Molecular Binding and Antigen Dynamics at the Onset of the Immune Response

UNIVERSITÄT
ZU KÖLN

Doctoral thesis
for
the award of the doctoral degree
of the Faculty of Mathematics and Natural Sciences
of the University of Cologne

submitted by

Roberto Morán Tovar
from Bogotá, Colombia.

accepted in the year 2025

Contents

Preface ix

I Introduction 1

1 Motivation 3

2 The foundations 5

II The Activation Of The B Cell Repertoire 27

3 Primary infection: A Luria-Delbrück activation model 31

4 Vaccination: A minimal spatiotemporal model 35

5 Memory response: *In vivo* protection 39

III Epidemiology of infectious outbreaks 63

6 Infectious outbreaks: Establishment size and surveillance 65

IV Closing 67

7 Conclusions and Perspectives 69

List of Figures 73

Bibliography 75

Para Maria y Roberto

Acknowledgements

First, I would like to acknowledge the contribution of my parents in my formative process. They taught me that education is sacred. They also taught me that the only thing more important than education is being a good person. This dissertation is dedicated to them. I would like to thank Angela, my life partner and best friend. Life is just incredible, every day, next to you. Thanks to my family and friends who made my life as an immigrant more pleasant and my visits to Colombia always revitalizing.

This thesis contains the results of hours and hours of never-ending scientific conversations. The main interlocutor has been Michael Lässig. Thank you for being my advisor and mentor. Your intuition and excitement for our research questions have been a great inspiration. Thanks also for allowing me to contradict you many times.

Of course, none of this would have been possible without the continuous input of my research group. In this case, I would like to thank life for giving me the opportunity to share this part of my life with such an incredible group of scientists, human beings, and after all this time, good friends. Denny, Matthijs, Leon, Laura, Dennis and Malancha. My enormous thanks to Christa, who made possible the impossible, and to Stephan, for being the best when Christa retired and nobody knew how things would go on. Special thanks to the CRC1310 - Predictability in Evolution, for funding my position and for creating such a stimulating environment to do science.

Abstract

B cells are a key component of adaptive immunity in vertebrates. They recognize pathogens and trigger an immune response, which involves antibody secretion and the formation of immunological memory. In the context of immune receptor repertoires, new experimental techniques have provided us with unprecedented high-throughput genomic and phenotypic data. For instance, it has been observed that upon immunization, the immune system produces potent, specific and fast recognition of antigens, while maintaining a spectrum of genetically distinct activated B cell lineages. Moreover, heavy-tailed clone size distributions have been observed in different populations of B cells. To explain some of these observations, I have developed a spatiotemporal model of molecular recognition of antigens by B cells. Applied to various immunization processes, such as infections or vaccination, the model studies the collective response of B cells, driven by proliferation and diffusion of the antigen. The fundamental molecular interaction between antigens and B cells is mediated by a general kinetic proofreading scheme. Furthermore, drawing inspiration from the density of states in statistical physics, I characterize the diversity of the B cell population in terms of its functionally-determined density of receptors. This approach allows us to study how the immune system encodes in large but limited receptor repertoires the capacity to recognize virtually any pathogen that threatens the host. The molecular recognition process, driven by exponential growth of the antigen, can be mapped onto a generalized Luria-Delbrück process, akin to the seminal fluctuation experiment in microbial evolution. Overall, this model predicts key biological and medical phenomena such as the existence of primary elite neutralizers and the age-related decline in *de novo* responses. Finally, application of the model to memory B cell responses can be used to construct mechanistic *in vivo* protective functions that have only been heuristically determined.

Preface

Five years ago, I was discussing with Michael Lässig in the coffee room of the Physics Institute about what would be my PhD project. His research group has been working on the evolution of viruses for the past decade, specially focused on rapidly evolving viruses like Influenza. For him, it was a completely open problem, why it is likely that we get an Influenza infection every few years, if each time we get sick, we generate protective immunological memory. What type of transition is this? From his research on the virus itself, he knew that each season, when Influenza produces a wave of new infections in the globe, it accumulates mutations that change the regions of the proteins that are targeted by immunological memory. After all, immunological memory relies on the molecular recognition of the viral proteins. But, what was happening on the immune system side?

In a nutshell, the adaptive immune system in humans faces viral infections in the following way. Before the infection, the host is equipped with an immense diversity of receptors that could potentially bind the virus. The diversity is so large that practically every virus will be recognized. After a primary infection, the right receptors amplify their copy number and some of them also get even better at binding. In this way, when the host is exposed to the virus again, those receptors will more readily neutralize the viral activity. Then, when the virus mutates, these previously selected receptors loose their recognition capacity, we can get sick again, and it is necessary to find new good receptors in the background diversity. In our minds, this problem, and in particular the transition between being protected and not, was an energy-entropy trade-off problem. When is the recognition capacity of a few good receptors better than the capacity of billions of mediocre ones? This could be a nice research problem for a PhD.

In this moment, a terrible pandemic of a newly emerging virus put the whole world in pause and for many, changed radically their lives. Now working almost exclusively from home and in front of video calls all the day, for obvious reasons our research group became occupied with the pandemic. Those were rare times. On one hand, it was exciting to see almost a unique viral evolutionary experiment on live. On the other hand, many people around the world suffered not only from infection, but also from the economic and social burdens of the pandemic. In retrospective, I learned two important things after the pandemic. One, basic research on viruses, immune system and epidemiology is extremely relevant. Second, the only way to overcome difficult times, when economic and social stability is at risk, is to stay together as a community. The pandemic triggered a collaboration with the group

of Florian Klein from the University Hospital in Cologne and I got involved with a massive project for daily preventive testing in daycare facilities in the whole region of North-Rhein Westfalen. Although this time put me a bit aside from our immune problem, it resulted in a very interesting publication about the role of stochasticity in small infectious outbreaks and its consequences for optimal preventive surveillance. I did not want to leave this epidemiological interlude outside this dissertation and, therefore, the last chapter is dedicated to it.

After a few months of reviewing the state of the art, I realized that the original question from that afternoon in the coffee room was still way beyond the scope. And the main reason was that a more fundamental question was also still an open problem. Namely, how are the good receptors selected in the first place? So, before understanding how we loose immunological memory, it was important to answer how do we create immunological memory. I could not find a satisfactory answer in current literature. However, the immune recognition problem brought us to the very foundations of some biophysical concepts.

The partial answer to the immune recognition problem presented in this dissertation goes back to the seminal fluctuation experiment proposed by M. Delbrück and S. Luria in the middle of the 20th century. We found the original idea of this experiment being relevant in the interaction between the immune system and viruses during acute infections. And because the interaction relies on molecular binding, very soon we realized that the idea of kinetic proofreading proposed by J. Hopfield in the 70s was also relevant. The interplay between this two important concepts, namely, evolution of expanding populations and nonequilibrium molecular recognition, mark the foundations of this dissertation. As it is always the case in science, we are standing in the shoulder of giants.

These 5 years of research have raised more new questions than answers, and I am glad about that. I believe, from my still poor experience and ingenuity, that the next decades will witness an explosion of new research at the interface between quantitative immunology and biophysics. Many of the questions have been raised already more than a century ago. But the access to new experimental technologies and biophysical frameworks will provide the right tools to finally start answering them. I hope that the ideas presented here will contribute to this enterprise.

This thesis consists of four parts. Part I, dedicated to the introduction of the research project, consist of two chapters. In Chapter 1, I provide a brief motivation to the field and the main problem. In Chapter 2, I introduce three important topics that together form the main conceptual toolbox of my research: The topics are the *Luria-Delbrück experiment*, *Kinetic proofreading* and an overview on the *Adaptive immune system*. In the first two topics, I present the main concepts and sketch the mathematical models behind them. In the last topic, I present a brief description of the biology of the adaptive immune response in humans (and mice, although it is actually shared by all vertebrates). I also review some of the basic history of immunology and, more specifically, of the problem of specific immune acquisition. In this way, I do not only place my research questions in the context of about a century of the scientific development, but I also motivate the relevance of the topics and questions discussed here.

Part II of the thesis contains three chapters in which I develop a quantitative theory of the activation of a repertoire of B cell lineages caused by exposition to antigens. Chapter 3 formulates a minimal model of antigen-B cell molecular interaction and its consequences for the rapid and efficient response at the repertoire level to proliferating antigens during infections. In Chapter 4 I extend this theory to a spatiotemporal theory in which antigens not only replicate but also diffuse. This extension of the theory allows for the treatment of the case of the vaccination. Chapter 5 discusses the activation model in the context of recall infections, in which acquired immunity is challenge by the same or similar antigens. Part III consists of a single chapter. In Chapter 6, I present my the epidemiological interlude on stochasticity in infectious outbreaks. Part IV is the closure. I summarized the main results of my investigation, I state the main conclusions and develop the perspectives that emerge from this thesis.

April, 2025, Cologne
Roberto Morán Tovar

La línea consta de un número infinito de puntos; el plano de una número infinito de lineas; el volumen de un número infinito de planos; el hipervolumen de un número infinito de volúmenes ... No, decididamente no es éste, *more geométrico*, el mejor modo de iniciar mi relato. Afirmar que es verídico es ahora una convención de todo relato fantástico; el mio, sin embargo, es verídico

---

El libro de arena, J. L. Borges

Part I

Introduction

Chapter 1

Motivation

1.1 A young physicist looks at the adaptive immune system

Life is a nonequilibrium state of matter in which the dissipation of energy is harvested by organisms to self-replicate. Moreover, since the first self-replication event occurred at the origin of life on Earth, evolution, driven by the force of natural selection (and also just by pure chance), has created an intricate network that connects all forms of life to that first event. In words of Max Delbrück, ≪...there are no "absolute phenomena" in biology. Everything is time bound and space bound. The animal or plant or micro-organism [a physicist] is working with is but a link in an evolutionary chain of changing forms, none of which has any permanent validity. Even the molecular species and the chemical reactions which [the physicist] encounters are the fashions of today to be replaced by others as evolution goes on.≫[1] Perhaps for this and related reasons, for a long time the interest of physicists in the study of biological systems remained limited to a few concrete examples.

At the beginning of the 20th century, following the successful development of the theories of special and general relativity, and quantum mechanics, many physicists turned their attention to biological systems, in the search for a source of new fundamental physical laws. There was the hope that living matter, with its distinctive capacity to evade its own relaxation to a state of equilibrium and to self-replicate, might safeguard hitherto unknown laws of nature. This interest was further fueled by the advent of the theory of genetics, which provided a microscopic and material basis for heredity, thereby establishing a physical entity, the gene, upon which evolution could act. The venture of physicists into the realm of biology was very diverse, including topics such as phage-bacteria interactions, crystallography and DNA structure. All of this eventually played a fundamental role in the development of molecular biology. An important topic, which we will return to in the next sections, was the problem of antibody-based immunity.

Evolution has been shown to generate complexity through various mecha-

nisms [2, 3, 4, 5]. This complexity, in turn, has created new specific demands for living organisms. Sensing the environment, recognizing signals, and processing information are critical tasks that organisms face all the time. As a consequence of increasing complexity, many of these processes involve the interaction of different levels of biological organization. From the coordinated molecular interactions that make up cells, to aggregates of cells that form multicellular organisms, to populations of phylogenetically related individuals. This vast complexity, combined with the nonequilibrium aspect of biological systems, has proven challenging to study [6]. On the other hand, our understanding of nonequilibrium systems subjected to constant energy flows has become well established in recent decades, with general principles for the analysis of stationary states, linear responses, and fluctuation- dissipation relations. Interestingly, this research has been stimulated by the study of simple biological systems, which often exhibit phenomena not found in inanimate matter.

