capture the phenomenology observed. In scheme 1 and 2 in particular both show the existence of an optimal intermediate dosage corresponding to

maximal affinity of the Ab-secreting cells recalled population. In the next section we will interpret this observation with an appropriate analysis of the deterministic model.
In scheme 3 we observe that experimental data show a slight increase in affinity for longer injection delays, and so does our model. This is presumably due to a combination of two effects. Firstly the fact that higher affinity cells are produced late in the response, and waiting more before harvesting cells allows for higher affinity cells to be created. Secondly, giving the first GC time to produce high-affinity MCs is beneficial since then these cells can then colonize the second GC and continue their maturation even further there.

2.7 Analysis of deterministic model offers insight on the effect of Ag dosage

In the previous section we showed how Bayesian inference allows for the matching of model and data. Part of the importance of inference resides in the fact that, after this matching, one can study the model, which is much more easy to access, to gain insight on phenomena observed in the data.
In this section in particular we tackle the problem of explaining the effect of Ag dosage on the outcome of maturation. We show that the existence of an intermediate optimal dosage, which is observed both in model and data (cf. fig. 22), can be interpreted as a tradeoff between growth rate and maturation speed, both controlled by Ag concentration. This interpretation, which was intuitively suggested in section 2.4.7, is made here more precise by studying the asymptotic limit of the deterministic model in condition of constant Ag concentration.

2.7.1 *Asymptotic travelling wave behavior under constant Ag concentration*

We begin our analysis by considering the asymptotic behavior of the deterministic model under condition of constant Ag concentration C. Studying this regime can provide insight on the role of Ag concentration in regulating maturation. The analysis is aided by two simplifications. First, to be able to observe asymptotic population expansion we relax the maximum carrying capacity constraint and set $N_{max} = \infty$. Second, we consider the population affinity to be high enough that Ag-binding selection plays no role. Notice that this second simplification is not necessary in variant C of the model, in which Ag-binding selection is absent. We will later discuss the effect of removing these approximations.

As a preliminary observation in fig. 23B we report the evolution of the binding energy density function $\rho_t(\epsilon)$ of the population under constant Ag concentration $C = 30$. Color encodes the number of rounds t from the beginning of the GC evolution. Notice that by definition the area under the curve of the density function is equal to the population size. In fig. 23A and C we display the evolutions of the population size and average binding energy. From here one notices that, as the number t of evolution rounds increases, the size of the population increases exponentially with a growth rate ϕ and the average binding energy shifts linearly, with velocity u. This leads to the hypothesis that the den-

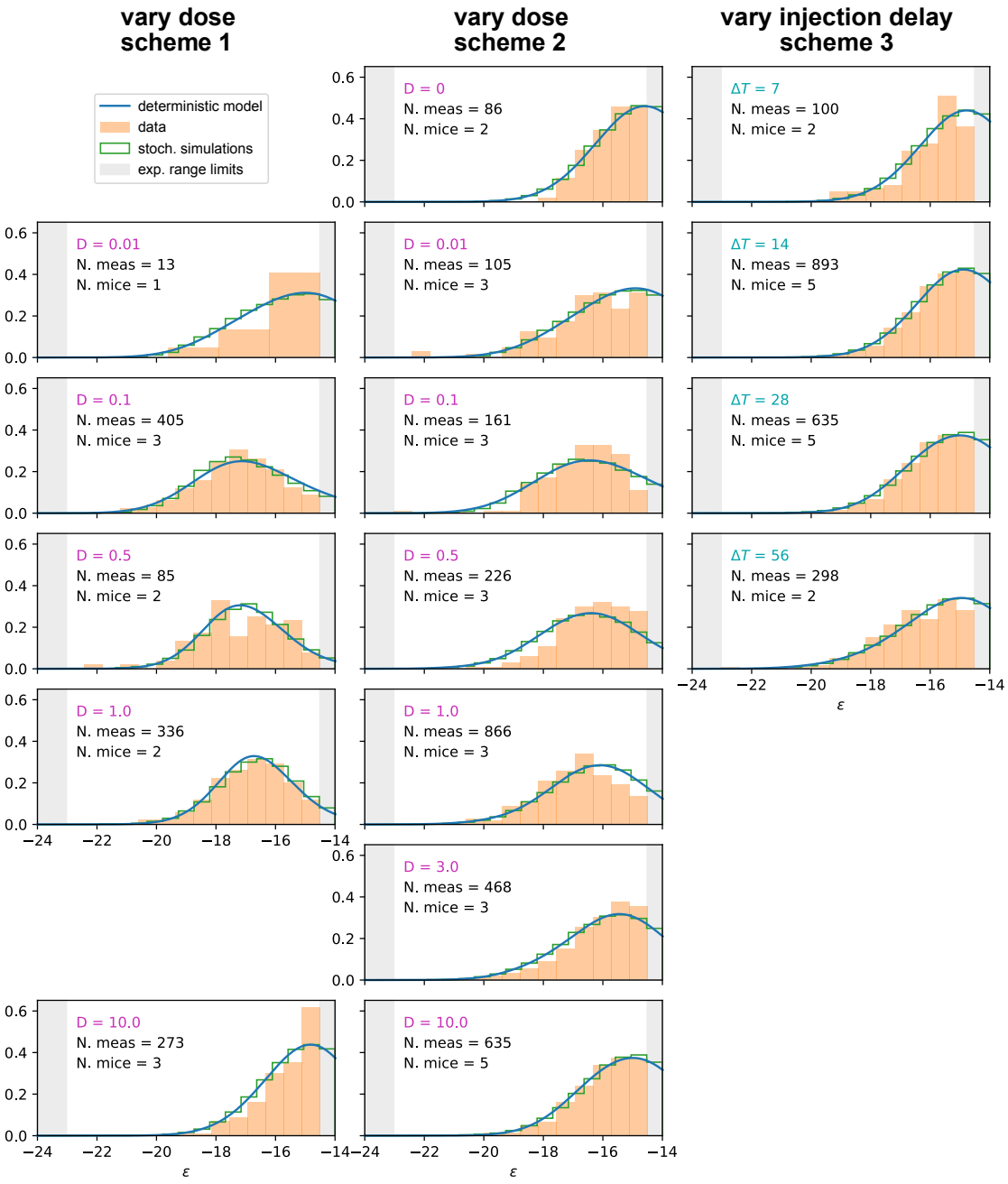

Figure 21: Comparison between experimental measurements (orange histograms), stochastic model (green histograms), and theoretical solution (blue curve) for affinity distributions of antibody-secreting cells (Ab-SCs) under all the different immunization protocols (see main text for the description of the schemes). Experimental data (orange histograms) consists in measurements of affinities of IgG-secreting cells extracted from mice spleen. The number of mice and single-cell measurements is indicated for each histogram (black). The experimental sensitivity range ($0.1\,\mathrm{nM} \leqslant K_d \leqslant 500\,\mathrm{nM}$, or equivalently $-23.03 \geqslant \epsilon \geqslant -14.51$) is delimited by the gray shaded area. Blue curves represent the expected binding energy distribution of the Ab-SCs population according to our theory under the same model conditions. For a good comparison all the distributions are normalized so that the area under the curve is unitary for the part inside the experimental sensitivity threshold. For every histogram we indicate the value of the varied immunization scheme parameter, corresponding to dosage D (pink) for the first two schemes and time delay ΔT (blue) for the third.

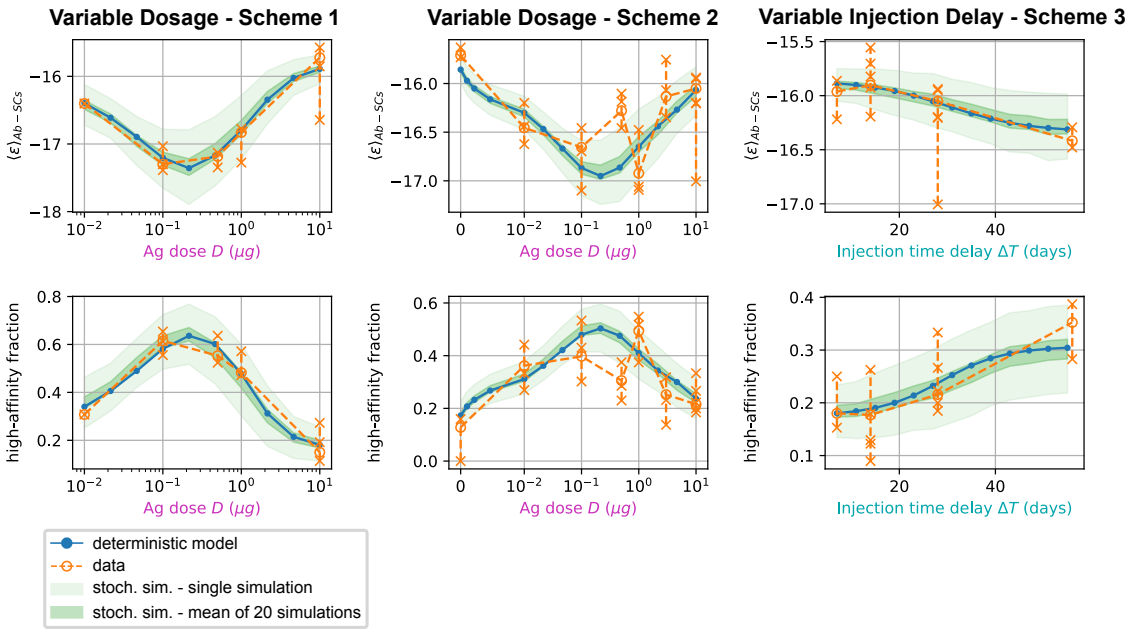

Figure 22: comparison between data and model prediction for the average binding energy (top) and high affinity fraction (bottom) of the Ab-secreting cell population under the three different immunization schemes (scheme 1 - left, scheme 2 - center, scheme 3 - right). The high affinity fraction corresponds to the fraction of measured cells having binding affinity $K_d < 50\,\text{nM}$, or equivalently binding energy $\epsilon < -16.8\ k_B T$. On the x axis we report the variable quantity in the scheme, which is administered dosage D for schemes 1 and 2 and delay between injection ΔT for scheme 3. Green shaded areas indicate the results of the stochastic model simulations. The light area covers one standard deviation around the average result for a single simulation, while the dark area corresponds to the standard deviation for the mean over 20 simulations. This quantifies the expected variation for populations of cells extracted from a spleen, that could potentially have been generated by many different GCs. Results are evaluated over 1000 different stochastic simulations per condition tested. The deterministic solution of the model, in blue, reproduces well the average over stochastic simulations in all the considered schemes. Data coming from experimental affinity measurement of IgG-secreting cells extracted from spleen of immunized mice are reported in orange. Orange empty dots represent averages over the data pooled from multiple mice immunized according to the same scheme, while orange crosses represent averages for measurements from a single mice. Crosses are connected with a vertical dashed line in order to convey a measure of individual variability. Notice that the number of mice per scheme considered can vary, see fig. 21. In order to compare these data with our model, both for the stochastic simulations and the theoretical solution we take into account the experimental sensitivity range when evaluating averages.

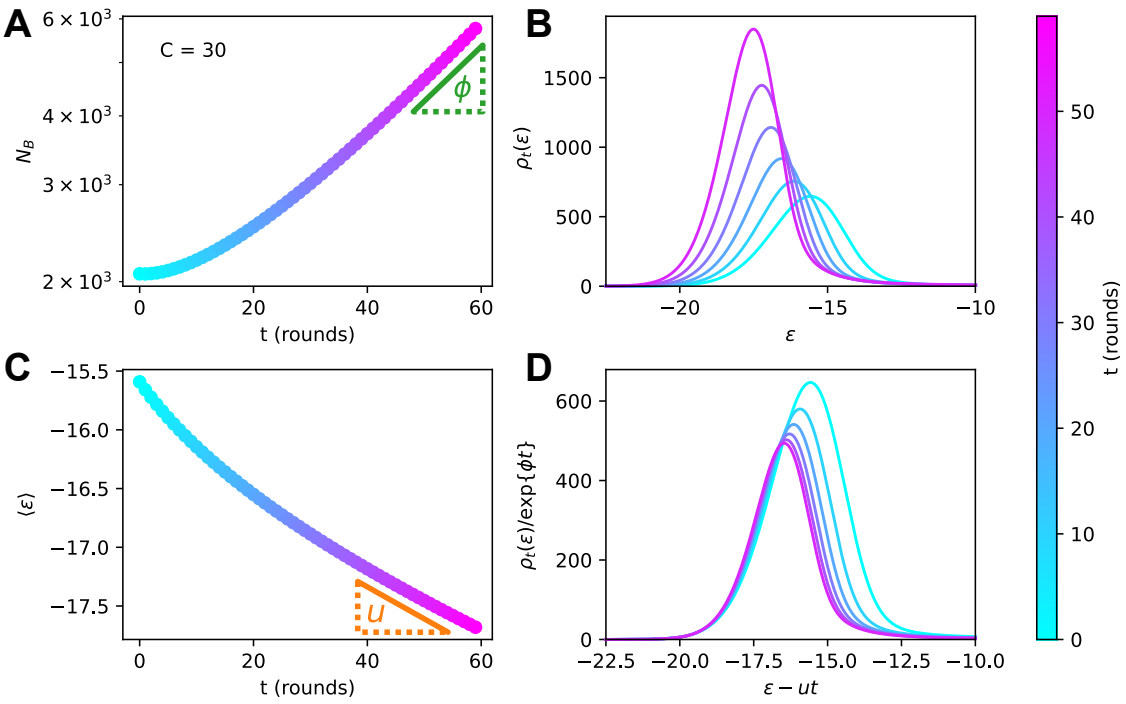

Figure 23: analysis of the asymptotic deterministic evolution for the large-size limit of the model, at constant available concentration $C = 30$. **A**: size of the population vs. number of maturation rounds, showing the exponential increase at rate ϕ. **C**: average binding energy of the B-cell population, decreasing linearly with speed u. **B**: evolution of the binding energy density function $\rho_t(\epsilon)$ shows a traveling-wave behavior. Different times t are represented with different colors, according to the color-scale on the right. **D**: distributions of binding energies, shifted by the time-dependent factor $-ut$ and rescaled by the exponential factor $\exp\{-\phi t\}$. Notice the convergence to the invariant distribution ρ^*.

sity function evolves asymptotically as a travelling wave with exponentially increasing size. In particular, if we name $\rho^*(\epsilon)$ the wave profile, this can be made explicit as:

$$\rho_t(\epsilon) \simeq \exp\{\phi\, t\} \times \rho^*(\epsilon - u\, t)\ . \tag{20}$$

In fig. 23D we plot the same density function, but we apply a shift $-ut$ and an exponential rescaling of factor $\exp\{-\phi t\}$. As expected the resulting function converges to a fixed shape.

2.7.2 *Eigenvalue equation*

The traveling-wave behavior can be mathematically established, and the growth rate ϕ and maturation speed u computed by solving an appropriate eigenvalue equation.
To do so we start by defining the restricted evolution operator $\mathbf{W} = \mathbf{D}\,\mathbf{S}_T\,\mathbf{R}$, that encodes the system evolution in our approximation in which carrying capacity and Ag-binding selection are neglected. In this approximation the density function evolves as $\rho_{t+1} = \mathbf{W}\,\rho_t$. For the definition of the operators that make up $\mathbf{W}$ one can refer to section 2.5.1. Notice that since these consist only of multiplications and convolutions the restricted evolution operator is a linear operator. Moreover all of its entries are non-negative.

Asymptotically in one round of maturation we expect the density function $\rho_t(\epsilon)$ to be shifted by u along the energy axis, and to be multiplied by e^{ϕ}. By again referring to the asymptotic form of the density function as ρ^* we can write this condition as:

$$e^{\phi}\,\rho^* = \mathbf{T}(-u)\,\mathbf{W}\,\rho^* \tag{21}$$

Where the operator $\mathbf{T}(x)$ implements a translation of the density function in energy space of magnitude x. Equation (21) is an eigenvalue equation for the combined operator $\mathbf{T}(-u)\,\mathbf{W}$, having eigenvector ρ^* and eigenvalue e^{ϕ}. Since this operator has all non-negative components Perron-Frobenius theorem guarantees that the eigenvector corresponding to the maximum eigenvalue has only non-negative components. Intuitively, evolution (i.e. repeated application of the operator) will select out this eigenvector since it is the one with maximal growth rate. When stating that $\mathbf{W}$ is a linear operator we overlooked one fact: selection for T-cell help is competitive and the expression of the survival probability depends non-trivially on the distribution itself through the value of $\bar{\epsilon}$ (see eq. (14)). This however can be simplified considering that the operator is translationally invariant, which leaves a translational degree of freedom in the determination of ρ^*. We remove this degree of freedom by operating the choice $\bar{\epsilon} = 0$. This gauge-fixing choice linearizes the operator.

To verify the correctness of our theoretical prediction we solve the eigenvalue problem at a given Ag concentration $C = 10$, graphically illustrating the procedure, and show that the result obtained from the eigenvalue equation matches the asymptotic behavior of the system. The solution procedure is as follows. Since we do not a priori know the value of the wave velocity u we solve the eigenvalue eq. (21) for different values of the shift. We name this shift Δ, and re-write the equation as $e^{\phi}\,\rho^* = \mathbf{T}(\Delta)\,\mathbf{W}\,\rho^*$. The resulting eigenvectors are plotted in fig. 24A (colors encode different values of Δ). The corresponding log-eigenvalues, representing the growth rate, are plotted in fig. 24C. In order for our solution to be consistent it must satisfy the gauge-fixing condition $\bar{\epsilon} = 0$. Therefore for each solution we evaluate $\bar{\epsilon}$,[7] as reported in fig. 24B, and we pick the one (Δ^*) for which this condition is satisfied as the eigenproblem solution.
If our theory is correct, upon repeated application of the restricted evolution operator we expect the population to asymptotically grow at rate ϕ^* and translate with velocity $u = -\Delta^*$, and its binding energy distribution to converge to the shape ρ^*. In fig. 24D we verify this by plotting the evolution of the normalized affinity distribution of the population, re-shifted on its mean, upon repeated application of the evolution operator, and compare it with the eigenproblem solution ρ^* (black dashed line). Moreover in E and F we compare the instantaneous growth rate and energy shift with the eigenproblem solution predictions ϕ^* and $-\Delta^*$. Color in the three plots encodes the evolution round. We observe that in all three cases the asymptotic prediction is matched.

2.7.3 *Ag concentration determines different maturation regimes*

The values of the growth rate ϕ and maturation speed u, which can be obtained from the eigenvalue equation previously described, depend themselves on the value of the

7 More precisely, since in our model T-cell help selection acts after duplication and mutations, $\bar{\epsilon}$ is evaluated on the density function ρ^* after application of these operators.

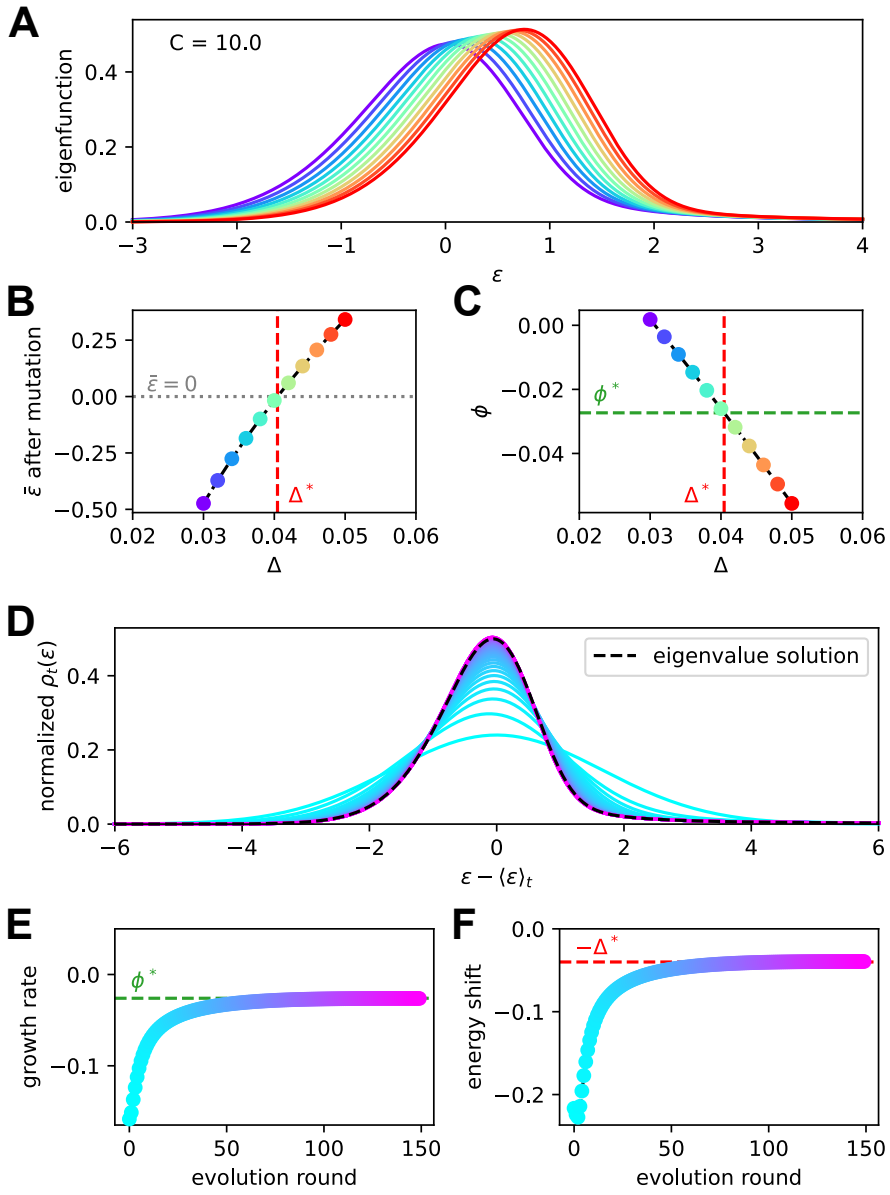

Figure 24: We check that upon repetition of the evolution operator the system converges at the eigenvalue equation solution. For a given constant Ag concentration (C = 10 in our case) we solve the eigenvalue equation $e^{\phi}\rho = \mathbf{T}(\Delta)\mathbf{W}\rho$ for various values of the shift Δ. In **A** we report the maximum eigenvalue eigenfunctions. By virtue of the Perron-Frobenius theorem these consist of only positive values. Color represents the value of the shift Δ for the corresponding solution. In **B** and **C** we plot the value of $\bar{\epsilon}$ after mutation and of the growth rate ϕ for every solution. The consistency condition requires us to pick the eigenfunction for whom the value of $\bar{\epsilon}$ after mutation is zero. This corresponds to the value Δ^* represented in vertical red dashed line and the value of the growth rate ϕ^* in horizontal green dashed line. In panel **D,E,F** we consider repeated application of the evolution operator to the binding energy distribution at constant Ag concentration C = 10. Color encodes the number of applications of this operator. In **D** and **E** we report respectively the growth rate and shift of the mean per evolution turn, and in **F** the full distribution of binding energies, normalized to the population size. All the quantities converge to their theoretical expectation given by the chosen solution of the eigenvalue equation, reported as green and red dashed lines.

Ag concentration C through the expression of the T-cell help survival probability eq. (7). In this equation Ag concentration controls the strength of selection, with more cells surviving when more Ag is available to bind.
To quantify this dependence we evaluate ϕ and u for different values of the Ag concentration. Since the numerical solution of eigenvalue eq. (21) is computationally expensive, requiring the instantiation and diagonalization of large matrices, in this case we simply evaluate ϕ and u as the asymptotic values of growth rate and maturation speed obtained upon repeated application of the restricted evolution operator $\mathbf{W}$. Results are shown in fig. 25. Two special values of the concentration are found. The first, C^*, is the concentration at which the growth rate ϕ vanishes, and the second C^{**} is the one at which the maturation speed u vanishes. These concentrations delimit different regimes of maturation:

- At low Ag concentration $C < C^*$, both ϕ and u are negative: the strong selection pressure produces high affinity clones and maturation is fast, but the number of cells decreases exponentially, leading to a quick extinction of the population.
- At high concentration $C > C^{**}$, the selection pressure is too weak to compensate the deleterious drift due to mutations, and binding energies increase on average at each round ($u > 0$). The growth rate ϕ is positive, hence an exponentially increasing number of poor-quality B-cells are produced.
- In the intermediate range of concentration, $C^* < C < C^{**}$, we have both population expansion (positive growth rate ϕ) and affinity maturation (negative maturation speed u). The most efficient maturations are obtained for values of C slightly exceeding C^*, as u is very close to 0 for values of C tending to C^{**}.

The above analysis provides a detailed picture of the effect of Ag concentration on population growth and maturation, even when the approximations introduced at the beginning of this analysis are removed. First, if we forbid the population from expanding indefinitely and enforce the maximum carrying capacity (N_{max}) again, the value of $u(C)$ is not modified, since this constraint has no effect on affinity. It also does not influences the regime $C < C^*$ in which the population contracts ($\phi(C) < 0$). However it prevents the population from expanding indefinitely, thus setting effectively the maximum asymptotic growth rate to $\phi(C) = 0$ if $C \geqslant C^*$. Second, if we reintroduce Ag-binding selection we observe no difference in asymptotic behavior when the population is maturating ($C < C^{**}$ and $u < 0$). However for high concentration $C > C^{**}$ a positive asymptotic velocity is not possible, since in this case the distribution will eventually reach the threshold Ag-binding energy ϵ_{Ag} and this selection will prevent further affinity decrease. This limits the maximum asymptotic velocity to 0 and maximum growth rate to $\phi(C^{**})$.
Moreover in reality Ag concentration is not constant but varies during immunization through consumption and decay (cf. fig. 12). The maturation behaviors observed during GC evolution (cf. fig. 15) can be understood depending on whether the value C of the concentration crosses the boundaries C^{**} or C^* over time.

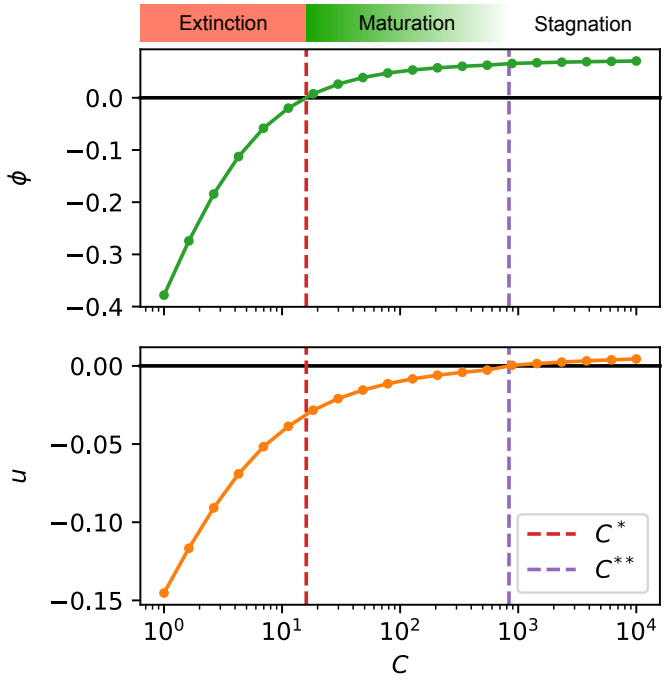

Figure 25: Values of the growth rate ϕ (top) and maturation speed u (bottom) as functions of the Ag concentration C. The points at which the two quantities are zeros define the two critical concentration C^* and C^{**} (red and purple vertical dashed lines). They split the asymptotic behavior of the system at constant Ag concentration in three different regimes: extinction for $C < C^*$ ($\phi < 0$), maturation for $C^* < C < C^{**}$ ($\phi > 0$ and $u < 0$) and finally stagnation for $C > C^{**}$ ($\phi > 0$ but $u \gtrsim 0$).

2.8 INFERENCE AS A TOOL TO INVESTIGATE AM MECHANISMS

In the previous section we showed an example of how, after fitting model and data through bayesian inference, the model can be used to explain phenomena observed in the data. In particular we focused on understanding the mechanism that generates the optimal intermediate Ag dosage.
In this section we carry the same approach further by studying other aspect of AM which would be hard to access experimentally, but can be easily studied in the model. While these results do not possess the same degree of certainty that experimental evidence would have, they still offer motivated support to some of the views already presented in literature. Their resolutive power is founded on the richness of information contained in our experimental dataset.

2.8.1 *Degree of permissiveness in GC selection*

The role and magnitude of permissiveness in GC selection is still an open question [13, 90, 148] (see section 1.3.4). Through phenomena such as bystander activation [55] and stochastic noise GC selection may also allow intermediate and low affinity clones to survive, rather than maturing exclusively via selection of the few best clones [74, 144]. These phenomena generate a wider diversity than previously appreciated, especially when considering complex Ags displaying different epitopes [70]. In [41] for example the authors try to characterize the GC response to complex Ags such as influenza vaccine, as opposed to simple ones such as haptens. While in the latter case a strong homogenizing selection and affinity maturation is observed, for complex Ags response is more polyclonal and a consistent part of the GC population (20-30%) is composed of low-affinity clones. This suggests a more permissive nature of the GC selection, in which even low-affinity clones have a non-zero probability of receiving T-cell help. This permissiveness might be important to maintain diversity in the repertoire, which in turn is useful when facing mutable pathogens [132].
In our model permissiveness is encoded by the two parameters a and b, that regulate the probability of surviving the competitive selection for receiving T-cell help, see eq. (7) and fig. 14D. a corresponds to the baseline probability for cells to survive a selection step, while b is equal to the probability for cells to fail selection even if they have high affinity. This could be due for example to the limited availability of T-cell help, which could increase the stochasticity of the selection process [69]. The role of the parameters a and b in controlling the population evolution is studied in more detail in appendix A.4. Our maximum likelihood fit of the data yields $a = 0.12$ and $b = 0.66$. These values imply that the probability that a high-affinity cell to survive the second step of selection is $1 - b = 34\%$, about two and a half times the probability for a low-affinity cell, given by $a = 12\%$. This observation is in support for the permissive and stochastic nature of selection, at least in our experimental conditions. In contrast, the non-permissive variant of our model with base-line levels $a = b = 0$, referred to as variant B, offers a much worse fit of the data, even when taking into account the smaller number of parameters of this variant (see section 2.6.3).

2.8.2 *Maturation with and without loss of clonality*

Recent experiments [144, 2, 70] have shown that maturation is accompanied by various degrees of homogenizing selection, that is, a reduction of clonality, leading in some cases to strong clonal dominance. We can assess the impact of homogenizing selection in our model by keeping trace of the offspring of each founder clone in the stochastic evolution of a single GC for a 4 weeks time-span. The evolution of clonality is reported for two representative simulations in fig. 26 A,B. The plot report the cumulative composition of the population as a function of time; the offspring of each founder clone is represented by a different color, associated to the binding energy of the founder clone according to the color scale on the right. In the simulation reported in fig. 26A a single clonal family ensued from a high-affinity clone progressively expands, and constitutes around 70 % of the total GC population at 4 weeks. Conversely, in fig. 26B no clone dominates the population, and the GC maintains its polyclonality throughout maturation, with many good affinity clones sharing substantial fractions of the GC.
To quantify the evolution of homogenization over time we estimated the fraction of the population constituted by the most expanded clone at each given time, where 100% would correspond to the GC being completely populated by the offspring of a single founder clone. In fig. 26C we plot the distribution of this most-expanded-clone fraction 1000 stochastic simulations at four different time-points (1,2,3,4 weeks after injection). All GCs in our simulations are highly polyclonal at the beginning, with each clone constituting 1% of the initial population. As time goes on, however, more and more GCs feature a dominant clone, sometimes with a very high population fraction. The median of the frequency distribution at week 4 is around 30%, meaning that in half of the simulated GCs a single clonal family makes up for more than 30% of the total B-cell population. Finally, in fig. 26 D we plot, for each simulation, the final (week 4) fraction of the population corresponding to the most-abundant clonal family against its initial binding energy. As expected homogenization correlates with the presence of a high-affinity founder precursor.
The absence of strong homogenization is also related in our model to the permissiveness of selection, which makes so that even low-affinity clones are not immediately removed.

2.8.3 *Maturation as combination of beneficial mutations and selection of high-affinity precursors*

The presence of various degrees of homogenization is linked to a second question: whether maturation in our GCs is a consequence of accumulation of beneficial mutations, or rather of selection of high affinity precursors. As discussed in section 1.3.4, experimental evidence suggests that maturation against complex Ag might rely more on the latter.
In fig. 48 we verified that in our model the initial choice of founder clones accounts for a large part of the stochasticity in the maturation outcome, indicating that selective expansion of high-affinity precursors plays an important role in affinity enhancement. This, coupled with the fact that the presence of a high-affinity precursor correlates with homogenizing selection (cf. fig. 26D) indicates that precursor selection plays an important role in our model.
To quantify the contribution of accumulation of beneficial mutations in appendix A.4.1

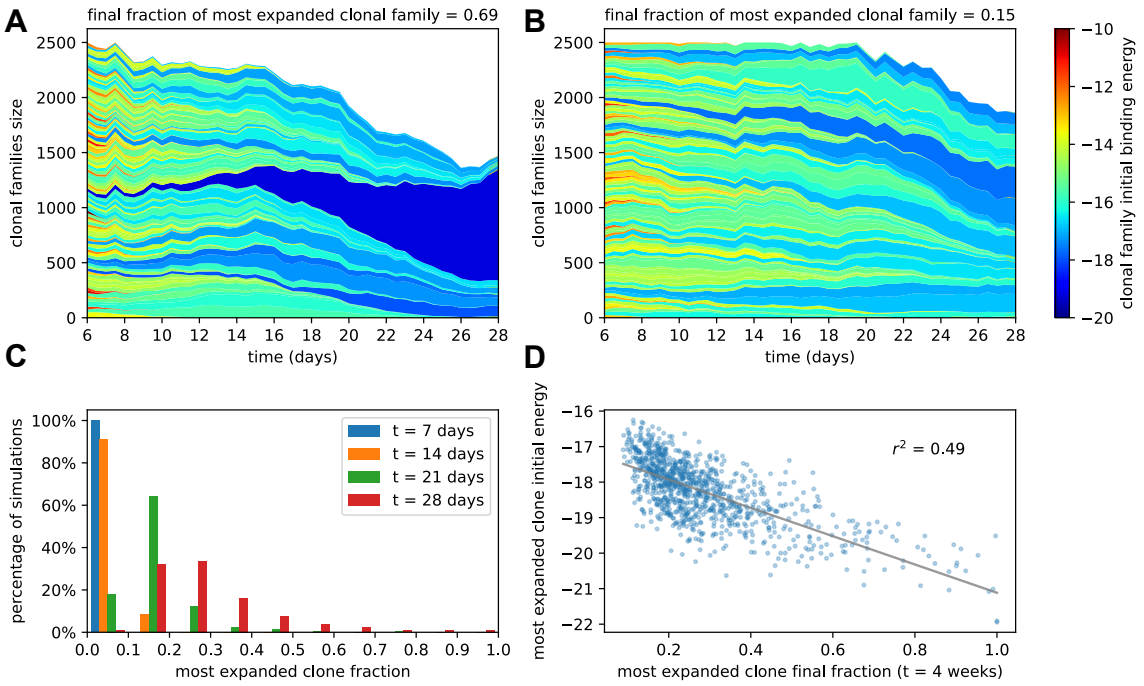

Figure 26: **A**: example of homogenizing selection in GC evolution. Population size as a function of time for each clonal family in stochastic simulations of a single GC. The GC were initiated with an injected antigen dosage of D = 1 µg. The color of the clonal family reflects the initial binding energy of the founder clone according to the color-scale on the right. On top we report the fraction of the final population composed by the most expanded clonal family. In this example the progeny of a single high-affinity founder clone (dark blue) progressively takes over the GC, and at week 4 constitutes around 70% of the GC B-cell population. **B**: example of heterogeneous GC evolution. Contrary to the previous example, many clonal families coexist, without one dominant clone taking over the GC. **C**: evolution of the distribution of the most-expanded clone fraction. We perform 1000 stochastic simulations and evaluate the fraction of the population constituted by the most-expanded clone at each time (cf. colors in the legend). Distributions show the percentage of simulations falling in 10 bins splitting equally the [0,1] interval according to the values of their dominant clone fractions. Notice the presence of heterogeneous and homogeneous GCs at week 4. **D**: scatter plot of final (week 4) population fraction versus initial binding energy for the most-expanded clone; the straight line shows the best linear fit ($r^2 \simeq 0.49$). The presence of a clone with high initial affinity favors the advent of a homogeneous GC.

and fig. 51 we measured the distribution of beneficial and deleterious mutations in MCs and PCs. This analysis shows that indeed selection favors the fixation of beneficial mutations and the disappearance of deleterious ones. Moreover, it shows that on average cells bear very few mutations, suggesting precursor selection plays a major role in our simulations. This is in agreement with experiments made in [102], in which the authors argue that for complex Ags selection of high-affinity precursors plays a bigger role in maturation than the accumulation of beneficial mutations. Similarly experiments performed in [43] on TT immunization found no relevant correlation between the number of mutations harbored by and Ab and its affinity, again indicating a major role of precursors affinity.

2.8.4 *The relative contribution of Ag-binding and competitive selection*

One of the open questions in the literature is whether selection in the GC is mainly *cell-intrinsic*, i.e. based solely on features of the cell such as its affinity, or *cell-extrinsic*, i.e. depending on features of other cells and thus competitive [132, 74] (see section 1.3.3). In our model these two modes of selection are represented respectively by Ag-binding and T-cell help selection. On the analyzed data our inference procedure supports the statistical prevalence of model variant C, featuring T-cell-based selection only, with respect to variant A, which included Ag-binding selection. This is compatible with experiments [129] showing that in absence of competition both high and low affinity cells perform equally well. Moreover the fact that Ag-binding selection does not seem to be a limiting step for GC colonization, at least in the range of our experimentally measurable affinities, is in agreement with experiments performed in [126], in which it is shown that in absence of high affinity competitors even clones with low affinity (as low as $K_d \sim 8\,\mu M$ or equivalently $\epsilon \sim -11.7$) can colonize GCs. This is also in accordance with the fact that selection in GCs should be relatively permissive [13, 149] in order not to limit the diversity of the repertoire. Let us emphasize, however, that the difference in the BIC of the two selection models is rather weak and that our conclusion is contingent on the data set collected and analyzed here.