In this context, the immune system is an excellent model to study the fundamental properties of non-steady state nonequilibrium physics, for which such general principles remain unknown. To recognize threats, immune cells integrate information from complex networks of signaling receptors driven by proliferating signals in the host, such as pathogens and cancer cells. It is therefore important to understand the mechanisms by which the immune system harnesses energy dissipation and time-varying input signals to enhance its recognition and protection capabilities, as well as the physical constraints under which it operates and evolves.

Finally, with the adaptive immune system, evolution has created a molecular machinery capable of effectively recognizing any foreign substance that enters our bodies, while at the same time being able to tolerate the substances that make up our own cells. Part of the solution is an immense diversity of receptors. The diversity is so large that it is fundamentally a statistical problem, and statistical physics offers a suitable initial toolbox to approach it. Concepts such as density of states, entropy, temperature, information, disorder, and fluctuations are cornerstones for studying this problem. After all, the proper functioning of the immune system relies on billions of different receptors that are always ready to bind and respond. Therefore, it is important to understand the molecular mechanisms by which energy dissipation can be harnessed by the immune system to achieve this exquisitely high specificity under the constraints imposed by the diversity of receptors itself.

Although much progress has been made in our understanding of adaptive immunity, many questions remain open. How do immune cells integrate information from multiple interacting receptors and multiple signals? How do they generate collective behavior and mount effective immune responses? I believe that answering these questions will shed light on new mechanisms by which living systems adapt their behavior to deal with transient driving forces, as is the case of the immune system during acute infections. Overall, I believe that the study of the adaptive immune system could greatly contribute to the advancement of fundamental physics.

Chapter 2

The foundations

This chapter contains three sections in which we discuss relevant topics that will serve to understand the models and the results of the main research part of the thesis. The first two sections discuss paradigmatic models in the field of biophysics: the Luria-Delbrück fluctuation experiment and kinetic proofreading. These sections do not contain exhaustive reviews of the topics, but rather introduce the main concepts, sketch the main calculations, and discuss their historical influence and relevance in the thesis. The third section introduces the adaptive immune system, with emphasis on the different ways in which B cells contribute to the immune response against viral antigens. In addition, it provides a historical revision of the problem of antigen recognition by the immune system. In this way, I introduce the main research questions of the dissertation.

2.1 The Luria-Delbrück experiment

2.1.1 The origin of natural variation

This beautiful three-day experiment was one of the milestones in the natural sciences of the 20th century. For «for their discoveries concerning the replication mechanism and the genetic structure of viruses»[7] Max Delbrück and Salvador E. Luria, together with Alfred Hershey, were awarded the Nobel Prize in Physiology or Medicine in 1969. To explain the main ideas and findings of this seminal work, let us come back to the original 1943 paper `MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE` [8]. Let us dive into the first paragraph and the first sentence of the second paragraph, where we will readily understand the question that the experiment was trying to answer. I have taken the liberty of quoting and commenting on the text literally, line by line.

- «WHEN a pure bacterial culture is attacked by a bacterial virus, the culture will clear after a few hours due to destruction of the sensitive cells by the virus. However, after further incubation for a few hours, or sometimes days, the culture will often become turbid again, due to the growth of a bacterial

variant which is resistant to the action of the virus.≫

Bacteria (*E. coli* B in this case) show at least two different traits regarding their ability to survive in the presence of a given virus (phage T1 in this case). This observation also suggest that the resistant trait is perhaps uncommon compared to the non-resistant one.

- ≪This variant can be isolated and freed from the virus and will in many cases retain its resistance to the action of the virus even if subcultured through many generations in the absence of the virus.≫

  Such traits, specially the resistant trait can be inherited to new generations and will persist even in the absence of the virus.

- ≪While the sensitive strain adsorbed the virus readily, the resistant variant will generally not show any affinity to it.≫

  This suggests that in this case, there are not a continuous set of traits from resistance and non-resistance, but that there are two discrete trait in the game.

- ≪The resistant bacterial variants appear readily in cultures grown from a single cell. They were, therefore, certainly not present when the culture was started.≫

  The resistant trait is acquired at some point during cell culture! How?

When I first read the original paper, I was struck by the simplicity with which the authors introduced the research question. And I was even more astonished to realize that this was one of the fundamental questions concerning the theory of evolution. (I was also impressed by the fact that the paper cited only 9 references, but that is a matter for another discussion). Some would even say that this was the question whose answer would support either a Lamarckian or a Darwinian view of evolution.

The alternative models Although DNA was only established as the hereditary material after the work of Franklin, Watson and Crick in 1953, the concepts of heredity and mutations were accepted in the scientific community by this time. Genes were known to be indivisible entities located in chromosomes. Two models on how the mutant trait is acquired were in force at that time. Again, in the words of of the authors [8]

- ≪D'HERELLE(1926) and many other investigators believed that the virus by direct action induced the resistant variants.≫

  This model assumes that mutations occur by some sort of pressure to change and adapt. In this case, the presence of a viral infection would induce bacteria to generate mutants to acquire resistance.

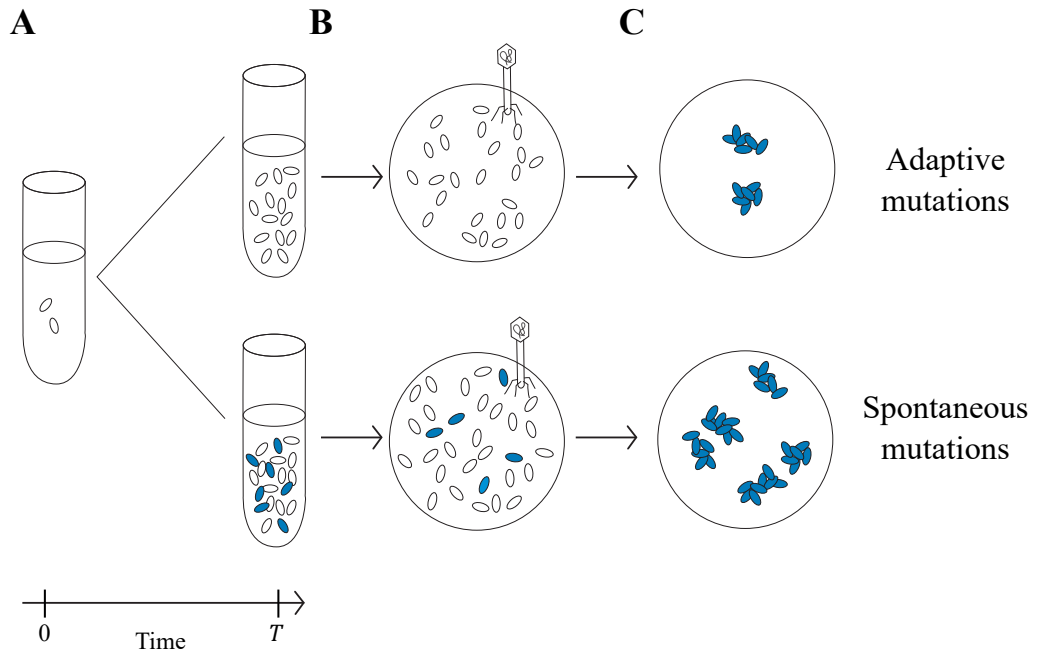


Fig. 2.1. Luria-Delbrück fluctuation experiment I Here we sketch the 1943 Luria-Delbrück experiment in three basic steps. **A** Bacteria are grown from an single cell and for a specific time T. **B** At time T, a sample from the culture is plated in dishes that contain phages. **C** Survival colonies are counted in the plates. Here the two alternative hypothesis are described. (top) Mutations occur in bacteria after the transfer to plates as an adaptation to the attack of phages. (bottom) Mutations occur spontaneously all the time during the experiment, including the time before the transfer to plates.

- ≪GRATIA(1921), BURNET(I929), and others, on the other hand, believed that the resistant bacterial variants are produced by mutation in the culture prior to the addition of virus. The virus merely brings the variants into prominence by eliminating all sensitive bacteria.≫

  This model assumes that mutations occur spontaneously all the time and not as an adaptive response to any pressure. They create natural variation within a population. When survival pressures occur, the response is determined by the existing variation. In this case, the presence of a viral infection selects existing resistant variants.

2.1.2 The experiment

When Max Delbrück and Salvador Luria first met in 1940, they were both actively working on bacterial growth under the attack of viral infection [1]. Soon after they began working together, they became interested in the rise of *secondary growth*, namely, the growth of resistant bacterial variants after phage infection. The possibility that secondary growth was the result of spontaneous mutants occurring prior to exposure to the phage attack was now on the table.

According to Luria [1], the first written statement about the experiment is found in one of his letter to Delbrück dated January 20, 1943: ≪I though that a clean cut experiment would be to find out how the fluctuations in the number of [T1 phage]-resistants depend on the culture from which they come. That is: If I plate with [T1 phage] ten samples of the *same* culture of B[acteria], I find numbers of resistants which fluctuates according to Poisson's Law. If I plate 10 samples of 10 *different* cultures of B[acteria], all containing the same amount of B[acteria], I find larger fluctuations. If the resistant were produced on the plate, after contact with [T1 phage], they should show the same fluctuations in both cases≫ [1].

Let us break down and sketch Luria's experiment proposal. The experimental setup is shown in Figure 2.1 and was the following:

1. Grow up bacteria starting from a single non-resistant cell for a controlled time T. The introduction of the time variable is fundamental here.
2. At time T, plate cultured bacteria together with the phage. Let them interact and wait until colonies from resistant cells may form.
3. Count the number of resistant colonies.
4. Repeat steps 1-3 and collect statistics about the number of resistant colonies in each plate.

The final statistics of the number of colonies in the plates should disentangle the debate between the two initial hypothesis. Let us see why.

2.1.3 The basic theoretical model

Here I sketch the logic of the theory of the problem as presented in the original paper by Luria and Delbrück, with some extra considerations discussed in the research group of M. Lässig.

Clonal dynamics of mutants in an expanding population Let us consider that bacteria in cell culture proliferate at a rate:

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = \lambda N(t) \tag{2.1}$$

where $N(t)$ denotes the total number of bacterial cells. Assuming that each cell culture starts from a single cell, we have that

$$N(t) = e^{\lambda t} \tag{2.2}$$

The basic assumption of the spontaneous mutations model is that, at any point in time during growth, bacteria have a mutation probability μ per cell per unit of time. Thus, the total number of new mutations occurring at time t is given by

$$\frac{\mathrm{d}m(t)}{\mathrm{d}t} = \mu N(t) \tag{2.3}$$

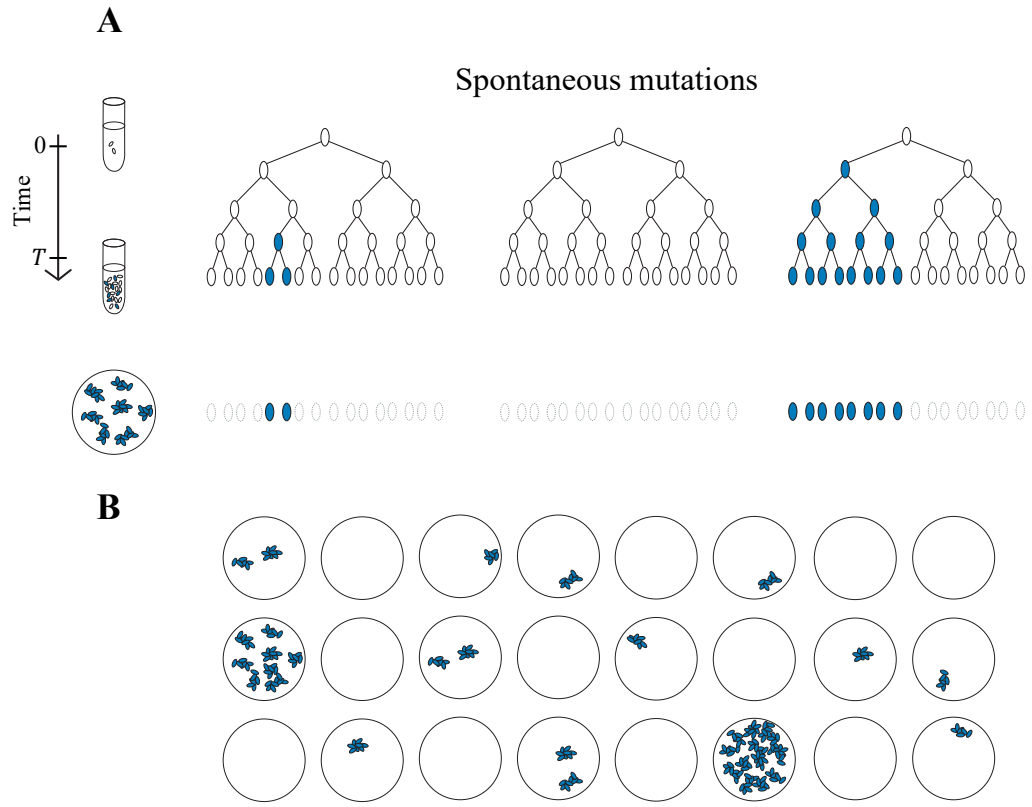


Fig. 2.2. Luria-Delbrück fluctuation experiment II A During the initial time of bacterial growth, mutants are likely to appear proportionally to the wild-type population size. The time of appearance will determine the final frequency on the population. **B** The spontaneous mutation model predicts large fluctuations in the number of growing survival colonies after exposure to phages. We expect many plates with a number of colonies closed to the expected value and a few jackpots with many colonies.

where $m(t)$ denotes the total number of mutation events. The total number of mutations that have occurred up to time t is given by

$$m(t) = \frac{\mu}{\lambda}\left(N(t) - 1\right) \approx \frac{\mu}{\lambda} N(t) \tag{2.4}$$

Let $M(t)$ be the average total number of mutant cells at time t. The rate of change of this number has two terms: i) the production of new mutants and ii) the growth of existing mutants. Assuming that mutants grow at the same rate as the unmutated cells (neutral mutation), we obtain for the growth of the average total number of mutants

$$\frac{\mathrm{d}M(t)}{\mathrm{d}t} = \mu N(t) + \lambda M(t) \tag{2.5}$$

using the integrating factor e^{-t} and assuming that $M(t = 0) = 0$, the solution to this equation is given by

$$M(t) = \mu N(t) t \tag{2.6}$$

From here it is interesting to see that the average total fraction of mutants in the population increases linearly in time

$$\frac{M(t)}{N(t)} = \mu t \tag{2.7}$$

A key observation at this point is that all mutants cells can be grouped in clones, which correspond to all cells deriving from a single mutation event. It is possible to have clones within clones defined by mutations occurring in an already mutated cell. For the moment, because of the nature of the problem investigated here (phage resistance mutation neglecting back mutation), we only consider mutants that are produce from the unmutated population. Additionally, we can group clone by their age, or time of appearance, τ. Let $n_\tau(t)$ be the expected clone size of a clone of age τ given by

$$n_\tau(t) = e^{\lambda(t-\tau)} \quad \text{for} \quad t > \tau \tag{2.8}$$

and let us calculate the expected total number of mutants cells of the same age (belonging to clones of the same age)

$$n_\tau(t) \cdot \frac{\mathrm{d}m(\tau)}{\mathrm{d}t} = \mu N(t) \tag{2.9}$$

Interestingly, this number does not depend on the age. This means that at a given observation time t, we would expect to find, on average, the same number of mutants from each past generation, either the first generation after the culture was started, or the last generation before the observation. This result follows from the fact that the expected clone size of each clone increases exponentially with the age, but the production of new clones decreases exponentially with the age. An intuitive explanation is the following: mutant cells produced early in the culture will produce large clones by the time of observation, but the probability of their occurrence is low. Conversely, mutant cells produced in later generations will produce smaller clones, but they are more likely to occur.

Equation 2.6 shows the average number of mutants at time t. Now, let us estimate the variance of the number of mutants. To do that, let us follow original Delbrück reasoning. Consider a time interval $[\tau, \tau + \mathrm{d}\tau]$ with $\tau < t$. The number of mutants produced in this time interval is a Poisson process with average and variance given by $\mu N(\tau)\mathrm{d}\tau$. But we are interested not in the number of new mutants but in the number of cells emerging from these mutation events. All mutants that emerged at time t' will expand a factor $e^{t-\tau}$ by time τ. We can then assume that the distribution of the number of mutants of age τ has average and variance given by

$$\mathrm{d}M = \mu e^{\lambda t}\mathrm{d}\tau \quad \text{and} \quad \mathrm{dVar}M = \mu e^{2\lambda t}e^{-\lambda\tau}\mathrm{d}\tau \tag{2.10}$$

We can now sum over the variable τ in the interval $[0, t]$. We obtain

$$M = \mu N(t)t \quad \text{and} \quad \mathrm{Var}M = \frac{\mu}{\lambda}N(t)[N(t) - 1] \tag{2.11}$$

The average number of mutants cells is the same as Equation 2.6. We can now compare variance and average as

$$\frac{\mathrm{Var}M}{M} = \frac{N(t)-1}{\lambda t}. \tag{2.12}$$

This calculation suggests that the spontaneous mutations hypothesis predicts a larger relative variance as the Poisson process of the adaptive mutations hypothesis.

The clone-size spectrum We can ask what is the expected number of mutant clones with clone size larger than n at time t, $\mathcal{M}(n,t)$. To answer this question we recall the clone size of a clone of age τ given by Equation 2.8. This can be rewritten as

$$\tau = t - \frac{1}{\lambda}\log n_\tau \tag{2.13}$$

All clones older than τ have on average clone size larger than n_τ. Therefore, to estimate the expected number of mutant clones with clone size larger than n_τ, we need to count the number of clones with age $\tau' < \tau$. This is given by Equation 2.4 evaluated at τ

$$\mathcal{M}(n_\tau, t) = \begin{cases} m(\tau) & \text{for} \quad \tau < t \\ m(t) & \text{for} \quad \tau > t \end{cases} \tag{2.14}$$

Using Equation 2.13 in the previous expression we obtain

$$\mathcal{M}(n,t) = \frac{\mu}{\lambda}N(t) \begin{cases} n^{-1} & \text{for} \quad n > 1 \\ 1 & \text{for} \quad n < 1 \end{cases} \tag{2.15}$$

where we have now omitted the dependence in τ. Note that here $n < N(t)$ for all t. From $\mathcal{M}$ we can compute the clone-size spectrum $P(n,t)$ of mutant clones as

$$P(n,t) = -\frac{\mathrm{d}\mathcal{M}(n,t)}{\mathrm{d}n} = \frac{\mu}{\lambda}N(t) \begin{cases} n^{-2} & \text{for} \quad n > 1 \\ 0 & \text{for} \quad n < 1 \end{cases} \tag{2.16}$$

2.1.4 Luria-Delbrück dynamics in the immune system

Although Luria and Delbrück, and many scientists since then, were more concerned about the total mutant population size spectrum, the clone-size spectrum results also useful for various reasons. New technological advances allows nowadays to track single mutations and their progeny in microbial populations[9]. With such a high resolution, we can now investigate $P(n,t)$ and learn about the evolution under expanding populations.

As shown in Equation 2.16, the distribution of clone size shows a heavy tail of large clone size that decays as a power law. This means that large clones are not exponentially rare, as is the case, for example, of Gaussian distributions. These large clones are often called *jackpot* events and are a hallmark of the Luria-Delbrück experiment. The existence of this heavy tail and of jackpots explain the large relative variance calculated in Equation 2.12.

Interestingly, power-law distributions have also been observed in the clonal composition of B cells in the immune system [10, 11, 12]. Although different explanations have been proposed for this property of the immune system [13, 14, 15], we argue in the work presented below that the origin of power laws in the clone-size distribution of B cells can have similar origin as the one of mutants in the fluctuation experiment described above. As we will study in the next chapters, the immune response to acute infections resembles the dynamics of mutants in the Luria-Delbrück setup. This will be one of the main topics in Chapter 3.

2.2 Kinetic proofreading

2.2.1 Discrimination in equilibrium

Living systems are constantly transforming energy far from thermodynamic equilibrium. Nevertheless, there are many examples where the assumption of thermodynamic equilibrium has been successfully applied to study problems in biological systems. Examples include the folding of proteins, the activity of ion channels, the regulation of DNA by transcription factors, and the catalytic activity of enzymes. Due to an appropriate separation of time scales, it is common to model these systems as consisting of a few meso-states in contact with a heat bath or other types of reservoirs. Despite the high complexity and diverse driving forces that affect such systems, those equilibrium models are not only good and handy descriptions of the real systems, but have also allowed to accurately predict quantities like, for example, the half-maximal inhibitory concentration (IC50) of antibody-antigen binding dynamics.

This section concerns about the problem of specificity in enzyme catalysis. Let us consider the perhaps simplest case of biochemical discrimination. An enzyme $\mathcal{E}$ catalyzes substrate S in solution. A nice way to picture the catalytic process is by the state of the enzyme

$$\mathcal{E} \quad \underset{k_{\text{on}}[S]}{\overset{k_{\text{off}}+\omega}{\rightleftharpoons}} \quad \mathcal{E}S \tag{2.17}$$

where $\mathcal{E}$ is the state of free enzyme, $\mathcal{E}S$ is the substrate-enzyme complex, $k_{\text{on}}[S]$ is complex formation rate, k_{off} the dissociation rate, and ω the catalysis rate. Here we are assuming that after catalysis, the enzyme comes back directly to its free state. In steady-state, the concentration of associated complex $[\mathcal{E}S]$ is given by

$$[\mathcal{E}S] = \frac{k_{\text{on}}[S]}{k_{\text{off}}+\omega}[\mathcal{E}] \tag{2.18}$$

To avoid cumbersome notation, all calculations are in the steady-state unless otherwise noted. The total catalytic rate is given by

$$[\mathcal{E}S]\cdot\omega \tag{2.19}$$

Now assume that a second undesired substrate S' is also present in solution at the same concentration as S, and that its properties are also similar to those of S. In particular, assume that the association rate with the enzyme and the further catalysis rate are the same as for S, but that the substrate S dwells longer in the complex with the enzyme than the substrate S'. So we have that $k_{\text{off}} < k'_{\text{off}}$. The complete catalytic process is described by

$$\mathcal{E}S' \underset{k_{\text{on}}[S']}{\overset{k'_{\text{off}}+\omega}{\rightleftharpoons}} \mathcal{E} \underset{k_{\text{on}}[S]}{\overset{k_{\text{off}}+\omega}{\rightleftharpoons}} \mathcal{E}S \tag{2.20}$$

and in analogy with the analysis for S, the total catalytic rate given by

$$[\mathcal{E}S'] \cdot \omega = \frac{k_{\text{on}}[S']}{k'_{\text{off}} + \omega}[\mathcal{E}] \tag{2.21}$$