2.8.5 *Fractions of PCs and MCs amongst Ab-SCs*

Our experimental setup does not allow us to identify whether the IgG-SCs we observe originate from reactivated MCs or residual PCs generated during previous immunizations. We therefore compared the experimental measurements with a weighted mixture of the MC and PC populations predicted by our model. This mixture, which we call the Ab-SC population, represents the population of cells that respond to antigenic challenge under particular conditions. We introduced the parameters g_{recall} and g_{imm}, corresponding to the fraction of reactivated MCs in the Ab-SC populations when measurement is performed one day after boost or four days after the second injection, and fit their value on the experimental measurements. The result of our inference procedure indicates that, when the system is probed 1 day after pure TT boost, most of the response consists in reactivated memory cells ($g_{recall} = 56\%$). This is in agreement with experimental obser-

vations performed in [39], in which the frequency of IgG-SCs increased from $0.6 \pm 0.1\%$ to $1.6 \pm 0.2\%$ one day after the boost, indicating that around 64% of IgG-SCs were not present before the boost. When the measurement is performed 4 days after the second injection then we predict that the vast majority of responders consist of residual PCs ($g_{imm} = 0\%$, with a confidence interval extending to 6%, cfr fig. 18B). This is consistent with experimental data (unpublished), which indicate that the majority of IgG-SCs are still active 28 days after CFA immunization, and will be secreting at +4 days.
Concerning the biological difference between the MC and PC populations, it has been observed that MCs show on average less maturation than PCs [58, 130, 131], a feature that is reproduced in our model (cf. fig. 15C) as a consequence of the temporal switch we introduced (fig. 14B) and might be important in maintaining diversity in the response, especially against highly mutable pathogens, and mitigating original antigenic sin [139, 99]. The results of our inference are in agreement with the fact that experimentally we observe a higher affinity of the responders if measurement is performed 4 days after the last injection (scheme 1) rather than 1 day after boost (scheme 2,3). This difference in affinity could also originate from some form of selection acting on the responder population during the first days of the response, which could selectively expand high-affinity clones in the time between Ag challenge and measurement. Including this selection in the model would result in a different estimate of the fractions g_{recall} and g_{imm}. However, for simplicity and lack of explicit experimental evidence we did not include this selection in the model.

2.9 CONCLUSION

2.9.1 *Summary and significance*

In this chapter we have investigated the relationship between Ag dosage and quality of immunization outcome. Several studies [149, 63, 38, 46, 112] report the fact that better affinity maturation is not always favored by higher doses of Ag, but can instead be enhanced by lower doses. Similarly, the strength of a response to a vaccine, usually measured through the count of responding cells, may show a bell-like curve at intermediate dosages, and understanding the mechanisms underlying this behavior and locating the optimal Ag dose are of crucial importance [121]. Our works provide quantitative theoretical and experimental support to these findings. In particular, the stochastic model for Affinity Maturation we consider here is capable of explaining and accounting for the existence of an intermediate optimal Ag dosage, that results in the highest average affinity of the recalled population. The initial inspiration for our model was provided by previous studies of maturation [155]. However these studies were oriented towards the understanding of AM in the presence of multiple Ags, and many model ingredients are different. For instance amongst the main ones, while not considering the presence of multiple Ag variants, our model includes a realistic dynamic for Ag concentration, and a prominent role for stochasticity and permissiveness in GC selection. Together these ingredients make the model predictions directly comparable to experimental data, leading to important results.

First, our model is amenable to detailed mathematical analysis. We show that the stochastic evolution of the distribution of binding energies can be accurately approximated by a deterministic dynamics (see fig. 22), which we resolve exactly. Under con-

stant Ag concentration, the distribution of binding energies behaves as a traveling wave, whose speed and growth rate can be recovered by solving an appropriate eigenvalue equation (21). The dependence of these two quantities on Ag concentration reveals the role Ag availability plays in controlling the strength of selection, both in the generated data and models. In particular, high Ag dosage results in low selection pressure and no maturation, and conversely too low Ag dosage in high selection pressure and population extinction. Only intermediate Ag concentration and intermediate selection pressure ensures both population survival and successful AM.

Secondly, we show that a single set of parameters of our model is able to reproduce quantitatively the many distributions of single-cell affinities measured on IgG-SC extracted from mice immunized against Tetanus Toxoid corresponding to multiple protocols largely varying in Ag dosages and delays between injections. To determine the best parameters, we introduce a maximum-likelihood-based inference method. Our inference method fully exploits the results of the experimental technique, developed in [40], giving access not to the average affinity, as titer measurement would, but to the complete affinity distribution of the recalled Ab-SC population. This population information is crucial for accurate inference of the model parameters and for a meaningful validation of the model.

Finally, we show how study of the inferred model can provide insights on the internal processes of affinity maturation, such as on the role of permissiveness, on the degree of homogenization in GC selection and on the relative contribution of Ag-binding and competitive selection. Inference techniques are powerful instruments in this respect, since they help us investigate experimentally unaccessible features of the system through their indirect but measurable effects. Our inference procedure is very flexible and can readily be applied to new datasets, providing *ad hoc* estimates of parameters for different antigens or even different organisms.

2.9.2 *Model limitations and discussion*

In building the model, we chose to only keep the minimal features that could allow us to understand the existence of an optimal dosage and be able to reproduce experimental observations, while still being mathematically tractable. Among the simplifications, the number of duplications per cell is considered independent of the cell affinity. It has been however shown that an affinity correlates with GC dark zone dwelling time and number of divisions [44]. This phenomenon introduces an effective fitness difference, which is in practice qualitatively accounted for by the selection terms in our model. Neglecting this differential amplification is also partially justified by the observation that under non-competitive conditions high and low affinity cells have similar rates of division, but different rates of cell death [132, 10].
Moreover we consider the distribution of affinity-affecting mutations $K_{aa}(\Delta\epsilon)$ (cf. eq. (132)) to be independent of the clone's affinity, similarly to [155, 154, 161]. In reality, independence holds only away from affinity peaks in the Ab sequence space; close to these peaks, affinity-increasing mutations become rare, and it is expected that Ag affinity of clones eventually saturate, while the binding energy can take arbitrarily low values in our model. However, in the regime defined by the values of the parameters inferred on our experimental data, MCs and PCs generated by our stochastic model accumulate on average very few mutations in the course of evolution, as showed in appendix A.4.1 and

fig. 51, with the maximum number of beneficial mutations accumulated being compatible with experimental evidence [43]. In this regime mutations account for only a part of the maturation, the rest being achieved through selection of high-affinity founder clones. This is in line with the limited maturation observed in our experiments. In cases where the saturation effect may become relevant, other approaches to model the effect of affinity-affecting mutations might be more appropriate, for example the introduction of a "shape space" representation [127, 153].
The model and results reported here do not include Ab-feedback [155], the phenomenon by which GC B-cells not only have to compete amongst themselves for Ag acquisition but also with Abs produced earlier in maturation [13, 90, 162], which could prevent B-cells from internalizing Ag by binding to it. We did not include Ab-feedback in our model, however preliminary investigations (not shown) suggest that it would not affect the existence of an optimal dosage range.
GC lifetimes reported in literature vary considerably, from 1-2 weeks for soluble protein boosting to several months or longer for certain infections [148, 90]. In alum immunizations GC lifetimes of 3-4 weeks have been observed [142]. In our simulations a long lifetime for GCs is observed and for a high dose of Ag they can have an effective lifetime lasting up to 3 months (cf. fig. 49).
The concentration of Ag is crucial in determining the strength of selection and the lifetime of the GC in our model. In reality, Ag dosage value also controls the initiation of the GC and AM. In particular one could expect that for very low dosages the GC reaction would not be initiated at all. For simplicity we avoid including this phenomenon in our model, and GC reaction takes place in our simulations even at very low Ag dosages, with the result that very few, highly affine MC are produced in this regime. To avoid a discontinuity with respect to the case of null Ag dosage, $D = 0$, in which we expect the measured B-cell population to originate directly from naive precursors, we perform differentiation at the beginning of the simulation round, before mutations and selection (as described in appendix A.1). This generates a core of low-affinity MCs keeping the average affinity of the population close to μ_{naive}, even when few additional high-affinity MCs are added. However, this might be an unnecessary caution, since when looking at the data we observe that even the lowest tested dosage ($D = 0.01\,\mu g$ TT, fig. 21) shows the hallmark of maturation when compared to the the case of zero dosage ($D = 0\,\mu g$ TT, fig. 21). This signals that in the dosage range considered in our experiments we expect maturation to occur.
Furthermore, in our model Ag inputs, e.g. resulting from a new injection, cannot enter a GC while the maturation process is ongoing. Our choice is partly justified by the observation that injecting an Ag bolus when a GC maturation process is in place mostly results in disruption of the ongoing GC reaction [149, 118, 133, 51, 150]. We only model a single "average" GC, whose output is assumed to be representative of the outcome of AM. In reality, MC and PC populations are generated by many parallel GC reactions, which could in principle weakly interact via invasion of clones from one GC to another [90, 148].
Last of all, to test the robustness of some of our hypothesis we performed the inference procedure under slightly different conditions. In particuar we considered the effect of increasing the Ag decay rate, of setting p_{mem} to be a constant and not depend on the number of MCs accumulated during evolution, and also of considering the MC/PC time-switch to be only partial, with a residual production of MCs all along the evolution.

We verified that even in these case the model is in good agreement with the data. The results are reported in appendix A.2.

2.9.3 *Outlooks*

As shown above our model for AM is simple enough to be amenable to detailed mathematical analysis and, yet, is able to accurately reproduce the full affinity distributions of Ab-SCs generated during the immunization process. This finding suggests several extensions to the current work. First our model could be used to predict the outcome of more complex immunization protocols than the ones investigated experimentally in this work. In particular, it would be interesting to consider the case of continuous delivery methods (osmotic pumps, repeated injections...) [143, 27], through which the Ag concentration can be precisely controlled over time, and make predictions for the optimal delivery process. Secondly, the quantitative fit of the model parameters was made here possible thanks to the maximum-likelihood algorithm we have introduced, which is flexible and robust. Our inference procedure, whose code is made available in a public repository, could be readily applied to to different measurements, as well as to variants of the present models, with extra parameters corresponding to features of the affinity maturation process that are hardly experimentally accessible, such as selection permissiveness. The combination of quantitative modelling with inference appears as a promising tool to understand the mechanisms governing the immune response and to guide the development of strategies to control and direct it.

3

STOCHASTIC EFFECTS IN MATURATION MODEL: SURVIVAL, LINEAGES, COMPETITION

3.1 INTRODUCTION AND CHAPTER OUTLINE

This chapter is dedicated to theoretical analysis of some aspect of maturation models, particularly concerning stochastic effects in evolution of B-cell lineages through a population bottleneck.

In section 1.3 we showed how theoretical models are invaluable tools to guide our understanding of AM, offering precious insight on maturation inner mechanisms and guiding vaccine design. Some of these models [155, 154, 161] present a phenomenology termed *population bottleneck* (see section 1.3.3), in which a high selection pressure exposes the B-cell population to a high risk of extinction and termination of the GCR. While a strong selection pressure enhances affinity increase, population extinction is an undesirable outcome. For optimal maturation one is therefore interested in tuning the selection pressure, for example by controlling Ag availability, so as to maximize affinity increase while at the same time avoid extinction. Here we are interested in studying what controls this extinction probability in B-cell lineages that undergo a population bottleneck.
Moreover a second question we tackle concerns how the bottleneck selection pressure shapes the evolutionary trajectory of these lineages in affinity space. This question bears analogies with what observed in reconstructed B-cell phylogenetic trees. The application of modern sequencing techniques to experiments of immunization allows for the reconstruction of the phylogenies of B-cell lineages [57]. If the B-cell population is subject to strong selection the reconstructed phylogenetic trees bear the signatures of selection [109, 57]. Inference techniques have been developed to extract the fitness of nodes in a phylogenetic tree purely from this phylogenetic information [106, 73].[1] Since for B-cells fitness is tightly correlated to Ag affinity, performing such inference can provide information on the evolution of lineages in affinity space [57]. With this approach a B-cell phylogenetic tree can show how for a particular lineage the affinity of a progenitor, driven by the stochastic accumulation of mutations and selection, evolves forward in time to a progeny with possibly very different affinities. Here we approach the problem from the opposite direction. We consider a set of cells with similar affinities that successfully passed the population bottleneck, which in spite of their similar affinities might have different evolutionary histories, and ask what is the most-likely evolutionary trajectory of their progenitors in affinity space through the bottleneck. Rather than a phylogenetic description, aimed at reconstructing the history of the population in genetic space, our approach provides a description of the evolution in phenotypic space, showing how different "affinity compartments" in the population are related over time.

1 Or in the case of the phylogeny of influenza virus also with the use of other relevant genetic information such as the location of mutations in the genome [78].

The chapter is organized in the following sections, whose description summarizes the steps of our analysis:

Simplified model for Affinity Maturation under a population bottleneck : we start by introducing a maturation model similar to the one described in the previous chapter which we use to simulate maturation in the population bottleneck regime. We qualitatively illustrate the effect of this bottleneck on the evolution of cell lineages and how progenitor affinity influences the outcome of evolution. Moreover two model limits are performed: the one of big population size, in which stochastic processes turn into deterministic operators, and the one of continuous time, in which the population density function evolves under the action of a drift-diffusion equation.

Maturation speed and growth rate : building from the continuous-time description, in this section we analyze the asymptotic model behavior. We show that similarly to the model presented in the previous chapter also in this case the density function asymptotically evolves as a traveling wave. However in this case the continuous-time limit allows us to derive an explicit expression for the population growth rate and maturation speed as a function of Ag concentration and other model parameters.

Probability of survival to bottleneck : with the use of the model we study the bottleneck survival probability. We first focus on single lineages. Through the use of a recursive relation we quantify the probability of lineage extinction as a function of the progenitor binding energy in absence of competition. For lineages that go extinct we also quantify stochastic properties such as extinction time and progeny size. Building on these results we then study the bottleneck survival probability for the full population. In this case competition can be accounted for in an effective manner, using the combination of the deterministic model limit and a finite-size correction.

Most-likely evolutionary trajectory : the drift-diffusion equation that describes evolution in the continuous-time limit admits a path-integral expansion. This expansion provides an expression for the action of a trajectory in energy space, that quantifies its probability. Using this result in combination with the method of characteristic trajectories, we derive equations for the most-likely evolutionary trajectories in affinity space, and compare these prediction with stochastic simulations.

Conclusion and perspectives : here we shortly recapitulate the main results described in the chapter, discuss their limitations and possible extensions.

Some of the results presented in this chapter will be the object of a publication, currently at the draft stage and available on arXiv [97].

3.2 simplified model for affinity maturation under a population bottleneck

In this section we define a simple stochastic model for AM which captures the bottleneck phenomenology. We then perform two model limits. First, the limit of big population

size, in which the model becomes deterministic. Second, the limit of continuous time, in which the evolution of the system is described by a drift-diffusion equation.
This model, which is against inspired to similar models [155, 154, 96], is a slightly modified and simplified versions of the one utilized in chapter 2. The differences are motivated by the fact that in this case we are interested in reproducing the bottleneck phenomenology, and thus set the parameters to have a more stringent selection. In addition to this, since we are not interested in reproducing experimental data but rather in obtaining theoretical results, we keep the model simple and consider only minimal ingredients. Amongst the main difference are the fact that in this case we do not consider permissiveness or stochasticity in selection (i.e. remove parameters a and b in eq. (7)), and we do not keep track of the dynamics of Ag concentration but rather consider it constant as done in [155]. This is partly justified by the fact that rather than studying the long-term evolution of the system or the accumulation of MCs and PCs here we are interested in the evolution under a population bottleneck, which is mainly resolved few evolution rounds.

3.2.1 *model definition*

We consider as before the evolution of a population of B-cells inside a Germinal Center. Over time the population increases its average affinity for the Antigen thanks to repeated cycles of mutation and selection. Each cell in the population is solely characterized by its affinity for the Ag, measured in terms of binding energy ϵ and expressed in units of $k_B T$.
The simulation starts when the GC is mature. We consider the initial population to be composed of N_i cells whose binding energy is independently extracted from a Gaussian distribution of naive responders, with mean μ_i and standard deviation σ_i. Cells then undergo iterative rounds of duplication, mutation and selection. These steps are schematized in fig. 27. At the beginning of the round cells duplicate once. Each daughter cell can then independently develop an affinity-affecting mutation with probability p_{aa}, which causes its binding energy to change, or a lethal mutation with probability p_{let}, in which case it is removed from the population. Alternatively, with probability p_{sil}, the cell either does not mutate or develops silent mutations. In both cases its affinity is unchanged. Affinity-affecting mutations change the binding energy of cells by adding a random contribution $\Delta\epsilon$ to the binding energy, extracted from a Gaussian distribution with mean μ_M and standard deviation σ_M (see fig. 28). Parameters are chosen such that only a small fraction of the mutations is beneficial, i.e. decreases the binding energy (cf. appendix B.1).
After mutations cells are selected for Ag binding. Only cells that are able to bind the Ag with sufficient affinity survive this step of selection. As in the previous chapter (cf. eq. (6)), the survival probability for a cell with binding energy ϵ is given by:

$$P_{Ag}(\epsilon) = \frac{Ce^{-\epsilon}}{Ce^{-\epsilon} + e^{-\epsilon_{Ag}}} \tag{22}$$

Where ϵ_{Ag} is a threshold binding energy and C represents the dimensionless concentration of Ag available for cells to bind. This concentration controls the strength of selection, making successful binding more likely when more Ag is available to bind. The functional dependence of this selection on ϵ and C is displayed in fig. 28C. As a second

selection step cells compete to receive a survival signal from T-follicular helper cells. The survival probability for a cell with binding energy ϵ is:

$$P_T(\epsilon, \bar{\epsilon}) = \frac{Ce^{-\epsilon}}{Ce^{-\epsilon} + e^{-\bar{\epsilon}}}, \qquad \text{with } e^{-\bar{\epsilon}} = \langle e^{-\epsilon} \rangle_{pop} \tag{23}$$

Where the term $\langle e^{-\epsilon} \rangle_{pop}$ represents the average of this quantity over the population and encodes for competition, cf. fig. 28 B and D. The surviving cells then can differentiate into Plasma or Memory Cells with total probability p_{diff}. We do not keep track of these differentiated cells in the simulation. After this step if the population size exceeds the maximum carrying capacity N_{max} cells are randomly removed until the threshold is met. The surviving cells start then the next round of selection. The standard value for all model parameters is reported in table 4, and their choice is discussed in appendix B.1.

parameter	value	description
μ_i, σ_i	3, 1.5	mean and variance of the population initial binding energy distribution
N_i	2500	initial population size
N_{max}	2500	maximum carrying capacity
p_{sil}, p_{let}, p_{aa}	0.75, 0.15, 0.1	effective probability of silent, lethal, affinity-affecting mutation.
μ_M, σ_M	0.3, 0.3	mean and variance of distribution of affinity-affecting mutations
ϵ_{Ag}	0	Ag-binding selection threshold energy
C	5	Ag concentration
p_{diff}	0.1	differentiation probability

Table 4: standard values of model parameters. Unless otherwise specified these are the values used in simulations. The choice of their value is discussed in appendix B.1

3.2.2 *Qualitative model behavior: bottleneck and lineages*

Similarly to other AM models [155, 161], for standard parameter values the population initially undergoes a bottleneck state. This is caused by the strong selection pressure initially imposed by Ag-binding selection, which later relaxes if the average population energy reaches values $\langle \epsilon \rangle_{pop} \sim \epsilon_{Ag}$. By controlling the selection pressure (cf. eqs. (22) and (23)) Ag concentration also controls the population survival probability. To illustrate this in fig. 29 we report the average evolution of 1000 stochastic simulations for three different values of the concentration C. In all three cases the population size initially decreases under the combined effect of the two selection steps (fig. 29A). This decrease lasts for few turns, and is accompanied by a quick increase in average affinity (fig. 29B). At this point surviving populations are composed of few high-affinity cells on which the main selection force acting is competitive selection. If this selection pressure is not too strong then the population can expand and mature. Through a mechanism analogous to the one studied in section 2.7 Ag concentration controls maturation speed, as can be seen by comparing the final maturation rate of cases $C = 10$ and $C = 2.7$ in fig. 29B. In

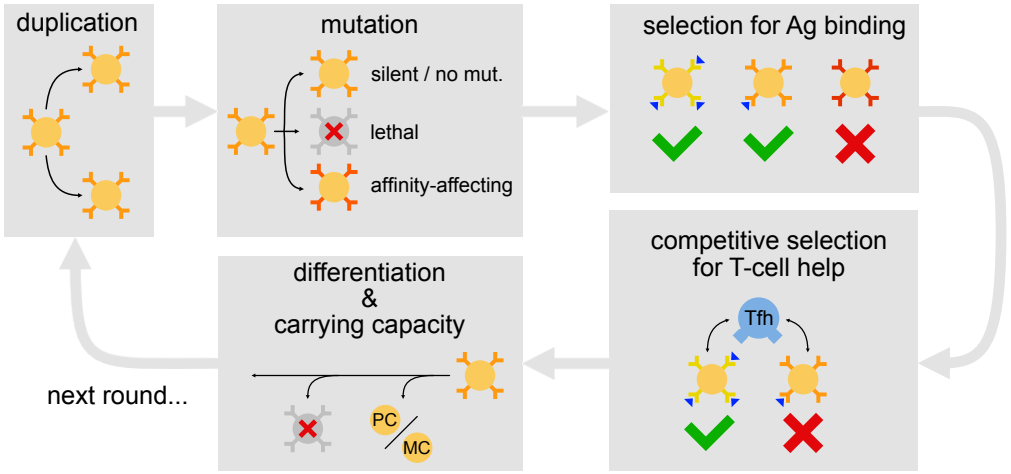

Figure 27: schematic representation of the processes that constitute an evolution round. Initially all cells in the population duplicate. Each daughter cell can develop affinity-affecting ($p_{aa} = 0.1$) or lethal ($p_{let} = 0.15$) mutations. Most of the mutations have a deleterious effect. Cells are then selected on the base of their ability to bind the Antigen, and compete to receive T-cell help. Surviving cells have a small probability ($p_{diff} = 0.1$) of differentiating into Memory or Plasma Cells (MC/PC) and exiting the cycle. Moreover if the size of the population at this step exceeds the maximum carrying capacity ($N_{max} = 2500$) then extra cells are removed randomly. The remaining cells enter then a new evolution round.

fig. 29C we display the fraction of surviving simulations as a function of time. In the case $C = 1$ all simulations end with population extinction. This is due to the fact that at this value of concentration competitive selection alone is sufficiently strong to prevent population growth. The intermediate concentration value $C = 2.7$ is instead sufficient to sustain population growth. In this case extinction can occur during the bottleneck state, when population size is small, but if some cells are able to survive the bottleneck then the population grows back to full size and continues maturation. Finally, at high concentration $C = 10$ the bottleneck pressure is not sufficient to significantly endanger population survival, and all simulations are able to overcome the low-population state without going extinct. In the next session we will try to quantify how Ag concentration controls this survival probability.

At the level of clonal families, survival and future expansion of is strongly dependent in our model on the progenitor affinity. In fig. 30A we display the evolution of lineages generated by three progenitors with different initial affinity. The lowest-affinity one (red) goes extinct in few evolution rounds. In the one with intermediate affinity (orange) only few individuals are able to survive the bottleneck. The high-affinity one (green) instead expands and eventually takes over the population. In fig. 30B we display genealogies for the three progenitors, in the form of trees in which each node corresponds to a different cell. In our analysis we will try to quantify the lineage survival probability as a function of progenitor affinity.

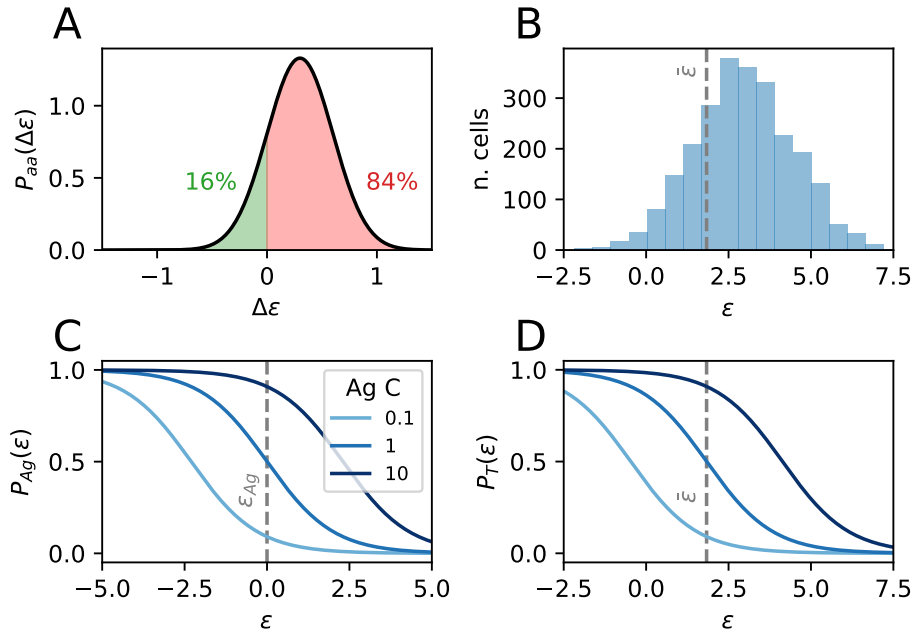

Figure 28: **A**: probability distribution of the energy change $\Delta\epsilon$ introduced by affinity-affecting mutations. Most of the mutations are deleterious. **B**: example of binding energy histogram for a population of B-cells and the corresponding value of $\bar{\epsilon}$. **C**: Probability for a cell of surviving Ag-binding selection, as a function of the energy of the cell ϵ and the antigen concentration C. The threshold value ϵ_{Ag} is represented with a gray dashed line. An higher concentration corresponds to an higher survival probability. **D**: same as panel C, but for T-cell selection survival probability. Competition is introduced by the fact that in this case the threshold $\bar{\epsilon}$ depends on the binding energy distribution of the population, as shown in panel B.

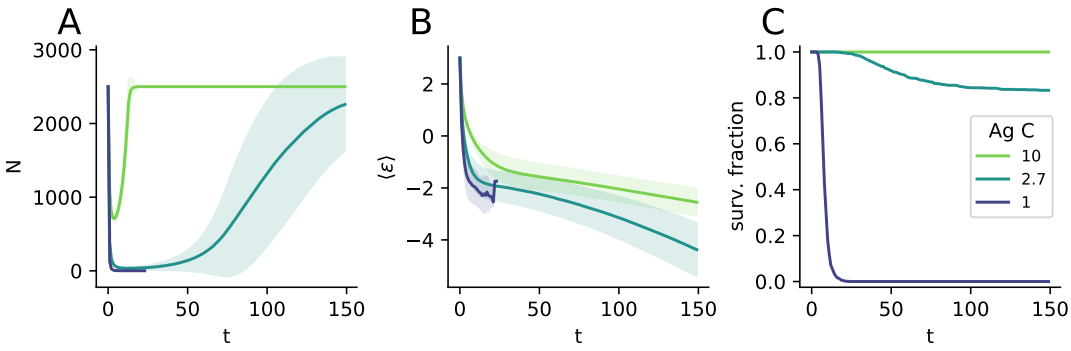

Figure 29: Average evolution for 1000 different stochastic simulations of the model at three different levels of Ag concentration $C = 1, 2.7, 10$, color-coded according to the legend on the right. **A**: population size N as a function of evolution round. Shaded area covers one standard deviation for surviving simulations. The minimum population size on the bottleneck depends strongly on Ag concentration **B**: same as panel A but for the average population binding energy $\langle\epsilon\rangle$. Notice how for surviving population the maturation speed depends on Ag concentration. **C**: Fraction of surviving simulations as a function of time. At low concentration the bottleneck drives all simulations to extinction, while at high concentration the population survives with high probability.

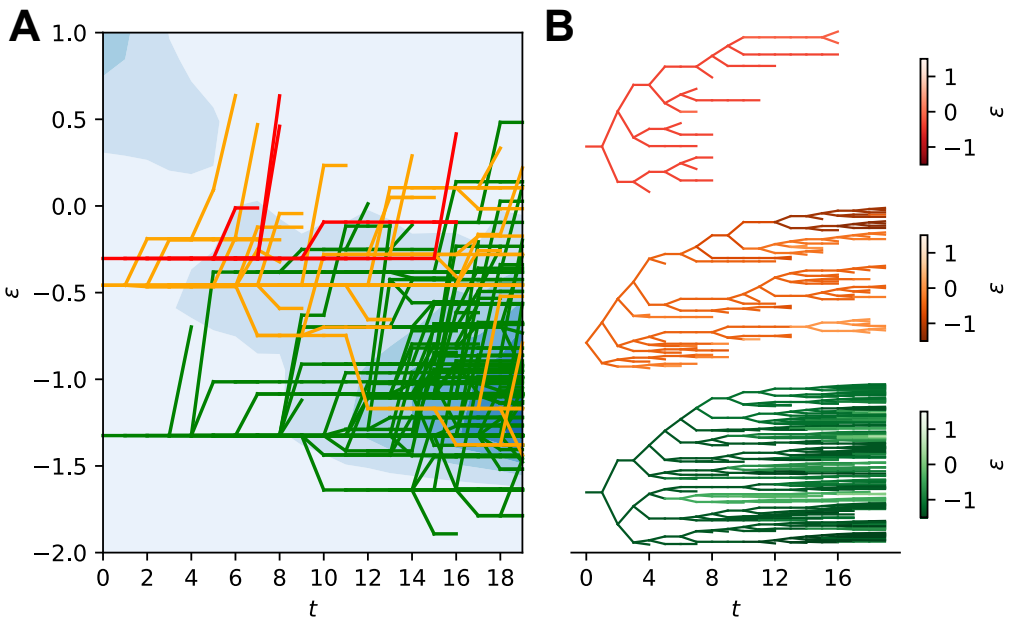

Figure 30: **A**: example of stochastic lineage evolution through a population bottleneck. For three different progenitors (red, orange and green) we plot the evolution of the progeny in energy space, as the population undergoes a bottleneck. Cell density is plotted on the background as reference (blue, darker shade corresponds to higher number of cells). The lineage of the red progenitor quickly goes extinct, while the lineage of the orange progenitor survives the bottleneck but only with few individuals, mainly having low affinity. The green progenitor lineage conversely survives the population bottleneck and its progeny populates the bulk of the final distribution. Notice how fate correlates with the initial progenitor affinity. **B**: Genealogies for the three progenitors displayed in panel A. Each node of the tree corresponds to a different cell. Affinity along a branch is indicated in colors according to the colorscale on the right.

3.2.3 *Limit of big population size*

As a preliminary step to later develop our theory we perform the deterministic limit of big population size. This is analogous to what done in section 2.5 for the model presented in the previous chapters, with few differences (mainly in the duplication, mutation and T-cell help selection operator) due to the differences in model definition. As observed in the previous chapter, in this limit the population binding energy can be approximated with a continuous distribution, and the state of the system is completely determined by the density function $\rho_t(\epsilon)$. This function expressed the density of cells having energy ϵ at evolution round t, so that its integral is equal to the size of the population, and if normalized it becomes equal to the cells binding energy distribution. Evolution is expressed in terms of operators acting on this function. In particular:

1. Cell duplication corresponds simply to doubling in size:

$$\mathbf{A}[\rho](\epsilon) = 2 \times \rho(\epsilon) \tag{24}$$

2. Mutations are represented as the convolution with a mutation kernel:

$$K(\Delta\epsilon) = p_{aa}\, K_{aa}(\Delta\epsilon) + p_{sil}\, \delta(\Delta\epsilon) \tag{25}$$

   Where $K_{aa}(\Delta\epsilon) = \mathcal{N}[\mu_M, \sigma_M](\Delta\epsilon)$ is the distribution of affinity-affecting mutations, consisting in Gaussian with mean μ_M and standard deviation σ_M, and $\delta(\Delta\epsilon)$ is Dirac delta distribution. Notice that the kernel is not normalized, to account for the contribution of lethal mutations. It acts on the distribution as:

$$\mathbf{M}[\rho](\epsilon) = \int d\Delta\epsilon\, \rho(\epsilon - \Delta\epsilon)\, K(\Delta\epsilon) \tag{26}$$

3. Ag-binding selection is implemented by in the product of the population function with the survival probability eq. (22):

$$\mathbf{S}_{Ag}[\rho](\epsilon) = P_{Ag}(\epsilon)\, \rho(\epsilon) \tag{27}$$

4. Similarly, T-cell help selection is given by the product with the survival probability eq. (23):

$$\mathbf{S}_T[\rho](\epsilon) = P_T(\epsilon, \bar{\epsilon})\, \rho(\epsilon), \quad \text{with } e^{-\bar{\epsilon}} = \int d\epsilon\, e^{-\epsilon}\, \rho(\epsilon)/N \tag{28}$$

   Where $N = \int d\epsilon \rho(\epsilon)$ is the current population size.

5. Differentiation consists simply in a product involving the differentiation probability:

$$\mathbf{D}[\rho](\epsilon) = (1 - p_{diff})\, \rho(\epsilon) \tag{29}$$

6. Finally, the carrying capacity constraint is implemented again by a product and is operated only if the size of the population exceeds the maximum limit:

$$\mathbf{C}[\rho](\epsilon) = \min\{1, N_{max}/N\}\, \rho(\epsilon) \tag{30}$$

   Where again $N = \int d\epsilon \rho(\epsilon)$ is the current population size.

3.2.4 *Continuous time description*

As a second preliminary step for our analysis we also provide a continuous-time description of the process, starting from the deterministic limit introduced in the previous section. In this description the density function $\rho(t, \epsilon)$ becomes continuous in the variable t, which we consider expressed in units of evolution rounds. We can describe the evolution of this function with a PDE. We start by considering duplication, lethal mutations and differentiation, which give an energy-independent growth rate:

$$\lambda = \ln\left[2\,(1-p_{let})\,(1-p_{diff})\right] \sim 0.42 \tag{31}$$

Where the numeric value is intended for the standard values of the parameters (see table 4). A second component of the total growth rate is given by Ag-binding and competitive selection, which introduce energy-dependent terms:

$$\ln P_{Ag}(\epsilon) + \ln P_T(\epsilon, \bar{\epsilon}_t)\,, \quad \text{with } e^{-\bar{\epsilon}_t} = \int d\epsilon\; e^{-\epsilon}\,\frac{\rho(t,\epsilon)}{N_t} \text{ and } N_t = \int d\epsilon\;\rho(t,\epsilon) \tag{32}$$

Where N_t represents the size of the population at time t. Finally, the growth rate also contains a term that enforces the finite carrying capacity constraint. This term is non-null only when the population size reaches the maximum allowed size $N_t = N_{max}$ and in this case is defined through the constraint equation.[2]

$$\Omega(t) = \begin{cases} 0 & \text{if } N_t < N_{max} \\ \text{defined through } \frac{d}{dt}N_t \leqslant 0 & \text{if } N_t = N_{max} \end{cases} \tag{33}$$

We group the total growth rate term in the following function:

$$\Gamma(t,\epsilon) = \lambda + \ln P_{Ag}(\epsilon) + \ln P_T(\epsilon, \bar{\epsilon}_t) + \Omega(t) \tag{34}$$

Even though we do not explicitly indicate it, this function depends on the full state of the distribution $\rho(t,\epsilon)$ through the definition of $\bar{\epsilon}_t$. This total growth rate includes the effects of duplication, differentiation, selection and lethal mutations. The missing part, i.e. silent and affinity-affecting mutations, will add a drift and diffusion term in the evolution of the density function. The drift and diffusion coefficient are given by the mean and variance of the distribution of silent and affinity affecting mutations $K(\Delta)$ (cf. eq. (25)):

$$\nu = \langle \Delta\epsilon \rangle_M = \frac{p_{aa}}{p_{aa}+p_{sil}}\,\mu_M \sim 3.5 \times 10^{-2} \tag{35}$$

$$D = \langle \Delta\epsilon^2 \rangle_M - \langle \Delta\epsilon \rangle_M^2 = \frac{p_{aa}}{p_{aa}+p_{sil}}\left[\sigma_M^2 + \frac{p_{sil}}{p_{aa}+p_{sil}}\,\mu_M^2\right] \sim 2.0 \times 10^{-2} \tag{36}$$

Where again numerical values[3] are to be intended as relative to the standard parameter choice (cf. table 4). Adding all of these contributions together results in the following drift-diffusion PDE equation that describes the evolution of the density function:

$$\partial_t \rho = \Gamma\rho - \nu\,\partial_\epsilon \rho + \frac{D}{2}\,\partial_\epsilon^2 \rho \tag{37}$$

2 We provide here the definition of $\Omega(t)$ for completeness, but subsequent analysis will be carried out for the most part in the assumption of infinite maximal population size, and will not be influenced by this term.

3 Notice that since both energy and time are expressed in dimensionless units, so are the drift and diffusion coefficients.

As a last remark we point out that when comparing this continuous-time description with discrete time simulations of the model we expect some discrepancies due to different sources. One source is the fact that the deterministic description obtained with the infinite-size limit poorly captures stochastic finite-size effects. A second source is the fact that the continuous-time description of mutations as a drift-diffusion process is accurate in the limit in which the mutational effect is small and many evolution steps are considered. Moreover the fact that the drift and diffusion coefficients depend only on the first two moments of the mutations distribution makes it accurate for Gaussian-like distributions of mutational effects, while in our case the mutation kernel (cf. eq. (25)) is the combination of a Gaussian and a Dirac delta.

3.3 MATURATION SPEED AND GROWTH RATE

As a first step in our analysis, starting from the continuous-time description introduced above, we study the asymptotic maturation regime in which the population evolves as a traveling wave. This is similar in spirit to what done in section 2.7. In this case however the simplicity of the model allows us to recover analytic expressions which explicit the role of model parameters in controlling population growth rate and maturation speed. It also shows how the existence of the maturation regime depends on the value of model parameters.

3.3.1 *Traveling-wave asymptotic solution*

We are interested in finding traveling-wave solutions for eq. (37), which describe the asymptotic evolution of the population affinity in our model. These solution will have the form $\rho(t, \epsilon) \sim \rho^*(\epsilon - ut) \exp\{\phi t\}$, where u represents the wave velocity and ϕ the population growth rate. We start by expressing eq. (37) in terms of the logarithm of the density function $\omega(t, \epsilon) = \log \rho(t, \epsilon)$:

$$\partial_t \omega = \Gamma - \nu \, \partial_\epsilon \omega + \frac{D}{2} \left[(\partial_\epsilon \omega)^2 + \partial_\epsilon^2 \omega \right] \tag{38}$$

This formulation admits an approximation, known as the WKB approximation from Wentzel-Kramers-Brillouin, which is commonly used in quantum mechanics as a semi-classical slow-varying phase approximation. In our case this approximation consists in dropping the second derivative, which leaves us with a first-order partial differential equation:

$$\partial_t \omega = \Gamma - \nu \, \partial_\epsilon \omega + \frac{D}{2} \left(\partial_\epsilon \omega \right)^2 \tag{39}$$

For the logarithm of the density function the traveling-wave ansatz reads $\omega(\epsilon, t) \sim \omega^*(\epsilon - ut) + \phi t$. Plugging this into eq. (39) restitutes an equation for the asymptotic wave shape:

$$-u \, p^* + \phi = \lambda + \ln P_{Ag} + \ln P_T + \Omega(t) - \nu \, p^* + \frac{D}{2} \, p^{*2} \tag{40}$$

Where for ease of notation we indicate with $p^*(x) = \partial_x \omega^*(x)$. Since we are interested in the asymptotic behavior, similarly to what done in section 2.7.1 we neglect both Ag-binding selection (term $\ln P_{Ag}$) and the carrying capacity constraint (term $\Omega(t)$), so as

to allow for positive values of the asymptotic velocity and growth rate. This results in a second degree equation for $p^*(\epsilon)$ which admits two solutions:

$$p^*_\pm(x) = \frac{\nu - u \pm \sqrt{(\nu-u)^2 - 2D[\lambda - \phi + \ln P_T(x)]}}{D} \tag{41}$$

With a slight abuse of notation, considering that for a traveling-wave solution the threshold binding energy for T-selection must evolve as $\bar{\epsilon}_t = \bar{\epsilon}_0 + ut$, we indicate with:

$$P_T(x) = \frac{C\, e^{-x}}{C\, e^{-x} + e^{-\bar{\epsilon}_0}} \tag{42}$$

To further ease the notation we will define $\theta = \nu - u$ and $\gamma = \lambda - \phi$. In order for the solution p^* to be valid on the whole real domain the term under square root in eq. (41) must be positive for any value of x. Considering that $\sup_x \ln P_T(x) = 0$ this condition translates into the following equation:

$$\theta^2 - 2\,D\,\gamma \geqslant 0 \tag{43}$$

Moreover, considering that $\omega^*(x)$ represents the logarithm of a probability distribution we must have that $\lim_{x\to\pm\infty} \omega(x) = -\infty$. At the same time we have that $\lim_{x\to+\infty} \ln P_T(x) = -\infty$. This forces us to accept as valid only the solution $p^*(x) = p_-(x)$.
The traveling-wave ansatz contains one extra translational degree of freedom that needs to be fixed with a choice of gauge. We pick the condition $\bar{\epsilon}_0 = 0$, which translates into $\int d\epsilon\, \rho^*(\epsilon)\, e^{-\epsilon} = 1$. Through a zero-order saddle-point approximation of this integral one can approximate this condition with $\partial_\epsilon \omega^*(\bar{\epsilon}_0) = p^*(\bar{\epsilon}_0) = 1$.[4] By plugging this into eq. (41) one obtains:

$$1 = \frac{\theta - \sqrt{\theta^2 - 2D[\gamma + \zeta]}}{D}\,, \quad \text{with } \zeta = \ln P_T(x = \bar{\epsilon}_0 = 0) = \ln \frac{C}{C+1} \tag{44}$$

This equation can be split into the following two conditions:

$$D - \theta \leqslant 0 \tag{45}$$

$$\theta = \gamma + \zeta + D/2 \tag{46}$$

Now we plug eq. (46) we obtained for θ into the condition for γ eq. (43). This results in the following second degree inequality for γ:

$$\gamma^2 + 2\gamma(\zeta - \frac{D}{2}) + (\zeta + \frac{D}{2})^2 \geqslant 0 \tag{47}$$

This inequality admits two sets of solutions, either $\gamma \geqslant \gamma_+$ or $\gamma \leqslant \gamma_-$, with:

$$\gamma_\pm = \frac{D}{2} - \zeta \pm \sqrt{-2\zeta D} \tag{48}$$

The existence of these two solutions is always granted since by definition $D > 0$ and $\zeta < 0$. However by plugging eq. (46) into condition eq. (45) one can easily verify that only the solutions $\gamma \geqslant \gamma_+$ are acceptable. Finally, since amongst all the possible solutions one can assume that the one with the greatest growth rate will dominate, we select the

4 Notice that the fact that the energy ϵ and time t are dimensionless introduces here a mixing of otherwise dimensional constants.

solution with minimum γ, that is $\gamma = \gamma_+$. By replacing the explicit expressions for γ, θ and ζ into this equality and into eq. (46) one obtains expressions for the population wave speed and growth rate:

$$\phi(C) = \lambda - \frac{D}{2} + \ln\frac{C}{C+1} - \sqrt{-2D\ln\frac{C}{C+1}} \tag{49}$$

$$u(C) = \nu - D - \sqrt{-2D\ln\frac{C}{C+1}} \tag{50}$$

From these equations we can derive expressions for the critical concentrations through the defining conditions $\phi(C^*) = 0$, $u(C^{**}) = 0$:

$$C^* = \frac{1}{\exp\{(\sqrt{\lambda} - \sqrt{D/2})^2\} - 1} \qquad \text{if} \quad \lambda > \frac{D}{2} \tag{51}$$

$$C^{**} = \frac{1}{\exp\{\frac{(\nu - D)^2}{2D}\} - 1} \qquad \text{if} \quad \nu > D \tag{52}$$

For standard values of the model parameters this gives $C^* \sim 2.8$, $C^{**} \sim 168$. Notice that as indicated not for any value of the model parameters λ, ν, D these critical concentrations exist. The existence conditions can easily be understood by considering that:

$$\lim_{C\to\infty} \phi(C) = \lambda - \frac{D}{2} \tag{53}$$

$$\lim_{C\to\infty} u(C) = \nu - D \tag{54}$$

As a side note we point out that this high concentration limit does not correspond to the regime of amplification and diffusion without selection, in which a traveling-wave ansatz does not describe well the asymptotic evolution. Rather, since we performed first the asymptotic limit and then the high-concentration limit, this corresponds to the case in which the value of $\bar{\epsilon}$ is determined by pure diffusion, and is situated on the left tail of the distribution, while selection acts on the bulk of the distribution. This can be better understood by considering that in the pure amplification, drift and diffusion case, with Gaussian initial conditions, the evolution of $\omega(t, \epsilon)$ is given by:

$$\omega(t, \epsilon) \propto -\frac{(\epsilon - \nu t)^2}{2Dt} + \lambda t \tag{55}$$

And the point in which $p(t, \epsilon) = 1$, which is the definition of $\bar{\epsilon}_t$ in the saddle-point approximation, evolves as:

$$\begin{cases} \frac{d}{dt}\omega(t, \bar{\epsilon}_t) = \lambda - \frac{D}{2} \\ \frac{d}{dt}\bar{\epsilon}_t = \nu - D \end{cases} \tag{56}$$

Moreover, through the explicit expressions for the critical concentrations we can derive the conditions of existence for the maturation regime. This regime is realized when $\phi > 0$ and $u < 0$, which is only possible when two conditions are met. The first condition is the existence of the critical concentration C^*. The second is the fact that it must be either $C^{**} = +\infty$ (i.e. this critical concentration does not exist) or if it exists then it must be $C^* < C^{**}$. This translates into:

$$\begin{cases} \lambda > D/2 \\ \nu < D \quad \text{or} \quad \nu < D/2 - \sqrt{2\lambda D} \end{cases} \tag{57}$$

In addition to the previous results from eq. (41) we can also obtain an expression for $p^*(\epsilon)$, which encodes the asymptotic wave shape:

$$p^*(\epsilon) = 1 + \sqrt{-\frac{2}{D}\ln\frac{C}{C+1}} - \sqrt{-\frac{2}{D}\ln\frac{C\exp\{-\epsilon\}}{C\exp\{-\epsilon\}+1}} \tag{58}$$

We conclude by pointing out that these results were obtained with the help of two approximations, namely the WKB approximation and the saddle-point expansion $p^*(\bar{\epsilon}) = 1$. However they have the advantage that they provide an explicit expression with which to interpret the role of the model parameters in controlling the maturation speed and population survival.

3.3.2 *Dependence of growth rate and maturation speed on model parameters*

Similarly to what was found in section 2.7, also in this model Ag concentration C qualitatively splits the model maturation in three regimes, depicted in fig. 31. These regimes are delimited by the two critical concentrations C^* and C^{**} (eqs. (51) and (52)). In the low concentration regime $C < C^*$ the growth rate is negative, and the population quickly goes extinct. At high concentration $C > C^{**}$ the growth rate is positive but so is the wave velocity, meaning that the selection pressure is not strong enough to compensate the effect of deleterious mutations. Finally, optimal maturation is obtained at intermediate concentration $C^* < C < C^{**}$, at which both the growth rate is positive and wave velocity is negative. Notice that the numerical values of C^* and C^{**} found here differ from the ones in section 2.7 (compare fig. 31 and fig. 25). This is due to difference in model definitions and parameters value between the two cases. For example the parameters a and b (cf. eq. (7)) encoding for permissiveness and stochasticity in selection, are not included here, and as discussed in appendix A.4 they influence the value of asymptotic growth rate and maturation speed.
The difference with what done in the previous chapter lays in the fact that in this case the effect of model parameters is made explicit in analytic expressions. Equations (49) and (50) for example show how a larger value for the diffusion coefficient results in a faster maturation, but also in a reduced growth rate. Moreover the expressions for the critical concentrations, combined with their existence conditions, allowed us to explicit the conditions for the existence of a maturation regime (cf. eq. (57)). These conditions indicate that effective maturation (i.e. the existence of a regime in which simultaneously $\phi > 0$ and $u < 0$) can be obtained only in a sub-part of the (λ, ν, D) parameters space. Loosely speaking these conditions state that the growth rate λ, encoding for the combined effect of duplication, differentiation and lethal mutations, must be big enough compared to the diffusion coefficient to allow the existence of a positive growth rate. Moreover to allow for affinity increase the drift coefficient ν must no be too big compared to the diffusion coefficient.

We conclude this section by pointing out some limitations of this analysis. First, this analysis is based on the infinite-size limit, which removes stochastic effects and makes analytic treatment possible. However this limit fails to account for finite-size effects[5]

5 One example is the value of $\bar{\epsilon} = -\log\langle e^{-\epsilon}\rangle_{pop}$, which depends strongly on the high-affinity tail of the population and is closer to the mean when the population size is small.

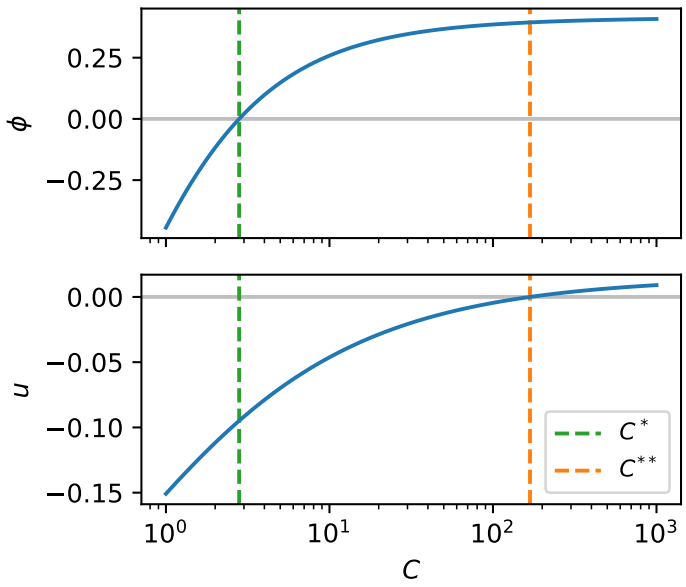

Figure 31: Value of the population growth rate ϕ (top) and maturation speed u (bottom) as a function of Ag concentration C, as predicted by the asymptotic solution of eq. (37) in the WKB approximation. Asymptotically the population size evolves as $N \sim \exp\{t\phi\}$, and the average energy of the population as $\langle \epsilon \rangle \sim ut$. The critical concentrations at which $\phi(C^*) = 0$ and $u(C^{**}) = 0$ are indicated with dashed vertical lines. They split the behavior of the model in three regimes: extinction for $C < C^*$, maturation for $C^* < C < C^{**}$ and stagnation for $C > C^{**}$

and stochasticity, that for evolving populations are important effects. We expect this to generate a discrepancy between the asymptotic velocity in stochastic simulations, where the population size is finite, with the prediction of the deterministic limit.
Another limitation of the asymptotic limit is that it relies on the fact that in our model positive mutations can accumulate indefinitely, with no upper limit to affinity. This is a consequence of the mutation mechanism in our model, in which affinity-affecting mutations introduce a change $\Delta\epsilon$ in binding energy with probability $P_{aa}(\Delta\epsilon)$ (cf. fig. 28 A) that do not depend on the current binding energy of the cell developing the mutation. In reality the probability of developing a beneficial mutation will be in general a decreasing function of the Ab affinity, eventually reaching zero when the cell is in a local maximum of the fitness space, thus preventing infinite maturation.
In spite of these approximations we still believe this analysis to be meaningful. It still represent the "attractor" asymptotic state in a regime in which maturation has not yet reached a fitness peak, and beneficial mutations can still be accumulated. Even though the exact values of the asymptotic velocity may quantitatively differ in theory and simulations, the theory is still useful to show the existence of qualitatively different maturation regimes, and how they depend on model parameters.

3.4 PROBABILITY OF SURVIVAL TO BOTTLENECK

In this next section we study the survival probability of B-cell lineages in a population bottleneck. This survival probability depends strongly on the affinity of the lineage progenitor. We also quantify other stochastic properties of lineages that go extinct, such as the average extinction time and total progeny size at extinction. We then make use of these results to evaluate the survival probability for the full population, also accounting for competition in an effective manner.

3.4.1 *Lineage extinction probability and extinction time in a population bottleneck*

Let us consider a progenitor cell with binding energy ϵ, present in the population at the beginning of evolution $t = 0$. At each evolution round this cell will divide and its offspring will have some probability of being removed from the population, either due to selection, differentiation or lethal mutations. In fig. 32A we report an example of lineage evolution for a progenitor with binding energy $\epsilon = 1$. Color indicates the binding energy of each cell, according to the color-scale on top. In this example cells accumulate deleterious mutations until the lineage eventually goes extinct after $t = 26$ evolution rounds.

We are interested in quantifying the probability $d_t(\epsilon)$ that *all of the offspring* of a progenitor with binding energy ϵ will be extinct be extinct by evolution round t.
The expression for $t = 1$ can easily be written as the probability that both daughter cells generated during the duplication phase will be removed by the end of the round. As stated above, this can occur either by lethal mutation, by failing selection or by differ-

entiation. This probability is more easily expressed as one minus the probability of not being removed:

$$d_1(\epsilon) = \left(1 - \int d\Delta\, K(\Delta)\, P_S(\epsilon + \Delta)\, (1 - p_{diff})\right)^2 \tag{59}$$

Where the expression for $K(\Delta)$ is the one given in eq. (25), and $P_S(\epsilon)$ is the probability for a cell with binding energy ϵ of surviving selection. In a population bottleneck most of the selection pressure is generated by Ag-binding selection (i.e. $\bar{\epsilon}_t < \epsilon_{Ag}$), and as a first approximation we therefore neglect competitive selection for T-cell help and consider simply $P_S(\epsilon) = P_{Ag}(\epsilon)$ (cf. eq. (22)). This introduces two important simplifications. Firstly, the expression of P_S does not depend on time. Secondly, removing the competitive selection uncouples the fate of all cells in the population. Later when considering the survival probability for the full population (section 3.4.3) we will extend this theory to add an effective competitive selection pressure.
The terms for $t > 1$ can be evaluated using an iterative relation that expresses the probability of extinction in t rounds as the probability for each daughter cell to either go extinct in one round, or to survive the first round but to have their respective offspring go extinct in $t-1$ rounds:

$$\begin{aligned} d_t(\epsilon) &= \left[\sqrt{d_1(\epsilon)} + \int d\Delta\, K(\Delta)\, P_S(\epsilon + \Delta)\, (1 - p_{diff})\; d_{t-1}(\epsilon + \Delta)\right]^2 \\ &= \left[1 - \int d\Delta\, K(\Delta)\, P_S(\epsilon + \Delta)\, (1 - p_{diff})\;\; (1 - d_{t-1}(\epsilon + \Delta))\right]^2 \end{aligned} \tag{60}$$

The idea behind the formula is the following: the probability that all of the offspring goes extinct in t rounds is the probability that each of the two daughter cells generated during the duplication phase of the first round independently are removed before the end of round t. Since division is symmetric this probability must be the same for each daughter cell, and it is the term in square brackets. Furthermore, the probability of offspring extinction for each daughter cell can be further decomposed in the probability that the daughter cell will be removed (by lethal mutation, failure of being selected or differentiation) in round one (term $\sqrt{d_1(\epsilon)}$), or that it will succeed selection, but its offspring will go extinct in the remaining $t-1$ rounds (integral containing d_{t-1}). One can also provide an interpretation for the simplified form that appears on the second line. Here the term in square brackets is equal to one minus the probability that the daughter cell survives, and some of its offspring also survives for other $t-1$ rounds.
In fig. 32B we plot the behavior of $d_t(\epsilon)$ as a function of evolution round t and binding energy ϵ (orange curves, color indicates extinction round t). As suggested by intuition, the extinction probability is a monotonically increasing function of time.[6] As t becomes bigger the extinction probability must therefore reach an asymptotic value $d_\infty(\epsilon)$. This asymptotic function tends to one for high energies, indicating that as expected high-energy (i.e. low-affinity) cells quickly go extinct, however it does not tend to zero for low energies. This is because lethal mutations and differentiation still confer a small chance of extinction, mainly in the first few evolution rounds when the number of offspring is still small. This small chance of extinction is $\lim_{\epsilon \to -\infty} d_\infty(\epsilon) = \min\{1, [1 - 1/\alpha]^2\}$, with $\alpha = (1 - p_{let})(1 - p_{diff})$ being the probability for a high-affinity cell to survive the round

6 This property can be easily proven by induction.

and not be removed by lethal mutations or differentiation.[7]
The probability that a lineage generated by a progenitor with energy ϵ goes extinct exactly at round t can then easily be expressed as $r_t(\epsilon) = d_t(\epsilon) - d_{t-1}(\epsilon)$. This allows us to evaluate the mean and variance for the extinction time probabilities (see fig. 32C) simply from the first two moments of the distribution:

$$\langle t \rangle_\epsilon = \sum_{t=0}^{\infty} t\, r_t(\epsilon), \qquad \langle t^2 \rangle_\epsilon = \sum_{t=0}^{\infty} t^2\, r_t(\epsilon), \tag{61}$$

In fig. 32 B and C We compare these predictions with stochastic simulations effectuated in the approximation of no competitive selection. In panel B for different values of the progenitor binding energy we report the mean extinction probability (blue), and as expected this coincides with the limit $t \to \infty$ of the function $d_t(\epsilon)$. Unsurprisingly, both theory and simulations predict that at lower affinity corresponds higher extinction probability. In panel C instead we compare the extinction time from the same simulations (blue, error bars indicate the standard deviation of extinction times for each progenitor affinity) with the predictions from the theory (orange, shaded area covers one standard deviations). We again find a very good match, with both showing a peak at intermediate affinities. This peak has the following interpretation. Low-affinity progenitors (i.e. having high binding energy) have a close-to-one probability of extinction, and on average go extinct in the first few rounds. High-affinity cells on the contrary have a small but non-zero probability of extinction (see the value of extinction probability in fig. 32). This comes mainly from affinity-independent terms such as the differentiation probability, that confer to the lineage a small chance of going extinct on the first selection rounds, when the population is still small. This is consistent with the fact that lineages that go extinct have a small size and go extinct in the first few rounds. For affinities close to $\epsilon = 1$ we have an intermediate probability of survival, and the maximum value for average extinction time and progeny size. This can be even better understood if we turn off mutations, in which case equations can be solved exactly (cf. appendix B.2). In this case we find a divergence in average extinction time and progeny size at a value of ϵ for which every cell on average has one daughter cell that survives selection.[8] The presence of mutations removes this divergence by pushing lineages to one or the other side of the divide: either to high affinities and therefore to survival or to lower affinities and therefore extinction.

3.4.2 *Lineage size at extinction*

A quantity related to extinction time is the total progeny size at extinction. For instance in the example displayed in fig. 32A the lineage consists of a total of 82 cells. Like extinction time, this quantity is well-defined only for lineages that go extinct. Similarly to what done above, we can also derive a recursive formula to quantify the total offspring

7 This last property can be shown by writing eq. (60) as a fixed point equation for $t \to \infty$, and considering that for $\epsilon \to -\infty$ mutations do not sensibly change the survival probability and can therefore be neglected. Notice that if $\alpha < 1/2$ then $d_\infty(\epsilon \ll 0) = 1$, as is to be expected if on average less of one individual in the offspring will survive. However this is a pathological case: in this regime irrespective of the energy the population will asymptotically always go extinct $d_\infty(\epsilon) = 1$.

8 Notice that in the absence of affinity-affecting mutations generated genealogies belong to the class of Galton-Watson trees.

size. However there is a difference: in the case of the extinction time the recursion is based on the fact that the extinction time for the progeny of the mother cell is a random variable equal to the maximum of two other random variables, namely the extinction times of the progeny of the two daughter cells. When considering the offspring size instead we need to consider the sum of two random variables representing the number of descendants of each daughter cell. In the latter case recursion includes a convolution, which is numerically harder to compute. Therefore instead of considering directly the probabilities ones must resort to using the probability generating functions. The recursive relations can be expressed in term of these functions, and can be used to evaluate the moments of the probability distribution without having to numerically perform the convolution.

We name $q_n(\epsilon)$ the probability that a progenitor with energy ϵ generates a total offspring of exactly n cells before extinction, not counting the progenitor itself. This probability can be better expressed if we separate the contribution of the two daughter cells. Considering genealogies encoded as binary trees (cf. fig. 32A), we call $b_m(\epsilon)$ the probability that along the branch corresponding to one of the daughter cells of a progenitor with energy ϵ we find a total of m descendants (including the daughter cell itself) before extinction. The expression for $m = 0$ is simply given by the probability that the daughter cell is removed before the end of the round:

$$b_0(\epsilon) = 1 - \int d\Delta\, K(\Delta)\, P_S(\epsilon + \Delta)\, (1 - p_{\text{diff}}) = \sqrt{d_1(\epsilon)} \tag{62}$$

The recursive relation in this case is composed of two equations. The first is a convolution that decomposes the probability of having n descendants as a sum over all possible repartitions of the descendant number along the two branches:

$$q_n(\epsilon) = \sum_{m=0}^{n} b_m(\epsilon)\; b_{n-m}(\epsilon) \tag{63}$$

The second expresses the probability to find m descend along a branch as the probability that the daughter cell survives and has $m - 1$ descendants:

$$b_m(\epsilon) = \int d\Delta\, K(\Delta)\, P_S(\epsilon + \Delta)\, (1 - p_{\text{diff}})\, q_{m-1}(\epsilon + \Delta) \tag{64}$$

We express these relations in terms of the generating functions $Q(z, \epsilon)$ and $B(z, \epsilon)$, defined as:

$$Q(z, \epsilon) = \sum_{n=0}^{\infty} q_n(\epsilon)\, z^n, \qquad B(z, \epsilon) = \sum_{m=0}^{\infty} b_m(\epsilon)\, z^m \tag{65}$$

In terms of these functions eqs. (63) and (64) read:

$$Q(z, \epsilon) = B(z, \epsilon)^2 \tag{66}$$

$$\frac{1}{z}[B(z, \epsilon) - b_0(\epsilon)] = \int d\Delta\, K(\Delta)\, P_S(\epsilon + \Delta)\, (1 - p_{\text{diff}})\, Q(\epsilon + \Delta, z) \tag{67}$$

These relations can be used to evaluate the moments of these distributions with two additional considerations. The first is that $\sum_{n=0}^{\infty} q_n(\epsilon) = d_\infty(\epsilon)$. This sum does not converge to one since it only considers lineages that eventually go extinct. For the functions Q and B this translates into:

$$Q(z = 1, \epsilon) = d_\infty(\epsilon), \qquad B(z = 1, \epsilon) = \sqrt{d_\infty(\epsilon)} \tag{68}$$

Secondly, the moments of the distributions can be evaluated from the generating functions as:

$$\langle n^k \rangle_\epsilon = \frac{1}{d_\infty(\epsilon)} \sum_{n=0}^{\infty} n^k \, q_n(\epsilon) = \frac{1}{d_\infty(\epsilon)} \, (z\partial_z)^k \, Q(z,\epsilon)|_{z=1} \tag{69}$$

$$\langle m^k \rangle_\epsilon = \frac{1}{\sqrt{d_\infty(\epsilon)}} \sum_{m=0}^{\infty} m^k \, b_m(\epsilon) = \frac{1}{\sqrt{d_\infty(\epsilon)}} \, (z\partial_z)^k \, B(z,\epsilon)|_{z=1} \tag{70}$$

Applying the operator $z\partial_z$ one and two times on eq. (66) restitutes the following relations between the first two moments:

$$\langle n \rangle_\epsilon = 2\langle m \rangle_\epsilon \,, \qquad \langle n^2 \rangle_\epsilon = 2\langle m^2 \rangle_\epsilon + 2\langle m \rangle_\epsilon^2 \tag{71}$$

This corresponds simply to the fact that the total number of descendants is the sum of the descendants along the two branches. Applying the same operator on eq. (67) gives:

$$\sqrt{d_\infty(\epsilon)} \, \langle m \rangle_\epsilon = \int d\Delta \, K(\Delta) \, P_S(\epsilon+\Delta) \, (1-p_{\text{diff}}) \, d_\infty(\epsilon+\Delta) \, [2\langle m \rangle_{\epsilon+\Delta} + 1] \tag{72}$$

$$\begin{aligned} \sqrt{d_\infty(\epsilon)} \, \langle (m-1)^2 \rangle_\epsilon = \int d\Delta \, K(\Delta) \, P_S(\epsilon+\Delta) \, (1-p_{\text{diff}}) \, d_\infty(\epsilon+\Delta) \times \\ \times [2\langle m^2 \rangle_{\epsilon+\Delta} + 2\langle m \rangle_{\epsilon+\Delta}^2 + 1] \end{aligned} \tag{73}$$

These equations can be solved numerically if we express them as fixed-point equations for the functions $\langle m \rangle_\epsilon$ and $\langle m^2 \rangle_\epsilon$:

$$\langle m \rangle_\epsilon = \frac{1}{\sqrt{d_\infty(\epsilon)}} \int d\Delta \, K(\Delta) \, P_S(\epsilon+\Delta) \, (1-p_{\text{diff}}) \, d_\infty(\epsilon+\Delta) \, [2\langle m \rangle_{\epsilon+\Delta} + 1] \tag{74}$$

$$\begin{aligned} \langle m^2 \rangle_\epsilon = \frac{1}{\sqrt{d_\infty(\epsilon)}} \int d\Delta \, K(\Delta) \, P_S(\epsilon+\Delta) \, (1-p_{\text{diff}}) \, d_\infty(\epsilon+\Delta) \times \\ \times [2\langle m^2 \rangle_{\epsilon+\Delta} + 2\langle m \rangle_{\epsilon+\Delta}^2 + 4\langle m \rangle_{\epsilon+\Delta} + 1] \end{aligned} \tag{75}$$

The moments for n can then easily be evaluated using eq. (71).

In fig. 32D we compare the theoretical prediction for the first two moments (orange line represents the mean and shaded area covers one standard deviation) with the corresponding quantities from stochastic simulations (blue, error bars cover one standard deviation). Once more we find a good match. The peak at intermediate affinities can be explained, as done above for the extinction time, considering the critical nature of this phenomenon at intermediate values of the binding energy (see appendix B.2).

3.4.3 *Probability of bottleneck survival for the full population*

Building on the results derived above, we now turn to the problem of quantifying the average probability of extinction for the whole population.
As a first approximation we do not consider competitive selection, since most of the selection pressure in a bottleneck is given by Ag-binding selection. Given a total of N_i cells in the initial population, having energies $\{\epsilon_k\}_{k=1\ldots N_i}$ the probability that all cells will be extinct by evolution round t is simply given by the product of extinction probabilities

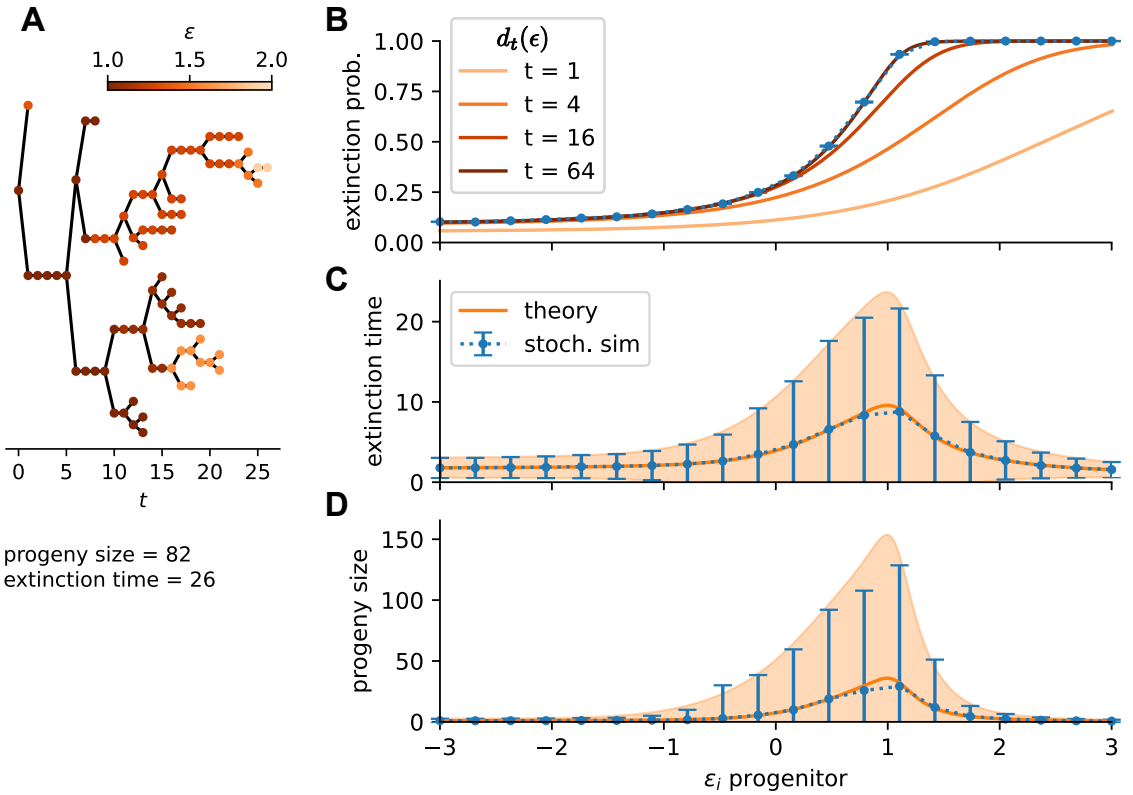

Figure 32: **A**: example of lineage evolution stemming from a progenitor with binding energy $\epsilon_i = 1$, from a stochastic simulation performed at Ag concentration $C = 7$ in the approximation of only Ag-binding selection. Each node in the tree represent a cell, its binding energy ϵ represented using the colorscale on top. In this example cells progressively accumulate deleterious mutations until after 26 evolution rounds the population eventually goes extinct. **B,C,D**: Comparison between stochastic simulations (blue) and theory (orange) for the probability of extinction (B), lineage extinction time (C) and average progeny size at extinction (D) as a function of the progenitor energy ϵ_i in absence of competitive selection. For each conditions we consider 5000 different stochastic simulations that terminate with extinction at Ag concentration $C = 7$. **B**: stochastic extinction probability (blue dots, error bar indicate the standard error of the mean) evaluated as the fraction of simulations that terminate with extinction over the total number of simulations performed. This is compared to the value of $d_t(\epsilon)$ as described by our theory. This represents the probability that the progeny of a progenitor with binding energy ϵ goes extinct by round t, and converges asymptotically ($t \to \infty$) to the total extinction probability. **C**: mean and standard deviation of extinction time (blue) over 5000 simulations terminating in extinction. This is compared to the theoretical prediction (orange) for the mean and standard deviation of this quantity, obtained using the time extinction probability $r_t(\epsilon)$. **D**: same as B but for the progeny size. In this case the theoretical predictions are obtained using the generating function theory.

for all cells $\prod_k d_t(\epsilon_k)$. Moreover, since the initial energies are independently extracted from a Gaussian distribution $\varphi(\epsilon)$ with mean μ_i and standard deviation σ_i, the average extinction probability by round t over all possible extractions of the initial population is given by:

$$P_{ext}(t) = \left(\int d\epsilon \; \varphi(\epsilon) \; d_t(\epsilon) \right)^{N_i} \tag{76}$$

With the help of this formula we evaluate the average survival probability as a function of Ag concentration C and initial population size N_i, and compare the prediction with stochastic simulations in which we turn off T-cell selection. The results, reported in fig. 33B and C (blue), match exactly.

When one adds competitive selection as expected the empirical survival probability evaluated from simulations decreases slightly (compare blue and orange dotted line in fig. 33B and C). The theory can be extended to account for T-selection in an effective manner. In practice, one needs first to extend the theory to include a time-dependence of the survival probability. At this point competitive selection can be included by introducing an effective coupling between cells in a "mean field" fashion, by estimating the average evolution of the term $\bar{\epsilon} = -\log\langle e^{-\epsilon}\rangle_{pop}$ contained in the expression for the T-selection survival probability eq. (23).
Proceeding with order we first extend the theory considering the case in which the survival probability $P_S(\epsilon, t)$ is time-dependent. In this case the probability of extinction does not depend anymore uniquely on the number of evolution rounds we consider, but also on the initial time at which the progenitor is considered. We define $d_{t,s}(\epsilon)$ as the probability that a cell, which at the end of round t has binding energy ϵ, will have all of its offspring extinct by the end of round $s > t$. For any value $t \geqslant 0$ we can write as before the probability of extinction in one round:

$$d_{t,t+1}(\epsilon) = \left(1 - \int d\Delta \, K(\Delta) \, P_S(\epsilon + \Delta, t) \, (1 - p_{diff}) \right)^2 \tag{77}$$

And for any pair of rounds $s > t \geqslant 0$, with $s - t > 1$, the following recursive relation, analogous to eq. (60), holds:

$$d_{t,s}(\epsilon) = \left[1 - \int d\Delta \, K(\Delta) \, P_S(\epsilon + \Delta, t) \, (1 - p_{diff}) \; (1 - d_{t+1,s}(\epsilon + \Delta)) \right]^2 \tag{78}$$

Finally, similarly to eq. (76), the probability that the full population goes extinct by evolution round t is given by:

$$P_{ext}(t) = \left(\int d\epsilon \; \varphi(\epsilon) \, d_{0,t}(\epsilon) \right)^{N_i} \tag{79}$$

Where as before $\varphi(\epsilon)$ is a gaussian distribution with mean μ_i and standard deviation σ_i.
At this point we explicit the time dependence of the survival probability including selection for T-cell help: $P_S(\epsilon, t) = P_{Ag}(\epsilon) \, P_T(\epsilon, \bar{\epsilon}_t)$ (cf. eq. (23)). Given the stochastic nature of our model, the variable $\bar{\epsilon}_t$ which quantifies selection pressure is in reality a stochastic variable. However with a crude approximation we can estimate its average evolution

using the big-population-size limit introduced in section 3.2.3, under which the model becomes deterministic. This allows us to numerically evaluate the extinction probability eq. (79). The resulting estimate however underestimates the real survival probability (compare red curve and orange dotted line in fig. 33B and C). This mismatch originates mainly from the fact that in the big-size approximation $\bar{\epsilon}$ evolves faster than in stochastic simulations (cf. blue and orange line in fig. 33A-left). In turn, this occurs because the value of $\bar{\epsilon}$ is strongly dependent on the high-affinity tail of the population, whose evolution is influenced by finite-size effects.

This discrepancy can however be reduced with a simple finite-size correction. This correction is based on the consideration that the big-size limit of the model approximate the population binding energy histogram with a continuous distribution, encoded in the density function $\rho_t(\epsilon)$ (cf. section 3.2.3). At the beginning of evolution this function takes the shape of a normal distribution (i.e. the initial binding energy distribution of naive responders) with tails extending indefinitely in both directions. However since in reality the population is finite, consisting of N_i individuals, we do not expect these tails to be populated. The correction procedure consists in removing these tails, by setting the initial distribution equal to zero outside a range delimited by two values $[\epsilon^-, \epsilon^+]$. These two values are chosen equal to the expected energy of respectively the highest and lowest affinity individuals of the population. The probability distribution for their binding energy can be expressed as a function of the naive binding energy distribution $\varphi(\epsilon)$ (as before a Gaussian with mean μ_i and variance σ_i^2) from which the energy of all cells is extracted. If we call $F(\epsilon) = \int_{-\infty}^{\epsilon} d\epsilon' \, \varphi(\epsilon')$ the cumulative distribution function, then these distributions can be expressed as:

$$\varphi^+(\epsilon) = \frac{d}{d\epsilon}[F(\epsilon)]^{N_i} \tag{80}$$

$$\varphi^-(\epsilon) = -\frac{d}{d\epsilon}[1 - F(\epsilon)]^{N_i} \tag{81}$$

The values $\epsilon^\pm$ correspond then simply to the means of these distributions.

Removing the tails to the initial distribution causes an initial slow-down in the evolution of $\bar{\epsilon}$ (cf. green line in fig. 33A-left). This slow-down is eventually lost, but the agreement remains for a time sufficient for most of the stochastic simulations to go extinct (cf. fig. 33A-right) which is the relevant timescale to capture bottleneck survival. By taking the value of $\bar{\epsilon}$ obtained by combining the big-size approximation with the cutoff correction, and using it to evaluate the population survival probability, we obtain a much better agreement of the theory with simulations (compare orange curve and orange dotted line in fig. 33B and C). The remaining discrepancy are partly to be attributed to the fact that the average evolution of $\bar{\epsilon}$ is still not exactly captured, and partly to the fact that with this theory we are only accounting for the coupling in a "mean-field" manner, and do not account for correlations between the stochastic evolution of $\bar{\epsilon}$ and of the energies in the population.