If we assume that the desired product is the one formed after catalysis of S and the undesired after catalysis of S', we define the error fraction, ς, as the ratio between the rate of undesired product formation over the rate of desired product formation

$$\varsigma = \frac{k_{\text{off}} + \omega}{k'_{\text{off}} + \omega} \tag{2.22}$$

The minimum error fraction is achievable when the catalytic rate is much smaller than the dissociation rates

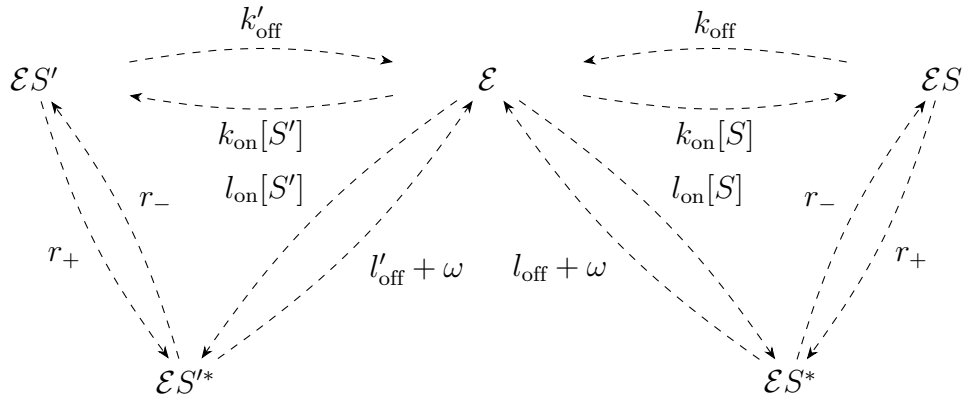
$$\varsigma \longrightarrow \varsigma_0 = \frac{k_{\text{off}}}{k'_{\text{off}}} = \frac{K}{K'} \equiv e^{-\Delta G_{SS'}} \quad \text{for} \quad \omega \ll k_{\text{off}}, k'_{\text{off}} \tag{2.23}$$

with $K \equiv k_{\text{off}}/k_{\text{on}}$ and $\Delta G_{SS'}$ the difference in the standard free energy change for the desired and undesired substrate-enzyme binding in $k_B T$ units. In this case, the error fraction is upper bounded by $\Delta G_{SS'}$. Interestingly, this error fraction shows a trade-off between specificity and catalytic speed. The minimum error, or maximum specificity, is achieved at the expense of a slower reaction.

The Pauling's problem In the late 50's of the last century, when the central dogma of molecular biology had just taken off, L. Pauling posed the problem of accuracy in the synthesis of proteins [16]. He knew that proteins are synthesized by the catalytic activity of enzymes. According to Pauling, this process must work by distinguishing different amino acids (the monomers that make up proteins) on the basis of molecular interaction. But he also knew that some amino acids can be very similar to each other and therefore interfere with the enzymatic reaction. This will promote errors in the process and affect in the final composition of the protein. Based on thermodynamic equilibrium assumptions about this chemical reaction, Pauling predicted errors that greatly exceeded experimental observations at the time [17]. In retrospect, this was a great, but not surprising observation. After all, the synthesis of proteins take place inside cells, and we know that living things are constantly challenging thermodynamic equilibrium.

2.2.2 The Hopfield-Ninio Model

In the 70s, John Hopfield [18] and Jaques Ninio [19] independently proposed a model that could beat the upper bound described above an achieved higher molecular discriminatory power. The Hopfield-Ninio Model consists of an enzymatic reaction with a two-step enzyme-substrate complex preceding catalysis. The complete enzymatic reaction is now given by



(2.24)

where it is assumed that the reactions between the two states of the complex, r_+ and r_-, are insensitive to the difference between S and S'. The expression for ς in this case is given by

$$\varsigma = \frac{[k'_{\text{off}} l_{\text{on}}[S'] + r_+(k_{\text{on}}[S'] + l_{\text{on}}[S'])]\,[r_+(l_{\text{off}} + \omega) + k_{\text{off}}(l_{\text{off}} + r_- + \omega)]}{[k_{\text{off}} l_{\text{on}}[S] + r_+(k_{\text{on}}[S] + l_{\text{on}}[S])]\,[r_+(l'_{\text{off}} + \omega) + k'_{\text{off}}(l'_{\text{off}} + r_- + \omega)]} \tag{2.25}$$

which is a more cumbersome expression as in the previous case. In addition, in order to satisfy detailed balance, we have the conditions

$$\frac{r_+}{r_-} = \frac{l_{\text{on}} k_{\text{off}}}{(l_{\text{off}} + \omega) k_{\text{on}}} = \frac{l_{\text{on}} k'_{\text{off}}}{(l'_{\text{off}} + \omega) k_{\text{on}}} \tag{2.26}$$

Using this conditions in the expression of ς we get

$$\varsigma_{\text{eq}} = \frac{[l'_{\text{off}} l_{\text{on}}[S'] + r_-(k_{\text{on}}[S'] + l_{\text{on}}[S'])]\,[l_{\text{off}}(l_{\text{off}} + 2r_-) + \omega(l_{\text{off}} + r_-)]}{[l_{\text{off}} l_{\text{on}}[S] + r_-(k_{\text{on}}[S] + l_{\text{on}}[S])]\,[l'_{\text{off}}(l'_{\text{off}} + 2r_-) + \omega(l'_{\text{off}} + r_-)]}. \tag{2.27}$$

which is never smaller than ς_0. Let see this more clearly. Let us assume for a moment $k_{\text{on}} = l_{\text{on}}$. We obtain

$$\varsigma_{\text{eq}} = \frac{[(l'_{\text{off}} + 2r_-)]\,[l_{\text{off}}(l_{\text{off}} + 2r_-) + \omega(l_{\text{off}} + r_-)]}{[(l_{\text{off}} + 2r_-)]\,[l'_{\text{off}}(l'_{\text{off}} + 2r_-) + \omega(l'_{\text{off}} + r_-)]}. \tag{2.28}$$

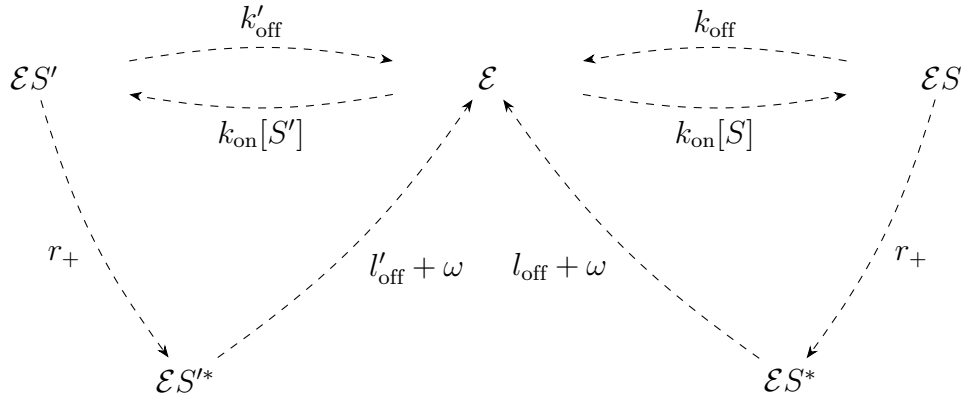
which in the limit of very slow catalysis rate is given by

$$\varsigma_{\text{eq}} \longrightarrow \frac{l_{\text{off}}}{l'_{\text{off}}} \quad \text{for} \quad \omega \ll l_{\text{off}}, l'_{\text{off}}. \tag{2.29}$$

Using the detail balance condition from Equation 2.26 in the same limit of slow catalysis rate we have that $l_{\text{off}}/l'_{\text{off}} = k_{\text{off}}/k'_{\text{off}}$ and therefore

$$\varsigma_{\text{eq}} = \frac{k_{\text{off}}}{k'_{\text{off}}} = \varsigma_0 \tag{2.30}$$

Equilibrium imposes a lower limit on the error fraction, even in this more sophisticated reaction scheme. Is there a change to do better? «Liberated from the constraints of equilibrium (detailed balance), new properties emerge»[17] in nature. This was the insight of Hopfield and Ninio. In this particular case, they proposed that the intermediate reaction that transforms $\mathcal{E}S$ into $\mathcal{E}S^*$ (and equivalent for $\mathcal{E}S'$ into $\mathcal{E}S'^*$) can be driven out of equilibrium by an external force. It can be assumed then that the new intermediate state is a high energetic state due to the external driving. In this case, most of the transition to the exited state comes from the driven reaction and one can neglect the reactions coming directly from the free enzyme (we neglect l_{on}). In this case, the complete enzymatic reaction can be approximated as



(2.31)

For this nonequilibrium model of enzymatic catalysis, the error ratio is given by

$$\varsigma = \frac{(k_{\text{off}} + r_+)(l_{\text{off}} + \omega)}{(k'_{\text{off}} + r_+)(l'_{\text{off}} + \omega)} \tag{2.32}$$

which in the case of slow catalysis rate gives

$$\varsigma = \frac{(k_{\text{off}} + r_+)l_{\text{off}}}{(k'_{\text{off}} + r_+)l'_{\text{off}}} \quad \text{for} \quad \omega \ll l_{\text{off}}, l'_{\text{off}}. \tag{2.33}$$

We saw above that in the slow catalysis rate regime, the simple enzymatic reaction setup of Equation 2.20 achieves its lowest error fraction ς_0. In this case, there is a new relevant time-scale defined by the driven reaction rate: $\sim r_+^{-1}$. If the driven rate is fast, then we obtain again

$$\varsigma = \frac{l_{\text{off}}}{l'_{\text{off}}} = \varsigma_{\text{eq}} \quad \text{for} \quad r_+ \gg k_{\text{off}}, k'_{\text{off}}. \tag{2.34}$$

which is the same as the equilibrium error rate. The lowest rate is achieved for slow driven reactions

$$\varsigma = \frac{k_{\text{off}} l_{\text{off}}}{k'_{\text{off}} l'_{\text{off}}} \approx \varsigma_0^2 \quad \text{for} \quad r_+ \ll k_{\text{off}}, k'_{\text{off}}. \tag{2.35}$$

An intuitive way to understand this is the following: once the first enzyme complex has been formed, the probability of forming the intermediate state is given by

$$p* = \frac{r_+}{r_+ + k_{\text{off}}}. \tag{2.36}$$

If the proofreading reaction is on average faster than the dissociation events of both substrates ($r_+ \gg k_{\text{off}}$), then it cannot be used to discriminate between them because $p^* \to 1$ for all k_{off}. Only if the proofreading reaction is on average slower than the dissociation events ($r_+ \ll k_{\text{off}}$), it can be used to discriminate between two different dissociation rates k_{off} and k'_{off} because $p^* \sim k_{\text{off}}^{-1}$. The trade-off between catalysis speed and specificity is again present here because of the delay caused by the slow proofreading reaction.