3.5 MOST-LIKELY EVOLUTIONARY TRAJECTORY

In this next section we study the evolutionary trajectories of B-cell lineages under a population bottleneck. The question we ask is the following. Let us imagine we observe a

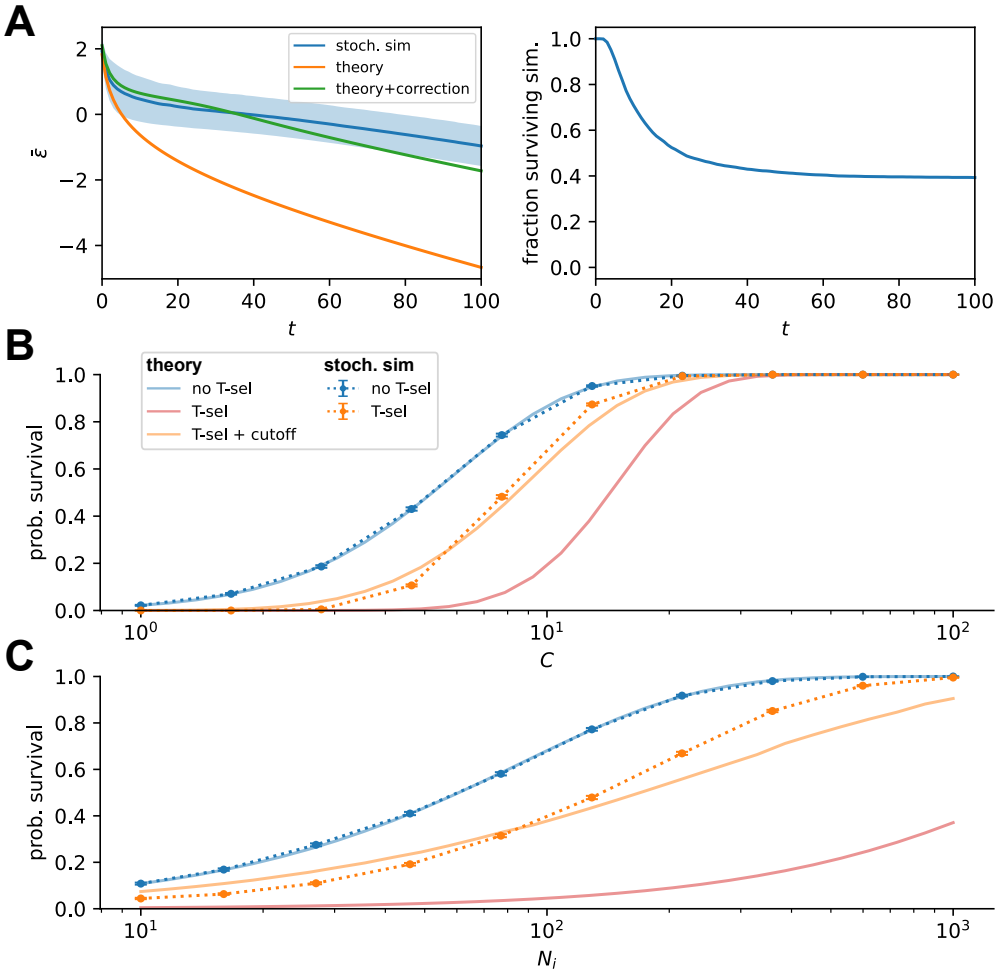

Figure 33: Probability of population survival in a bottleneck condition. **A**: Left: comparison between the evolution of $\bar{\epsilon}$ in stochastic simulations (blue, mean and standard deviation over 5000 simulations) and prediction of our theory without (orange) and with (green) finite-size correction, consisting in cutting the tails of the initial distribution in proximity of the expected value for the highest-affinity individual. The correction improves the prediction for the evolution of $\bar{\epsilon}$ at short times. Right: fraction of surviving simulations as a function of evolution round. The value of $\bar{\epsilon}$ is well-approximated during the time it takes for most of the simulations to go extinct. **B**: bottleneck survival probability as a function of antigen concentration C. Comparison between stochastic simulations (dotted line, error bars indicate the standard error of the mean) and the predictions our theory (full lines). Stochastic simulations are reported both without (blue) and with (orange) T-cell selection. For the theory instead we consider the case without T-selection (blue), with T-selection (red) and with T-selection plus finite-size correction (orange). In the absence of T-selection all cells evolve independently, and the theory and simulations match exactly. The inclusion of T-selection slightly decreases the survival probability in stochastic simulations. Accounting for this contribution by using the infinite-size estimate for the evolution of $\bar{\epsilon}$ overestimates the selection pressure. Adding the finite-size correction results in a much better estimate. **C**: same as B, but the survival probability is evaluated as a function of the initial size of the population N_i.

cell with binding energy ϵ_f at time t_f which successfully survived the population bottleneck. What is the most likely evolutionary trajectory in affinity space that its ancestry followed? Can we estimate the binding energy ϵ_i of its progenitor at time t_i, before the population bottleneck?
To answer these questions we apply the path integral formalism on eq. (37). By combining this formalism with the method of characteristic curves one can find an expression for the *action* of trajectories in affinity space, which quantifies the log-probability of trajectories. To verify its validity we compare theoretical maximum-action trajectories with the average trajectories of lineages from simulations.

3.5.1 *Path integral formulation*

We start by applying the path integral formalism to the PDE eq. (37) that describes the evolution of the population density function. The first step consists in performing an infinitesimal expansion of the equation for a small time-increment δt. In particular, we want to recover an expression of the form $\rho(t+\delta, \epsilon) = \exp\{\delta t\, O\}\, \rho(t, \epsilon)$, where ρ is the cell density function, whose evolution is described by the PDE equation, and O is an operator responsible for its infinitesimal evolution. This expansion reads:

$$\begin{aligned}\rho(t+\delta t, \epsilon) &= \rho(t, \epsilon) + \delta t\, \partial_t \rho(t, \epsilon) + o(\delta t) = \\ &= \left\{1 + \delta t \left[\Gamma - \nu \partial_\epsilon + \frac{D}{2}\partial_\epsilon^2\right]\right\} \rho(t, \epsilon) + o(\delta t) = \\ &= \exp\left\{\delta t \left[\Gamma - \nu \partial_\epsilon + \frac{D}{2}\partial_\epsilon^2\right]\right\} \rho(t, \epsilon) + o(\delta t)\end{aligned} \tag{82}$$

Where we considered up to first order in δt and used the expansion of the exponential function for small values of the argument. At this point one can use the Fourier representation of Dirac delta function to express $\rho(\epsilon, t)$ as:

$$\rho(t, \epsilon) = \int d\epsilon'\, \rho(t, \epsilon')\, \delta(\epsilon - \epsilon') = \int d\epsilon'\, \rho(t, \epsilon') \int_{i\mathbb{R}} \frac{d\widetilde{\epsilon}}{2\pi i}\, e^{-\widetilde{\epsilon}(\epsilon - \epsilon')} \tag{83}$$

Substituting this expression in eq. (82) gives:

$$\rho(t+\delta t, \epsilon) = \int d\epsilon' \int_{i\mathbb{R}} \frac{d\widetilde{\epsilon}}{2\pi i}\, \rho(t, \epsilon')\, \exp\left\{\delta t \left[\Gamma + \nu\widetilde{\epsilon} + \frac{D}{2}\widetilde{\epsilon}^2 - \widetilde{\epsilon}\, \frac{\epsilon - \epsilon'}{\delta t}\right]\right\} + o(\delta t) \tag{84}$$

By repeating iteratively the procedure for many small increments δt, and sending the increment to zero, one obtains the following path integral expression for the system's evolution:

$$\rho(t_f, \epsilon_f) = \int d\epsilon_i\, \rho(t_i, \epsilon_i) \int_{(t_i, \epsilon_i)}^{(T, \epsilon_f)} \mathcal{D}[\epsilon] \int_{i\mathbb{R}} \frac{\mathcal{D}[\widetilde{\epsilon}]}{2\pi i}\, \exp\left\{\int dt \left[\Gamma + \nu\widetilde{\epsilon} + \frac{D}{2}\widetilde{\epsilon}^2 - \widetilde{\epsilon}\, \partial_t \epsilon\right]\right\} \tag{85}$$

Here ϵ and $\widetilde{\epsilon}$ represent trajectories respectively in energy and response variable. This formula has the following interpretation. The value of the density function at a given final energy ϵ_f at the final time t_f is given by a sum of contributions coming from all possible trajectories linking an initial point ϵ_i at initial time t_i to the final point. Not all trajectories are equal, with their relative weight being determined by the *action*:

$$S[\epsilon, \widetilde{\epsilon}] = \int dt \left[\Gamma(\epsilon, t) + \nu\, \widetilde{\epsilon} + \frac{D}{2}\, \widetilde{\epsilon}^2 - \widetilde{\epsilon}\, \frac{d\epsilon}{dt}\right] \tag{86}$$

The trajectories that maximize this term are the ones that contribute the most to the density at final time. These trajectories must satisfy the stationary action equations:

$$\begin{cases} \delta S/\delta \epsilon = 0 \implies d\widetilde{\epsilon}/dt = -\partial_\epsilon \Gamma \\ \delta S/\delta \widetilde{\epsilon} = 0 \implies d\epsilon/dt = \nu + D\widetilde{\epsilon} \end{cases} \tag{87}$$

However these equations lack an interpretation for the value of the response variable $\widetilde{\epsilon}$. This can be obtained through the method of characteristic trajectories.

3.5.2 *Method of characteristic trajectories*

The method of characteristic trajectories is a standard technique for solving PDEs. It works by transforming a first-order PDE into a set of ordinary differential equations. In our case this technique can be applied to eq. (39), which describes the evolution of $\omega = \log \rho$ in the WKB approximation. We start by deriving both members of the equation by the binding energy ϵ:

$$\partial_t p = \partial_\epsilon \Gamma - \nu\, \partial_\epsilon p + D\, p\, \partial_\epsilon p \tag{88}$$

Where we indicate with $p(t, \epsilon) = \partial_\epsilon \omega(t, \epsilon)$. The technique consists in defining a trajectory in (ϵ, p)-space as a pair of functions of time $\hat{p}(t)$, $\hat{\epsilon}(t)$, and constrain these function to represent the solution of our equation through the relation $p(t, \hat{\epsilon}(t)) = \hat{p}(t)$. The total time derivative of this relation reads:

$$\frac{d\hat{p}}{dt} = \partial_t p + \frac{d\hat{\epsilon}}{dt}\, \partial_\epsilon p = \partial_\epsilon \Gamma + \partial_\epsilon p \left(-\nu + D\, p + \frac{d\hat{\epsilon}}{dt} \right) \tag{89}$$

By defining the value of the time derivative of $\hat{\epsilon}(t)$ so that the term in brackets is zero one is left with the following pair of coupled ODEs:

$$\begin{cases} d\hat{\epsilon}/dt = \nu - D\, \hat{p} \\ d\hat{p}/dt = \partial_\epsilon \Gamma \end{cases} \tag{90}$$

By comparing these equations with the stationary action eq. (87) one notices that the two are identical up to the identification $\hat{p} = -\widetilde{\epsilon}$. Indeed, one can identify the conjugated variable as $\widetilde{\epsilon} = -\partial_\epsilon \log \rho$, and the most-likely evolutionary trajectories will satisfy the equation:

$$\frac{d\epsilon_t}{dt} = \nu - D\, \partial_\epsilon \log \rho(t, \epsilon_t) \tag{91}$$

To find these trajectories in practice one first evaluates the evolution of $\rho(t, \epsilon)$ using a simple numerical integration scheme on eq. (37). At this point, for any given value of ϵ_f one can back-propagate the most-likely trajectory in time using eq. (91) up to the initial point ϵ_i at t_i.

3.5.3 *Action and trajectories in a simplified case: no competition and no silent mutations*

At this point the theoretical prediction for the most-likely trajectories can be compared to stochastic simulations. Before proceeding with the comparison we consider that there

are three main reasons why the stochastic simulations and the theoretical prediction could differ, linked to the way we defined our infinite-size and continuous-time limit (see discussion at the end of section 3.2.4). The first one is the fact that, as shown in the previous section, the infinite-size population limit does not capture well the stochastic evolution of the competitive selection pressure, encoded in the threshold energy $\bar{\epsilon} = -\log\langle e^{-\epsilon}\rangle_{\text{pop}}$ (cf. eq. (23)). A second reason is that the mutation kernel (cf. eq. (25)) is not gaussian but is the composition of a Dirac delta and a Gaussian. This is not well-captured by a drift-diffusion equation such as eq. (37) which works better for Gaussian-like mutation distributions. Finally the approximation of continuous time is usually accurate only if we consider a big number of evolution rounds, each one contributing very little to the total system evolution. Here instead we consider a limited number of rounds (around 30, consistent with the timescale of the population bottleneck) each one adding a non-negligible contribution to the system evolution.

Following these considerations, as a first check for our theory we set ourselves in simplified conditions and remove the first two sources of discrepancies, namely we turn off competitive selection and silent mutations. This makes the fate of cells uncoupled, and simplifies the mutation kernel eq. (25) into a pure Gaussian distribution. Under this conditions we compare the predictions of the theory with stochastic simulations.
We begin by performing 1000 stochastic simulations of the system for a number $t_f = 30$ of rounds. We pool together all cells in the final populations, and split them in different groups depending on the value of their energy (20 such bins, splitting equally the interval $[-10, 5]$). The number of cells per bin divided by the bin size and the total number of simulations gives us an estimate of the average expected cell density for a given value of the final energy ϵ_f. In fig. 34B we compare this value with the prediction $\rho(t_f, \epsilon_f)$ obtained with eq. (37). The two are in very good agreement.
We then consider the path integral formulation eq. (85). As discussed above this equation expresses the final density as a sum of contributions corresponding to all possible trajectories from any initial energy ϵ_i to the given final energy ϵ_f, with the action quantifying the log-probability of the trajectory. With a very crude approximation we can simplify this sum considering only the most important contribution, which is given by the most-likely trajectory $\epsilon_i \to \epsilon_f$, i.e. the one that satisfies the stationary action equations:

$$\log \rho(t_f, \epsilon_f) \sim S[\epsilon_i \to \epsilon_f] + \log \rho(t_i, \epsilon_i) \tag{92}$$

To test this we take the prediction for the evolution of $\rho(t, \epsilon)$ obtained above from eq. (37), and for each of the final energy bins considered above we back-propagate the most-likely trajectory leading there (i.e. to the energy ϵ_f corresponding to the middle of the bin) using eq. (91). This gives us a prediction for the most likely trajectory $\epsilon_i \to \epsilon_f$. In fig. 34A we compare the sum of the action over these trajectory $S[\epsilon_i \to \epsilon_f]$ and the initial log-density $\log \rho(t_i, \epsilon_i)$ with the value of the final log-density $\log\rho(t_f, \epsilon_f)$ evaluate as described above from stochastic simulations. The match attests that indeed the most-likely trajectory can explain most of the density observed.

Lastly, we extend the comparison to trajectories by comparing the theoretical most-likely trajectories, obtained as described above, with their counterparts evaluated from stochastic simulations. The latter are obtained by performing the average over all evo-

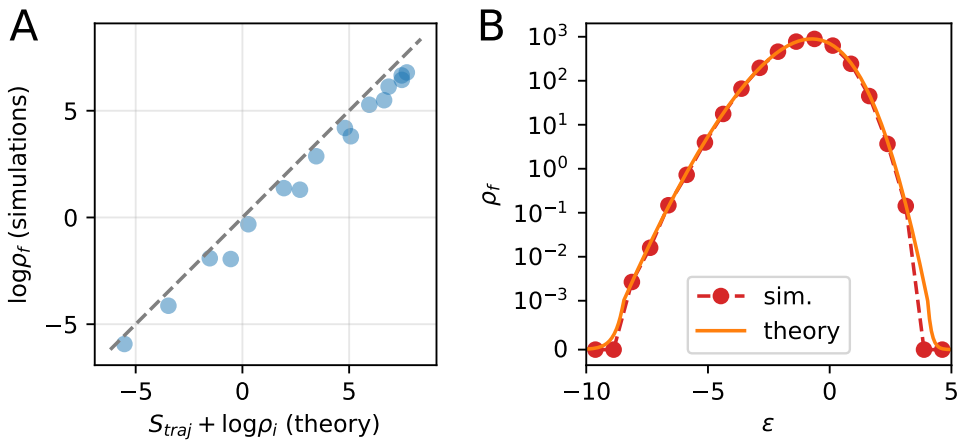

Figure 34: Analysis of trajectories probabilities and action in the simplified case of no T-cell selection and no silent mutations. With respect to the standard case we change the values of the following parameters: $p_{sil} = 0$, $p_{let} = 0.2$, $p_{aa} = 0.8$, $\mu_M = 0.1$. We consider a total evolution time of $t_f = 30$, and compare the results of 1000 stochastic simulations with our theory. **A**: comparison between the value of the action plus the logarithm of the initial cell density $S + \log \rho_i$ along the least-action trajectory as evaluated from our theory, and the logarithm of the final cell density $\log \rho_f$, derived from stochastic simulations. Each dot corresponds to a different value of the final energy ϵ_f. The fact that the result is close to the diagonal (gray dashed line) signifies that, at least in this simplified case, the theory correctly predicts the final cell density as a function of the probability of the most-likely trajectory leading to a particular final energy. **B**: comparison between the final ($t_f = 30$) cell density ρ_f, as estimated from simulations (red dots) or as predicted by the theory (orange line). The good match between the two attests that in this simplified case the theory can correctly predict the evolution of the binding energy probability distribution.

lutionary trajectories of cells in the same final energy bin. In fig. 35 we compare the two, reporting for reference on the background in shades of blue the evolution of $\rho(t, \epsilon)$, with darker shades corresponding to a higher density. The density initially undergoes a decrease in the population bottleneck phase, and then expands again. Theory and simulations are in very good agreement, both showing as expected that ancestors of the final population are located in the high-affinity tail of the initial distribution.

The good match between theory and experiments in these simplified conditions testifies that possible discrepancies must be explained by either of the two removed effects, namely competitive selection or silent mutations.

3.5.4 *Evolutionary trajectories with and without competition*

Next we move the comparison from the simplified version of the model to the standard version, complete with competitive selection and silent mutations. In fig. 36A we display the same quantities of fig. 35 in this standard case. While maintaining a qualitative agreement on the fact that trajectories must originate from the high-affinity tail of the initial distribution, here we find a less good agreement between theory and simulation. To explain this discrepancy we analyzed the evolution of the competitive selection pres-

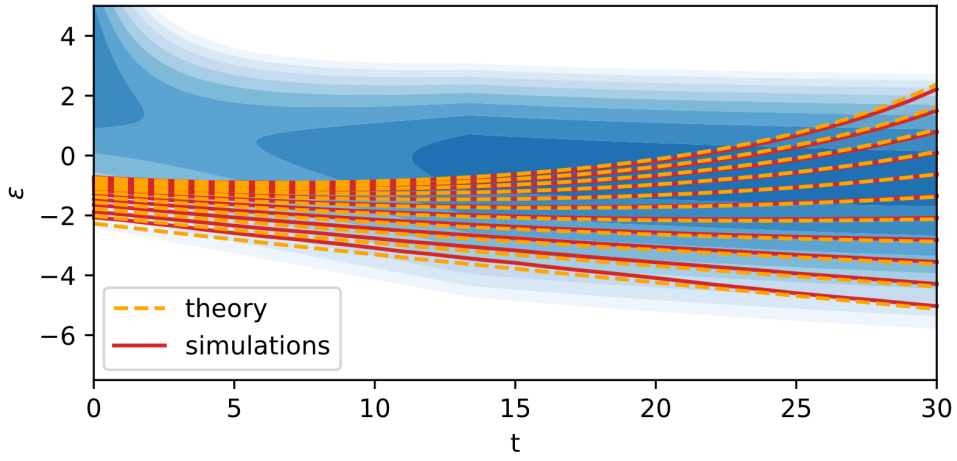

Figure 35: Average evolutionary trajectories, comparison between theory and simulations in the simplified case in which competitive selection and silent mutations are removed ($p_{sil} = 0$, $p_{let} = 0.2$, $p_{aa} = 0.8$, $\mu_M = 0.1$). 1000 different simulations were performed. For each simulation cells in the final population ($t_f = 30$) were pooled according to their binding energy, by dividing the energy interval $(-10, 5)$ in bins of width 0.25. For each pool the average evolutionary trajectory of cells was evaluated (red lines). To minimize noise only trajectories for pools grouping more than 1000 cells are displayed. Theoretical prediction are obtained by simulating the evolution of the cell density function ρ using eq. (37). This evolution is plotted for reference on the background, with darker color corresponding to higher density. Theoretical trajectories (orange dashed lines) were then evaluated by evolving the stationary action eq. (91) back in time starting from the same final condition as simulated average trajectories.

sure, quantified by the threshold energy $\bar{\epsilon}$ (cf. eq. (23)). This analysis is performed in appendix B.3. The result is that, similarly to what observed in section 3.4.3, the discrepancy can be explained by the fact that the theoretical infinte-size approximation poorly captures the evolution of this selection pressure, overestimating its evolution speed and not capturing its stochastic nature (cf. fig. 53A). This stochasticity depends strongly on initial conditions, and is greatly reduced if we start all simulations from the same initial population (cf. fig. 53B). By coupling this with an appropriate finite-size correction on the initial distribution, described in appendix B.3, we can improve the agreement between theory and stochastic simulations, also in the presence of competition.
In fig. 36B we compare stochastic trajectories (in red) obtained by performing 1000 stochastic simulations starting from the same initial population (cf. fig. 53C) to the theoretical counterparts evaluated with this finite-size cutoff correction on the initial distribution $\rho(t_i, \epsilon_i)$ (see appendix B.3). Both show that all individuals in the final population trace back to very few high-affinity ancestors in the initial population, as expected when the system is subject to strong selection pressure [105].

One interesting difference between the case of fixed or variable initial population is the fact that in the latter case the trajectories, both in stochastic simulations and in the theory, stay parallel and lead to very different initial conditions, while in the case of fixed initial population they all coalesce to the few highest-affinity individuals. This coalescence must also occur when the initial population is varied, but it is masked by the variation in energies of the initial population. This shows that the outcome of evolution

in a bottleneck depends strongly on initial conditions, and the presence of a high-affinity progenitor can lead to very different outcome. In particular for our values of model parameters very high-affinity cells in the final population are only found when a high-affinity progenitor was already present in the initial population.

3.6 CONCLUSION AND PERSPECTIVES

We conclude this chapter by summarizing the main results obtained and drafting some possible future directions.

In this chapter we focused on the study of the evolution of B-cell lineages that undergo a population bottleneck. We started by introducing a simplified version of the maturation model presented in the previous chapter, maintaining only minimal ingredients and setting the value of model parameters so as to impose a strong initial selection force to the population.
By analyzing the asymptotic limit of the model in the continuous-time description we derived explicit equations for the value of the asymptotic growth rate and maturation speed as a function of model parameters. These equations reveal a tradeoff phenomenon involving Ag concentration analogous to the one described in section 2.7, and have the advantage of explicitly encoding the effect of different model parameters in controlling the asymptotic maturation regime. Thanks to this we could also find conditions on the value of these parameters for the existence of an effective maturation regime.
We moved then to the study of the survival probability of B-cell lineages that undergo the population bottleneck. Through a recursive relation that links the probability of bottleneck survival of a cell to the one of its daughter cells we were able to retrieve the dependence of a lineage extinction probability on its progenitor affinity. For lineages that go extinct we also evaluated the mean and variance of extinction time and progeny size, revealing a peak in extinction time corresponding to average affinity progenitors. Lineages stemming from these progenitor spawn in equilibrium between extinction and survival, and persist in this state until mutations drive the lineage either to survival or to extinction. Building on these results we then evaluate the survival probability for the full population as a function of Ag concentration and population size. We also included the effect of competition in an effective manner, using the deterministic model limit combined with a finite-size correction. The bottleneck phenomenology was included in different maturation models [155, 161], which considered as optimal the maturation regime in which the population was subject to a strong enough selection force to grant good affinity enhancement, while at the same time not strong enough to cause population extinction. Part on the importance of this work is in the fact that in these models the survival probability was numerically evaluated from stochastic simulations, while here it is evaluated through our theory.
Finally we tackled the study of evolutionary trajectories of B-cell lineages in affinity space. With the combination of the path integral formalism and the method of characteristic trajectories we derived an equation for the most likely evolutionary trajectory of a lineage in affinity space, and compared it with simulations both with and without competitive selection. Our approach is somewhat complementary to genealogical trajectories, that illustrate the evolution of lineages in genetic space. A phylogenetic tree uses genetic information to trace back different individuals, usually with different phe-

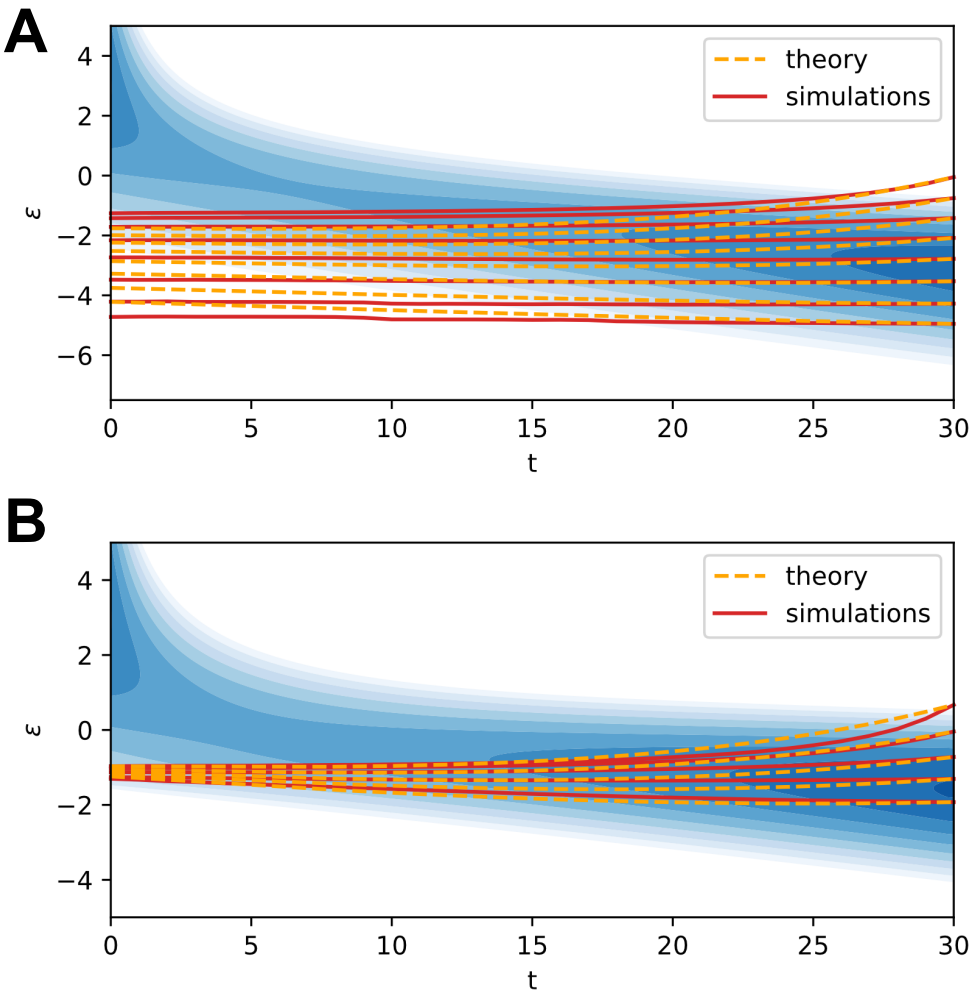

Figure 36: Average evolutionary trajectories, comparison between theory and simulations. Plots were obtained in the same way as fig. 35. **A**: standard value of model parameters, including competitive selection for T-cell help. **B**: standard value of model parameters and competitive selection, but stochastic simulations were performed using the same initial populations in all trials. Theoretical predictions are evaluated using the finite-size correction procedure described in appendix B.3. Notice how in this case all of the trajectories in the final populations can be traced back to few high-affinity ancestors in the initial population.

notypes, to a common ancestor. In our case rather than in genetic space we follow the evolution of lineages in phenotypic space, with our main variable being Ag binding energy, which for B-cells is strongly related to fitness. While the two aspects are certainly related, our trajectories connect "affinity compartments", meaning cells that have potentially different histories but similar affinity, at different times, indicating that for the most part cells that are located in that compartment must have similar phenotypic histories. It would indeed be interesting to combine genetic information with phenotypic information to see how the two are related in maturation experiments. Some modern experimental techniques are indeed capable of collecting both genetic and phenotypic information at the single-cell level [43], even though in a single experiment phenotypic information are only collected for cells at the end of the maturation and not at multiple time-intervals.

These results could be extended in different directions. For instance it would be interesting to investigate how stochasticity and permissiveness in selection could influence both survival probability and maturation trajectories. Another possible improvement to our work would be to study the relationship between our trajectory and coalescence time, quantifying how much in our simulations proximity in phenotypic space also implies genetic coalescence to a common ancestor. Finally, an important research direction would be the extension of the results obtained in this section to the case of multiple Ags. This is better discussed in section 5.3.

4

PERSPECTIVES - MICROSCOPIC B-T CELL INTERACTIONS

4.1 INTRODUCTION

In this short chapter we delineate one of the possible future directions of the work we presented, along with some preliminary results.

Through mathematical analysis of the models presented in chapter 2 and chapter 3 we showed how Ag concentration controls both population maturation speed and growth rate. This can be traced back to its role in tuning competitive selection pressure. In the models we presented this is encoded in the formula for the T-cell selection survival probability eq. (23). Here the selection threshold energy $\bar{\epsilon}$ plays the role of an average population fitness, increasing the pressure as the population evolves. Ag concentration acts as a chemical potential by applying a shift of magnitude $\log C$ to this threshold energy. This formula, used also in other maturation models [155, 154], has the advantage of effectively accounting for the effect of Ag concentration in competitive selection, while at the same time remaining sufficiently simple for analytical treatment.

The simplicity of this formula hides however the complexity of the underlying competition mechanisms. A deeper look at the selection process reveals that, as discussed in section 1.3.3, competition for T-cell help is a dynamical process, involving repeated interactions between B and T cells [7, 134]. Experiments show that the contact duration correlates with the affinity of the cell for the Ag [134]. This might be related to the fact that BCR affinity has been shown to regulate the expression of adhesion molecules [71]. Moreover, B-T cell interactions might have a co-stimulatory role, and signals from B-cells might be important to maintain the T-cell pool [152]. For example a positive feedback loop of activation signals between B and T cells involving ICOS and CD40 was proposed in [152, 75]. Moreover ICOS stimulation by non-specific bystander B-cells might be an important signal for recruiting and maintaining Tfh cells [132, 160].

From these considerations a natural question ensues: how does the observed effect of Ag concentration emerges from these repeated microscopic interactions between B and T cells? Similar questions were already tackled using mathematical models, for example the model introduced in [69] showed how repeated interactions between B and T cells can result in improved affinity discrimination. With the aim of trying to bridge the gap between the observed effect of Ag dosage and the microscopic cellular interactions in GCs, in this section we introduce a microscopic B-T cell interaction mechanism that we integrate in the model presented in chapter 3. With the hypothesis that the magnitude of the T-cell response depends both on the dose of administered Ag and on co-stimulatory effects delivered by B-cells, the model shows indeed different asymptotic maturation regimes in which Ag concentration controls both maturation speed and growth rate of the population. At the same time, the model remains amenable to mathematical analysis.

4.2.1 *Microscopic mechanism of B-T cell interaction*

We consider a population of N_T T-cells and N_B B-cells which interact through repeated and short-lived contacts [7, 134]. In particular in our model an unbound T-cell can bind a free B-cell with uniform probability at a rate ω_b, which is the same for all B-cells. To reproduce the experimental observation that contacts duration correlates with the amount of Ag displayed on the B-cell, and thus indirectly with its affinity [134], we assume the unbinding rate to be equal to $\omega_u\, e^{\epsilon}$, where again ϵ is the binding energy between the Ag and the BCR. For simplicity we only consider one-to-one contacts. In fig. 37 we provide a schematic depiction of this setup. We call N_T^{free} the number of T-cells which are not bound to a B-cell. Given these definitions we can write a master equation for the probability $p(t, \epsilon)$ of a B-cell with binding energy ϵ to be bound to a T-cell at time t:

$$\frac{dp}{dt}(t, \epsilon) = (1 - p(t, \epsilon))\, \omega_b\, N_T^{\text{free}}(t) - p(t, \epsilon)\, \omega_u\, e^{\epsilon} \tag{93}$$

The conservation of the total number of T-cells gives a further condition:

$$N_T - N_T^{\text{free}}(t) = N_B \int d\epsilon\, \varphi(\epsilon)\, p(t, \epsilon) \tag{94}$$

Where $\varphi(\epsilon)$ is the normalized binding energy distribution of the B-cell population. Differentiating both terms w.r.t. time gives us an equation for the evolution of the number of free T-cells:

$$\frac{dN_T^{\text{free}}}{dt}(t) = -N_T^{\text{free}}(t)\, \omega_b\, N_B \int d\epsilon\, \varphi(\epsilon)\, (1 - p(t, \epsilon)) + \omega_u\, N_B \int d\epsilon\, \varphi(\epsilon)\, p(t, \epsilon)\, e^{\epsilon} \tag{95}$$

Experiments [134] show that the typical time for binding and unbinding events is of the order of minutes. On longer timescale, of the order of the GC cycling time (around 12 hours) we can consider this binding probability to be approximately at equilibrium. Equilibrium convergence times are discussed in appendix C.1, which contains a simple perturbative analysis of the master equation. The equilibrium value of the probability p_{eq} can easily be recovered from eq. (93):

$$p_{eq}(\epsilon) = \frac{e^{-\epsilon}}{e^{-\epsilon} + e^{-\bar{\epsilon}}} \quad \text{with} \quad e^{\bar{\epsilon}} = \frac{\omega_b}{\omega_u} N_T^{\text{free}} \tag{96}$$

By expressing N_T^{free} as a function of $\bar{\epsilon}$ and substituting in eq. (94) we obtain an equation for the equilibrium value of $\bar{\epsilon}$:

$$N_T - \frac{\omega_u}{\omega_b} e^{\bar{\epsilon}} = N_B \int d\epsilon\, \varphi(\epsilon)\, \frac{e^{-\epsilon}}{e^{-\epsilon} + e^{-\bar{\epsilon}}} \tag{97}$$

We embed this interaction mechanism in the deterministic version of the maturation model presented in section 3.2.3 by assuming that the probability for a B-cell to successfully survive selection is proportional to its equilibrium probability of being involved in a contact with a T-cell. In practice we keep the definitions of all the evolution operators

obtained in the deterministic limit of big population size, as defined in section 3.2.3,[1] but we substitute the probability of surviving competitive selection for T-cell help (cf. eq. (28)) with $P_T = p_{eq}(\epsilon)$. Competition in this case is included in eq. (97), which contains the integral over the population binding energy distribution $\varphi(\epsilon)$.[2] This equation also encodes the dependence of $\bar{\epsilon}$ on the other parameters. On the l.h.s. of the equation we find the total number of T-cells minus the ones that are free, and on the r.h.s. the number of B cells that are in a contact. If we divide both sides of the equation by N_B both side will display the fraction of B-cells that is currently undergoing binding:

$$\frac{N_T}{N_B} - \frac{\omega_u}{\omega_b}\frac{e^{\bar{\epsilon}}}{N_B} = \int d\epsilon\, \varphi(\epsilon)\, \frac{e^{-\epsilon}}{e^{-\epsilon} + e^{-\bar{\epsilon}}} \tag{98}$$

At a fixed population size N_B, the main parameter controlling the value of $\bar{\epsilon}$ is the number of T-cells N_T. Rather than a true number of cells it should be interpreted as the total amount of help signals available. To provide some intuition in fig. 38A we plot the r.h.s and l.h.s. of the equation as a function of $\bar{\epsilon}$. The r.h.s. is a sigmoid function that interpolates between 0 and 1, its flex point being at values of $\bar{\epsilon}$ close to the mean of the distribution $\varphi(\epsilon)$. The l.h.s is a negative exponential which converges to N_T/N_B for small values of $\bar{\epsilon}$. The equilibrium value of $\bar{\epsilon}$ is the one at which the two curves intersect. Its y-axis value is the fraction of B-cells currently involved in binding, and under our assumptions also the fraction surviving selection at every round. By increasing the amount of T-cells available the value of $\bar{\epsilon}$ increases, and so does the number of cells that survive selection. In this sense $\bar{\epsilon}$ controls the selection pressure, as evident from the similarity between the current expression for the survival probability eq. (96) and the original form of the T-cell survival probability in our model eq. (23).
A second quantity controlling the value of $\bar{\epsilon}$ is the ratio between the binding and unbinding rates. By moving these terms to the exponent one notices that changing the ratio of the binding and unbinding rates is equivalent to applying an horizontal shift to the orange curve equal to $-\log \omega_u/\omega_b$. These rates however should both have critical times of the order of minutes, and the logarithm of their ratio should be small and not impact much the solution, compared for example to the term $\log N_B$ which imparts the same translation and can vary of orders of magnitude. For simplicity in numerical simulations we set $\omega_u/\omega_b = 1$.

The last model specification concerns the determination of the parameter N_T, which as discussed above quantifies the magnitude of the total T-cell help potentially available to cells. The effect of Ag concentration C can be encoded in this term, in particular one could suppose that a higher amount of Ag available for B cells to bind and display would entail also a higher total number of survival signals. Moreover, as discussed in section 1.3.3, non-Ag-specific interactions with bystander B cells might be important for Tfh cell recruitment and maintenance [132, 160]. B cells might constitute a field in which T cells are immersed and that continually provide signals that supports the Tfh phenotype. This might make the total number of survival signals available depend on

1 We also keep the same value for all model parameters, as reported in table 4.
2 This quantity relates to the population binding energy density function $\rho(\epsilon)$, employed in the previous chapter, as $\rho(\epsilon) = N_B\, \varphi(\epsilon)$.

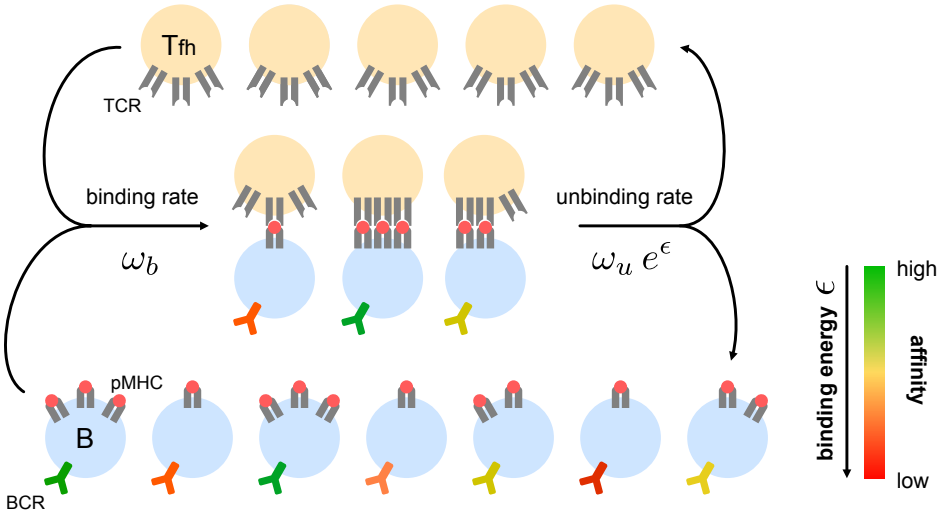

Figure 37: schematic representation of the microscopic interaction model between B and T cells. B-cells (blue) collect Ag in an amount that depends on the affinity of their BCR for the Ag (see colorbar on the right). Ag is then internalized, processed and displayed in the form of peptide-MHC complexes (pMHC). Tfh cells (orange) that are not involved in binding can randomly encounter a free B-cell at a constant rate ω_b forming a contact in which TCRs on the Tfh cell surface bind pMHC molecules on the B cell. We reproduce the experimental observation that the contact duration depends on the affinity of the B-cell by setting the unbinding rate to $\omega_u\, e^{\epsilon}$, where ϵ is the cell BCR binding energy. In this way cells with higher affinity have longer contact duration.

the number of B cells present. To explore different scenarios we consider three different possibilities:

$$N_T = \alpha_T\, C\, f(N_B)\,, \qquad \begin{cases} \text{case } \mathcal{A} \text{ (independent)}: & f(N_B) = 1 \\ \text{case } \mathcal{B} \text{ (linear)}: & f(N_B) = \frac{N_B}{N_B^0} \\ \text{case } \mathcal{C} \text{ (mixed)}: & f(N_B) = \frac{N_B}{N_B + N_B^0} \end{cases} \tag{99}$$

In all three possibilities we consider the number of T cells to be proportional to the Ag concentration C through a proportionality constant α_T. This proportionality constant can be interpreted as the typical number of T cells at Ag concentration $C = 1$. We choose $\alpha_T = 1000$. The difference between the three cases is the dependence of N_T on the number of B cells N_B. In case $\mathcal{A}$ we do not include any such dependence. Case $\mathcal{B}$ features instead a linear dependence, so that the number of T cells is also proportional to the number of B cells. For good comparison with the other cases we normalize this number by N_B^0, which represents a typical B cell population size. We take $N_B^0 = 1000$. This case however could lead potentially to an infinite amount of survival signals, if the B cell population continues to expand. In reality this increase must eventually stop, therefore in case $\mathcal{C}$ we interpolate between the two previous cases, considering a function that increases almost linearly for small values of N_B and saturates for big values, with the transition between the two regimes occurring at $N_B \sim N_0^B = 1000$. In fig. 38B we compare the dependence of N_T on the number of B cells in the three different cases.