2.2.3 Proofreading in the immune system

As it has been shown above, enzymes can be driven by external work to increase the specificity of their catalytic activity. The problem of specificity is ubiquitous in nature and is a fundamental aspect when we study information processing in biological systems [20, 21, 22, 17]. In this context, 20 years after the kinetic proofreading model was proposed, Timothy W. McKeithan realized that it could also play an important role in the study of the immune system. McKeithan proposed a kinetic proofreading scheme for the recognition of epitopes by T cells [23]. His model, based of the auto-phosphorylation of intra-membrane regions of T cell receptors, has been extended with a focus in the phenotype [24]. Moreover, some recent experiments have claimed to have measured the strength of kinetic proofreading in T cells, estimating 2.6 effective steps taking place in the discrimination process [25].

2.3 The adaptive immune system

2.3.1 The multifaceted B cell response to viral antigens

We begin this brief review of the adaptive immune system at the end of the story. During the 20th and 21st centuries, our understanding of the mechanisms by which the immune system fights off infection has increased enormously. Thanks to the advent of new technologies, combined with the great dedication and astuteness of many scientists, we now have very detailed descriptions of the biology of immune responses. In this section, we describe the main biological paradigms necessary to understand the subject of this dissertation: the immune response of B cells during immune challenges.

The immune system of vertebrates consists of a complex network of agents that respond in a coordinated manner to all types of threats and protect the host. In the literature, the immune system is divided into at least two subsystems: the innate and the adaptive immune system. After physical barriers such as the skin,

the innate immune system is the first line of defense against pathogens. It is responsible for protecting us from very well known pathogens. And by "very well known", I mean pathogens with which vertebrates have co-evolved over very long time scales and have developed mechanisms to recognize them well. A key point here is that we have learned to recognize features of these pathogens that have not changed much in all that time. In addition, the innate immune system is also responsible for directing important immunological processes such as inflammation and the spread of immune alert signals throughout the host.

The adaptive immune system is also very complex and consists of different types of cells. The two most important are B cells and T cells. B from Bone marrow, and T from Thymus, the places where they mature. The bone marrow and the thymus are the primary lymphoid organs in humans. B and T cells are able to recognize a special type of pathogens and threats. They can fight out infections caused by pathogens that the host has never seen before. An example of such pathogens is SARS-CoV-2, the coronavirus that caused the 2019 pandemic. Similarly, they can also protect us against pathogens that evolve fast. So fast that each time we get infected with them, they look very different from each other. This is the case of the seasonal Influenza A virus, that has been circulating for decades in the world populations. B and T cells can also protect us from corrupted versions of ourselves: cancer. They can recognize very small differences between mutated cancer cells and destroy them. We see that the main power of B and T cells is that they can protect us from unknown threats.

T cells have two main functions: to kill infected or malignant cells and to regulate B cells. The former is done by cytotoxic T cells and later by helper T cells. The world of T cells is a fascinating one. T cells are not only important for protection against pathogens, but also for protection against cancer cells, which they can also recognize as threats. But here I will focus on B cells, which are responsible for producing antibodies and much more.

B cells B cells recognize pathogens using specialized receptors in their membrane called B cell receptors (BCRs). Roughly speaking, these receptors are the bound version of the soluble antibodies. Some B cells have them attached in their membrane, others secrete them. BCRs bind to certain regions of pathogens called antigens. Moreover, pathogenic antigens are known to have localized sub-regions on their surface that are more immunogenic to B cells [26]. These regions are known as epitopes. What makes one region an epitope and others not is not clear. Some intrinsic properties of the antigen may play a role. For example, the amino acid composition or steric and geometric effects [27]. On the other hand, as we will discuss later, intrinsic properties of the immune system may also determine which regions of the antigen are efficiently targeted. Overall, this remains one of the fundamental open problems in immunology.

As we discussed above, each host is able to produce a very large amount of different BCRs. This is know as the BCR repertoire. The rule is that each different B cell expresses in its membrane a unique type of BCRs, and therefore some time we would refer to the BCR repertoire as the B cell repertoire. In each

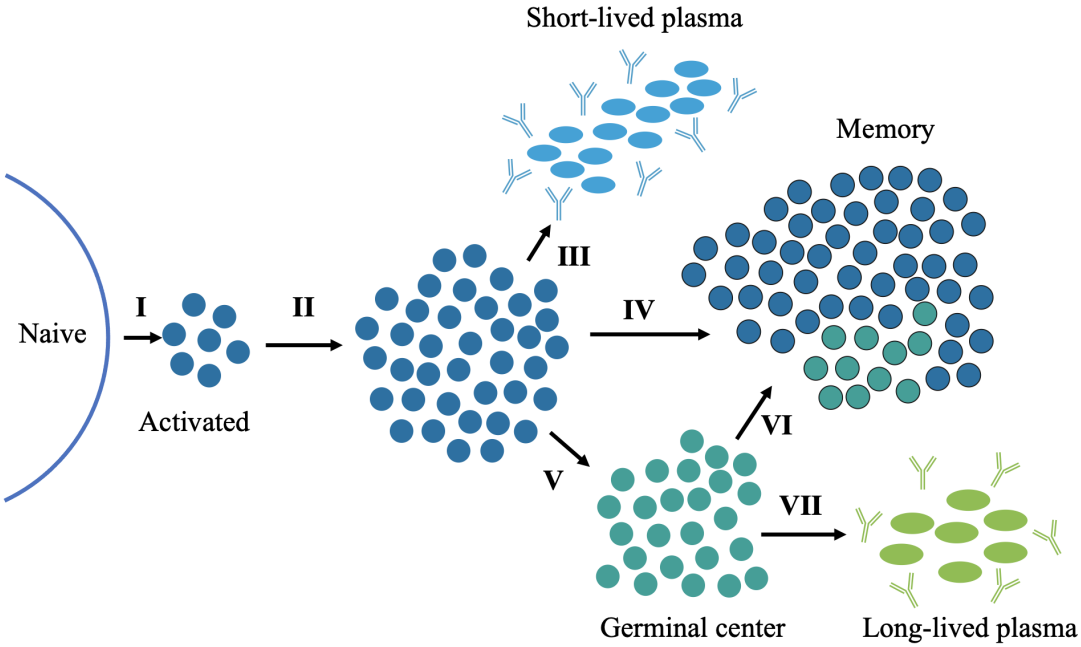


Fig. 2.3. B cell immune response Summary of the different subpopulations of B cells as explained in the main text.

host, a very large number of circulating naive B cells are present all the time. Naive means that after they were produced, they have not yet engaged in any immune response. This number can be on the order of 10^8 in mice or 10^{11} in humans, with a similar number of different BRCs [10, 28, 29, 30].

Viral infections In primary viral infections, B cells can encounter viral antigens in the affected tissue. However, there is a draining system, the lymphatic system, that can transport antigens from one place to another. It turns out that B cells are most likely to encounter antigens in lymph nodes or secondary lymphoid organs, such as the spleen and tonsils. The high concentration of antigens and B cells in these regions of the lymphatic system increases the chances that B cells will recognize threats. The antigens that bind to BCRs may be soluble antigens floating in the tissue where the B cells reside. However, there are also specialized cells that can unspecifically retain antigens on their surface and present them to B cells, facilitating BCR-antigen encounters.

When B cells successfully recognize a pathogenic antigen, they begin to mount an immune response (I in Figure 2.3). A first step is the proliferation of B cells that have recognized the antigen (II in Figure 2.3). Each B cell undergoes multiple cell divisions to form a larger clonal population. As discussed in this dissertation, this is a fundamental step in our understanding of how the immune system works and, in my opinion, remains understudied. Of the billions of different B cells that that make up the naive repertoire, it has recently been observed that about $10^2 - 10^3$ different B cells respond [31, 32]. This is a larger number compared to the handful of B cells that were originally thought to respond. Still, this is a large

bottleneck of diversity. On average, only one in a million B cells will respond to an infection.

B cell fate During or after proliferation, B cells differentiate into new phenotypes. Initially, B cells choose between three different fates: i) plasmablasts (III in Figure 2.3), ii) memory B cells (IV in Figure 2.3), or iii) germinal center (GC) B cells (V in Figure 2.3). The mechanisms by which this decision is made are not fully understood and are beyond the scope of this thesis. Plasmablasts are antibody-secreting cells and are rapidly generated upon recognition of an ongoing infection. They are capable of producing up to tens of thousands of antibodies per second during their lifetime. In the literature it is usually said that the antibodies produced by plasmablasts are of low affinity. We note that this depends on what we compare them to. These antibodies help to clear viruses from the body by neutralizing their ability to infect new cells or by marking them as a threat for other arms of the immune system. Plasmablasts survive for a few weeks after infection. For this reason, they are sometimes called short-lived plasma cells.

The second of the possible fates for B cells is to become memory B cells. Two main features distinguish memory from naive B cells. First, memory B cells are part of larger clones compared to naive B cells. This is a consequence of the initial proliferation during the primary immune response. Second, memory B cells are more ready to engage in an immune response than naive B cells. They are very sensitive to exposure to the antigen that generated them in the first place and undergo rapid proliferation and differentiation into plasmablasts. As we discuss later, this two features are crucial for the protection they confer against recurrent infections.

The third and final fate that B cells can take is to become GCs B cells. GCs are temporary tissues formed by responding B cells in lymph nodes after an immunization event. In GCs, B cells undergo a process called affinity maturation.

First, B cells migrate to the so-called *dark zone* of the GC. There, B cells accumulate somatic mutations in the genes that encode for the BCR. This process is carried out by a specialized enzyme called activation induced cytidine deaminase (AID). Once modifications in the BCR have occurred, the B cells migrate to the so-called *light zone*. The light zone is populated by follicular dendritic cells and helper T cells. The former carry antigen in their membrane to be presented to B cells, and the latter are ready to provide survival signal to B cells. Mutated B cells coming from the dark zone will capture antigen from follicular dendritic cells and, based on this, will request help from T cells in order to survive and proliferate. Most B cells go through several rounds of dark and light zones: several rounds of mutation and selection. At the end of the process, B cells that have acquired mutations that increase their affinity with the antigen are able survive and further proliferate. Affinity-maturated B cells leave GCs in two different forms. They can become memory B cells (VI in Figure 2.3) or long-lived plasma cells (VII in Figure 2.3). Memory B cells, similar to those produced before the formation of GCs, will circulate between lymph nodes and other tissues awaiting reactivation. Long-lived plasma cells will migrate to the bone marrow and secrete high-affinity

antibodies for up to decades. This is the beautiful process of affinity maturation, a Darwinian process of very rapid evolution of BCRs.

A note on time scale The multifaceted immune response of B cells occurs on different time scale. Production of plasmablasts and secretion of non-affinity maturated antibodies begins a few days after the onset of the infection. Plasmablasts then survive in the host for several weeks. They then die of exhaustion. This part of the immune response is called the extrafollicular response. As we can see, this does not provide long-term protection for the host against reinfection with the same pathogen. At the same time, many memory B cells are produced directly from the primary responding B cells [32, 33, 34]. Although it is unclear for how long will those memory B cells circulate in the host, they probably stay survive in the host its lifetime. It is believed that in aged individuals, up to 20% of all B cells are memory B cells. In this sense, they provide long-term immunity against the infecting pathogen.

For several weeks or up to a few months after resolution of the primary infection, affinity maturation takes place in the previously formed GCs. The reason for the termination of GCs is not completely clear, but it may be a combination of several factors, including exhaustion of antigen in the GC, extinction of all participating B cells, and a pre-defined period of existence. During all this time, memory B cells are exported from GCs. It has been observed that there is no particular bias as to the point in the GC reaction at which memory B cells are exported. On the other hand, long-lived plasma cells are typically exported later in the life of a GC. Both GC-derived memory B cells and long-lived plasma cells provide long-term protection for the host. It is interesting to notice that GCs do not seem to play an important role in the primary infection. Affinity maturation is usually understood as an adaptation for long-term immunity.