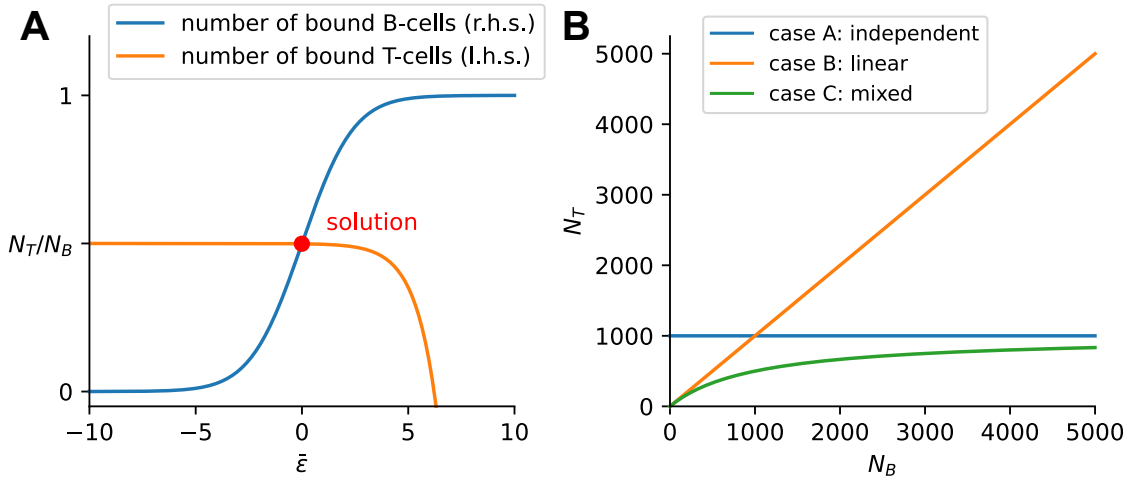

Figure 38: **A**: graphical solution of eq. (98). We plot both sides of the equation (r.h.s. in blue and l.h.s. in orange) as a function of $\bar{\epsilon}$. The value of $\bar{\epsilon}$ that solves the equation is the one at the intersection between the curves. This was obtained with $N_B = 2N_T$, $\omega_u/\omega_b = 1$ and considering the binding energy distribution $\varphi(\epsilon)$ to be a Gaussian with zero mean and unit variance. **B**: dependence of the number of T cells N_T on the number of B cells N_B in the three cases described in the text.

4.2.2 *Maturation in the independent case*

We start by studying maturation in case $\mathcal{A}$, in which the magnitude of T-cell help available only depends on Ag concentration $N_T = \alpha_T C$. We aim here at performing an analysis similar to the one presented in section 3.3, understanding the way in which Ag concentration affects the population growth rate and maturation speed. Similarly to what done there, in order to observe asymptotic growth and maturation we therefore neglect the carrying capacity constraint and Ag-binding selection, only considering duplication, mutations and competitive selection for T-cell help.

In fig. 39 we report the evolution of the system for different values of the Ag concentration C (see color-scale on the lower part). In panels A and B we display the evolution of respectively the population size and average binding energy. Remarkably, irrespective of the value of C, the population quickly converges to a fixed size, and matures at the same rate. The final size depends linearly on the Ag concentration.
The key to understand this behavior resides in the value of $\bar{\epsilon}$ and its displacement from the mean of the distribution $\langle\epsilon\rangle$. It is easy to argument that the value of $\bar{\epsilon}$ must be close to the population mean. In fact, if the value were much higher than the mean $\bar{\epsilon} \gg \langle\epsilon\rangle$ (equivalent to a low selection pressure regime) then the integral on the r.h.s. of eq. (97) would approximate to 1:

$$\bar{\epsilon} \gg \langle\epsilon\rangle \quad \Longrightarrow \quad N_T - \frac{\omega_u}{\omega_b} e^{\bar{\epsilon}} \sim N_B \tag{100}$$

From which, neglecting the binding rates, we have $\bar{\epsilon} \sim \log(N_T - N_B)$. At this point we find an inconsistency: the low selection pressure that we assumed would generate an exponential increase of the population size, until eventually $N_B \sim N_T$, which invalidates our hypothesis. This can also be visualized from the graphical solution fig. 38A. Con-

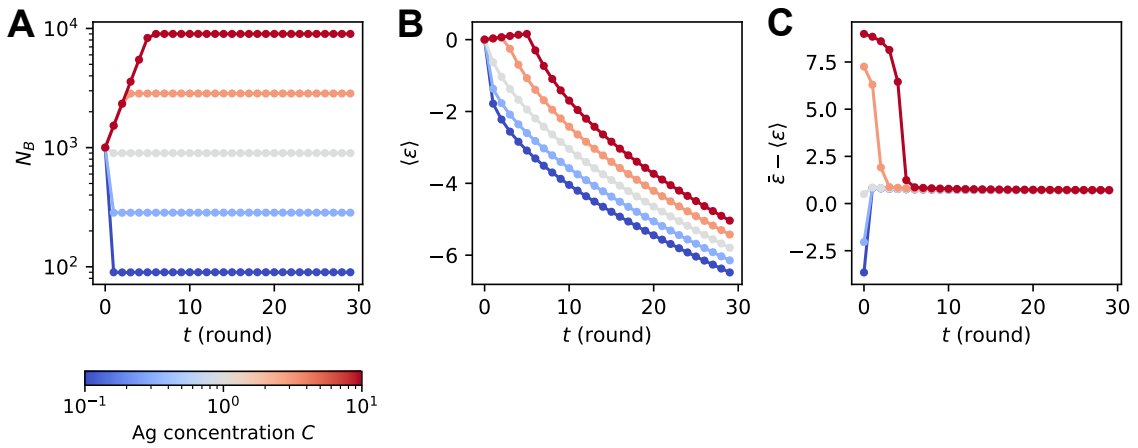

Figure 39: Population size (**A**), average binding energy (**B**) and displacement of $\bar{\epsilon}$ from the average binding energy (**C**) as a function of the evolution round, and for different values of the Ag concentration C according to the colorscale on the bottom. Simulations are performed under model variant $\mathcal{A}$.

sider that the sigmoid transition is in proximity of the distribution mean. Having a low selection pressure $\bar{\epsilon} \gg \langle\epsilon\rangle$ is therefore only possible if $N_T > N_B$. However the exponential increase of the population in the absence of selection pressure eventually invalidates this assumption.
In a similar way, the selection pressure cannot be too high, or equivalently $\bar{\epsilon} \ll \langle\epsilon\rangle$. In this case in fact the r.h.s. of eq. (97) approximates to zero.[3] At the same time the strong selection pressure also implies $\bar{\epsilon} \ll 0$,[4] which also removes the second term on the l.h.s. of the equation. This leaves us only with the term N_T which by definition is non-infinitesimal, which again invalidates the assumption. More simply, the same intuition can be obtained from the graphical solution. The case $\bar{\epsilon} \ll \langle\epsilon\rangle$ can be realized only if $N_B \gg N_T$. However this is not compatible with the fact that the strong selection pressure will exponentially decrease the population size, eventually invalidating the assumption.

From these considerations we deduce that the value of $\bar{\epsilon}$ must be close to the mean of the population binding energy distribution $\bar{\epsilon} \sim \langle\epsilon\rangle$ (i.e. their difference must be comparable to the distribution standard deviation). This imparts a sufficient selection pressure to decrease the population average binding energy, and with it the value of $\bar{\epsilon}$. This will eventually make the second term on the l.h.s. of eq. (97) negligible, reducing it to:

$$\alpha_T\, C \sim N_B \int d\epsilon\, \varphi(\epsilon)\, \frac{e^{-\epsilon}}{e^{-\epsilon} + e^{-\bar{\epsilon}}} \tag{101}$$

From the definition of T-cell selection the term on the r.h.s. of this equation equals the number of B cells that successfully survive selection at every round. This number is proportional to the Ag concentration C. This both explains the fixed size observed in simulations and the fact that this size is proportional to Ag concentration. In particular,

3 More precisely, if we take $\varphi(\epsilon)$ to be a Gaussian with mean μ and variance σ^2 then the r.h.s. can be approximated with $N_B \exp\{\bar{\epsilon} - \mu + \sigma^2/2\}$.

4 This derives simply from the assumption $\bar{\epsilon} \ll \langle\epsilon\rangle$ and the consideration that the strong selection pressure will progressively lower the distribution binding energy.

since after T-cell help selection we perform differentiation, the asymptotic population size is equal to $N_B^* = \alpha_T\, C\, (1 - p_{diff})$.
In practice when setting $N_T = \alpha_T\, C$ our mechanisms asymptotically behaves as a simple population size control, tuning the selection pressure so as to allow for the survival of the same fixed number of B-cells. This pressure collaterally also generates the population maturation. The fact that all the evolution operators are invariant under population size rescaling makes so that changing the concentration will impact the asymptotic size of the population, but not the maturation speed or the asymptotic form of $\varphi(\epsilon)$.
As stated before, the maturation speed is a consequence of the pressure imparted by the population constraint. Using this knowledge the maturation rate can be derived using an eigenvalue equation similar in spirit to eq. (21), but using the operators defined in section 3.2.3. The idea behind the equation is that asymptotically the binding energy density function $\rho(\epsilon)$ (equal to the product between the binding energy distribution $\varphi(\epsilon)$ and the population size N_B) must evolve as a traveling wave under the action of the evolution operator. This ansatz reads in general:

$$e^{\phi}\, \rho = \mathbf{T}(-\Delta)\, \mathbf{D}\, \mathbf{S}_T\, \mathbf{M}\, \mathbf{A}\, \rho \tag{102}$$

Where as for eq. (21) $\mathbf{T}(x)$ represents the operator that imparts a translation of magnitude x to the density function. As stated above, in this case the T-cell help selection operator acts by multiplying the density function by the survival probability eq. (96). All the operators are invariant under translation, adding a translational degree of freedom to the determination of the egigenfunction $\rho(\epsilon)$. We fix this degree of freedom by imposing $\bar{\epsilon} = 0$ in eq. (96). The values of Δ and ϕ represent respectively the asymptotic shift and growth rate of the population in a single evolution round. In fig. 40A we report the relationship between these quantities obtained by a numerical solution of eq. (102) for different values of the shift Δ. As stated above, in case $\mathcal{A}$ the microscopic selection mechanism asymptotically acts as a population control, fixing the population size to a finite value N_B^*. As a consequence, the only acceptable asymptotic value of the growth rate is $\phi = 0$. This collaterally fixes the maturation speed to the corresponding value Δ^*. To verify this result in fig. 40B we compare the instantaneous maturation speed of model simulations performed at different value of the Ag concentration C (see colorscale) with the theoretical asymptotic prediction. For all values of C the maturation speed converges to its theoretical expectation.

In conclusion, in model variant $\mathcal{A}$ (i.e. when the amount of available T-cell help is independent from the population size) we observed how including this selection mechanism in our model fixes the asymptotic population size to a value N_B^*, proportional to the Ag concentration. If one were to include in the model a dynamics for Ag concentration then the population would eventually go extinct as a consequence of Ag depletion. However in this case the value of C does not impact the maturation speed, that is locked to a constant value Δ^* by the fixed population size constraint.

4.2.3 *Maturation in the linear case*

In model variant $\mathcal{B}$ we add a linear dependence between the number of available T cell help signals and the size of the B-cell population $N_T = \alpha_T\, C\, N_B / N_B^0$. This modification changes the phenomenology of the model. In fig. 41 we plot the evolution of the sys-

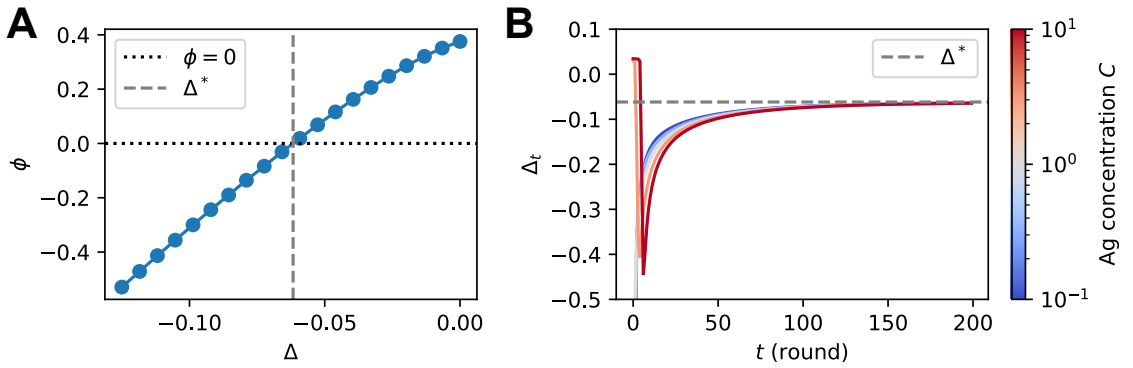

Figure 40: **A**: relationship between asymptotic growth rate ϕ and maturation speed Δ obtained from the numerical solution of eigenvalue eq. (102) for different values of the shift Δ. In model variant $\mathcal{A}$ the microscopic selection mechanism imposes a fixed population size, locking the value of the growth rate to $\phi = 0$. This also locks the value of the maturation speed to Δ^*. **B**: evolution of the instantaneous maturation speed, defined as the displacement of the average binding energy of the population $\Delta_t = \langle\epsilon\rangle_t - \langle\epsilon\rangle_{t-1}$, for simulations performed at different values of Ag concentration C (see colorscale on the right). For all values of C the maturation speed converges to the value Δ^* predicted by the theory.

tem for simulations performed at different values of the Ag concentration C. As in the previous case, to observe asymptotic growth and maturation we remove the carrying capacity constraint and Ag-binding selection. Looking at panels A and B, which represent the evolution of the population size and average binding energy, we distinguish three kinds of behaviors. At high values of the Ag concentration all simulations follow the same trajectory, with positive growth rate and constant drift in the positive direction of the energy axis, i.e. a constant decrease affinity. At intermediate values of C we find a constant growth rate (positive or negative) and a constant drift in energy space. Finally, for small values of C, after an initial transition phase the population undergoes a super-exponential decrease in size.

Also in this case the different regimes can be understood by looking at the position of the threshold selection energy $\bar{\epsilon}$ relative to the mean of the binding energy distribution $\langle\epsilon\rangle$, reported in fig. 41C. To understand the high concentration regime we set ourselves in the low-selection-pressure regime $\bar{\epsilon} \gg \langle\epsilon\rangle$, in which the value of $\bar{\epsilon}$ can be approximated as in eq. (100). By substituting here the expression for the number of T cells we obtain:

$$\bar{\epsilon} \sim \log\{N_B(1 - \alpha'_T C)\} \tag{103}$$

For simplicity of notation we absorbed the term N_B^0 in the constant α_T, writing $\alpha'_T = \alpha_T / N_B^0$. From this expression we identify a critical concentration $C^{\dagger\dagger}$ such that $\alpha'_T C^{\dagger\dagger} = 1$. For values of concentration $C > C^{\dagger\dagger}$ this approximation is consistent, since in the weak selection regime the population will increase exponentially at a rate $\phi^{max} = \log[2(1 - p_{let})(1 - p_{diff})]$ and drift under the effect of mutations at a speed $\Delta^{max} = \mu_M p_{aa}/(p_{aa} + p_{sil}) < \phi^{max}$.[5] This would increment over time the relative distance between the selection energy $\bar{\epsilon}$ and the average binding energy of the population $\langle\epsilon\rangle$, at a

5 For the standard values of model parameters reported in table 4 it is $\phi^{max} \sim 0.42$ and $\Delta^{max} \sim 3.5 \times 10^{-2}$. These are the population growth rate and drift velocity in absence of selection.

constant rate independent of the Ag concentration C, which in turn would make competitive selection more and more weak. These observations are consistent with what observed in simulations (cf. fig. 42).
At intermediate values of Ag concentration the population asymptotically evolves according to the traveling-wave ansatz, with the population size increasing at an exponential rate ϕ and the average binding energy translating with velocity Δ. In this regime the threshold binding energy for competition is close to the mean of the population $\bar{\epsilon} \sim \langle \epsilon \rangle$, and the integral on the r.h.s of eq. (98) takes non-infinitesimal values in the interval $[0, 1]$. Under model variant B this equation reads:

$$\alpha'_T C - \frac{\omega_u}{\omega_b} \exp\{\bar{\epsilon} - \log N_B\} = \int d\epsilon \, \varphi(\epsilon) \frac{e^{-\epsilon}}{e^{-\epsilon} + e^{-\bar{\epsilon}}} \tag{104}$$

Notice that under these hypothesis the term $\bar{\epsilon} - \log N_B$ must decrease at every round by a quantity $\Delta - \phi$. If $\Delta - \phi < 0$ then this term becomes asymptotically negligible, and the previous equation rewrites into eq. (101) (with the substitution $\alpha_T \to \alpha'_T$). Similarly to what done for this equation one can numerically derive the relation between Δ and ϕ by solving the eigenvalue eq. (102) for different values of the shift. The eigenfunctions can then be used, in combination with eq. (101), to derive the corresponding value of Ag concentration, de facto obtaining the functions $\Delta(C)$ and $\phi(C)$. These functions allows us to derive a second critical concentration $C^\dagger$, which is defined by $\Delta(C^\dagger) = \phi(C^\dagger)$. For values of concentration $C^\dagger < C < C^{\dagger\dagger}$ the system asymptotically evolves as a traveling wave. Values of $C < C^\dagger$ instead result in $\Delta(C) > \phi(C)$, invalidating our hypothesis and leading to a different evolution regime.
To understand this third low-concentration regime we set ourselves in the hypothesis of high selection pressure $\bar{\epsilon} \ll \langle \epsilon \rangle$. In this case, if we approximate the population binding energy distribution $\varphi(\epsilon)$ with a Gaussian having mean μ and standard deviation σ then the integral on the r.h.s. of eq. (104) can be approximated to:

$$\alpha'_T C - \frac{\omega_u}{\omega_b} \exp\{\bar{\epsilon} - \log N_B\} \sim \exp\{\bar{\epsilon} - \mu + \sigma^2/2\} \tag{105}$$

In this case the term on the r.h.s can be neglected by hypothesis. Since the first term on the l.h.s. is non-infinitesimal at the leading order it must be $\bar{\epsilon} \sim \log N_B$. Moreover in the high-selection-pressure regime a round of selection causes a population decrease of the order $\Delta \log N_B = \bar{\epsilon} - \langle \epsilon \rangle$. As a consequence the variation in the value of $\bar{\epsilon}$ between two evolution rounds must be of the order $\Delta\bar{\epsilon}_t \sim \bar{\epsilon}_{t-1} - \langle \epsilon \rangle_{t-1}$. This mechanism is at the base of the super-exponential decrease in the population size, and non-linear decrease of $\bar{\epsilon} - \langle \epsilon \rangle$, for values of Ag concentration $C < C^\dagger$.

In fig. 42 we recapitulate the previous considerations by drawing a phase diagram for the behavior of the population asymptotic growth rate ϕ and maturation velocity Δ as a function of Ag concentration C. This diagram explains the behavior observed in simulations. The two critical concentrations $C^\dagger$ and $C^{\dagger\dagger}$ separates three different evolution regimes in which the population, in order of increasing concentration, undergoes super-exponential decrease, evolves according to the traveling-wave ansatz, or diffuses uniquely under the effect of mutations. In the first case the values of Δ and ϕ are not well-defined. From this diagram two additional critical concentration values can be identified. The first, defined by the relation $\phi(C^*) = 0$, divides concentration intervals in which the

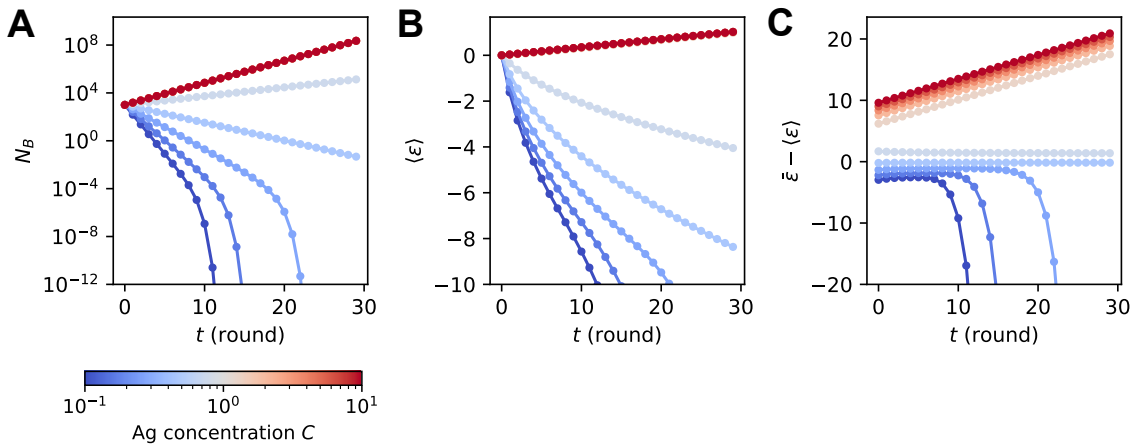

Figure 41: Population size (**A**), average binding energy (**B**) and displacement of $\bar{\epsilon}$ from the average binding energy (**C**) as a function of the evolution round, and for different values of the Ag concentration C according to the colorscale on the bottom. Simulations are performed under model variant $\mathcal{B}$.

population is either increasing or decreasing, separating the regime of extinction and maturation. The second, defined by $\Delta(C^{**}) = 0$, separates the regime of maturation, in which selection pressure is high enough to grant affinity increase, to the regime of stagnation and affinity decrease, in which the system evolves mainly under the effect of mutations. The concentrations C^* and C^{**} plays the same role of the analogous critical concentrations described in section 2.7, separating the behavior of the model in three difference regimes of extinction, maturation or stagnation. The difference with what presented in previous chapters lays in the fact that in this case this behavior arises naturally as consequence of our microscopic selection model, rather than by making the selection pressure explicitly depend on Ag concentration (cf. eq. (7)).

As a last observation we point out that if one were to re-include Ag-binding selection, adding a fixed-threshold selection pressure, then asymptotic positive values of the shift Δ would be impossible. This would set $\Delta(C) = 0$ and $\phi(C) = \phi(C^{**})$ for values $C > C^{**}$.

4.2.4 *Maturation in the mixed case*

Model variant $\mathcal{C}$ is an interpolation between the two variants previously described. The value of N_T is linearly dependent on the number of B-cells when the population size is small, but saturates to a constant value as the size of the population increases. As a consequence also the behaviors we observe in simulations can be interpreted with what previously observed. In fig. 43 we report the evolution of simulations, performed in the absence of Ag-binding selection and carrying capacity, at different values of Ag concentration C. For values of $C < C^*$ simulations behave qualitatively as in model variant $\mathcal{B}$, with a super-exponential decrease in population size at low concentration $C < C^\dagger$, and a traveling-wave behavior for $C^\dagger < C < C^*$, with Ag concentration influencing growth rate and maturation speed. Conversely, for values of $C > C^\dagger$ the population size saturates and maturation proceeds at a constant velocity $\Delta(C^*)$. These behaviors can be easily understood by the fact that when the population is exponentially decreasing then

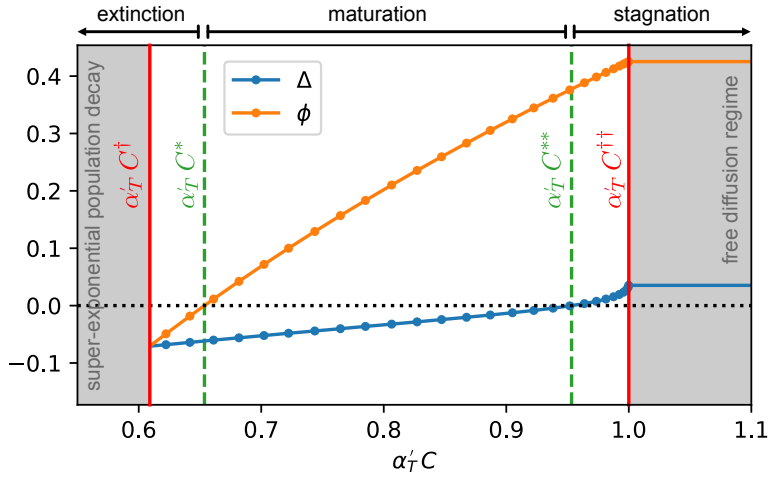

Figure 42: phase-diagram of the asymptotic maturation regimes as a function of the rescaled Ag concentration $\alpha'_T C$. The diagram reports the values of the asymptotic population growth rate ϕ, and shift in average binding energy Δ for an evolution round. For values of $C < C^\dagger$ these values are not well-defined, since the population undergoes super-exponential decrease. For values $C^\dagger < C < C^{\dagger\dagger}$ the population binding energy distribution evolves as a traveling-wave, and the function $\Delta(C)$ and $\phi(C)$ are numerically obtained by solving the eigenvalue eq. (102). Finally, for values $C > C^{\dagger\dagger}$ the population asymptotically evolves uniquely under the action of mutations, amplification and differentiation, which effectively constitute a drift and diffusion process, with drift velocity Δ^{max} and growth rate ϕ^{max} which do not depend on Ag concentration. The values at which $\phi(C^*) = 0$ and $\Delta(C^{**}) = 0$ identify two critical concentrations that separates regimes of extinction (the population size is quickly decreasing), maturation (population size increase and affinity increase), and stagnation (population size increase but affinity decrease).

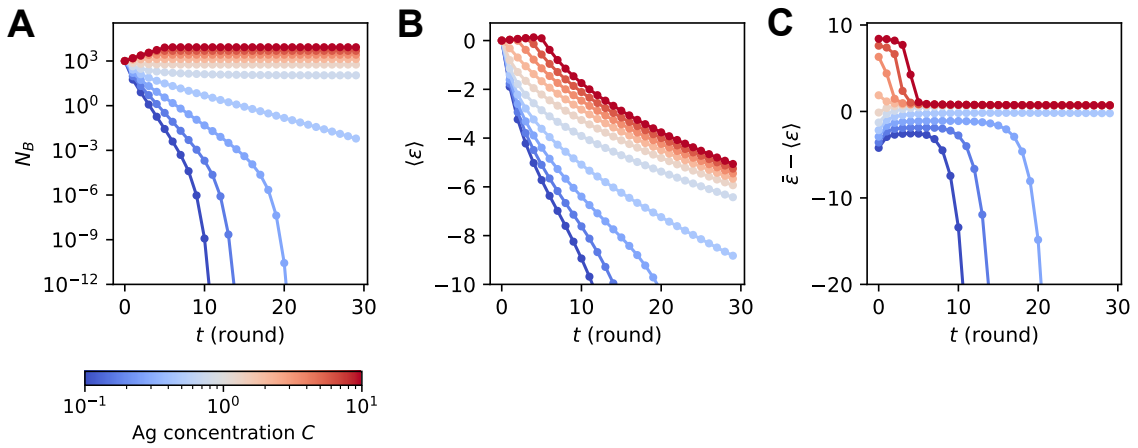

Figure 43: Population size (**A**), average binding energy (**B**) and displacement of $\bar{\epsilon}$ from the average binding energy (**C**) as a function of the evolution round, and for different values of the Ag concentration C according to the colorscale on the bottom. Simulations are performed under model variant $\mathcal{C}$.

N_T effectively becomes a linear function of N_B, as in model variant $\mathcal{B}$. When the concentration is high enough to support population survival instead the value of N_T will not increase indefinitely but rather saturate to a finite value, at which point the model behavior is analogous to the one of model variant $\mathcal{A}$.

The merit of this last variant is that our microscopic selection mechanism both implements a dependence of the growth rate and maturation speed on Ag concentration, while at the same time providing a mechanism for population size control that removes the need for explicitly adding a finite carrying capacity to the model. The presence of these effects is explained in our case by the fact that for small population size the availability of survival signals increases with population size (a fact supported by the hypothesis of a co-stimulatory role of B-T cell signals [152]), but this availability eventually saturates at big population size.

4.3 Conclusion and perspectives

In this short chapter we introduced a microscopic model for B-T cell interaction and integrated it with the maturation model presented in the previous chapter. We did this with the aim of understanding how the effect that Ag concentration plays in controlling maturation, namely influencing maturation speed and population growth rate, can emerge from the microscopic interactions between the cellular agents of the GC. In particular we found that a feedback loop by which the number of available Tfh help signals is proportional to the B-cell population is sufficient to explain this effect. This hypothesis is justified by experimental observations on how stimulations by non-cognate B-cells help maintain the T-cell pool [132, 160]. Moreover this same mechanism also removes the need for an artificial enforcement of the carrying capacity if one assumes that the total number of survival signals available eventually saturates at a maximum. In addition to this the model remains analytically tractable, allowing us to draw an asymptotic phase diagram.

These preliminary results could be continued in different directions. For example it would be interesting to study stochastic effects, that when considering microscopic interactions between cellular agents might play a major role. With some model modifications they might account for the experimental observation of selection permissiveness, that was otherwise artificially introduced in the model we employed in chapter 2. In addition to this one could study the model evolution when the hypothesis of equilibrium in the binding-unbinding process (cf. eq. (93)) is dropped.
This bridging of different scales, from microscopic interaction between cells to macroscopic effects on population maturation, is in our opinion an important research area, that could assist experiments in improving our understanding of AM mechanisms. One interesting direction in this sense might be to study how the ICOS/CD40 signaling feedback mechanism described in [152, 76] (see section 1.3.3 and fig. 6) would affect affinity discrimination and maturation if included in a model similar to ours.

5

PERSPECTIVES - BREADTH ACQUISITION IN SINGLE AG IMMUNIZATION

5.1 INTRODUCTION

In this last short chapter we present some preliminary results concerning the study of breadth acquisition, pathogen mutability and Ag concentration.

As discussed in section 1.3.5, one of the main frontiers in vaccination research is represented by developing immunization techniques capable of conferring protection against highly-mutable pathogens. Acquiring such protection is made difficult by the fact that when facing these pathogens the IS naturally tends to focus the Ab response on easily accessible but mutable residues. The pathogen can then easily escape recognition by developing mutations on these residues. Acquiring protection is nonetheless possible, as testified by the existence of *Broadly-Neutralizing Abs* (BNAbs). These Abs target less-accessible but conserved part of the pathogen, and have therefore high affinity for many different pathogen variants.
For HIV infections in rare cases BNAbs are also naturally developed by patients some years after infection. At this point the viral population features a wide diversity of mutants. This suggests that the presence of many Ag mutated variants must also be an essential component of vaccines directed at artificially eliciting these Abs.

Starting from these considerations we study the effect of single Ag immunization on breadth acquisition. In particular we consider a B-cell population that undergoes maturation to bind the *wild-type* Ag. At the same time we keep track of the population affinity for a *mutant* Ag, that is not administered. Even though the population is not directly selected for binding this mutant, the natural correlation of the binding energies of Abs for these two Ag variants allow the acquisition of some affinity for the mutant through selection for the wild-type. The pathogen mutability, which can be interpreted as a measure of how conserved the targeted epitope is between variants, is a parameter of the model and it controls the magnitude of this correlation. It also controls how correlated the effect of affinity-affecting mutations is on the binding energy of an Ab for the two Ag variants. This is because mutations on residues that bind conserved parts of the epitopes are expected to have similar effects, while for variable residues they could have uncorrelated effects.
We study the effect of Ag concentration and pathogen mutability on the acquisition of affinity for the mutant, and find that low Ag concentration can help breadth acquisition by increasing selection pressure. Our theory also indicates the existence of a critical level of mutability above which wild-type immunization is insufficient for breadth acquisition. In this cases additional selection pressure is necessary, which could be obtained by the administration of multiple Ag variants.

5.2.1 *Model extension to multiple antigens*

Let us consider a population of B-cells and two different Ags, which we will label Ag1 and Ag2. Each B-cell will have a different binding energy for the two Ag variants ϵ_1 and ϵ_2. We can consider these two values to be random variables that take different values for each B-cell. Given the similarity between the two Ag variants these two random variables will be correlated. In particular we consider that they can be decomposed in two parts, relative to the binding of the conserved and variable part of the Ag:

$$\epsilon_1 \sim \epsilon_c + \epsilon_{v1}, \quad \epsilon_2 \sim \epsilon_c + \epsilon_{v2} \tag{106}$$

Where the binding energy for the conserved part ϵ_c is the same for the two Ag variants, while the energy for the variable part $\epsilon_{v1/v2}$ is different. As a first approximation we can suppose these three random variables to be independent. Moreover since we are not interested in energy values but only in energy differences after a round of maturation for simplicity we set the means of all three random variables to zero. The important quantities in this case are the variances σ_c^2, σ_{v1}^2, σ_{v2}^2. These variances are a proxy for the energy difference between good and bad binders for each of the regions in question. If we suppose $\sigma_{v1} = \sigma_{v2} = \sigma_v$ then the correlation matrix between ϵ_1 and ϵ_2 has elements:

$$\mathrm{Var}[\epsilon_1] = \mathrm{Var}[\epsilon_2] = \sigma_c^2 + \sigma_v^2 = \sigma^2, \quad \mathrm{Cov}[\epsilon_1, \epsilon_2] = \sigma_c^2 = \alpha\,\sigma^2 \tag{107}$$

Where we define $\sigma^2 = \sigma_c^2 + \sigma_v^2$ as the variance of the total binding energy, and $\alpha = \sigma_c^2/(\sigma_c^2 + \sigma_v^2)$ the relative contribution of the conserved part to the total binding energy variance. This quantity is key in our model extension. A value $\alpha \sim 1$ would correspond to the conserved region controlling most of the binding energy, which results in $\epsilon_1 \sim \epsilon_2$. Conversely, in case $\alpha \sim 0$ most of the binding energy difference between cells is given by the binding of the variable region, which makes the two binding energies ϵ_1 and ϵ_2 nearly independent. As such α can be interpreted as a measure of how many of the binding residues are conserved in the two considered Ag variants, and by extension also as a measure of pathogen variability. Intuitively one could think of the total binding energies ϵ_1 and ϵ_2 to be a sum of $L \gg 1$ random variables, representing the contribution of multiple residues, each one with the same variance σ^2/L. The same covariance matrix is obtained if one consider that α is the fraction of random variables shared between ϵ_1 and ϵ_2. In this interpretation α represents the fraction of conserved sites amongst all the binding residues.
In the approximation that ϵ_c, ϵ_{v1} and ϵ_{v2} are normally distributed then the expression of the conditional probability reads:

$$P(\epsilon_2|\epsilon_1) = \mathcal{N}[\alpha\,\epsilon_1, \sigma\sqrt{1-\alpha^2}](\epsilon_2) \tag{108}$$

Where $\mathcal{N}[\mu, \sigma]$ indicates a Gaussian distribution with mean μ and variance σ^2. As expected the case $\alpha = 1$ reduces to $\epsilon_2 = \epsilon_1$, while for $\alpha = 0$ the two variables are independent $P(\epsilon_2|\epsilon_1) = P(\epsilon_2)$.

After these preliminary considerations we extend the deterministic model described in section 3.2.3 to the case of two Ags, and maintaining the values of parameters specified in table 4. In particular we consider the case of immunization against an antigen, that we will call w for *wild-type*. While the population of B-cells matures under the effect of mutations and selection for binding this Ag, we study the evolution of the population binding energy for a second Ag mutant, which we label m, not administered in the vaccine. The idea is that the affinity for the mutant is a measure of immunization breadth. Indeed, the population will be able to bind a generic non-encountered mutant, which it was not directly selected to bind, only if it retains sufficiently good binding affinity for the conserved part of the Ag. If this happens then it has a good chance of being protected against future encounters with many different Ag mutants.
The density function ρ depends in this case on the two binding energies. As before this function is the product between the population size N_B and its binding energy distribution φ:

$$\rho(\epsilon_w, \epsilon_m) = N_B \, \varphi(\epsilon_w, \epsilon_m) \tag{109}$$

Amplification and differentiation operators are simple to extend, consisting in simple multiplications:

$$\mathbf{A}[\rho] = 2\,\rho, \quad \mathbf{D}[\rho] = (1 - p_{diff})\,\rho \tag{110}$$

Since only one Ag is encountered, competitive selection for T-cell help has a natural extension, depending only one the value of ϵ_w and not on ϵ_m:

$$\mathbf{S}_T[\rho](\epsilon_w, \epsilon_m) = \frac{Ce^{-\epsilon_w}}{Ce^{-\epsilon_w} + e^{-\bar{\epsilon}_w}}\,\rho(\epsilon_w, \epsilon_m), \quad \text{with } e^{-\bar{\epsilon}_w} = \langle e^{-\epsilon_w} \rangle \tag{111}$$

The average $\langle e^{-\bar{\epsilon}_w} \rangle$ is to be considered on the whole population, i.e. on the distribution $\varphi(\epsilon_w, \epsilon_m)$. Encoding mutations instead requires a non-trivial extension. While lethal and silent mutations have the simple effect of killing the cell or not changing the binding energy, affinity-affecting mutations can instead occur in two separate ways. If they change a region that binds a conserved part of the Ag then they introduce the same effect on the two binding energies. If on the other hand they occur on the variable part they can then have two potentially different effect on the two binding energy. The probability of a mutation occurring on one of the two parts depends on their relative size. Following the above discussed analogy, which identifies the parameter α as the fraction of conserved sites, we set the probability of a mutation occurring on the conserved part equal to α. As a consequence the probability for an affinity-affecting mutation to have an additive effect (Δ_w, Δ_m) on the two energies is:

$$K_{aa}(\Delta_w, \Delta_m) = \alpha\,\delta(\Delta_w - \Delta_m)\,K_{aa}(\Delta_w) + (1 - \alpha)\,K_{aa}(\Delta_w)\,K_{aa}(\Delta_m) \tag{112}$$

Where δ represents Dirac delta function and $K_{aa}(\Delta)$ is the single-Ag affinity-affecting mutations effect distribution, that appears in eq. (25), and consists of a Gaussian with mean μ_M and standard deviation σ_M. The total mutation kernel is a combination of silent, lethal and affinity-affecting mutations (occurring with probabilities p_{sil}, p_l, p_{aa}), as is defined as:

$$K(\Delta_w, \Delta_m) = p_{aa}\,K_{aa}(\Delta_w, \Delta_m) + p_{sil}\,\delta(\Delta_w)\,\delta(\Delta_m) \tag{113}$$

Lethal mutations are accounted for in the fact that this kernel is not normalized, its integral being equal to $1 - p_l$. Finally, the mutation operator acts as:

$$\mathbf{M}[\rho](\epsilon_w, \epsilon_m) = \int d\Delta_w \int d\Delta_m \, K(\Delta_w, \Delta_m) \, \rho(\epsilon_w - \Delta_w, \epsilon_m - \Delta_m) \tag{114}$$

As a last remark we point out that if marginalized over the mutant energy, then we are left with operators that are identical to the ones described in section 3.2.3. This is consistent with the fact that the *mutant* Ag is not administered, and evolution must proceed independently from it. Indeed, the marginalized evolution is identical to the one described in the single Ag case.

5.2.2 *Effect of one maturation round on breadth*

We use this model extension to study the change in average affinity for the *mutant* Ag, when vaccinating against the *wild-type*.
As for the case of single Ag we perform our analysis in the *asymptotic* regime, in which evolution is driven by amplification, mutation, competitive selection and differentiation. As before, this is done to observe both asymptotic expansion and maturation in both positive and negative energy direction. In the single Ag case, similarly to what observed in section 2.7, under the action of these operators the density function asymptotically evolves as a traveling wave with exponentially increasing or decreasing size $\rho_t(\epsilon_w) \sim \exp\{\phi t\} \, \widetilde{\rho}(\epsilon_w - u_w \, t)$, where $\widetilde{\rho}$ is the asymptotic form of the wave function. The growth rate ϕ and the wave velocity u_w depend on the value of Ag concentration, and as shown in section 2.7.2 can be recovered with an appropriate eigenvalue equation $e^{\phi} \rho = \mathbf{T}(-u_w) \, \mathbf{D} \, \mathbf{S}_T \, \mathbf{M} \, \mathbf{A} \, \rho$, where $\mathbf{T}(\Delta)$ is the operator that implements a translation of the density function of magnitude Δ. In short, since the marginalized evolution is identical to the single-Ag case, asymptotically the marginal density function $\rho(\epsilon_w)$ must evolve as a traveling wave with speed $u_w(C)$ and growth rate $\phi(C)$. We indicate with $\widetilde{\varphi}(\epsilon_w)$ the normalized asymptotic binding energy distribution. Since evolution is invariant for translation we set its mean equal to zero. In fig. 44 A and B we display the value of ϕ and u_w as a function of Ag concentration. As in section 2.7, these functions define two critical concentrations $\phi(C^*) = 0$ and $u_w(C^{**}) = 0$.[1] Efficient maturation can occur only for values of $C^* < C < C^{**}$, for which we have both positive growth rate and negative maturation speed.

In our analysis we will consider a single evolution round, which evolves our density function from an initial state $\rho_i(\epsilon_w, \epsilon_m)$ to a final state $\rho_f(\epsilon_w, \epsilon_m)$, according to:

$$\rho_f(\epsilon_w, \epsilon_m) = \mathbf{D} \, \mathbf{S}_T \, \mathbf{M} \, \mathbf{A} \, \rho_i(\epsilon_w, \epsilon_m) \tag{115}$$

The aim of our analysis is to quantify the average binding energy increase or decrease for the *mutant* Ag conferred by this evolution round:

$$u_m = \langle \epsilon_m \rangle_f - \langle \epsilon_m \rangle_i \tag{116}$$

1 Notice that since the model used here is different from the one employed in section 2.7 the numerical values of the critical concentrations is different, and is more similar to the one found in section 3.3 with the use of the WKB approximation.

We start by defining the initial state $\rho_i(\epsilon_w, \epsilon_m)$. We can decompose the density function in the product between the population size and its binding energy distribution, and use Bayes rule to separate the probability distribution into a product between the marginal and conditional distributions:

$$\rho_i(\epsilon_w, \epsilon_m) = N_i \, \varphi(\epsilon_w) \, \varphi(\epsilon_m|\epsilon_w) \tag{117}$$

We consider the marginal distribution to have converged to the asymptotic state and set $\varphi(\epsilon_w) = \widetilde{\varphi}(\epsilon_w)$. Moreover based on the above considerations (cf. eq. (108)) we set the conditional distribution equal to:

$$\varphi(\epsilon_m|\epsilon_w) = \mathcal{N}[\alpha \, \epsilon_w, \sigma_w \sqrt{1-\alpha^2}](\epsilon_m) \tag{118}$$

Where as above $\mathcal{N}[\mu, \sigma]$ represents a Gaussian with mean μ and standard deviation σ, and σ_w is the standard deviation of the distribution $\widetilde{\varphi}(\epsilon_w)$.

Having given these definitions, we can now proceed with the evaluations of $\langle \epsilon_m \rangle_i$ and $\langle \epsilon_m \rangle_f$. The former is simply given by:

$$\langle \epsilon_m \rangle_i = \int d\epsilon_w \, d\epsilon_m \; \widetilde{\varphi}(\epsilon_w) \, \varphi(\epsilon_m|\epsilon_w) \, \epsilon_m = \alpha \int d\epsilon_w \; \widetilde{\varphi}(\epsilon_w) \, \epsilon_w = 0 \tag{119}$$

Which derives from the fact that, as discussed above, we removed the translational degree of freedom in the eigenvalue equation solution by setting the mean of $\widetilde{\varphi}(\epsilon_w)$ equal to zero. Calculations for $\langle \epsilon_m \rangle_f$ are less straightforward. Its expression reads:

$$\begin{aligned}\langle \epsilon_m \rangle_f &= \frac{1}{N_f} \int d\epsilon_w \, d\epsilon_m \; \rho_f(\epsilon_w, \epsilon_m) \, \epsilon_m = \\ &= \frac{2(1-p_{diff})}{N_f} \int d\epsilon_w \, d\epsilon_m \frac{C e^{-\epsilon_w}}{C e^{-\epsilon_w} + e^{-\bar{\epsilon}_w}} \\ &\qquad \int d\Delta_w \, d\Delta_m \; K(\Delta_w, \Delta_m) \, \rho_i(\epsilon_w - \Delta_w, \epsilon_m - \Delta_m) \, \epsilon_m \end{aligned} \tag{120}$$

Where we made explicit the operators expression in the definition of ρ_f eq. (115). As a first step we decompose the initial density function using eq. (117), and then perform the integration over ϵ_m:

$$\int d\epsilon_m \; \varphi(\epsilon_m - \Delta_m|\epsilon_w - \Delta_w) \, \epsilon_m = \Delta_m + \alpha(\epsilon_w - \Delta_w) \tag{121}$$

Where we made use of eq. (118) for the definition of the conditional probability. This reduces the expression for $\langle \epsilon_m \rangle_f$ to:

$$\begin{aligned}\langle \epsilon_m \rangle_f = 2(1-p_{diff}) \frac{N_i}{N_f} &\int d\epsilon_w \frac{C e^{-\epsilon_w}}{C e^{-\epsilon_w} + e^{-\bar{\epsilon}_w}} \\ &\int d\Delta_w \, d\Delta_m \; K(\Delta_w, \Delta_m) \, \widetilde{\varphi}(\epsilon_w - \Delta_w) \, [\Delta_m + \alpha \, (\epsilon_w - \Delta_w)] \end{aligned} \tag{122}$$

We now perform integration over Δ_m. Using the expressions for the mutation kernel eqs. (112) and (113) we can write:

$$\begin{aligned}&\int d\Delta_m \; K(\Delta_w, \Delta_m) \, [\Delta_m + \alpha \, (\epsilon_w - \Delta_w)] = \\ &\qquad = \underbrace{\alpha \, \epsilon_w \, [p_{sil} \, \delta(\Delta_w) + p_{aa} \, K_{aa}(\Delta_w)]}_{①} + \underbrace{p_{aa} \, (1-\alpha) \, \mu_M \, K_{aa}(\Delta_w)}_{②} \end{aligned} \tag{123}$$

Where as a reminder $\mu_M > 0$ is the mean of the mutation kernel K_{aa}, representing the average effect of mutations. Now we evaluate separately the two contributions coming from the terms ① and ② in the previous equation. For term ① we have:

$$① \to \alpha \frac{N_i}{N_f} 2(1-p_{diff}) \int d\epsilon_w \; \epsilon_w \; \frac{Ce^{-\epsilon_w}}{Ce^{-\epsilon_w} + e^{-\bar{\epsilon}_w}} \int d\Delta_w \; [p_{sil}\, \delta(\Delta_w) + p_{aa}\, K_{aa}(\Delta_w)]\, \widetilde{\varphi}(\epsilon_w - \Delta_w) \quad (124)$$

Now we point out that the term in the integrals are the exact expression of the mutation and selection operators in the single-Ag case, that act on the marginalized distribution. Since this distribution is in the asymptotic state we know that their combined effect, together with the amplification and selection operators, is a simple translation of u_w and amplification of e^{Φ}:

$$2(1-p_{diff}) \frac{Ce^{-\epsilon_w}}{Ce^{-\epsilon_w} + e^{-\bar{\epsilon}_w}} \int d\Delta_w \; [p_{sil}\, \delta(\Delta_w) + p_{aa}\, K_{aa}(\Delta_w)]\, \widetilde{\varphi}(\epsilon_w - \Delta_w) = = e^{\Phi}\, \widetilde{\varphi}(\epsilon_w - u_w) \quad (125)$$

This reduces term ① to simply:

$$① \to \alpha \frac{N_i e^{\Phi}}{N_f} \int d\epsilon_w \; \epsilon_w \, \widetilde{\varphi}(\epsilon_w - u_w) = \alpha\, u_w \quad (126)$$

Where we used the fact that since the marginal evolution is the same it must be $N_f = N_i e^{\Phi}$, and the fact the mean of $\widetilde{\varphi}$ is zero. Now we turn to the evaluation of term ②:

$$② \to (1-\alpha)\, \mu_M \left\{ \frac{N_i}{N_f} 2(1-p_{diff}) \int d\epsilon_w \; \frac{Ce^{-\epsilon_w}}{Ce^{-\epsilon_w} + e^{-\bar{\epsilon}_w}} \int d\Delta_w \; [p_{aa}\, K_{aa}(\Delta_w)] \; \widetilde{\varphi}(\epsilon_w - \Delta_w) \right\} \quad (127)$$

The term in curly brackets represents a fraction. This can be understood by considering that if we were to add $p_{sil}\, \delta(\Delta_w)$ to the term in square brackets then using eq. (125) it would become equal to $\int d\epsilon \widetilde{\varphi}(\epsilon_w - u_w) = 1$. Without this missing term the quantity corresponds to the fraction of cells in the final population that have undergone an affinity-affecting mutation in this round. We name this fraction f_{aa}, which allows us to write the final form for the final average binding energy:

$$\langle \epsilon_m \rangle_f = \alpha\, u_w(C) + (1-\alpha)\, \mu_M\, f_{aa}(C) \quad (128)$$

Notice that f_{aa} depends on Ag concentration through the asymptotic form $\widetilde{\varphi}$ Before proceeding with the analysis of this result we try to provide more intuition into the meaning of the fraction f_{aa}. This quantity depends on the probability of affinity-affecting mutations p_{aa} (e.g. if $p_{aa}/(p_{aa} + p_{sil}) = 1$ then $f_{aa} = 1$) however, because of selection, it is not equal to it. Interestingly, this term originates from a particular kind of coupling in the mutation fate between the two energies ϵ_w and ϵ_m. These energies are coupled in two ways. The first is that mutations are either silent or affinity-affecting for both energies. The second is that affinity-affecting mutations have a probability α to have the same effect on both energies. The term f_{aa} is linked to the first kind of coupling.

To demonstrate this one can consider the case in which this first coupling type is removed. That is to say, cells will have a probability α of developing a mutation that has the same effect on the two energies (silent or affinity-affecting), and a probability $1-\alpha$ of developing independent mutations (silent or affinity-affecting, independently on each cell). This is obtained by redefining the mutation kernel (cf. eq. (113)) as:

$$\begin{aligned} &K(\Delta_w, \Delta_m) = (1-p_l)\,\left[\alpha\, K_1(\Delta_w)\,\delta(\Delta_w-\Delta_m) + (1-\alpha)\, K_1(\Delta_w)\, K_1(\Delta_m)\right], \\ &\text{with } K_1(\Delta) = \frac{p_{aa}}{p_{aa}+p_{sil}}\, K_{aa}(\Delta) + \frac{p_{sil}}{p_{aa}+p_{sil}}\,\delta(\Delta) \end{aligned} \tag{129}$$

Notice that if Δ_m is marginalized out then the remaining kernel is again the same of the single-Ag case. Using this new definition one can perform the same calculations done above, obtaining this time:

$$\langle \epsilon_m \rangle_f = \alpha\, u_w(C) + (1-\alpha)\,\mu_M\, f^*_{aa} \tag{130}$$

With $f^*_{aa} = p_{aa}/(p_{aa}+p_{sil})$, which in this case does not depend on concentration. In fig. 44C we plot the dependence of f_{aa} on concentration and compare it with f^*_{aa}. Notice how for any value of Ag concentration it is $f_{aa}(C) < f^*_{aa}$. Since $\mu_M > 0$ (mutations are on average deleterious) a smaller value of this fraction is beneficial, since it decreases the final average mutation energy $\langle \epsilon_m \rangle$. This occurs because, as stated above, in the first case the fate of energies is more correlated, and maturation on the mutant energy leverages on indirect selection through this correlation.
Moreover, the second term in both eq. (128) and eq. (130) represents the effect of the affinity-affecting mutations on ϵ_m that are not subject to selection, not even in an indirect way. These are by definition all of the mutations that have independent effects on the two energies. In the second case, since silent and affinity-affecting mutations can occur independently, the probability of an independent affinity-affecting mutation is simply the probability of a mutation on the variable part (i.e. $1-\alpha$) times the relative probability of an affinity-affecting mutation rather than a silent one (i.e. f^*_{aa}), resulting in the second term in eq. (130). This quantity does not depend on Ag concentration. On the other hand in the original case such mutations are a fraction $1-\alpha$ of all affinity-affecting mutations (i.e. $f_{aa}(C)$). Since affinity-affecting mutations here always affect both energy components, the fraction of surviving cells depends here on the strength of selection. However, in the limit of weak selection $C \to \infty$, in which all cells survive, the two must be equal, so that $f_{aa}(C \to \infty) \to f^*_{aa}$ (cf. fig. 44 C).

5.2.3 *Effect of Ag concentration and Ag mutability on breadth acquisition*

In the previous section we showed how in our setting a single evolution round modifies the average binding energy for the *mutant* Ag of a quantity:

$$u_m(C,\alpha) = \langle \epsilon_m \rangle_f - \langle \epsilon_m \rangle_i = \alpha\, u_w(C) + (1-\alpha)\,\mu_M\, f_{aa}(C) \tag{131}$$

This quantity depends both on Ag concentration C, which controls the strength of selection, and on the conserved fraction α, which controls the correlation of the effect of mutations on the two binding energies. This fraction quantifies how much of the binding energy depends on residues that are conserved amongst virus variants, and thus is also a measure of virus mutability

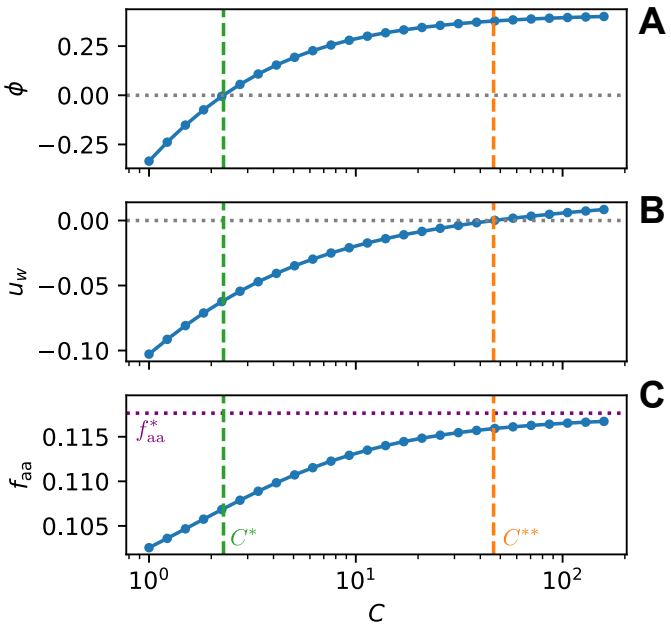

Figure 44: values of the asymptotic growth rate ϕ (**A**), maturation speed u_w (**B**) and fraction f_{aa} of surviving cells in an evolution round that during the round have accumulated an affinity-affecting mutation (**C**). These values depend on Ag concentration C. Values at which $\phi(C^*) = 0$ and $u_w(C^{**}) = 0$. Define two critical concentrations. Efficient affinity maturation occurs only for values of $C^* < C < C^{**}$ at which both growth rate is positive and maturation speed is negative. Values of the fraction f_{aa} are compared to $f^*_{aa} = p_{aa}/(p_{aa} + p_{sil})$, which is the value that such fraction would take if no selection was acting on the population.

In fig. 45 we display the value of u_m as a function of these two quantities. For some values of α and C the value of u_m is negative (blue region in the diagram), meaning that even though the mutant Ag is not administered, in these regimes the population is still developing affinity for it. This occurs because, even though selection (whose strength is controlled by C) only acts on ϵ_w, mutations on the two energies are correlated (with the correlations depending on parameter α). As a result, both increasing the selection strength (i.e. decreasing C) and increasing the mutation correlation (i.e. increasing α) increases affinity maturation for the mutant.
We can inspect eq. (131) in the two extreme regimes $\alpha = 1$ and $\alpha = 0$. The first corresponds to the two Ags being identical. Therefore maturation is the same in both directions $u_m = u_w$. In the second case the effect of affinity-affecting mutations is always uncorrelated. Since selection has no correlation to build affinity on, deleterious mutations dominate making $u_m > 0$.
When analyzing fig. 45 one must take into account the limit on maturation imposed by the values of $\phi(C)$ and $u_w(C)$ (cf. fig. 44 A and B). The joint requirement of a positive growth rate $\phi > 0$ to sustain the population, and a negative velocity $u_w < 0$ to guarantee affinity increase makes so that maturation can only occur for intermediate concentration values $C^* < C < C^{**}$. With this considerations in mind it is natural to identify a critical value of the conserved fraction α^* such that $u_m(C^*, \alpha^*) = 0$. This value separates the region $C^* < C < C^{**}$ of the diagram in two subparts, defining two classes of Ags. For Ags with a high enough conserved fraction ($\alpha > \alpha^*$) increasing the selection strength by lowering Ag concentration C can lead to the acquisition of affinity for the non-administered mutant. For this Ags administering the wild-type in an optimal immunization strategy is sufficient to acquire some protection against mutants. For highly mutable pathogens instead ($\alpha < \alpha^*$) even a high selection pressure is not sufficient to retain enough affinity-enhancing mutations on the conserved region to contrast the accumulation of deleterious mutations. For these Ags, our theory indicates that development of protection against a generic mutant cannot be achieved by simple wild-type immunization. Additional selection pressure is required, that could be obtained by administering multiple Ag mutants, as done for example in [155, 127].

As a final remark, we point out that these results were obtained by considering a single evolution round, in which the marginal distribution $\varphi(\epsilon_w)$ is considered to be in the asymptotic state, while the conditional distribution $\varphi(\epsilon_m|\epsilon_w)$ is initialized with some reasonable assumptions on the correlation between the two binding energies in the population. Notice that while the marginal distribution $\varphi(\epsilon_w)$ is stable, evolving in the asymptotic state as a traveling wave, the other marginal distribution $\varphi(\epsilon_m)$ will not reach a similar stable asymptotic state without adding an explicit form of selection. Instead, the repeated application of the evolution operator will over time increase the variance, and reduce the correlation between the two energies. It would be interesting to extend the theory to the study of the population evolution when both energies are subject to selection, as for example is the case when immunizing with a mixture of different Ag mutants.

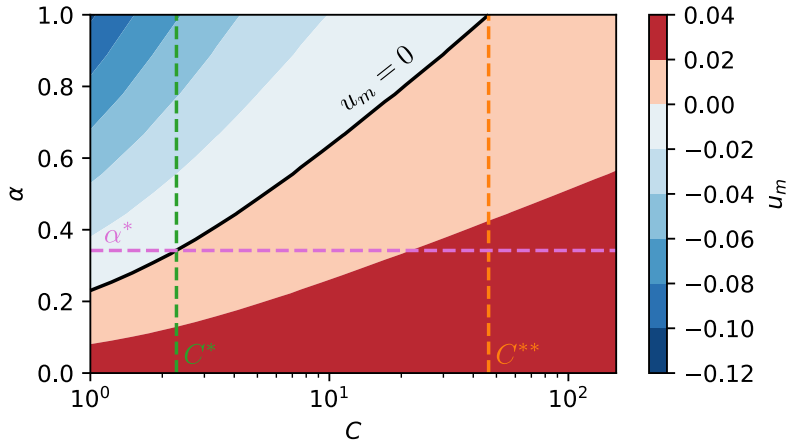

Figure 45: value of the average binding energy difference $u_m = \langle \epsilon_m \rangle_f - \langle \epsilon_m \rangle_i$ provided by an evolution round for the *mutant* Ag, not administered in immunization. This quantity depends on the value of Ag concentration C and on the Ag conserved fraction α. The line $u_m = 0$ (black) separates region of the diagram in which evolution causes an affinity increase (blue) or an affinity-decrease (red) for this Ag. We report on the diagram two lines identifying the two critical concentrations C^* (green) and C^{**} (orange). Efficient affinity maturation for the *wild-type* Ag can only occur for values of C between these two extremes. This defines a critical conserved fraction α^* for which $u_m * (C^*, \alpha^*) = 0$, which splits the diagram in two parts. For Ags which have a high conserved fraction $\alpha > \alpha^*$ (i.e. small mutation rate) lowering the Ag concentration C has a beneficial effect on breadth acquisition. Affinity for a mutant Ag can be acquired from immunization with the wild-type, provided that selection pressure is strong enough. For highly mutable pathogen instead ($\alpha < \alpha^*$) there is no favorable region of the diagram in which maturation for the mutant can occur. Breadth acquisition for these Ags cannot occur from simple vaccination with a wild-type, and requires instead additional selection pressure (e.g. vaccination with multiple Ag mutants).

Our theory suggests that for pathogens with limited mutability some level of breadth can be acquired also through wild-type immunization. In this case raising the selection pressure by lowering Ag concentration can favor breadth acquisition. This strategy is however no longer viable for highly-mutable pathogens. For these pathogens breadth acquisition requires applying additional selection pressure by administering multiple Ag mutants.

A natural continuation to our theory would be to repeat these calculations when multiple mutants are administered, studying how Ag concentration asymptotically controls breadth acquisition. How selection occurs in the presence of multiple Ag variants is however not experimentally known, and depends on how cells encounter the Ag and on the relative concentration of different mutants. If they manage to simultaneously encounter multiple Ag variants cells might be positively selected if they manage to have a good affinity for at least a single Ag variant. Alternatively, if they stochastically encounter different variants in different evolution rounds, they might be required to have a good affinity for most of the variants in order to remain in the GC for multiple rounds. These different scenarios, considered in [155], generate very different outcomes and are hard to discriminate experimentally.
These scenarios could however be investigated using inference, with an approach similar to the one we used in chapter 2 of this work. One could define a maturation model for immunization with multiple Ag variants which includes both scenarios, and which produces a likelihood function that could quantify the probability of experimental measurements as a function of model parameters. Our likelihood maximization procedure could then be used to infer model parameters under both scenarios, and the resulting maximum likelihood values could be compared. Likelihood differences could be used to determine the prevalence of a scenario on the other.

Another possible extension of our work to multiple Ag variants involves the calculation of bottleneck survival probability. Once defined how selection works when multiple variants are administered one could repeat the calculations of section 3.4 in the presence of multiple variants. Adding one Ag variant is equivalent in this case to raising of one unit the dimension of the extinction probability functions $d_t(\epsilon_1, \epsilon_2, \dots)$ (cf. eq. (60)) and of the evolution operators. Unless other simplifications are applied, this could limit the total number of variants for which functions can be numerically evaluated. A similar extension could also be conceived for the theory of evolutionary trajectories discussed in section 3.5. In this case one-dimensional trajectories would be embedded in the higher-dimensional space $(\epsilon_1, \epsilon_2, \dots)$ of binding energies for multiple variants, and applying selection forces for different variants would change the total probability of a trajectory and deviate its path. With an expression for the action of trajectories one could for example study wether sequential application of different selection forces for different variants, which deviate each time the trajectory direction, results in a higher final trajectory probability than if simultaneous selection pressure for all variants is applied.

A

APPENDIX - CHAPTER 2

A.1 NUMERICAL MODEL IMPLEMENTATION AND PARAMETERS CHOICE

Mature GCs usually appear 5-7 days after Ag administration. During this time a population of up to hundreds of different founder clones colonizes the GC and expands to a total size of a few thousand B-cells. The first mutations in the repertoire are observed around day 6 [60, 88]. Early GCs are highly polyclonal and contain 50 to 200 clones according to [144]. In agreement with these experimental findings at the time of Ag injection we pick a population of $N_{found} = 100$ founder clones. The affinities of these clones are extracted independently from an initial gaussian distribution whose mean and variance are chosen via the maximum-likelihood procedure described in section 2.6 and it matches the experimental distribution of germline responders (i.e. splenic IgG-SCs that are observed 1 day after boost of pure Ag, cf. fig. 21 in scheme 2 and Ag dosage $D = 0$). During the time of GC formation the founder clones expand uniformly without mutating. We chose to start our simulation at $T_{GC} = 6$ days after Ag injection. At this point the GCs are almost fully formed [34]. The simulation starts with the GC at its maximal size, set to $N_i = N_{max} = 2500$ clones. The maximal size is in agreement with [37] which reports around 3000 cells per GC, or [144] in which GCs are said to contain up to a few thousands B-cells. However we stress that GCs are heterogeneous in size [158].

From here the model proceeds in evolution rounds. Similarly to [155] we set the duration of a round to $T_{turn} = 12\,h$. This number is consistent with timing of cell migration [150, 90]. We neglect the fact that high affinity cells are found to dwell longer in the GC dark zone [45] undergoing additional divisions. In addition to this the fact that the average cell-cycle time is 12 hours or longer [7] indicates that 12h is probably a lower limit for the round duration.

As described in the main text, each round consists in cell division with somatic hypermutation, selection for Ag binding, selection for T-cell help, differentiation, and if necessary enforcement of a maximum carrying capacity. In our simulations before starting the first round we perform only once differentiation. This is done in order to recover the good average energy limit at low Ag concentrations. In fact when Ag dosage is small the population quickly goes extinct, while at the same time maturating very quickly. Performing differentiation first provides a nucleus of low-affinity germline-like clones whose binding energy controls the average binding energy of the MC population, even if few high affinity clones are added later. Notice that this does not change the asymptotic behavior of the model, since it would be equivalent to simply changing the order of operations in the round.

Proceeding with the standard turn order then the first operation performed is cell division and somatic hypermutation. During a round we consider cells to divide twice [90]. In GC dark zone cells up-regulate their expression of Activation-Induced Cytidine Deaminase. This enzyme increases the DNA mutation rate, inducing mutations in the region coding for the BCR and possibly changing the affinity for the Ag. Mutation rate has been estimated to an average of 10^{-3} mutations per base pair per division [89, 66].

Similarly to [155] in which the total binding energy consisted in the sum of contributions from 46 different residues, we consider $N_{res} = 50$ residues to contribute to the binding energy. The probability that upon division at least one mutation occurs in any of the 150 bp coding for these residues can be estimated as $p_{mut} = 1 - (1 - 10^{-3})^{150} \sim 0.14$. As done in [155, 161, 154] at every division and for each daughter cell independently we consider a $p_{sil} = .5$ probability of developing a silent mutation, in which case the binding energy of the daughter cell remains unchanged, a probability $p_{let} = .3$ of undergoing a lethal mutation, in which case the cell is removed, and finally a probability $p_{aa} = .2$ of developing an affinity-affecting mutation. These change the binding energy of the daughter cell by adding a variation $\epsilon \to \epsilon + \Delta\epsilon$. As done in [154] the variation follows a lognormal distribution $K_{aa}(\Delta\epsilon)$ defined as:

$$K_{aa}(\Delta\epsilon) = \text{Lognorm}[\mu = 1.9, \sigma = 0.5](\Delta\epsilon + 3) \tag{132}$$

where:

$$\text{Lognorm}[\mu, \sigma](x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\} & x > 0 \\ 0 & x \leqslant 0 \end{cases} \tag{133}$$

The parameters of the distribution, which are the same used in [154], are chosen so that only 5% of the mutations confer an increase in affinity, while the vast majority causes an affinity decrease. As a result of this process after the two mutations the population size increases almost 4-fold in size (two duplications but some cells are eliminated due to lethal mutations) and the average affinity decreases slightly due to the mainly negative effect of mutations (cf. fig. 13 histograms 1 to 2).

After duplication we implement selection. In order to avoid apoptosis cells must bind and internalize a sufficient amount of Ag. The amount of Ag internalized depends both on the affinity of the BCR and on the availability of Ag on the surface of the FDC [111, 67, 15]. We model this process by expressing the probability of survival of a cell with BCR having binding energy ϵ as described in the main text eq. (6). The value of the threshold binding energy has been obtained via maximum likelihood fit of the data, which yields for example $\epsilon_{Ag} = -13.59$ for variant A. This selection is not in present variant C of the model.

At the second step of selection, the one leading maturation in our model, B-cells compete to receive a survival signal from T-cells. T-cells in GCs are motile and continuously scan the surface of B-cells, sensing the density of pMHC-II complexes [134]. Cells with the highest pMHC-II density receive survival signal preferentially [35, 149]. We again express the probability of survival through eq. (7). The parameters a and b in this equation represent, respectively, the probability of survival at very high energy and the deficit in probability of survival at very low binding energy. Their effect is better discussed in appendix A.4. The formula interpolates smoothly between these two values, as depicted in fig. 14D. The threshold binding energy $\bar{\epsilon}$ depends on the population's binding energy distribution, introducing competition between the cells.

Cells that are able to survive selection can either re-enter the dark zone and start a new round of evolution or differentiate into Ab-producing PCs or quiescent MCs that can be reactivated upon future Ag injection. There is evidence that MC/PC output undergoes a temporal switch: MCs are preferentially produced early in the response [157].

Moreover there seems to be an affinity bias in differentiation [131]. Even though experiments show that affinity plays a role in deciding fate [130] simply by implementing a time-switch in the MC/PC differentiation probability (respectively μ_{MC}, μ_{PC}, cf. eqs. (8) and (9) and fig. 14B) we effectively recover both of these observations. The parameters of these functions ($\tau_{diff} = 11d$, $\Delta\tau_{diff} = 2d$) are chosen so as to be compatible with [157]. Notice that the sum of the two is constant $\mu_{MC}(t) + \mu_{PC}(t) = p_{diff} = 10\%$, compatible with seminal studies [114] that estimated that around 90% of the cells recirculate in the dark zone. In the model we consider for simplicity a complete switch, meaning that for $t \gg \tau_{diff}$ the probability of generating MCs decreases asymptotically to zero. However in appendix A.2 we discuss the more realistic case of a partial switch, in which there is a residual probability of MC production even for $t \gg \tau_{diff}$.

If new Ag is administered we consider a new GCR to start. The new GC is colonized partly by new B-cells coming from the naive pool and partly by reactivated MCs [58]. We allow only MCs that have already been generated at time of the second injection to colonize the new GC. This is done by picking a set of $N_i = 2500$ cells from the naive pool, with binding energies extracted from the same initial Gaussian distribution, and adding to them all the MCs generated up to the time of second injection. The founder clones of the new GC will consist of $N_{found} = 100$ cells randomly extracted from this cumulative population. Notice that the probability of extracting a MC from the cumulative population is an increasing function of the number N_{MC} of MCs extracted at time of injection: $p = N_{MC}/(N_i + N_{MC})$. In appendix A.2 we discuss instead the case in which the probability of extracting a seeder clone from the memory pool is set to a constant $p = 0.3$.

The concentration of Ag evolves as explained in the main text according to the differential eqs. (4) and (5). These equations account for Ag release, decay and consumption. The release rate was evaluated considering a half-life of 17h for Ag in CFA [82], which gives a value for the release rate of $k^+ = \ln 2/\tau_{1/2} \sim 0.98\,d^{-1}$. Ag on FDCs can be maintained for a long time, up to a year [53], through a mechanism of endocytosis and recycling of immune complexes [52]. To reproduce this long clearance time we take Ag lifetime to be 8.1 weeks, as measured in popliteal lymph nodes of mice [145]. This results in a Ag decay rate of $k^-_\varnothing = \ln 2/\tau_{1/2} \sim 0.012\,d^{-1}$. The case of a faster Ag-decay is discussed in appendix A.2. Finally, the consumption rate per B-cell $k^-_B = 2.07 \times 10^{-5}\,d^{-1}$ (variant C, see main text) is obtained via the maximum likelihood fit procedure described in section 2.6. This quantity controls both the GC lifetime and the extent of AM at the end of evolution. For the range of Ag dosages considered simulated GCs have an effective lifetime that vary between 1-2 weeks and 3 months (see fig. 49), compatible with lifetimes of real GCs [148]. Equations (4) and (5) are continuous in time. To include them in our discrete timestep model we perform an update of the values of the reservoir and available concentrations $C_{av}(t)$, $C_{res}(t)$ at each round $t = 0, 1, \ldots$ after selection for T-cell help and before differentiation. The Ag removal rate is given by the cumulative effect of decay and consumption: $k^-_t = k^-_\varnothing + N^B_t k^-_B$, and changes at each evolution round due to its dependence on the number of B-cells N^B_t at this stage of the round. The values

of the concentrations at the next round $t+1$ are obtained by evolving the corresponding quantities at round t for a time $T = 12\,h$ equivalent to the duration of the round:

$$C_{res}(t+1) = C_{res}(t)\, e^{-k^+ T} \tag{134}$$

$$C_{av}(t+1) = C_{av}(t)\, e^{-k_t^- T} + C_{res}(t)\, \frac{k^+}{k^+ - k_t^-} \left(e^{-k_t^- T} - e^{-k^+ T} \right) \tag{135}$$

At times smaller than the GC formation time $T_{GC} = 6\,d$ we do not account for GC evolution but we account for the evolution of concentration. This is done as in the previous equations but in this case the total consumption rate is evaluated considering an exponentially increasing number of cells, that starting from one at the time of injection exponentially grows to N_{max} at the time of GC formation. In particular concentration update at the end of round $t = 0, 1, \ldots, 11$ is done considering the following number of B-cells consuming Ag:

$$N_t^B = {N_{max}}^{t \times T_{turn}/T_{GC}} \quad \text{for } t < 12 = T_{GC}/T_{turn} \tag{136}$$

In our simulations GC evolution stops either naturally when Ag depletion leads to population extinction, or when the total simulation time is elapsed and cells are harvested, in which case the simulation is stopped irrespective of the population size and only cells produced up to that point are considered. The total simulation time depends on the immunization scheme considered and is set to match the time elapsed between injection and experimental measurement.

A.2 SMALL VARIATIONS OF THE STANDARD MODEL

In order to test the relative importance of different model parameters we performed the inference procedure using three different variants for selection, described in section 2.4.5. Variant A corresponds to the inference of all 9 inferred model parameters (μ_{naive}, σ_{naive}, k_B^-, α, a, b, g_{recall}, g_{imm}, ϵ_{Ag}). Variant B corresponds to the case in which stochasticity and permissiveness parameters a and b are set to zero and only the remaining 7 parameters are inferred. In variant C instead Ag-binding selection is neglected, and all 8 parameters with the exclusion of ϵ_{Ag} are inferred. The resulting maximum likelihood estimate (MLE) of the parameters, along with the corresponding value of the likelihood, which was presented in table 2, is also reported for the three cases in fig. 46. As described in section 2.6.3, the result of the inference procedure show that the removal of Ag-binding selection (variant A vs C, 9 vs 8 parameters) causes only a very modest decrease in log-likelihood, while the removal of stochasticity in T-cell help selection (variant A vs B, 9 vs 7 parameters) generates a consistent log-likelihood decrease. Both the Bayesian Information Criterion (BIC) [125] and Akaike Information Criterion (AIC) [3] suggest variant C as the preferred variant.

Moreover, to show that the model is robust under minor modifications of the hypotheses we consider three minor variations of model variant C, and show that they generate similar MLE predictions for the model parameters. First, we test the effect of introducing a residual asymptotic rate of MC/PC production. This case is labelled *soft MC/PC timeswitch*. For simplicity in the model we introduced a complete time-switch between MC and PC production in GCs, and therefore eventually almost only PCs are produced

for $t \gg \tau_{diff}$ (see eqs. (8) and (9) and fig. 14B). In this modification instead we introduce a residual rate of MC/PC production $\mu_{res} = 10\%$ modifying eqs. (8) and (9) as:

$$\mu_{MC}(t) = p_{diff} \left[\mu_{res} + (1 - \mu_{res}) \frac{1}{1 + \exp\{+\frac{t-\tau_{diff}}{\Delta\tau_{diff}}\}} \right] \tag{137}$$

$$\mu_{PC}(t) = p_{diff} \left[\mu_{res} + (1 - \mu_{res}) \frac{1}{1 + \exp\{-\frac{t-\tau_{diff}}{\Delta\tau_{diff}}\}} \right] \tag{138}$$

This makes so that the fraction of MCs in the differentiated population interpolates between ~ 90% at small times and ~ 10% at big times, granting some residual production of MCs at all times. Applying the inference procedure on this more realistic version of the model results in a better final likelihood than the standard (variant C) version. The inferred values of the parameters are on average similar to the ones obtained with the standard version of the model, with the difference of g_{recall} and g_{imm}. While the inequality $g_{recall} > g_{imm}$ still holds, the MLE for these parameter is higher than in the standard case. This can be expected since these parameters control the fraction of MCs in the elicited Ab-SC population, and in this version of the model the MC population contains differentiated cells that would have belonged to the PC compartment in the standard version.

A further modification involves the fraction of seeder clones extracted from the MC population when colonizing a GC. At the second injection some of the seeder clones for the new GC are extracted from the MC population generated following the first injection. The probability of extracting a seeder clone from the MC pool and not from the initial germline distribution depends in the standard version of the model on the number of accumulated memory cells N_{mem} as $p_{mem} = N_{mem}/(N_{mem} + N_i)$ (see appendix A.1 for details). This should account for the fact that intuitively if more MCs were produced in the previous maturation then also more should be recalled. However one could more simply suppose this probability to be constant. We test this case by setting $p_{mem} = 0.3$. This change generates only a very small likelihood decrease with respect to the standard version of the model, while the MLE for all parameters is almost unchanged.

Finally, we also test the effect of increasing the rate of Ag decay, multiplying it by a factor three (case labelled *faster Ag decay*). This results in a slight increase in the maximum likelihood, and the values of all the model parameters are again compatible with the one of the standard version of the model, with the exception of the Ag consumption rate k_B^- which decreases to compensate the faster decay rate.