Now that we have a better picture of how B cells work, let us go back a century to the discovery of acquired immunity. From this point, we will explore the history behind the paradigm shifts that have occurred in our understanding of the nature of antibody production. In doing so, we will recapture and contextualize some of the key issues explored in this thesis.

2.3.2 The birth of immunology and the antibody specificity problem

Since the discovery by E. von Behring and S. Kitasato in the late 19th century that sera from animals infected with diphtheria could be used to protect or treat non-immunized animals, immunology became a new and important scientific field in its own right. In the 20th and 21st centuries, it has had a major impact not only on medical applications, such as vaccination or serological therapy, but also on the development of new tools for research and disease diagnosis, such as the hemagglutination [35] and ELISA assays [36].

In the decades following the work of Behring and Kitasato, it was shown that

the so-called antibodies (at that time hypothetical entities present in the sera of immunized animals) are proteins [37, 38], that they are Y-shaped [39, 40], and that they bind to antigens through specific molecular interactions [41].

At the same time that the physicochemical properties of antibodies were being characterized, accumulating evidence showed that they were produced by the host after immunization with exquisite specificity against many different antigens. These observations set off a fundamental scientific debate that, in my opinion, still puzzles researchers today. Conceptually, the big question was what are the mechanisms common to all vertebrates that allow organisms to produce highly specific antibodies after immunization.

The first type of theories about antibody production suggested that antigens serve as templates for antibody production [42, 43]. These theories proposed that immune cells were initially able of producing generic antibodies that lacked any particular specificity. During immunization, foreign antigens could physically interfere with the process of antibody formation and change their shape to one compatible with the antigen. These types of theories were called *instructional theories*, because they assumed that the organisms actually used the contents of injected antigens as instructions to produce specific antibodies. Instructional theories were specially motivated by a specific type of experimental observation: under the right conditions, antibodies were elicited by virtually any chemical structure, even synthetic ones [44]. In this way, they became the first paradigm on antibody production during the first decades of the 20th century [45]

However, instructional theories could not properly explain at least three key observations about immunization processes. First, under this theories, the delay between inoculation and antibody production should be shorter for larger antigen doses because [43], but this contradicted experimental observations [44] In addition, they could not explain the exponential rise in antibody concentration soon after immunization. This would require a self-replicating template, which was not a feasible assumption for the theories. Without a self-replicating template, the increase in antibody concentration could not be faster than linear. Finally, experiments suggested discrepancies between template acquisition and antibody production. Foreign substances were taken up by macrophages, while antibodies were produced by plasma cells. Under these circumstances, how could a antigens be used as a template for antibody production?

The clonal selection theory Interestingly, even before the development of the instructional paradigm, P. Ehrlich first reported a different type of theory of antibody formation in 1900 [46]. Based on the initial observations of the high specificity shown by antitoxins elicited after immunization, and inspired by his own work in organic chemistry, Ehrlich proposed that host cells carry *side chains* in their membranes that can specifically interact with soluble substances. When such an interaction occurs, the host cells are stimulated to produce more of the specific side chain. In this case, the central idea was that the exposure to a foreign substance induces the production of antitoxins (antibodies) in the host from a set of pre-existing side chains. The difference between this theory and

the instructional theories resembles the difference between the Lamarck's and Darwin's theories of evolution discussed above. Although Ehrlich's theory could, in principle, explain the exponential increase in antibody levels after immunization, his theory lost acceptance as more and more chemical structures were found that could induce the production of new specific antibodies.

With no further evidence supporting instructional theories, in 1955 Niels Jerne reformulated a theory similar to that proposed by Ehrlich more than half a century earlier [47]. N. Jerne published his *Natural-Selection Theory Of Antibody Formation*, which challenged the instructional paradigm [43]. Instead of using foreign antigens as templates to produce specific antibodies, Jerne postulated that there is always a large diversity of antibodies with all kinds of specificities in a host organism. The theory further proposed that, in the presence of a foreign antigens, only a fraction of the best-fitting antibodies were chosen from this underlying diversity. These chosen antibodies would somehow instruct the cells to replicate them.

Jerne's theory itself had some problems. One of them was the need for a mechanism that could produce all this pre-existing diversity in the first place. It also contradicted the already established idea that proteins were not self-replicating units. However, the theory was rich and, under certain assumptions, could provide explanations for central tolerance and affinity maturation. Thus, the theory resonated with the scientific community and began a paradigm shift.

Two years after the publication of Jerne's idea, David Talmage [48] and Frank M. Burnet [49] made an important conceptual leap (in the direction of Ehrlich's model). Instead of antibodies replicating themselves, they proposed that antibodies might be found in the membrane of immune cells, and that upon binding to foreign antigens present in the host, would trigger the proliferation of the cell that produces the antibodies.

This idea was not isolated, but was inspired by the ≪current concept that the configuration of a protein is determined solely by information contained in the heredity units of the cell, the nucleic acids≫ [48]. Burnet's seminal paper, titled "A Modification of Jerne's Theory of Antibody Production units the Concept of Clonal Selection" speeded up the paradigm shift (and also rose a controversy, not uncommon in the history of science, about who had been the first who had the idea. In this case Burnet or Talmage [50]) Burnet theory was later termed the *clonal selection theory* of antibody production.

In the years after the first publication, Burnet explored the implication of his theory. He argued that the clonal selection theory could explain immunological memory (expanded clones), original antigenic sin (undesirable expansion of low-affinity clones), mucosal immunity (presence of expanded clones), natural antibodies (spontaneous production of antibodies by some cells), and autoimmune disease (expansion of self-reactive clones) [51, 43]. He also proposed that since it is unlikely that the huge diversity of antibodies was encoded in the germline, it must be produced by somatic mutations in early stages of lymphocyte development [51, 43]. The discovery of VDJ recombination in the 70's by Susumu Tonegawa [52,

53] will prove Burnet's ideas right.

In the middle of the century, a paradigm shift took place [45] and new ideas began to emerge from it. As correctly proposed by the work of N. Jerne, M. Burnet and D. W. Talmage, and others, in order to recognize the unknown, the immune system has come to a solution: create enough diversity in the form of a large repertoire of lymphocyte receptors. However, since the formulation of the clonal selection theory, an important question has been how this *selection* process occurs. At the heart of the theory, there is the central question of how antigen-specific, high-affinity immune cells are selected during immunization. It appears that the larger the repertoire, the smaller the antigen-specific subset of receptors specific for a given antigen. This is reminiscent of the energy-entropy trade-off commonly found in statistical physics. This question will become particularly relevant in the context of expanding antigen signals, like is the case in acute infections.

First mathematical models In the 70's, physicists became interested in immunology, and in particular, in the theory of clonal selection. Perhaps the first mathematical model was proposed by George I. Bell [54]. Bell proposed a sophisticated model of clonal dynamics in which naive lymphocytes (or target cells, as they used to be called) were stimulated for proliferation by the interaction with a specific foreign antigen. The interaction between the cells and the foreign antigens was determined by the equilibrium occupancy of receptor sites in the membrane of the cells. The dynamics of proliferating cells was described by [54]

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = (1-R)F(R)N_T(t) + \lambda H(R)N(t) \tag{2.37}$$

where $N(t)$ denotes the population size of proliferating cells and N_T the population size of target cells. The first term corresponds to a source of cells from stimulated target cells, and the second term corresponds to the growth of proliferating cells. Here $F(R)$ and $H(R)$ are functions of the average number of occupied receptor sites in the cells, R, and were used to model the dynamics of target and proliferating cells as a function of the foreign antigen concentration. For instance, they chose

$$F(R) = \frac{R}{1+R} \tag{2.38}$$

so that when $R \ll 1$ and $F(R) \ll 1$, no target cell begins to proliferate, and when $R \gg 1$ and $F(R) \approx 1$, target cells are likely to become proliferating cells.

Bell's model produced an exponential increase in the average binding constant of antibodies after immunization. By correctly setting the initial affinity of the target cells in the model, this increase was comparable to that observed in experiments. The model also allowed for the incorporation of tolerance mechanisms and considered different protocols of foreign antigen administration (although he did not consider exponential growth as in acute infections). An important limitation of Bell's model was that it did not account for the real diversity of receptors in the population of target cells. However, this is not surprising given the lack of experimental estimates of the diversity at the time.

The model proposed by Bell served as inspiration and guided the production of new models in the coming years [55]. An important extension to Bell's model was proposed by Alan S. Perelson, Majdedin Mirmirani and George F. Oster in 1976 [56, 57]. They included a new important experimental observation regarding the interaction of immune receptors with foreign antigens. Namely, that cells were more efficiently triggered to proliferate whenever their receptors were cross-linked. Receptor cross-linking consists of the ability of some multivalent foreign antigens to bind to different receptors in the same cell at the same time. This forced the receptors to remain close to each other. The cross-linked receptors were then able to interact with each other and further signal the cell to proliferate.

An important consequence of cross-linking was that, if bivalent antigens were considered in the model, this could explain the "log-bell-shaped" curve of antibody levels as a function of administered antigen observed in experiments [55]. Although this prediction failed when antigens with valency greater than two were considered, "log-bell-shaped" curves remained to be used as an input for further mathematical models [55].

New directions for the theory In the following decades, mathematical models focused on the dynamics of the affinity maturation process described by Herman N. Eisen and Gregory W. Siskind in 1964 [58]. This was probably fueled by exciting new discoveries, such as somatic hypermutation [53] and class-switch recombination [59]. These models include [60, 61]. Moreover, because affinity maturation allows for very rapid evolution of antibodies, new models also include long-term clonal dynamics under recurrent immunization with the same or related foreign agents [13, 14, 30]. However, some central aspects of the clonal selection theory were put on the side without a proper answer.

The problem of antigen recognition in acute infections As it has been discussed above, the study of host-pathogen interactions is a rich and complex field. In this dissertation, I have identified a concrete problem related B cell antigen recognition. The problem concerns immune recognition of non-steady signals by large receptor repertoires. In the following paragraphs, I will develop in more detail what are the contradictions or conundrums that exist in this case.

In the majority of cases, if not all, the initial acute phase of an infection is characterized by the self-replication of pathogens within the host. This process, whether local or systemic infections, results in the exponential growth of pathogenic antigen concentration inside the host. Now consider the large pre-existing diversity of B cell receptors ready to recognize the foreign antigen and respond against the infection. We can assume that a very small number of receptors bind to the antigen with high affinity (often referred to as specific binding), and that this number grows exponentially when we consider lower binding affinities (often referred to as nonspecific binding).

If the difference in affinity between specific and nonspecific B cells is not sufficient to compensate for the difference in their numbers, it is possible that the

activation of nonspecific B cells occurs before the activation of antigen-specific B cells. This would have a significant impact on the quality of antibodies subsequently secreted. It would also contradict the assumption that highly diverse repertoires have evolved in order to produce specific receptors for each unknown antigen. Are there any conditions under which specific binding predominates over nonspecific binding? In other words, what are the conditions under which specific binders are "chosen" from the underlying existing diversity?

I have recognized this open problem as a cornerstone of the clonal selection theory. Many other questions arise from it. What are the molecular mechanisms by which B cells translate their interaction with non-steady antigens into a coordinated and controlled activation of only the fraction of the repertoire with the best-matching receptors? What are the consequences for the resulting clonal composition of memory B cell clones after a primary infection? How do these mechanisms affect memory B cells and determine how do they prevent reinfections? How is this protection affected by antigenic evolution? I hope to provide partial answers to some of these questions in this dissertation.

Part II

The Activation Of The B Cell Repertoire

Can the truth (*the capability to synthesize an antibody*) be learned? If so, it must be assumed not to pre-exist; to be learned, it must be acquired. We are thus confronted with the difficulty to which Socrates calls attention in *Memo* (Socrates, 375 B.C.), namely that it makes as little sense to search for what one does not know as to search for what one knows, what one knows one cannot search for, since one know it already, and what one does not know one cannot search for, since one does not even know what to search for. Socrates resoles this difficulty by postulating that learning is nothing but recollecting. The truth (*the capability to synthesize an antibody*) cannot be brought on, but was already inherent.

A quote of the Philosophical Bits or a Bit of Philosophy (Kierkegaard, 1844) by N. K. Jerne [1]

Chapter 3

Primary infection: A Luria-Delbrück activation model

The content of this chapter has appeared as part of,

- Morán-Tovar, R. & Lässig, M. *Nonequilibrium Antigen Recognition during Infections and Vaccinations.* Phys. Rev. X **14**, 031026 (2024)[62].

On this work, I am first author and I have performed the research and analysis under supervision of the last author. The paper is co-written with the last author.

In accordance with the doctoral regulations, this article, already published in a peer-reviewed scientific journal, is not attached in the published version of this thesis.

3.1 Summary

During acute viral infections, the immune system of vertebrates is capable of mounting effective and rapid responses to eliminate the pathogen from the host. This includes the activation, proliferation and differentiation of B cells, which not only secrete specific antibodies to neutralize the pathogen activity, but also produce immunological memory in the form of memory B cells and long-lived plasma B cells. The number of naive B cells can be on the order of 10^8 in mice or 10^{12} in humans, with a similar number of B cell clones. However, the number of responding B cell clones has been observed to be on the order of 10^2 in mice. This means that only 1 in a million B cell clones participate in the response. This is a puzzle in itself. Why hundreds and not tens or thousands?, has it to do with the diversity of the naive repertoire?, are the responding clones the ones with high affinity for the viral antigen?, and even more intriguingly, why do these hundreds of B cell clones respond and not the other hundreds of millions? This last question is one of the central building blocks of the clonal selection theory discussed in the introduction. The mechanism by which specific responding B cell clones are chosen from the naive repertoire remains an open problem in immunology.

The problem becomes clearer (and even more puzzling) when we seriously consider the dynamics of the antigen. During an acute infection, the viral antigen concentration in the host changes dramatically. Starting from a few viral particles, the viral load can reach up to $> 10^{11}$ viral particles at its peak. Although this is a short time from the perspective of the host, it is a much longer time scale for the binding kinetics that take place between BCRs and viral antigens. Thus, we can assume that the change in antigen concentration is adiabatic. Now imagine that at some point during the acute phase of the infection, the B cell clones with the highest affinity with the viral antigen in the repertoire recognize the antigen and begin to mount an immune response. In the next few hours, when the antigen concentration would have increased a hundredfold, many more B cell clones with lower affinity would have a chance to participate in the response. Assuming that the number of low affinity B cell clones increases exponentially, this creates a scenario where a B cell activation catastrophe can occur, in which many low affinity B cells dominate the response, compromising its quality.

We consider a minimal model of the B cell immune response. It consists of three main aspects: i) the exponentially increasing concentration of antigen, ii) the molecular recognition of the antigen by the B cells based on kinetic proofreading, and iii) the proliferation of the B cells after activation. The increasing concentration of antigen adiabatically changes the probability of each B cell of being activated, by increasing the average number of association events, which are initially rare at the single B cell level. BCRs associated with antigens must undergo a series of irreversible steps before triggering full activation of the B cell. When an engaged BCR successfully completes all the required steps, it triggers activation and proliferation of the B cell. In this way, activated B cells form clonal populations. Under this framework, we calculate the activation probability of all B cells in the repertoire as a function of its binding affinity with the antigen.

The work presented in this chapter shows that the immune response governed by this dynamics has two distinct regimes of low and high specificity, dictated by the complexity of the immune repertoire and the strength of the proofreading mechanism. In the low specificity regime, low affinity B cells are activated before high affinity ones. This occurs because the difference in activation probability does not compensate for the difference in numbers: entropy dominates the response. In the high specificity regime, a larger difference in activation probability between low and high affinity B cells gives high affinity B cells a chance to be activated first, even though they are completely outnumbered by low affinity B cells. The amount of proofreading necessary to ensure an immune response in the high specificity regime is dictated by the complexity of the immune repertoire.

The resulting clonal dynamics of B cells resembles that of the classical Luria-Delbrück model of microbial evolution in expanding populations. In such a model, an exponentially proliferating wild-type population generates mutants. Each new mutant cell itself proliferates exponentially. A hallmark of this dynamics is the existence of jackpot mutation events: mutants that appear exceptionally early can overtake a large fraction of the total population. In the case of immune cells, the exponentially growing concentration of antigen leads to an exponentially growing

activation rate of B cells, which in turn proliferate exponentially upon activation. Independently of the specificity regime, this theory predicts jackpot activation events in the immune system. More remarkably, in the high specificity regime, the activation time of each B cell is correlated with its affinity: high affinity B cells are activated earlier. This correlation, together with the existence of jackpot clones, predicts the existence of so-called elite neutralizers in primary responses: high-affinity B cells that generate exceptionally large B cell clones during the immune response.

The mechanistic model of B cell repertoire activation presented in this chapter bridges different levels of biological complexity. Starting from the molecular details of the recognition process of single antigens by BCRs, we have established quantitative laws for the clonal composition of the activated repertoire during acute infections. Interestingly, such laws also depend on the complexity of the naive repertoire itself, here characterized by the density of B cell clones in a biophysically motivated binding affinity space. This multi-scale approach opens new way to investigate the adaptive immune system and its design principles.

Chapter 4

Vaccination: A minimal spatiotemporal model

The content of this chapter has appeared as the subsections **II.H. Spatiotemporal antigen dynamics**, **II.I. Immune response to vaccination**, and **III.C. Activated B-cell repertoires in mice**, as part of[1],

- Morán-Tovar, R. & Lässig, M. *Nonequilibrium Antigen Recognition during Infections and Vaccinations*. Phys. Rev. X **14**, 031026 (2024)[62].

On this work, I am first author and I have performed the research and analysis under supervision of the last author. The paper is co-written with the last author.

In accordance with the doctoral regulations, this article, already published in a peer-reviewed scientific journal, is not attached in the published version of this thesis.

4.1 Summary

The exponential proliferation of antigens during acute infections investigated in the previous chapter may have played an important role in the evolution of the adaptive immune system in all vertebrates. One can speculate that some of the key features of the immune response are optimized to protect the host against primary and recurrent acute infections. It is important to recognize that viral antigens also diffuse within the host and can also be subjected to active transport from peripheral tissues to lymph nodes. Vaccination is an interesting case of study in this sense. The different types of vaccines developed to date provide a richer set of antigen dynamics, including exponential proliferation and diffusion of antigens in live attenuated vaccines, only diffusion of antigens in dead/inactivated vaccines, or linear growth and diffusion of antigens in mRNA vaccines. As we

[1]This project was originally formulated and developed as a separate publication from the previous chapter. However, due to the peer-review process, we decided to include it in a compact form in the same publication.

have shown in the previous chapter, antigen dynamics is an important driving force that ultimately shapes the clonal composition of responding B cells and thus the quality of the overall immune response.

In this chapter, we develop a minimal spatiotemporal model that includes antigen diffusion and proliferation within the host and that accounts for two important aspects of an immunization event: i) antigen particles begin to spread within the host from a highly localized point in space, and ii) B cells are located at a given effective distance, r_0, from the starting point of the antigens. In this case, the antigen concentration is generally described by a reaction-diffusion equation. Furthermore, the number of antigen particles that can interact with B cells and could potentially trigger their activation is now limited to those that have traveled beyond r_0.

Our spatiotemporal model shows two distinct regimes for the activation dynamics of B cells, determined by the ratio between the effective distance, r_0, and the average distance traveled by antigen particles due to diffusion before the first B cells are activated $(Dt^*)^{1/2}$. If the ratio is smaller than a given threshold, corresponding to antigen particles diffusing rapidly beyond r_0, B cell activation occurs within the deterministic diffusion front. In this case, diffusion does not strongly limit antigen recognition of by B cells. When the ratio is greater than the threshold, diffusion strongly constrains antigen recognition. In this case, B cell activation occurs through rare diffusive paths, which significantly affects the quality of the immune response. The model presented here shows that, under physiologically relevant parameters, acute infections occur in the former regime, which is not limited by diffusion. Thus, the homogeneous theory presented in the previous chapter provides a good approximation of the B cell activation dynamics during acute infections.

In the case of vaccination with inactivated particles, the system is always limited by diffusion. In this case, the initial dose of antigen in the vaccination determines the scaled recognition radius of antigen diffusion, which in turn defines a new immune response onset and an effective growth rate of the antigen at r_0. Surprisingly, our model showed that under certain physiologically relevant parameters, the generalized Luria-Delbrück dynamics is also valid in the case of vaccination. In this case, observables such as the clone size exponent derived for the homogeneous model in the previous chapter become diffusion dependent. The model predicts a window of antigen dose in which the immune response is optimal. On one hand, for a too low antigen dose, the onset of B cell activation is delayed and restricted to rare events. On the other hand, for too high dose, the resulting activated repertoire is less focused on the high-affinity B cell clones, jeopardizing the quality of the overall response.

Finally, we analyzed data collected from vaccination in mice. As a proxy for the initial activation of B cell clones, we used only data from early germinal centers before high levels of somatic hypermutation and affinity maturation have occurred. We compute the clonal entropy as well as the rank-size relation, as presented in the previous chapter. Both measures suggest that B cells in mice use about three steps of proofreading for activation. Moreover, the clone size exponent

found in the data set is remarkably similar to a clone size exponent measured in human B cell repertoire data that included memory B cells. Power laws in the rank-size relation of human data have been attributed to long-term interaction of memory B cell clones with antigens during recurrent infections. Preliminarily, our analysis suggests that these two similar exponents in two different subpopulations of B cells may have a similar dynamical origin. Overall, this chapter provides a minimal spatiotemporal model of B cell repertoire activation that could improve our understanding of the immune response in different vaccination protocols that have been heuristically optimized in this sense in the past.

Chapter 5

Memory response: *In vivo* protection

Some of the content of this chapter has appeared as the section **VI. COMPLEXITY OF IMMUNE RECOGNITION**, as part of the preprint,

- Röschinger, T., Morán-Tovar, R., Pompei, S. & Lässig, M. *Adaptive ratchets and the evolution of molecular complexity* 2025. arXiv: 2111.09981 [62].

currently under peer-review process. On this work, I am second author and I have performed the research and analysis of the section in collaboration with the rest of the authors. The paper is co-written with the rest of the authors.

5.1 Introduction

This chapter contains the work in progress of the last project I worked on during the last part of my PhD time. The current version presents the mathematical foundations of the model and some preliminary simulations. From basic assumptions about how memory B cell clones are formed, we aim to develop an *in vivo* response function against recurrent infections.