A.3 INITIAL CONDITIONS AND STOCHASTICITY

Recent experiments estimated the number of different clones in early GC to be between 50 and 200 [144]. In our stochastic model we consider the population of founder clones to be composed of 100 cells. The limited number of founders controls the diversity of the initial population and increases the stochasticity in evolution. However, it does not strongly influence the average outcome. To verify this we compare 1000 stochastic simulations of the standard model (single Ag injection of $D = 1\,\mu g$ of Ag, model variant C) with a modified version in which the number of founder clones was set equal to the number of cells in the initial population (2500 cells). Results are reported in fig. 47. We observe that limiting the initial population diversity increases stochasticity in evolu-

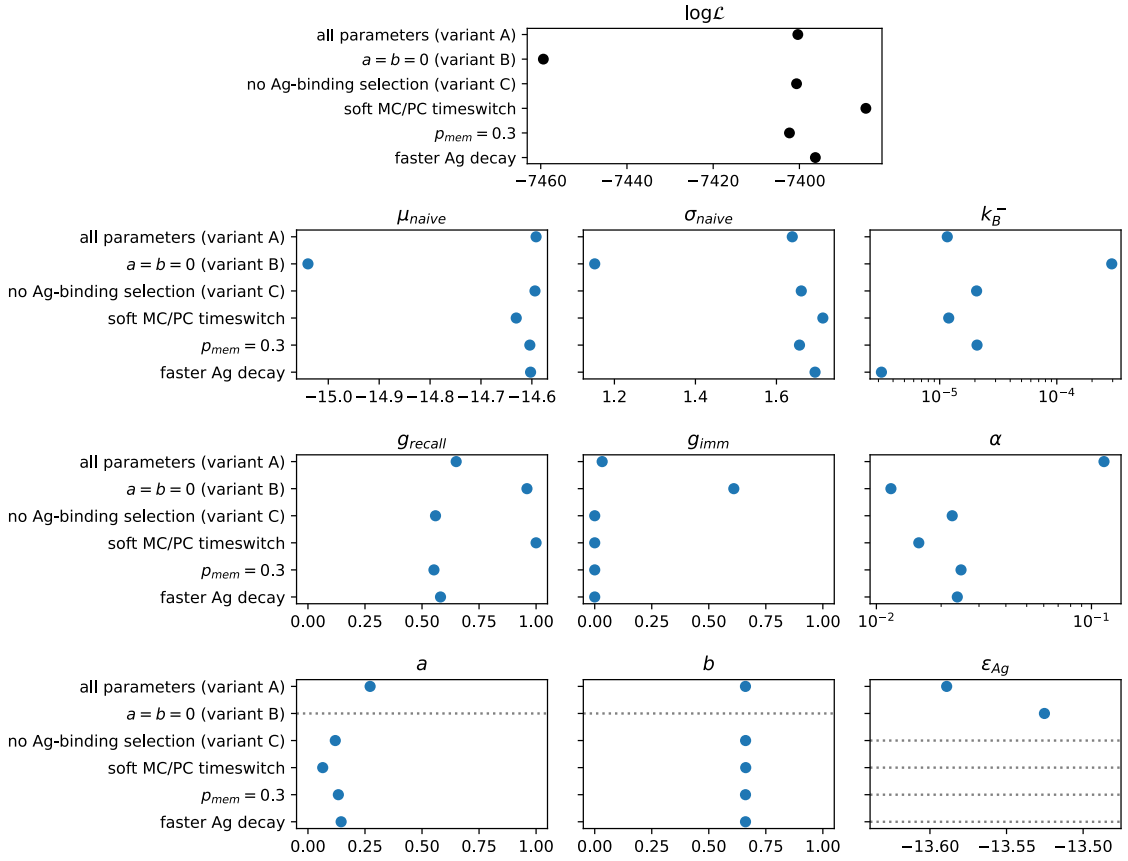

Figure 46: result of the inference procedure on six different variations of the model, as described in appendix A.2. The top plot displays the final maximum value of the log-likelihood obtained with the inference procedure (black). The rest of the plots feature the maximum-likelihood estimate of all of the inferred parameters (blue). Horizontal grey dashed lines mark parameters that are absent in the variant of the model considered.

tion, but does not impact much the average evolution trajectory and outcome. This is especially evident when observing MC/PC population evolution (panels B and D). The final average binding energies of these populations are very similar, but the standard deviation around the mean is halved in for the initial population with more founders. This fact was exploited in the artificial dataset generation, as described in section 2.6.4.

This observation raises the question of how much the outcome of evolution is controlled by the particular initial choice of the founder clones. In fig. 48 we quantify this by comparing 1000 stochastic GC evolutions of the standard model (injected Ag dosage $D = 1\,\mu g$, model scenario C), in which the founder population was re-extracted every time, with a modified version in which the founder population was kept the same amongst all stochastic trials. In the latter case we observe a considerable reduction in stochasticity, indicating that the outcome depends strongly on the initial founder clones choice. This is also in line with the observations made in section 2.8.2 and fig. 26D, where we show that the presence of a high-affinity founder clone correlates with a stronger homogenizing selection. This also suggests that in our model maturation occurs mainly by selection of high-affinity precursors, rather than by accumulation of beneficial mutations, a fact that is confirmed also by the limited number of mutations accumulated (see fig. 51).

Finally we estimate the lifetime of GCs in our model by evaluating the average and standard deviation of lifetimes of 1000 independent GC simulations for different values of injected Ag dosage. Since in the simulations GCs can also have a more or less long period of small population size prior to extinction (fig. 15A) we also evaluate an "effective" lifetime, considering the GC effectively extinguished when its size reaches 1% of its original size. These lifetimes, reported in fig. 49, depend on the amount of Ag administered and can vary between few weeks to some months. As discussed in section 2.5.2 the deterministic model slightly overestimates selection pressure at small population sizes. This overestimation also leads the theory to slightly underestimate the lifetime of stochastic simulations (compare theory and simulations in fig. 49).

A.4 ROLE OF a, b PARAMETERS

Parameters a and b in eq. (7) represent respectively the probability for a B-cell to pass or to fail a selection step irrespective from their affinity. Parameter a can be related to the "permissiveness" of selection, quantifying how likely is for a cell to be positively selected even if its affinity is small. Parameter b conversely encodes the stochasticity in selection, by virtue of which even high affinity cells are not granted survival (e.g. if they don't manage to encounter enough T-cells). For simplicity we define these parameters as constants, but one could imagine that their value may change over time, for example it may be related to the availability of T-cell help. Here we investigate the effect of these parameters on the asymptotic wave-like behavior of the system, discussed in section 2.7. This asymptotic behavior is characterized by two quantities: the asymptotic growth rate $\phi(C)$ and the asymptotic maturation speed $u(C)$, as defined in the main text eq. 20. These quantities are functions of the Ag concentration C. In fig. 50 we report how these functions change when the parameters a and b are progressively increased. In these tests we consider only the effect of one parameter at a time and we set the value of the other to its standard value $a = 0.12$, $b = 0.66$. By making the selection more permissive and allowing for survival of even low-affinity cells, parameter a has a double effect

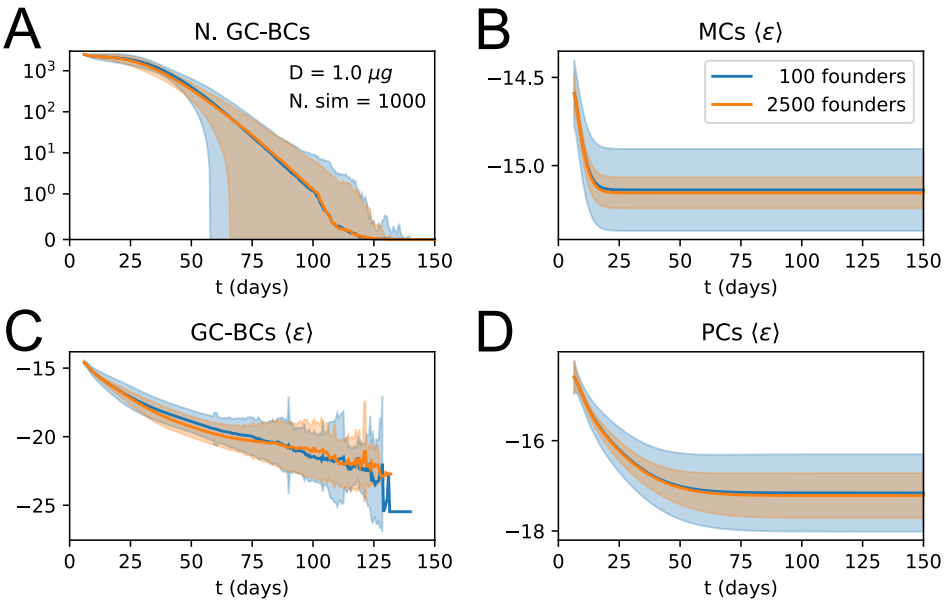

Figure 47: effect of increasing the number of founder clones in the population. We compare 1000 stochastic simulations of the standard version of the model (blue) with a modified version in which every cell in the initial population (2500 cells total) originates from a different founder clone, and has therefore a different affinity (orange). Solid lines represent average trajectories and the shaded area covers one standard deviation from the mean. The plots represent the number of GC B-cell (A), their average binding energy (C), and the average binding energy of MC (B) and PC (D) population as a function of time from Ag injection.

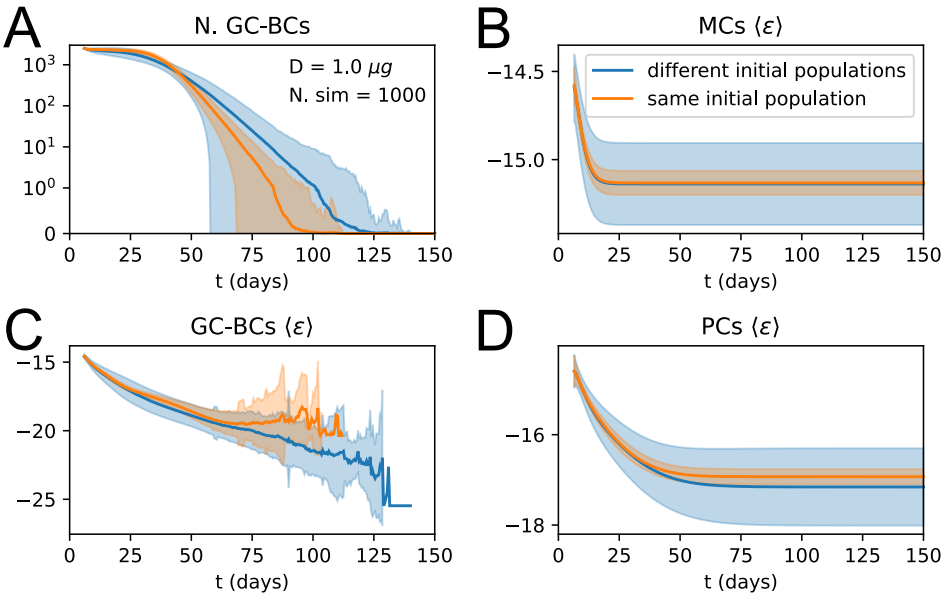

Figure 48: stochastic contribution of the initial founder clones population. We compare 1000 stochastic simulations of the standard version of the model (blue) with a modified version in which GCs are initialzied with the same 100 founder clones (orange). We observe that the initial choice of founder clones plays an important role in evolution and explains most of the variation observed in the outcome. The plots represent the number of GC B-cell (A), their average binding energy (C), and the average binding energy of MC (B) and PC (D) population as a function of time from Ag injection.

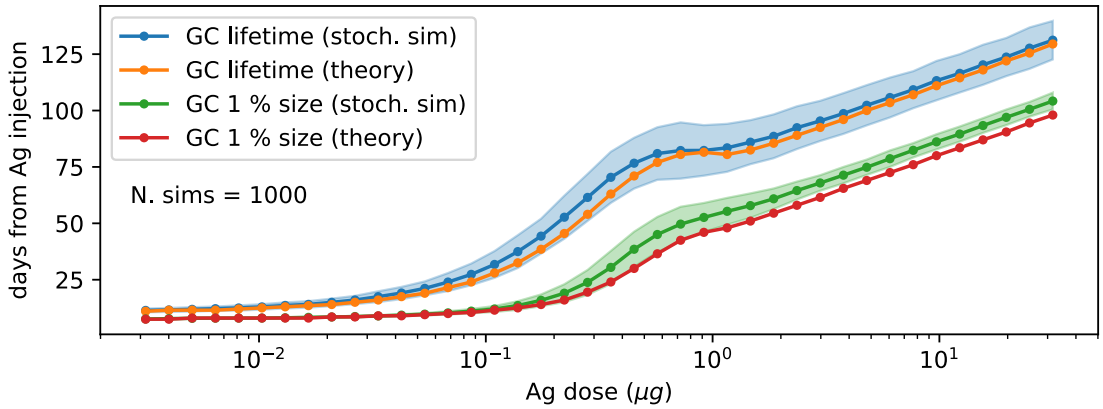

Figure 49: average lifetime and effective lifetime of GCs as a function of administered Ag dosage. The latter corresponds to the time at which the GC decays to 1% of its initial size. We compare the average over 1000 stochastic simulations (blue and green plots, shaded area corresponds to one standard deviation from the mean) to the theoretical prediction (orange and red lines). Due to finite-size effects, the theory in general slightly underestimates the lifetime of the GC.

on the asymptotic behavior: on the one hand it decreases the asymptotic wave velocity and slows down maturation (fig. 50E), and on the other hand it increases the growth rate of the population (fig. 50C). On the contrary, increasing parameter b corresponds to increasing the chance that high-affinity cells are selected out of the population. This both decreases the growth rate (fig. 50D) and also the maturation speed (fig. 50F).

The selection process in affinity maturation has both a *purifying* and *promoting* effect. On the one hand it must negatively select clones that accumulated negative mutations, purifying the population from low-fitness individuals. On the other hand it must also grant the survival and amplification of the clones that developed affinity-improving mutations. These two properties of selection are weakened by parameters a and b in our model, since they respectively grant survival of low-affinity clones and can cause the removal of high-affinity ones. According to our inference procedure parameters a and b together seem to account for 78% of the selection probability, making so that affinity can make the survival probability vary of only about 22%. This slows down maturation considerably, since it removes any deterrent against accumulating deleterious mutations. The fact that the inference procedure indicated a high value for these parameters suggests that selection in GCs may be permissive, at least for complex Ags, as suggested in [102].

A.4.1 *Number of accumulated mutations*

To quantify the number and impact of mutation events in our simulations we execute 1000 stochastic simulations of a single GC, at an injected Ag dosage of $D = 1\,\mu g$. In each simulation and for each cell we keep track of the number of beneficial and deleterious mutations that each cell accumulates during the course of evolution on the residues we consider ($N_{res} = 50$, see appendix A.1). Results are reported in fig. 51. In the top row we display the average number of cells for any value of beneficial and deleterious mutations number and at different times: 10, 30 and 50 days after Ag injection. To have an idea of

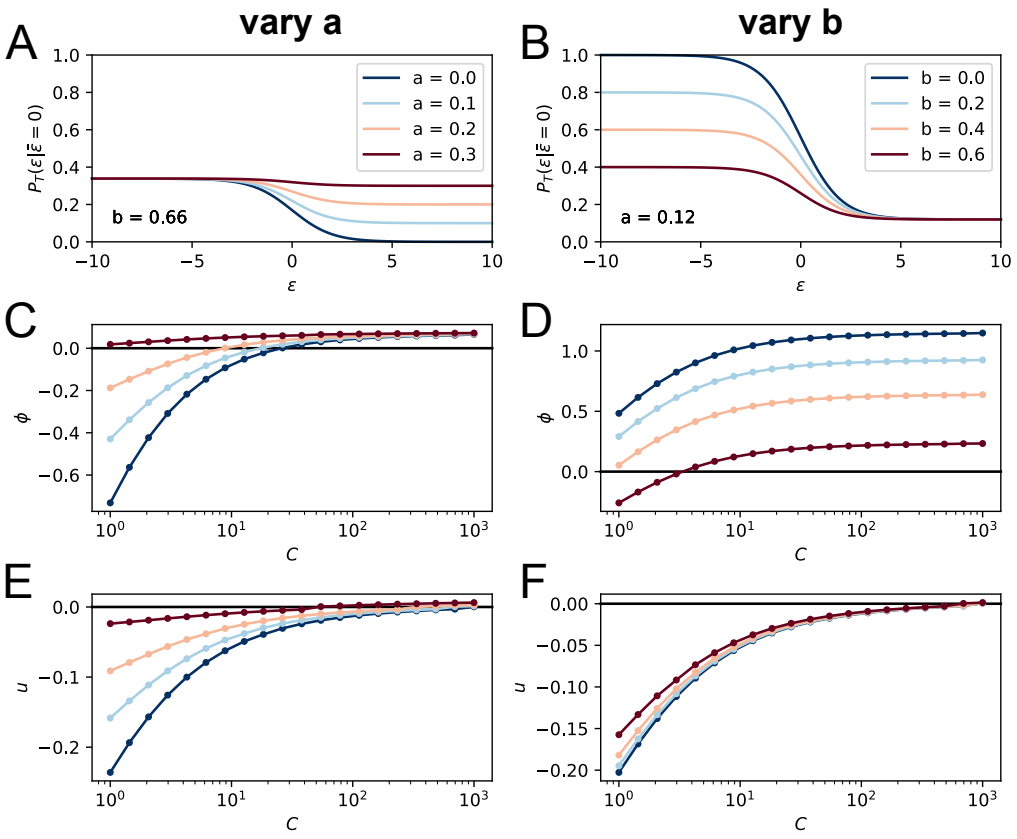

Figure 50: effect of parameters a, b on the model asymptotic behavior. On the top row we plot the corresponding T-cell selection survival probability (setting for simplicity $\bar{\epsilon} = 0$ and $C = 1$) respectively in the case $b = 0.66$ and a varying from 0 to 0.3 (**A**) and $a = 0.12$, b varying from 0 to 0.6 (**B**). Values of the parameters a and b are color-coded as showed in the legend. In **C** and **E** we show the effect of varying a on the asymptotic growth rate ϕ and evolution speed u (we set as before $b = 0.66$). Notice how increasing a both slows down evolution and increases the growth rate. In **D** and **F** we report the same quantities for variation of the parameter b (while $a = 0.12$). Increasing b decreases both the growth rate and the maturation speed.

the population size and maturation state at these timepoints one can refer to fig. 16, in which stochastic simulations are performed under the same conditions. In our simulations after the first days of maturation deleterious mutations start to appear (see fig. 51, t = 10). These are the first mutations to appear since they are much more likely than beneficial mutations (95% vs 5%, see fig. 14 A). However during the course of evolution these are gradually removed by selection, until eventually beneficial mutations, despite being much rarer, start to dominate (see fig. 51, t = 50).
In the bottom row of fig. 51 we display in the same way the number of mutations in the MC and PC population. The former is composed of cells that differentiate early (cf. fig. 14B and eqs. (8) and (9)) and therefore bear less mutations than the PC population. However, in both cases the vast majority of cells harbor very few mutations, with the average number of mutations per cell being 0.13 for MCs and 0.54 for PCs. The accumulation of more than 4-5 beneficial mutations in a single cell is very rare. These numbers are compatible with experiments performed in a recent work [43], in which mice were immunized against Tetanus Toxoid following a protocol similar to the one used in our experiments. The analysis of high-affinity binders showed an average of 6 non-synonymous mutations on the antibody heavy-chain variable region V_H and 3 mutations in the light-chain variable region V_L.
Our model neglects saturation of beneficial mutations, i.e. the phenomenon by which beneficial mutations cannot be accumulated indefinitely but become rarer as the cell approaches maximum possible affinity. This is partly justified here by the observation that, at least for the inferred value of model parameters, few beneficial mutations are found to accumulate in our simulations. Even when considering clones with the highest number of beneficial mutations, the number of mutations accumulated in our model is compatible with experiments [43].
As a final remark, notice that even though MCs are not as strongly skewed towards beneficial mutations as PCs, their average affinity is higher than the one of the starting population (cf.fig. 16B). This is because amongst the founder clones selection will expand the ones having higher affinity (cf.fig. 26), which will then be overly-represented in the MC population. This shows that in our model maturation is achieved only partially by accumulation of beneficial mutations, the rest being obtained through selective expansion of high-affinity precursors, as also showed in section 2.8.2, and confirmed by the strong dependence of the maturation outcome on the initial founder clones population (see fig. 48).

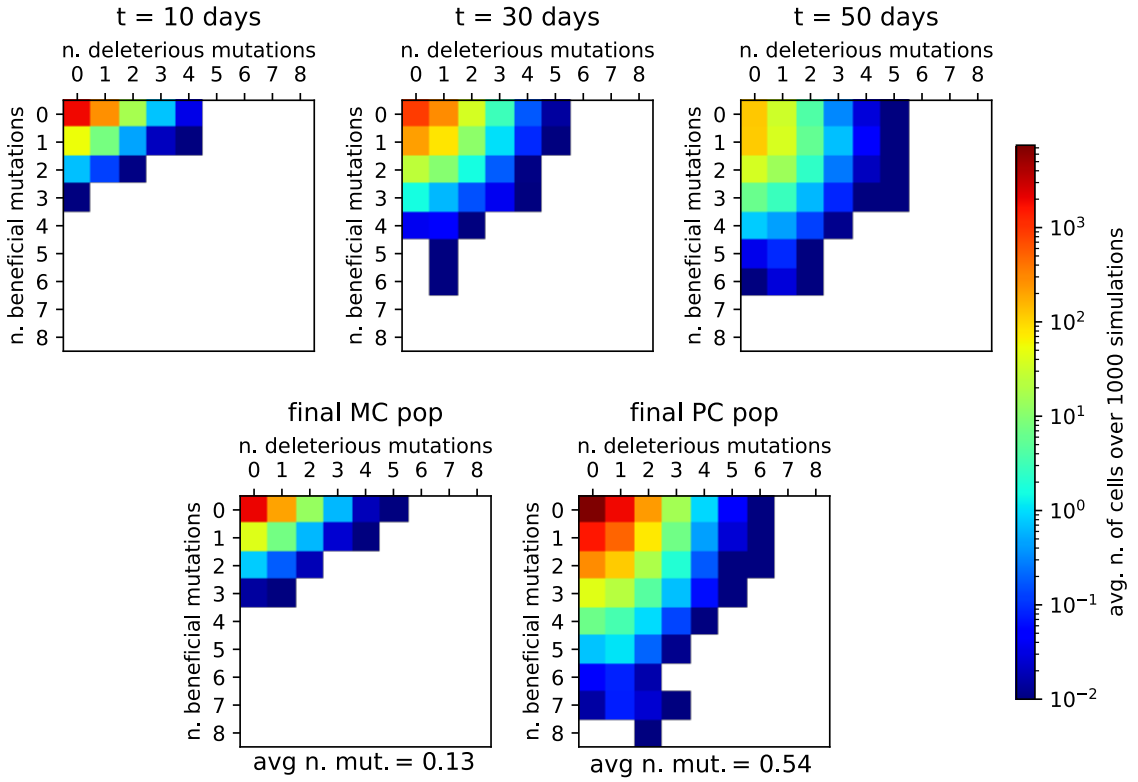

Figure 51: Distribution of beneficial and deleterious mutations over 1000 stochastic germinal center simulations at injected Ag dosage D = 1 µg. Color represents the average number of cells that have developed the specified number of beneficial and deleterious mutations for any examined population, according to the color-scale on the right. **Top**: distribution of mutation numbers in the GC population taken at different times (respectively, 10, 30 or 50 days after Ag injection from left to right). Notice that not all of the populations have the same total size. **Bottom**: distribution of mutation numbers for the final MC and PC distributions. The average number of total mutations accumulated is 0.13 for MCs and 0.54 for PCs. This number has to be compared to the total number of residues considered ($N_{res} = 50$).

B

APPENDIX - CHAPTER 3

B.1 PARAMETERS CHOICE

We pick the standard values of the parameters so as to be compatible with the literature, albeit with some more freedom granted by the fact that here we are more interested in a theoretical analysis, rather than in reproducing data.
Mature GCs have a B-cells population consisting of a few thousands cell [60, 88]. We therefore set the initial and maximum size of the population equal to $N_i = N_{max} = 2500$. This is in agreement with [37] which reports around 3000 cells per GC, or [144] in which GCs are said to contain up to a few thousands B-cells. However we stress that GCs are heterogeneous in size [158]. In [155] and [96] the duration of a turn of evolution is set to 12h. Here the authors consider two cell divisions per round. This is consistent with timing of cell migration [150, 90]. For simplicity we rescale the round so that the duration is set to $T_{round} = 6h$, and consider a single cell division per round. Time in our model will be rescaled by this standard quantity, so that the variable t has no dimension. Similarly, also the binding energy ϵ will be adimensional, expressed in standard units of $k_B T$.
In [154, 161] mutations occur at a rate of 0.5 per sequence per division, and are silent, lethal or affinity affecting with probabilities of respectively $0.5, 0.3, 0.2$. This fixes our effective mutation probabilities to $p_{sil} = 0.75, p_{let} = 0.15, p_{aa} = 0.1$. To reproduce the fact that most of the mutations are deleterious we pick for simplicity $\mu_M = \sigma_M$. This fixes the amount of beneficial mutations to $\sim 16\%$, which is somewhat higher but still compatible with other models [161, 155, 96] in which this fraction is set to 5%. Moreover we set $\mu_M = 0.3$ so as to set the mean effect of beneficial mutations to $\langle \Delta\epsilon \rangle_{beneficial} \sim -0.15$. This value is slightly smaller than $\langle \Delta\epsilon \rangle_{beneficial} \sim -0.53$ used in [96], but this is compensated by the higher rate of beneficial mutations in our model. The shape of the initial population is set to a gaussian with standard deviation $\sigma_i = 1.5$, which is compatible with experimental data [96]. Since evolution of the population is invariant for shifts of the energy space we set $\epsilon_{Ag} = 0$. Under this choice of gauge the zero in the energy space is the threshold energy for Ag-binding selection. Moreover we pick $\mu_i = 3$ so that the difference $\mu_i - \epsilon_{Ag} = 2\,\sigma_i$ and on average only around 2% of cells from the initial population meet this threshold. Finally, The differentiation probability is set to $p_{diff} = 0.1$ [96, 155].

B.2 CRITICAL TREE SIZE AND EXTINCTION TIME IN THE ABSENCE OF MUTATIONS

In sections 3.4.1 and 3.4.2, and fig. 32 C and D we observed how extinction times and lineage sizes present a peak at intermediate values of the binding energy. In these sections we gave an intuitive explanation for this phenomenon, based on the observation that low-affinity cells quickly go extinct while high-affinity cells will survive the bottleneck most of the time, only going extinct if random affinity-independents events (e.g. differentiation or lethal mutations) kill the population when it is still small. Only at

intermediate affinities (close to ϵ_{Ag}) we observe extinction events at long time and big population size.
For ease of notations we call $\gamma(\epsilon) = (1 - p_{let})(1 - p_{diff})\, P_S(\epsilon)$ the total round survival probability for a cell with energy ϵ. In this case the infinite-time extinction probability $d_\infty(\epsilon)$ can be found by rewriting eq. (60) in the limit $t \to \infty$:

$$\sqrt{d_\infty(\epsilon)} = \begin{cases} 1 & \text{if } \gamma(\epsilon) \leqslant 1/2 \\ \frac{1-\gamma(\epsilon)}{\gamma(\epsilon)} & \text{if } \gamma(\epsilon) > 1/2 \end{cases} \tag{139}$$

As expected, lineages will always go extinct if the average number of surviving offspring at division is not greater than one (i.e. $2\gamma(\epsilon) \leqslant 1$). In fig. 52A d_∞ is plotted as a function of γ.
Similarly, by rewriting eqs. (72) and (73) in the case of no mutations, we can also find an explicit expression for the mean and variance of the offspring size for a progenitor with binding energy ϵ:

$$\langle n \rangle_\epsilon = \begin{cases} \frac{2\gamma(\epsilon)}{1-2\gamma(\epsilon)} & \text{if } \gamma(\epsilon) < 1/2 \\ \frac{2-2\gamma(\epsilon)}{2\gamma(\epsilon)-1} & \text{if } \gamma(\epsilon) > 1/2 \end{cases} \tag{140}$$

$$\langle n^2 \rangle_\epsilon - \langle n \rangle_\epsilon^2 = \langle n \rangle_\epsilon (\langle n \rangle_\epsilon + 1)(\langle n \rangle_\epsilon / 2 + 1) \tag{141}$$

The mean and variance are reported in fig. 52C. Both of these quantities diverge for the critical value $\gamma = 1/2$. Having an explicit expression for the average extinction time is harder, but one can verify that in the particular case $\gamma = 1/2$ the extinction time probability distribution behaves as a power law $P(t|\gamma = 0.5) \sim 4t^{-2} + o(t^{-2})$, with infinite mean and variance. This can be proven by inserting the ansatz $d_t \sim 1 - \alpha t^{-1} + o(t^2)$ in eq. (60), together with the simplified form of the mutation kernel and the assumption that $\gamma = 1/2$. In this case one find that the only admissible solution is $\alpha = 4$ which, considering that $P(t) = d_t - d_{t-1}$, proves the statement.
This criticality is confirmed by the following consideration: in absence of mutations our model is a Galton-Watson process, and in the particular case $\gamma = 1/2$ the generated genealogies are critical Galton-Watson trees.
Interestingly, the divergence is removed when we add affinity-affecting mutations to the system. This happens because mutations will drive lineages away from the critical line, either to higher affinities (and thus to survival) or to lower affinities, and thus to extinction.

B.3 Stochastic evolution of the competitive selection pressure and finite-size correction for evolutionary trajectories

In section 3.5.4 we perform the comparison between average stochastic trajectories with their theoretical prediction obtained from the path integral formulation. Although the two agree qualitatively, stochastic trajectories still show some degree of deviation from their theoretical counterpart (cf. fig. 36A).
The main reason behind this discrepancy lays in the fact that, similarly to what observed in section 3.4.3, the theoretical model poorly captures the evolution of the competitive selection pressure, encoded in the threshold binding energy $\bar{\epsilon} = -\log\langle e^{-\epsilon} \rangle_{pop}$. In particular, the big population size approximation overestimates the size of the high-affinity

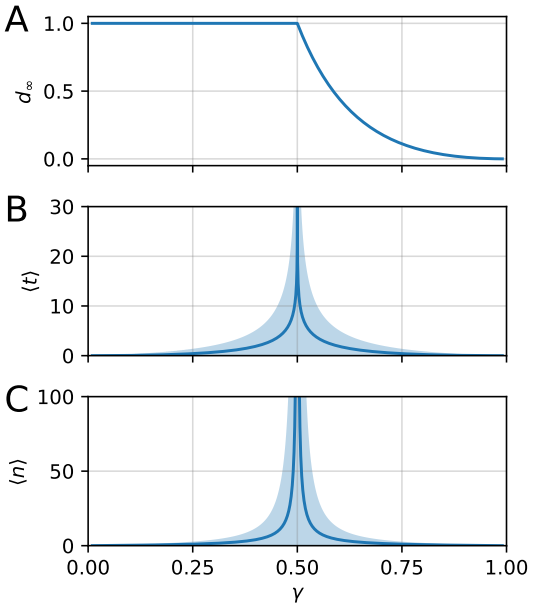

Figure 52: Value of the extinction probability d_∞ (**A**), average extinction time $\langle t \rangle$ (**B**) and average progeny size $\langle n \rangle$ (**C**) as a function of the survival probability $\gamma(\epsilon) = P_S(\epsilon)(1-p_{let})(1-p_{diff})$ in the approximation of no affinity-affecting mutations. In B and C shaded area covers one standard deviation. Notice how extinction times and genealogy sizes diverge at $\gamma = 0.5$.

tail of the binding energy distribution, to which the value of $\bar{\epsilon}$ is sensitive, thus overestimating its evolution speed. Moreover in this limit evolution becomes deterministic, and does not capture the stochastic nature of $\bar{\epsilon}$. To prove both these points in fig. 53A we perform 10 000 simulations of the standard stochastic model and display the density of $\bar{\epsilon}$ trajectories in time-energy space (shade of green, according to the colorscale on the right), together with the average trajectory (black dashed line). Comparing this to the theoretical prediction for the evolution of $\bar{\epsilon}$ (blue line) we see that indeed the maturation speed of the population is overestimated by the theory. Moreover, the average trajectory is only a limited representation for the wide spectrum of different stochastic trajectories observed.

To reduce the discrepancy between theory and simulations we perform two modifications. First, we initialize all stochastic simulations using the same initial population (see fig. 53D). This greatly reduces the stochastic variation of $\bar{\epsilon}$ trajectories (cf. fig. 53B), making the average stochastic evolution a much better approximation for the ensemble of stochastic trajectories. Second, with an approach similar to the one used in section 3.4.3, we can introduce a finite-size correction on the initial density function $\rho(t_i, \epsilon_i)$. This correction, consisting on a cutoff on the high-affinity tail of the distribution, accounts for the finite nature of the cell density function in stochastic simulations. In this case however instead of employing a hard cutoff, which would introduce instabilities in the numerical solution of eq. (37), we use a sigmoid cutoff, which leaves the function and all

of its derivatives continuous. In practice, we correct the initial density function $\rho(t_i, \epsilon_i)$ by multiplying it by the cutoff function $\zeta(\epsilon_i)$:

$$\rho^{coff}(t_i, \epsilon_i) = \zeta(\epsilon_i)\ \rho(t_i, \epsilon_i)\,, \qquad \text{with } \zeta(\epsilon) = \frac{\exp\left\{\frac{\epsilon - \epsilon_{coff}}{\sigma_{coff}}\right\}}{\exp\left\{\frac{\epsilon - \epsilon_{coff}}{\sigma_{coff}}\right\} + 1} \tag{142}$$

The cutoff parameter ϵ_{coff} is set equal to the minimum energy in the stochastic population ($\epsilon_{min} \sim -1.4$ in our case) and the parameter σ_{coff}, controlling the sharpness of the cutoff, is set equal to 0.1. This needs to be small enough to ensure a sharp cutoff, while at the same time not causing instabilities in the numerical integration. The initial density function with and without cutoff is displayed in fig. 53C. If we modify the theory by adding this correction to the initial distribution we obtain a theoretical prediction for the trajectory of $\bar{\epsilon}$ that better approximates the average evolution observed in stochastic simulations (cf. orange and black line in fig. 53B). As a result, also theoretical predictions for the most-likley evolutionary trajectories better match the ones obtained from stochastic simulations (see fig. 36B).

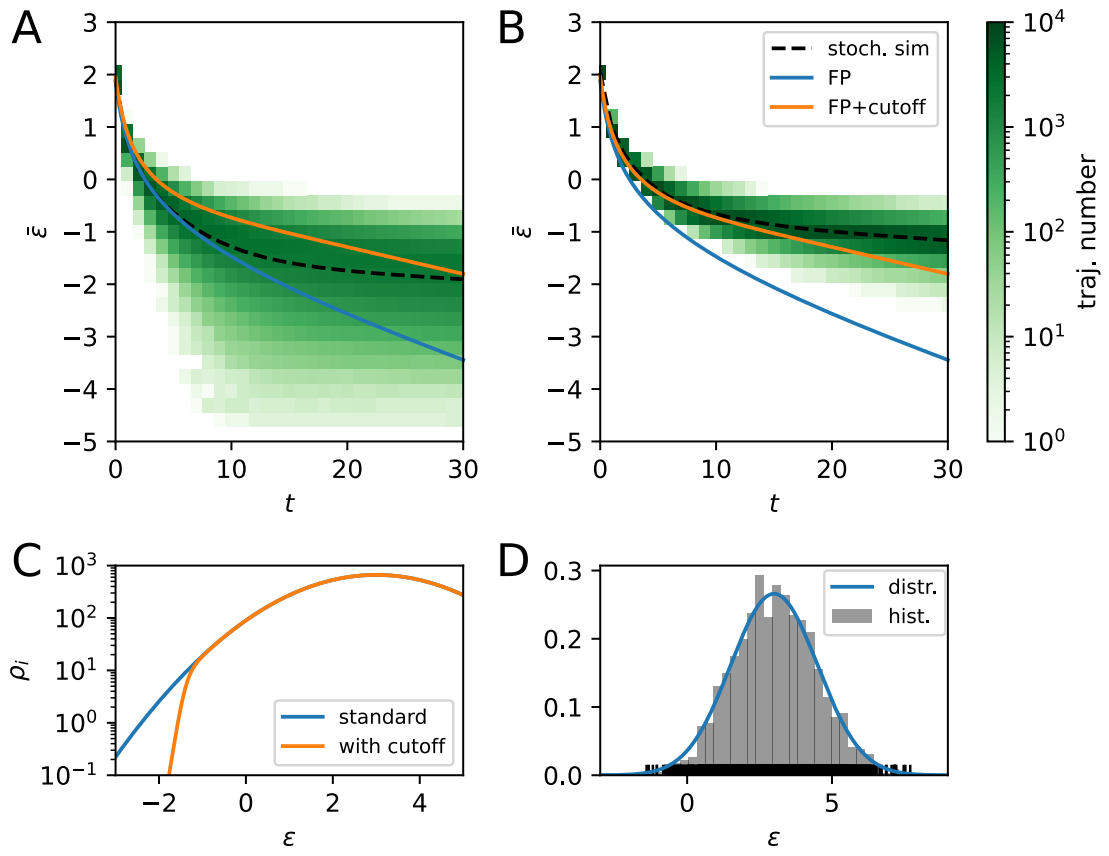

Figure 53: **A**: number of trajectories for $\bar{\epsilon} = -\log\langle e^{-\epsilon}\rangle_{pop}$ over 10 000 different stochastic simulations per energy-time bin (green, see colorscale on the right). The average stochastic trajectory is plotted as a black dotted line. In the prediction for the evolution of the same quantity obtained using FP eq. (37) (blue line) the evolution speed is overestimated. **B**: same as A but the initial population is always the same over all of the 10000 stochastic simulations. In this case the evolution is much less stochastic, which attests the importance of initial conditions in the model evolution. Moreover by applying a cutoff correction on the initial distribution we can improve the prediction of the deterministic model on the evolution of $\bar{\epsilon}$ (orange line). **C**: effect of the finite-size correction procedure (orange) on the cell density of the initial population ρ_i, compared to the case without correction (blue). The correction acts by exponentially decreasing the density of the tail in correspondence of the energy value $\epsilon_{min} \sim -1.4$ of the highest-affinity cell in the population. **D**: normalized histogram (gray) of the initial population used in the simulations displayed in B and for the cutoff procedure in C. Black ticks on the bottom correspond to single cell. The lowest-energy cell has energy $\epsilon_{min} \sim -1.4$. The histogram is compared to the Gaussian distribution from which the energy of each cell in the initial population is extracted (blue).

C

APPENDIX - CHAPTER 4

C.1 PERTUBATIVE ANALYSIS OF BINDING PROBABILITY MASTER EQUATION

Here we perform a simple perturbative analysis of eq. (93) to estimate the equilibrium convergence times.
We consider therefore a small perturbation to the equilibrium solution $p(t,\epsilon) \sim p_{eq}(\epsilon) + \delta p(t,\epsilon)$. The perturbation in binding probability is linked to the perturbation in number of free T-cells by eq. (94):

$$\delta N_T^{free}(t) = -N_B \int d\epsilon\, \varphi(\epsilon)\, \delta p(t,\epsilon) \tag{143}$$

By using this relation and linearizing eq. (93) w.r.t. the perturbation we obtain:

$$\frac{d\delta p}{dt}(t,\epsilon) = -\delta p(t,\epsilon)\left[\omega_u(e^{\epsilon} + e^{\bar{\epsilon}})\right] + \delta N_T^{free}(t)\,\omega_b\,(1 - p_{eq}(\epsilon)) \tag{144}$$

The first term gives us a local relaxation time $\tau = 1/(\omega_u(e^{\epsilon} + e^{\bar{\epsilon}}))$, which is maximum at small energies and behaves like $\tau \sim e^{-\bar{\epsilon}}$. If we consider $\bar{\epsilon}$ to be somehow linked with the average binding energy of the population, we see that our model converges exponentially slower to equilibrium the smaller is this average energy. In this regime the hypothesis of equilibrium will eventually fail. Intuitively according to our model this would corresponds to the regime in which the average unbinding time for T-cells become exponentially big and in a single round of selection a single T-cell will only have the time to bind randomly to a single B-cell, and not sample the whole population. In this regime the probability of being bound becomes uniform and T-cell selection will not be able to discriminate between good and bad binders, since all binders are on average good. In this limit our model is not realistic, however we consider this limit to correspond to a successful AM.
The second term in the r.h.s. represents instead a global adjustment of the fluctuations which acts in the same positive or negative direction for all energies, against the global perturbation effect. By combining the two previous equations one can explicit the time-derivative of this perturbation as:

$$\frac{d\delta N_T^{free}}{dt}(t) = -\delta N_T^{free}(t)\,\omega_b\,(N_T^{free} + N_B^{free}) + N_B\,\omega_u \int d\epsilon\, \varphi(\epsilon)\,\delta p(\epsilon,t)\,e^{\epsilon} \tag{145}$$

Again we have two terms. The first gives a relaxation time of order $\tau \sim 1/\omega_b(N_T^{free} + N_B^{free})$, which is a decreasing function of the equilibrium number of free B and T cells. In general this time will be small when the population size is big. It becomes small only if the size of both B and T cell populations become small, but in this case one could consider the population to be close to extinction. The second term instead derives from the first term of the equation for δp. It represent an average of the perturbation times the exponent of the binding energy ϵ, and again it acts against the global perturbation, therefore shortening convergence time.

BIBLIOGRAPHY

[1] C. Aagaard, T. T. K. T. Hoang, A. Izzo, R. Billeskov, J. Troudt, K. Arnett, A. Keyser, T. Elvang, P. Andersen, and J. Dietrich. Protection and polyfunctional t cells induced by ag85b-tb10. 4/ic31® against mycobacterium tuberculosis is highly dependent on the antigen dose. *PloS one*, 4(6):e5930, 2009.

[2] R. K. Abbott, J. H. Lee, S. Menis, P. Skog, M. Rossi, T. Ota, D. W. Kulp, D. Bhullar, O. Kalyuzhniy, C. Havenar-Daughton, et al. Precursor frequency and affinity determine b cell competitive fitness in germinal centers, tested with germline-targeting hiv vaccine immunogens. *Immunity*, 48(1):133–146, 2018.

[3] H. Akaike. A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, pages 215–222. Springer, 1974.

[4] M. Akkaya, B. Akkaya, A. S. Kim, P. Miozzo, H. Sohn, M. Pena, A. S. Roesler, B. P. Theall, T. Henke, J. Kabat, et al. Toll-like receptor 9 antagonizes antibody affinity maturation. *Nature immunology*, 19(3):255–266, 2018.

[5] M. Akkaya, K. Kwak, and S. K. Pierce. B cell memory: building two walls of protection against pathogens. *Nature Reviews Immunology*, pages 1–10, 2019.

[6] M. Akkaya, J. Traba, A. S. Roesler, P. Miozzo, B. Akkaya, B. P. Theall, H. Sohn, M. Pena, M. Smelkinson, J. Kabat, et al. Second signals rescue b cells from activation-induced mitochondrial dysfunction and death. *Nature immunology*, 19(8):871–884, 2018.

[7] C. D. Allen, T. Okada, H. L. Tang, and J. G. Cyster. Imaging of germinal center selection events during affinity maturation. *Science*, 315(5811):528–531, 2007.

[8] D. Allen, T. Simon, F. Sablitzky, K. Rajewsky, and A. Cumano. Antibody engineering for the analysis of affinity maturation of an anti-hapten response. *The EMBO journal*, 7(7):1995–2001, 1988.

[9] A. Amitai, L. Mesin, G. D. Victora, M. Kardar, and A. K. Chakraborty. A population dynamics model for clonal diversity in a germinal center. *Frontiers in microbiology*, 8:1693, 2017.

[10] S. M. Anderson, A. Khalil, M. Uduman, U. Hershberg, Y. Louzoun, A. M. Haberman, S. H. Kleinstein, and M. J. Shlomchik. Taking advantage: high-affinity b cells in the germinal center have lower death rates, but similar rates of division, compared to low-affinity cells. *The Journal of Immunology*, 183(11):7314–7325, 2009.

[11] T. Arulraj, M. Meyer-Hermann, S. C. Binder, and P. A. Robert. Synchronous germinal centre onset impacts the efficiency of antibody responses. *Frontiers in immunology*, 10:2116, 2019.

[12] S. Awate, L. A. B. Babiuk, and G. Mutwiri. Mechanisms of action of adjuvants. *Frontiers in immunology*, 4:114, 2013.

[13] O. Bannard and J. G. Cyster. Germinal centers: programmed for affinity maturation and antibody diversification. *Current Opinion in Immunology*, 45:21–30, Apr. 2017.

[14] F. D. Batista and M. S. Neuberger. Affinity dependence of the b cell response to antigen: a threshold, a ceiling, and the importance of off-rate. *Immunity*, 8(6):751–759, 1998.

[15] F. D. Batista and M. S. Neuberger. B cells extract and present immobilized antigen: implications for affinity discrimination. *The EMBO journal*, 19(4):513–520, 2000.

[16] D. Baumjohann, S. Preite, A. Reboldi, F. Ronchi, K. M. Ansel, A. Lanzavecchia, and F. Sallusto. Persistent antigen and germinal center b cells sustain t follicular helper cell responses and phenotype. *Immunity*, 38(3):596–605, 2013.

[17] C. Berek, A. Berger, and M. Apel. Maturation of the immune response in germinal centers. *Cell*, 67(6):1121–1129, 1991.

[18] C. Berek and C. Milstein. Mutation drift and repertoire shift in the maturation of the immune response. *Immunological reviews*, 96(1):23–41, 1987.

[19] N. L. Bernasconi, E. Traggiai, and A. Lanzavecchia. Maintenance of serological memory by polyclonal activation of human memory B cells. *Science*, 298(5601):2199–2202, 2002.

[20] S. C. Binder and M. Meyer-Hermann. Implications of intravital imaging of murine germinal centers on the control of b cell selection and division. *Frontiers in immunology*, 7:593, 2016.

[21] R. Brink and T. G. Phan. Self-reactive b cells in the germinal center reaction. *Annual review of immunology*, 36:339–357, 2018.

[22] L. Buchauer and H. Wardemann. Calculating germinal centre reactions. *Current Opinion in Systems Biology*, 18:1–8, 2019.

[23] D. L. Burnett, D. B. Langley, P. Schofield, J. R. Hermes, T. D. Chan, J. Jackson, K. Bourne, J. H. Reed, K. Patterson, B. T. Porebski, et al. Germinal center antibody mutation trajectories are determined by rapid self/foreign discrimination. *Science*, 360(6385):223–226, 2018.

[24] D. L. Burnett, J. H. Reed, D. Christ, and C. C. Goodnow. Clonal redemption and clonal anergy as mechanisms to balance b cell tolerance and immunity. *Immunological Reviews*, 292(1):61–75, 2019.

[25] A. C. Campi-Azevedo, P. de Almeida Estevam, J. G. Coelho-dos Reis, V. Peruhype-Magalhães, G. Villela-Rezende, P. F. Quaresma, M. d. L. S. Maia, R. H. G. Farias, L. A. B. Camacho, M. da Silva Freire, et al. Subdoses of 17dd yellow fever vaccine elicit equivalent virological/immunological kinetics timeline. *BMC infectious diseases*, 14(1):1–12, 2014.

[26] A. K. Chakraborty. A perspective on the role of computational models in immunology. *Annual review of immunology*, 35:403–439, 2017.

[27] K. M. Cirelli, D. G. Carnathan, B. Nogal, J. T. Martin, O. L. Rodriguez, A. A. Upadhyay, C. A. Enemuo, E. H. Gebru, Y. Choe, F. Viviano, et al. Slow delivery immunization enhances hiv neutralizing antibody and germinal center responses via modulation of immunodominance. *Cell*, 177(5):1153–1171, 2019.

[28] K. M. Cirelli and S. Crotty. Germinal center enhancement by extended antigen availability. *Current opinion in immunology*, 47:64–69, 2017.

[29] S. H. Clarke, K. Huppi, D. Ruezinsky, L. Staudt, W. Gerhard, and M. Weigert. Inter- and intraclonal diversity in the antibody response to influenza hemagglutinin. *The Journal of experimental medicine*, 161(4):687–704, 1985.

[30] R. L. Coffman, A. Sher, and R. A. Seder. Vaccine adjuvants: putting innate immunity to work. *Immunity*, 33(4):492–503, 2010.

[31] R. T. Cox. Probability, frequency and reasonable expectation. *American journal of physics*, 14(1):1–13, 1946.

[32] S. Crotty, P. Felgner, H. Davies, J. Glidewell, L. Villarreal, and R. Ahmed. Cutting edge: long-term B cell memory in humans after smallpox vaccination. *The Journal of Immunology*, 171(10):4969–4973, 2003.

[33] J. M. Dal Porto, A. M. Haberman, G. Kelsoe, and M. J. Shlomchik. Very low affinity b cells form germinal centers, become memory b cells, and participate in secondary immune responses when higher affinity competition is reduced. *The Journal of experimental medicine*, 195(9):1215–1221, 2002.

[34] N. S. De Silva and U. Klein. Dynamics of b cells in germinal centres. *Nature reviews immunology*, 15(3):137–148, 2015.

[35] D. Depoil, R. Zaru, M. Guiraud, A. Chauveau, J. Harriague, G. Bismuth, C. Utzny, S. Müller, and S. Valitutti. Immunological synapses are versatile structures enabling selective T cell polarization. *Immunity*, 22(2):185–194, 2005.

[36] I. Dogan, B. Bertocci, V. Vilmont, F. Delbos, J. Mégret, S. Storck, C.-A. Reynaud, and J.-C. Weill. Multiple layers of B cell memory with different effector functions. *Nature immunology*, 10(12):1292, 2009.

[37] H. N. Eisen. Affinity enhancement of antibodies: how low-affinity antibodies produced early in immune responses are followed by high-affinity antibodies later and in memory B-cell responses. *Cancer immunology research*, 2(5):381–392, 2014.

[38] H. N. Eisen and G. W. Siskind. Variations in affinities of antibodies during the immune response. *Biochemistry*, 3(7):996–1008, 1964.

[39] K. Eyer, C. Castrillon, G. Chenon, J. Bibette, P. Bruhns, A. D. Griffiths, and J. Baudry. The quantitative assessment of the secreted igg repertoire after recall to evaluate the quality of immunizations. *The Journal of Immunology*, 205(4):1176–1184, 2020.

[40] K. Eyer, R. C. Doineau, C. E. Castrillon, L. Briseño-Roa, V. Menrath, G. Mottet, P. England, A. Godina, E. Brient-Litzler, C. Nizak, et al. Single-cell deep phenotyping of IgG-secreting cells for high-resolution immune monitoring. *Nature biotechnology*, 35(10):977, 2017.

[41] J. Finney, C.-H. Yeh, G. Kelsoe, and M. Kuraoka. Germinal center responses to complex antigens. *Immunological reviews*, 284(1):42–50, 2018.

[42] G. Georgiou, G. C. Ippolito, J. Beausang, C. E. Busse, H. Wardemann, and S. R. Quake. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature biotechnology*, 32(2):158–168, 2014.

[43] A. Gérard, A. Woolfe, G. Mottet, M. Reichen, C. Castrillon, V. Menrath, S. Ellouze, A. Poitou, R. Doineau, L. Briseno-Roa, et al. High-throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics. *Nature Biotechnology*, 38(6):715–721, 2020.

[44] A. D. Gitlin, C. T. Mayer, T. Y. Oliveira, Z. Shulman, M. J. Jones, A. Koren, and M. C. Nussenzweig. T cell help controls the speed of the cell cycle in germinal center b cells. *Science*, 349(6248):643–646, 2015.

[45] A. D. Gitlin, Z. Shulman, and M. C. Nussenzweig. Clonal selection in the germinal centre by regulated proliferation and hypermutation. *Nature*, 509(7502):637–640, 2014.

[46] E. A. Goidl, W. E. Paul, G. W. Siskind, and B. Benacerraf. The effect of antigen dose and time after immunization on the amount and affinity of anti-hapten antibody. *The Journal of Immunology*, 100(2):371–375, 1968.

[47] A. Gonzalez-Fernandez and C. Milstein. Low antigen dose favours selection of somatic mutants with hallmarks of antibody affinity maturation. *Immunology*, 93(2):149, 1998.

[48] B. Greenwood. The contribution of vaccination to global health: past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1645):20130433, 2014.

[49] P. J. Guerin, L. M. Næss, C. Fogg, E. Rosenqvist, L. Pinoges, F. Bajunirwe, C. Nabasumba, R. Borrow, L. O. Frøholm, S. Ghabri, et al. Immunogenicity of fractional doses of tetravalent a/c/y/w135 meningococcal polysaccharide vaccine: results from a randomized non-inferiority controlled trial in uganda. *PLoS Negl Trop Dis*, 2(12):e342, 2008.

[50] S. Han, K. Hathcock, B. Zheng, T. B. Kepler, R. Hodes, and G. Kelsoe. Cellular interaction in germinal centers. roles of cd40 ligand and b7-2 in established germinal centers. *The Journal of Immunology*, 155(2):556–567, 1995.

[51] S. Han, B. Zheng, J. Dal Porto, and G. Kelsoe. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl) acetyl. iv. affinity-dependent, antigen-driven b cell apoptosis in germinal centers as a mechanism for maintaining self-tolerance. *The Journal of experimental medicine*, 182(6):1635–1644, 1995.

[52] B. A. Heesters, P. Chatterjee, Y.-A. Kim, S. F. Gonzalez, M. P. Kuligowski, T. Kirchhausen, and M. C. Carroll. Endocytosis and recycling of immune complexes by follicular dendritic cells enhances b cell antigen binding and activation. *Immunity*, 38(6):1164–1175, 2013.

[53] B. A. Heesters, R. C. Myers, and M. C. Carroll. Follicular dendritic cells: dynamic antigen libraries. *Nature Reviews Immunology*, 14(7):495–504, 2014.

[54] J. J. Hopfield. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences*, 71(10):4135–4139, 1974.

[55] F. Horns, C. L. Dekker, and S. R. Quake. Transcriptional program of memory B cell activation, broadly binding anti-influenza antibodies, and bystander activation after vaccination revealed by single-cell transcriptomics. *bioRxiv*, page 709337, 2019.

[56] F. Horns, C. L. Dekker, and S. R. Quake. Memory b cell activation, broad anti-influenza antibodies, and bystander activation revealed by single-cell transcriptomics. *Cell Reports*, 30(3):905–913.e6, Jan. 2020.

[57] F. Horns, C. Vollmers, C. L. Dekker, and S. R. Quake. Signatures of selection in the human antibody repertoire: Selective sweeps, competing subclones, and neutral drift. *Proceedings of the National Academy of Sciences*, 116(4):1261–1266, 2019.

[58] T. Inoue, I. Moran, R. Shinnakasu, T. G. Phan, and T. Kurosaki. Generation of memory B cells and their reactivation. *Immunological reviews*, 283(1):138–149, 2018.

[59] J. Jacob, G. Kelsoe, K. Rajewsky, and U. Weiss. Intraclonal generation of antibody mutants in germinal centres. *Nature*, 354(6352):389–392, 1991.

[60] J. Jacob, J. Przylepa, C. Miller, and G. Kelsoe. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl) acetyl. III. the kinetics of V region mutation and selection in germinal center B cells. *Journal of Experimental Medicine*, 178(4):1293–1307, 1993.

[61] E. Jacobson, L. Caporale, and G. Thorbecke. Effect of thymus cell injections on germinal center formation in lymphoid tissues of nude (thymusless) mice. *Cellular immunology*, 13(3):416–430, 1974.

[62] E. T. Jaynes. *Probability theory: The logic of science*. Cambridge university press, 2003.

[63] M. Kang, T. J. Eisen, E. A. Eisen, A. K. Chakraborty, and H. N. Eisen. Affinity inequality among serum antibodies that originate in lymphoid germinal centers. *PLOS ONE*, 10(10):e0139222, Oct. 2015.

[64] T. B. Kepler and A. S. Perelson. Cyclic re-entry of germinal center b cells and the efficiency of affinity maturation. *Immunology today*, 14(8):412–415, 1993.

[65] F. Klein, R. Diskin, J. F. Scheid, C. Gaebler, H. Mouquet, I. S. Georgiev, M. Pancera, T. Zhou, R.-B. Incesu, B. Z. Fu, et al. Somatic mutations of the immunoglobulin

framework are generally required for broad and potent hiv-1 neutralization. *Cell*, 153(1):126–138, 2013.

[66] S. H. Kleinstein, Y. Louzoun, and M. J. Shlomchik. Estimating hypermutation rates from clonal tree data. *The Journal of Immunology*, 171(9):4639–4649, Oct. 2003.

[67] M. Knežević, H. Jiang, and S. Wang. Active tuning of synaptic patterns enhances immune discrimination. *Physical Review Letters*, 121(23), Dec. 2018.

[68] N. J. Kräutler, D. Suan, D. Butt, K. Bourne, J. R. Hermes, T. D. Chan, C. Sundling, W. Kaplan, P. Schofield, J. Jackson, et al. Differentiation of germinal center b cells into plasma cells is initiated by high-affinity antigen and completed by tfh cells. *Journal of Experimental Medicine*, 214(5):1259–1267, 2017.

[69] V. Krishna and K. E. Bachman. A mechanism of t cell dependent selection of antigen engaged germinal center b cells. *PloS one*, 13(8):e0200241, 2018.

[70] M. Kuraoka, A. G. Schmidt, T. Nojima, F. Feng, A. Watanabe, D. Kitamura, S. C. Harrison, T. B. Kepler, and G. Kelsoe. Complex antigens drive permissive clonal selection in germinal centers. *Immunity*, 44(3):542–552, 2016.

[71] K. Kwak, M. Akkaya, and S. K. Pierce. B cell signaling in context. *Nature immunology*, 20(8):963–969, 2019.

[72] K. Kwak, N. Quizon, H. Sohn, A. Saniee, J. Manzella-Lapeira, P. Holla, J. Brzostowski, J. Lu, H. Xie, C. Xu, et al. Intrinsic properties of human germinal center b cells set antigen affinity thresholds. *Science immunology*, 3(29), 2018.

[73] M. Lässig and M. Łuksza. Adaptive evolution: Can we read the future from a tree? *Elife*, 3:e05060, 2014.

[74] A. W. Lau and R. Brink. Selection in the germinal center. *Current Opinion in Immunology*, 63:29–34, 2020.

[75] D. Liu, H. Xu, C. Shih, Z. Wan, X. Ma, W. Ma, D. Luo, and H. Qi. T–b-cell entanglement and icosl-driven feed-forward regulation of germinal centre reaction. *Nature*, 517(7533):214–218, 2015.

[76] Y. Liu, D. Joshua, G. Williams, C. Smith, J. Gordon, and I. MacLennan. Mechanism of antigen-driven selection in germinal centres. *Nature*, 342(6252):929–931, 1989.

[77] L. L. Lu, T. J. Suscovich, S. M. Fortune, and G. Alter. Beyond binding: antibody effector functions in infectious diseases. *Nature Reviews Immunology*, 18(1):46, 2018.

[78] M. Łuksza and M. Lässig. A predictive fitness model for influenza. *Nature*, 507(7490):57–61, 2014.

[79] S. Luo and A. S. Perelson. Competitive exclusion by autologous antibodies can prevent broad hiv-1 antibodies from arising. *Proceedings of the National Academy of Sciences*, 112(37):11654–11659, 2015.

[80] W. Luo, F. Weisel, and M. J. Shlomchik. B cell receptor and cd40 signaling are rewired for synergistic induction of the c-myc transcription factor in germinal center b cells. *Immunity*, 48(2):313–326, 2018.

[81] D. J. MacKay and D. J. Mac Kay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[82] D. S. MacLean, J. D. Robertson, and M. Jay. Monitoring the retention of a protein antigen in complete freund's adjuvant, alum, and pluronic F-127 gel formulations by X-ray fluorescence. *Pharmaceutical development and technology*, 6(2):241–246, 2001.

[83] T. Mandel, R. Phipps, A. Abbot, and J. Tew. Long-term antigen retention by dendritic cells in the popliteal lymph node of immunized mice. *Immunology*, 43(2):353, 1981.

[84] T. Mandels, R. Phippsi, A. Abbot, and J. Tew. The follicular dendritic cell: long term antigen retention during immunity. *Immunological reviews*, 53(1):29–59, 1980.

[85] J. R. Mascola and D. C. Montefiori. Hiv-1: nature's master of disguise. *Nature medicine*, 9(4):393–394, 2003.

[86] C. T. Mayer, A. Gazumyan, E. E. Kara, A. D. Gitlin, J. Golijanin, C. Viant, J. Pai, T. Y. Oliveira, Q. Wang, A. Escolano, M. Medina-Ramirez, R. W. Sanders, and M. C. Nussenzweig. The microanatomic segregation of selection by apoptosis in the germinal center. *Science*, 358(6360):eaao2602, Sept. 2017.

[87] L. J. McHeyzer-Williams, P. J. Milpied, S. L. Okitsu, and M. G. McHeyzer-Williams. Class-switched memory b cells remodel BCRs within secondary germinal centers. *Nature Immunology*, 16(3):296–305, Feb. 2015.

[88] M. McHeyzer-Williams, M. McLean, P. Lalor, and G. Nossal. Antigen-driven B cell differentiation in vivo. *Journal of Experimental Medicine*, 178(1):295–307, 1993.

[89] D. McKean, K. Huppi, M. Bell, L. Staudt, W. Gerhard, and M. Weigert. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proceedings of the National Academy of Sciences*, 81(10):3180–3184, May 1984.

[90] L. Mesin, J. Ersching, and G. D. Victora. Germinal center B cell dynamics. *Immunity*, 45(3):471–482, 2016.

[91] L. Mesin, A. Schiepers, J. Ersching, A. Barbulescu, C. B. Cavazzoni, A. Angelini, T. Okada, T. Kurosaki, and G. D. Victora. Restricted clonality and limited germinal center reentry characterize memory b cell reactivation by boosting. *Cell*, 180(1):92–106, 2020.

[92] M. Meyer-Hermann. Overcoming the dichotomy of quantity and quality in antibody responses. *The Journal of Immunology*, 193(11):5414–5419, 2014.

[93] M. Meyer-Hermann. Injection of antibodies against immunodominant epitopes tunes germinal centers to generate broadly neutralizing antibodies. *Cell reports*, 29(5):1066–1073, 2019.

[94] M. Meyer-Hermann, E. Mohr, N. Pelletier, Y. Zhang, G. D. Victora, and K.-M. Toellner. A theory of germinal center B cell selection, division, and exit. *Cell reports*, 2(1):162–174, 2012.

[95] T. Mohan, Z. Berman, S.-M. Kang, and B.-Z. Wang. Sequential immunizations with a panel of hiv-1 env virus-like particles coach immune system to make broadly neutralizing antibodies. *Scientific reports*, 8(1):1–12, 2018.

[96] M. Molari, K. Eyer, J. Baudry, S. Cocco, and R. Monasson. Quantitative modeling of the effect of antigen dosage on B-cell affinity distributions in maturating germinal centers. *eLife*, 9:e55678, 2020.

[97] M. Molari, R. Monasson, and S. Cocco. Survival probability and size of lineages in antibody affinity maturation. *arXiv preprint arXiv:2010.11580*, 2020.

[98] K. Moor, M. Diard, M. E. Sellin, B. Felmy, S. Y. Wotzka, A. Toska, E. Bakkeren, M. Arnoldini, F. Bansept, A. Dal Co, et al. High-avidity iga protects the intestine by enchaining growing bacteria. *Nature*, 544(7651):498–502, 2017.

[99] D. M. Morens, D. S. Burke, and S. B. Halstead. The wages of original antigenic sin. *Emerging infectious diseases*, 16(6):1023, 2010.

[100] H. Mouquet. Antibody b cell responses in hiv-1 infection. *Trends in immunology*, 35(11):549–561, 2014.

[101] K. Murphy and C. Weaver. *Janeway's immunobiology*. Garland science, 2016.

[102] R. Murugan, L. Buchauer, G. Triller, C. Kreschel, G. Costa, G. P. Martí, K. Imkeller, C. E. Busse, S. Chakravarty, B. K. L. Sim, et al. Clonal selection drives protective memory b cell responses in controlled human malaria infection. *Science immunology*, 3(20), 2018.

[103] A. Nanni, S. Meredith, S. Gati, K. Holm, T. Harmon, and A. Ginsberg. Strengthening global vaccine access for adolescents and adults. *Vaccine*, 35(49):6823–6827, Dec. 2017.

[104] E. Natkanski, W.-Y. Lee, B. Mistry, A. Casal, J. E. Molloy, and P. Tolar. B cells use mechanical energy to discriminate antigen affinities. *Science*, 340(6140):1587–1590, May 2013.

[105] R. A. Neher and O. Hallatschek. Genealogies of rapidly adapting populations. *Proceedings of the National Academy of Sciences*, 110(2):437–442, 2013.

[106] R. A. Neher, C. A. Russell, and B. I. Shraiman. Predicting evolution from the shape of genealogical trees. *Elife*, 3:e03568, 2014.

[107] R. A. Neher and A. M. Walczak. Progress and open problems in evolutionary dynamics. *arXiv preprint arXiv:1804.07720*, 2018.

[108] J. Ninio. Kinetic amplification of enzyme discrimination. *Biochimie*, 57(5):587–595, 1975.

[109] A. Nourmohammad, J. Otwinowski, M. Łuksza, T. Mora, and A. M. Walczak. Fierce selection and interference in b-cell repertoire response to chronic hiv-1. *Molecular biology and evolution*, 36(10):2184–2194, 2019.

[110] A. Nourmohammad, J. Otwinowski, and J. B. Plotkin. Host-pathogen coevolution and the emergence of broadly neutralizing antibodies in chronic infections. *PLoS genetics*, 12(7):e1006171, 2016.

[111] C. R. Nowosad, K. M. Spillane, and P. Tolar. Germinal center b cells recognize antigen through a specialized immune synapse architecture. *Nature Immunology*, 17(7):870–877, May 2016.

[112] V. Nussenzweig and B. Benacerraf. Antihapten antibody specificity and L chain type. *Journal of Experimental Medicine*, 126(4):727–743, 1967.

[113] T. V. Obukhanych and M. C. Nussenzweig. T-independent type ii immune responses generate memory b cells. *The Journal of experimental medicine*, 203(2):305–310, 2006.

[114] M. Oprea and A. S. Perelson. Somatic mutation leads to efficient affinity maturation when centrocytes recycle back to centroblasts. *The Journal of Immunology*, 158(11):5155–5162, 1997.

[115] V. Ovchinnikov, J. E. Louveau, J. P. Barton, M. Karplus, and A. K. Chakraborty. Role of framework mutations and antibody flexibility in the evolution of broadly neutralizing antibodies. *Elife*, 7:e33038, 2018.

[116] S. Ozawa, S. Clark, A. Portnoy, S. Grewal, M. L. Stack, A. Sinha, A. Mirelman, H. Franklin, I. K. Friberg, Y. Tam, et al. Estimated economic impact of vaccinations in 73 low-and middle-income countries, 2001–2020. *Bulletin of the World Health Organization*, 95(9):629, 2017.

[117] I. Papa and C. G. Vinuesa. Synaptic interactions in germinal centers. *Frontiers in Immunology*, 9:1858, 2018.

[118] B. Pulendran, G. Kannourakis, S. Nouri, K. G. Smith, and G. Nossal. Soluble antigen can cause enhanced apoptosis of germinal-centre B cells. *Nature*, 375(6529):331, 1995.

[119] B. J. Quah, V. P. Barlow, V. McPhun, K. I. Matthaei, M. D. Hulett, and C. R. Parish. Bystander b cells rapidly acquire antigen receptors from activated b cells by membrane transfer. *Proceedings of the National Academy of Sciences*, 105(11):4259–4264, 2008.

[120] J. A. Regules, S. B. Cicatelli, J. W. Bennett, K. M. Paolino, P. S. Twomey, J. E. Moon, A. K. Kathcart, K. D. Hauns, J. L. Komisar, A. N. Qabar, et al. Fractional third and fourth dose of rts, s/as01 malaria candidate vaccine: a phase 2a controlled human malaria parasite infection and immunogenicity study. *The Journal of infectious diseases*, 214(5):762–771, 2016.

[121] S. J. Rhodes, G. M. Knight, D. E. Kirschner, R. G. White, and T. G. Evans. Dose finding for new vaccines: the role for immunostimulation/immunodynamic modelling. *Journal of theoretical biology*, 465:51–55, 2019.

[122] S. J. Rhodes, A. Zelmer, G. M. Knight, S. A. Prabowo, L. Stockdale, T. G. Evans, T. Lindenstrøm, R. G. White, and H. Fletcher. The tb vaccine h56+ ic31 dose-response curve is peaked not saturating: data generation for new mathematical modelling methods to inform vaccine dose decisions. *Vaccine*, 34(50):6285–6291, 2016.

[123] P. A. Robert, A. L. Marschall, and M. Meyer-Hermann. Induction of broadly neutralizing antibodies in germinal centre simulations. *Current opinion in biotechnology*, 51:137–145, 2018.

[124] J. Rolf, S. E. Bell, D. Kovesdi, M. L. Janas, D. R. Soond, L. M. Webb, S. Santinelli, T. Saunders, B. Hebeis, N. Killeen, et al. Phosphoinositide 3-kinase activity in t cells regulates the magnitude of the germinal center reaction. *The Journal of Immunology*, 185(7):4042–4052, 2010.

[125] G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[126] T. A. Schwickert, G. D. Victora, D. R. Fooksman, A. O. Kamphorst, M. R. Mugnier, A. D. Gitlin, M. L. Dustin, and M. C. Nussenzweig. A dynamic T cell–limited checkpoint regulates affinity-dependent B cell entry into the germinal center. *Journal of Experimental Medicine*, 208(6):1243–1252, 2011.

[127] J. S. Shaffer, P. L. Moore, M. Kardar, and A. K. Chakraborty. Optimal immunization cocktails can promote induction of broadly neutralizing Abs against highly mutable pathogens. *Proceedings of the National Academy of Sciences*, 113(45):E7039–E7048, 2016.

[128] M. Shapiro-Shelef and K. Calame. Regulation of plasma-cell development. *Nature Reviews Immunology*, 5(3):230–242, 2005.

[129] T.-A. Y. Shih, E. Meffre, M. Roederer, and M. C. Nussenzweig. Role of bcr affinity in t cell–dependent antibody responses in vivo. *Nature immunology*, 3(6):570–575, 2002.

[130] R. Shinnakasu, T. Inoue, K. Kometani, S. Moriyama, Y. Adachi, M. Nakayama, Y. Takahashi, H. Fukuyama, T. Okada, and T. Kurosaki. Regulated selection of germinal-center cells into the memory b cell compartment. *Nature immunology*, 17(7):861–869, 2016.

[131] R. Shinnakasu and T. Kurosaki. Regulation of memory B and plasma cell differentiation. *Current opinion in immunology*, 45:126–131, 2017.

[132] M. J. Shlomchik, W. Luo, and F. Weisel. Linking signaling and selection in the germinal center. *Immunological reviews*, 288(1):49–63, 2019.

[133] K. M. Shokat and C. C. Goodnow. Antigen-induced B-cell death and elimination during germinal-centre immune responses. *Nature*, 375(6529):334, 1995.

[134] Z. Shulman, A. D. Gitlin, J. S. Weinstein, B. Lainez, E. Esplugues, R. A. Flavell, J. E. Craft, and M. C. Nussenzweig. Dynamic signaling by t follicular helper cells during germinal center b cell selection. *Science*, 345(6200):1058–1062, 2014.

[135] M. K. Slifka, R. Antia, J. K. Whitmire, and R. Ahmed. Humoral immunity due to long-lived plasma cells. *Immunity*, 8(3):363–372, 1998.

[136] L. M. Sompayrac. *How the immune system works*. John Wiley & Sons, 2019.

[137] K. M. Spillane and P. Tolar. Mechanics of antigen extraction in the b cell synapse. *Molecular immunology*, 101:319–328, 2018.

[138] I. Stewart, D. Radtke, B. Phillips, S. J. McGowan, and O. Bannard. Germinal center b cells replace their antigen receptors in dark zones and fail light zone entry when immunoglobulin gene mutations are damaging. *Immunity*, 49(3):477–489, 2018.

[139] D. Suan, C. Sundling, and R. Brink. Plasma cell and memory b cell differentiation from the germinal center. *Current opinion in immunology*, 45:97–102, 2017.

[140] Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314(1-2):141–151, 1999.

[141] D. M.-Y. Sze, K.-M. Toellner, C. G. De Vinuesa, D. R. Taylor, and I. C. MacLennan. Intrinsic constraint on plasmablast growth and extrinsic limits of plasma cell survival. *The Journal of experimental medicine*, 192(6):813–822, 2000.

[142] Y. Takahashi, P. R. Dutta, D. M. Cerasoli, and G. Kelsoe. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl) acetyl. V. affinity maturation develops in two stages of clonal selection. *Journal of Experimental Medicine*, 187(6):885–895, 1998.

[143] H. H. Tam, M. B. Melo, M. Kang, J. M. Pelet, V. M. Ruda, M. H. Foley, J. K. Hu, S. Kumari, J. Crampton, A. D. Baldeon, R. W. Sanders, J. P. Moore, S. Crotty, R. Langer, D. G. Anderson, A. K. Chakraborty, and D. J. Irvine. Sustained antigen availability during germinal center initiation enhances antibody responses to vaccination. *Proceedings of the National Academy of Sciences*, 113(43):E6639–E6648, Oct. 2016.

[144] J. M. Tas, L. Mesin, G. Pasqual, S. Targ, J. T. Jacobsen, Y. M. Mano, C. S. Chen, J.-C. Weill, C.-A. Reynaud, E. P. Browne, et al. Visualizing antibody affinity maturation in germinal centers. *Science*, 351(6277):1048–1054, 2016.

[145] J. Tew and T. Mandel. Prolonged antigen half-life in the lymphoid follicles of specifically immunized mice. *Immunology*, 37(1):69, 1979.

[146] H. Toyama, S. Okada, M. Hatano, Y. Takahashi, N. Takeda, H. Ichii, T. Takemori, Y. Kuroda, and T. Tokuhisa. Memory b cells without somatic hypermutation are generated from bcl6-deficient b cells. *Immunity*, 17(3):329–339, 2002.

[147] G. D. Victora and L. Mesin. Clonal and cellular dynamics in germinal centers. *Current Opinion in Immunology*, 28:90–96, June 2014.

[148] G. D. Victora and H. Mouquet. What are the primary limitations in b-cell affinity maturation, and how much affinity maturation can we drive with vaccination? lessons from the antibody response to hiv-1. *Cold Spring Harbor perspectives in biology*, 10(5):a029389, 2018.

[149] G. D. Victora and M. C. Nussenzweig. Germinal centers. *Annual review of immunology*, 30:429–457, 2012.

[150] G. D. Victora, T. A. Schwickert, D. R. Fooksman, A. O. Kamphorst, M. Meyer-Hermann, M. L. Dustin, and M. C. Nussenzweig. Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. *Cell*, 143(4):592–605, 2010.

[151] G. D. Victora and P. C. Wilson. Germinal center selection and the antibody response to influenza. *Cell*, 163(3):545–548, 2015.

[152] Z. Wan, Y. Lin, Y. Zhao, and H. Qi. Tfh cells in bystander and cognate interactions with b cells. *Immunological reviews*, 288(1):28–36, 2019.

[153] P. Wang, C.-m. Shih, H. Qi, and Y.-h. Lan. A stochastic model of the germinal center integrating local antigen competition, individualistic t–b interactions, and b cell receptor signaling. *The Journal of Immunology*, 197(4):1169–1182, 2016.

[154] S. Wang. Optimal sequential immunization can focus antibody responses against diversity loss and distraction. *PLoS computational biology*, 13(1):e1005336, 2017.

[155] S. Wang, J. Mata-Fink, B. Kriegsman, M. Hanson, D. J. Irvine, H. N. Eisen, D. R. Burton, K. D. Wittrup, M. Kardar, and A. K. Chakraborty. Manipulating the selection forces during affinity maturation to generate cross-reactive hiv antibodies. *Cell*, 160(4):785–797, 2015.

[156] F. Weisel and M. Shlomchik. Memory b cells of mice and humans. *Annual review of immunology*, 35:255–284, 2017.

[157] F. J. Weisel, G. V. Zuccarino-Catania, M. Chikina, and M. J. Shlomchik. A temporal switch in the germinal center determines differential output of memory b and plasma cells. *Immunity*, 44(1):116–130, 2016.

[158] N. Wittenbrink, A. Klein, A. A. Weiser, J. Schuchhardt, and M. Or-Guil. Is there a typical germinal center? a large-scale immunohistological study on the cellular composition of germinal centers during the hapten-carrier–driven primary immune response in mice. *The Journal of Immunology*, 187(12):6185–6196, 2011.

[159] R. Wong and D. Bhattacharya. Basics of memory B-cell responses: lessons from and for the real world. *Immunology*, 156(2):120–129, 2019.

[160] H. Xu, X. Li, D. Liu, J. Li, X. Zhang, X. Chen, S. Hou, L. Peng, C. Xu, W. Liu, et al. Follicular t-helper cell recruitment governed by bystander b cells and icos-driven motility. *Nature*, 496(7446):523–527, 2013.

[161] J. Zhang and E. I. Shakhnovich. Optimality of mutation and selection in germinal centers. *PLoS computational biology*, 6(6):e1000800, 2010.

[162] Y. Zhang, M. Meyer-Hermann, L. A. George, M. T. Figge, M. Khan, M. Goodall, S. P. Young, A. Reynolds, F. Falciani, A. Waisman, et al. Germinal center B cells govern their own fate via antibody feedback. *Journal of Experimental Medicine*, 210(3):457–464, 2013.

RÉSUMÉ

La Maturation d'Affinité (MA) est le processus biologique grâce auquel notre système immunitaire génère de puissants anticorps contre les nouveaux agents pathogènes rencontrés. Ce processus est également à la base de la vaccination, l'une des procédures médicales les plus efficaces jamais mises au point, qui permet de sauver des millions de vies chaque année. La MA présentent encore de nombreuses questions ouvertes, dont les réponses peuvent améliorer la manière dont nous vaccinons. Les mécanismes à la base de la MA sont extrêmement complexes, avec des interactions non linéaires entre nombreux cellules différentes. Dans ce contexte, les modèles théoriques et l'inférence Bayésienne sont des outils précieux pour relier les hypothèses qualitatives aux descriptions quantitatives et extraire informations des données expérimentales. Dans ce manuscrit, nous utilisons ces outils pour aborder certaines questions ouvertes, comme l'effet du dosage de l'antigène sur la qualité de la vaccination.

MOTS CLÉS

Physique Statistique, Biophysique, Immunologie, Inférence Bayésienne, Maturation d'Affinité

ABSTRACT

Affinity Maturation (AM) is the biological process through which our Immune System generates potent Antibodies (Abs) against newly encountered pathogens. This process is also at the base of vaccination, one of the most successful and cost-effective medical procedures ever developed, responsible for saving millions of lives every year. AM still present many open questions, whose answers have the potential of improving the way we vaccinate. The mechanisms at the base of AM are extremely complex, involving non-linear interactions between many different cellular agents. In this context theoretical models and Bayesian Inference are invaluable tools, respectively to link qualitative hypothesis to quantitative descriptions and to extract information from experimental data. In this manuscript we make use of these tools to tackle some of the open questions, such as the non-trivial effect of Ag dosage on the outcome of vaccination.

KEYWORDS

Statistical Physics, Bayesian Inference, Biophysics, Immunology, Affinity Maturation