As discussed in Chapter 2, immune protection conferred by B cells after a primary infection comes in two flavors: antiserum antibodies and memory B cells. Antiserum antibodies are maintained at homeostatic levels by affinity-matured long-lived plasma cells residing in the bone marrow. If the formation of long-lived plasma cells is corrupted, even affinity-matured plasma cells cannot maintain high levels of antibodies and they would have a decay time on the order of months or a few years [63]. Memory B cells reside in the various lymphoid organs as well as in other tissues, waiting to be reactivated. Although both plasma cells and memory B cells are derived from the same set of activated B cells (like the one studied in Chapters 3 and 4) they have an important difference. According to recent findings, the level of affinity maturation of long-lived plasma cells is significantly higher than that of memory B cells [32, 34]. Some studies have even observed that most of the B cell clones populating the memory compartment do not undergo affinity maturation at all [33]

What is the role of each arm of protection in preventing reinfection with the same pathogen that triggered the primary response? To answer this question, we must first understand the mechanism of action of each arm of protection. In the case of antiserum antibodies, they protect against reinfection by binding, in quasi-equilibrium conditions, to any viral particle that arrives in the host and neutralizing its ability to infect cells. (Although it is the case that some antibodies bind but do not neutralize the action of viruses, we assume here that all binding results in full neutralization). In the case of memory B cells, they recognize viral particles again and, similar to naive B cells, go through a round of activation-proliferation, differentiation into plasma cells, and finally antibody shedding. In this case, the initial recognition of the viral antigen by memory B cells requires that the infection has taken off with an exponential increase in viral load.

These different mechanisms of action suggest that B cell-acquired immunity operates in at least two layers (See Figure 5.1). First, antiserum antibodies prevent proliferation of the pathogen by binding. When binding between viral particles and circulating antibodies becomes inefficient, memory B cells recognize the growing virus and massively produce new antibodies to stop growth. Whenever memory B cells are unable to rapidly stop viral growth, a *de novo* naïve response is triggered and full breakthrough infection is expected.

This view, common among immunologists, is supported by the observation that immune protection correlates with antibody titer drops. Let us break down this statement. Antibody titers are measurements from *in vitro* essays that quantify the binding probability, P_{binding}, of antibodies in serum with a given antigen. Let

us start from the basic equilibrium form of the binding probability

$$P_{\text{binding}} = \frac{1}{1 + \frac{K}{[\mathbf{Ab}]}} \tag{5.1}$$

where K is the effective dissociation constant and $[\mathbf{Ab}]$ the concentration of antibodies in the serum. In these assays, a series of twofold dilutions of serum are performed to find the transition from bound to unbound states by a change in the chemical potential of the reaction. In this case, the binding probability as a function of the number of dilutions x is given by

$$P_{\text{binding}}(x) = \frac{1}{1 + \frac{K}{[\mathbf{Ab}]2^{-x}}} \tag{5.2}$$

In this case, the number of twofold dilutions that a serum can afford before losing binding for a given antigen is called the titer.

$$P_{\text{binding}}(T) = \frac{1}{2} \longrightarrow T = -\log_2 K + const \tag{5.3}$$

The titer measured with the antigen that elicited the antibodies is called the homologous titer.

$$T_{\text{homo}} = -\log_2 K_{\text{homo}} + const \tag{5.4}$$

A titer measured with an antigen different from the one that elicited the antibodies is called an heterologous titer.

$$T_{\text{hete}} = -\log_2 K_{\text{hete}} + const \tag{5.5}$$

The difference between an heterologous and the homologous titer is called a titer drop.

$$\Delta T = T_{\text{homo}} - T_{\text{hete}} = \log_2 K_{\text{hete}}/K_{\text{homo}} \sim \Delta\Delta G \tag{5.6}$$

where $\Delta\Delta G$ is an effective difference in free energy in units of $k_B T$ and measure a distance between the homologous and the heterologous antigens. What is important here is at constant antibody concentration, binding probably gets reduced between different antigens by increasing $\Delta\Delta E$. This can occur, for example, by mutations in the region targeted by antibodies. We will discuss this in more detail below. We can then define a binding probability that depends on the titer drop as

$$P_{\text{binding}}(\Delta T) = \frac{1}{1 + \frac{K_{\mathbf{homo}} 2^{\Delta T}}{[\mathbf{Ab}]}} \tag{5.7}$$

Difference studies have found a correlation between titer drops and the immunological protection [64, 65]. For example, as shown in Figure 5.1, at a titer drop ΔT_1, the chance of the virus being able to grow in a host becomes of 50%. However, it is also reported that at a second titer drop $\Delta T_2 > \Delta T_1$, the chance for the host of getting a severe infection is of 50%.

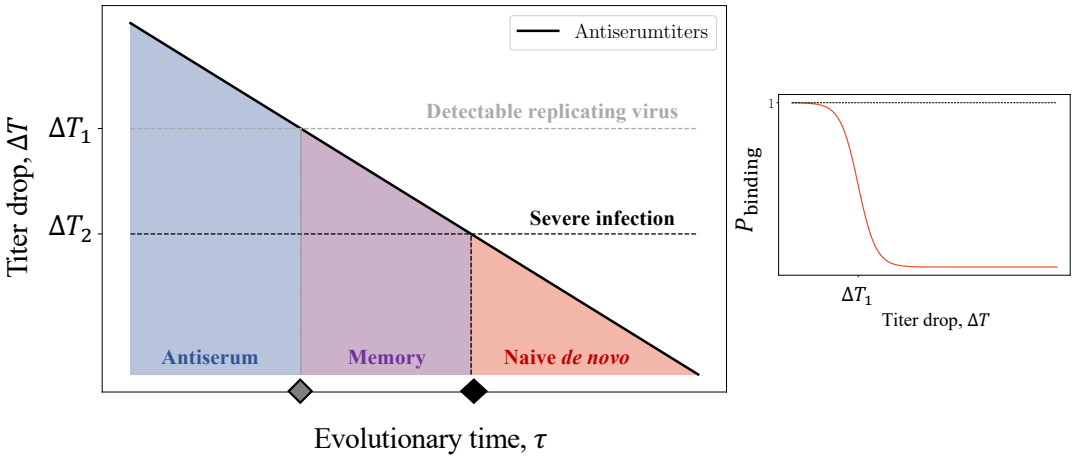


Fig. 5.1. Layers of Immune protection As described in the main text, it has been found that titer drops correlate with different stages of immune protection. Here we show a layer model in which antibody protect for a first range of titer drop. After that, when the binding capacity of antibodies has decrease enough (see right), memory B cells confer an *in vivo* protection against severe infections. When memory protection fails, a *de novo* response will occur.

Assuming that ΔT_1 marks the transition point at which the neutralization quality of antibodies in serum has decreased enough to properly impede growth of the virus

$$\frac{1}{2} = \frac{1}{1 + \frac{K_{\mathbf{homo}} 2^{\Delta T_1}}{[\mathbf{Ab}]}} \tag{5.8}$$

and further assuming that this transition occurs in a sharp way, it is unlikely that the severity of the infection is determined by the quality of the antibodies. Therefore, we hypothesize that memory B cells provide an *in vivo* protection beyond equilibrium antibody binding, extending immune protection over a longer range of antigenic viral evolution, and explaining why severe breakthrough infections are likely to occur at $\Delta T_2 > \Delta T_1$.

In this chapter, we study the *in vivo* protection conferred by memory B cells during reinfection. We analyze how such *in vivo* protection depends on how the memory compartment is formed, how memory B cells are reactivated, and finally how much the antigen has changed by evolution between primary and recurrent infections.

5.1.1 A note on disorder and fluctuations in the Immune System

As discussed in the previous section, the immune system is equipped with a very large and diverse repertoire of receptors ready to recognize any potential threat that enters the host. An important aspect of this process is that during the time scale of a single infection, we can assume that the repertoire is quenched, with a much longer turnover time scale. In this case, what we assume to be quenched

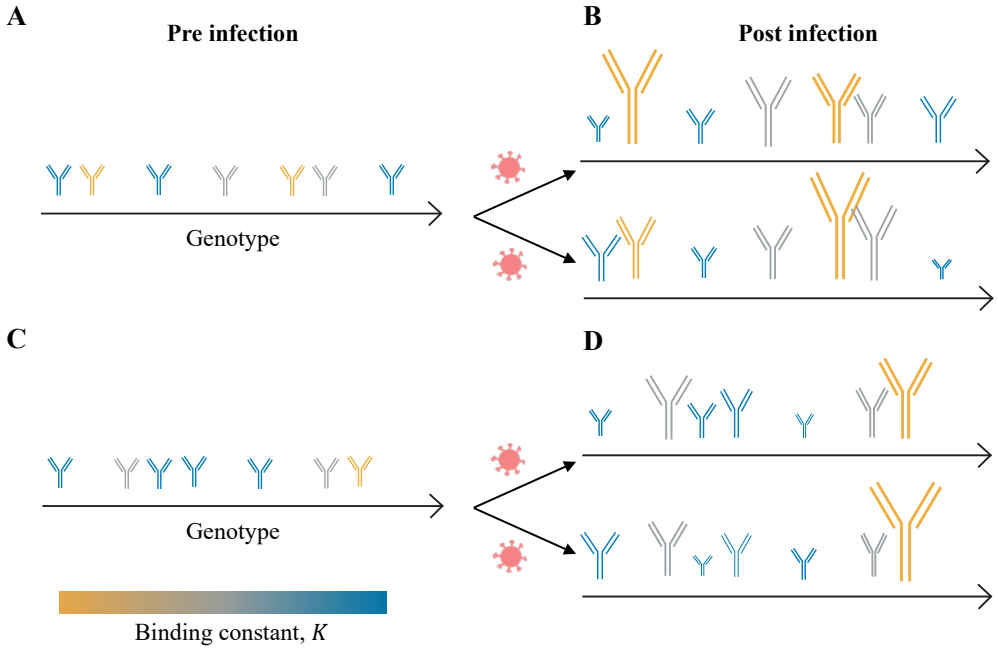


Fig. 5.2. Different levels of disorder in the immune system. A For a given realization of the B cell repertoire, the genotypes of BCRs are fixed (quenched) during the relevant time scale of an infection. Here, this is shown by fixed positions in the genotype axis. For a given viral antigen, each B cell clone is also characterized by a binding affinity, here represented by the color. **B** After an infection, responding clones change their clones size, here represented by the size of the antibody figure. Hypothetical different infections in the same host at the same time would fluctuate due to the intrinsic stochastic nature of the activation-proliferation process. **C** A different realization of the B cell repertoire than in **A**. **D** Two different infections occurring in the naive repertoire in **C**.

are the genotypes of all the BCRs in the repertoire. Now, if we consider a given antigen, say α, and map the genotype of each BCR to its binding affinity for α, then we can say that the affinities of the B cell repertoire are quenched.

As a *gedanken* experiment, we can imagine repeated infections with the same pathogen in the same host at the same time (in the same realization of the repertoire disorder). In Figures 5.2A and B we sketch different infections occurring in the same realization of the naive disorder by the same viral strain. In this case, the fluctuations in the different responses are due to the stochastic nature of the activation of the different B cells.

Although this *gedanken* experiment is useful for understanding the collective activation of the B cell repertoire, it is not very biologically relevant. Given the stochastic generation of the B cell repertoire and the separation of time scales, it is reasonable to assume that each host represents a different realization of the naive disorder. We show in Figures 5.2A and C two different realizations of the B cell repertoire and in Figures 5.2B and D, the result of two independent infections occurring with the same virus in each of the repertoire realizations. It can also be assumed that, even in the same host, different infections with the same pathogen occur with a time interval greater than the turnover rate of the disorder. In other words, each infection experiences a different *disorder* of the repertoire, even in the

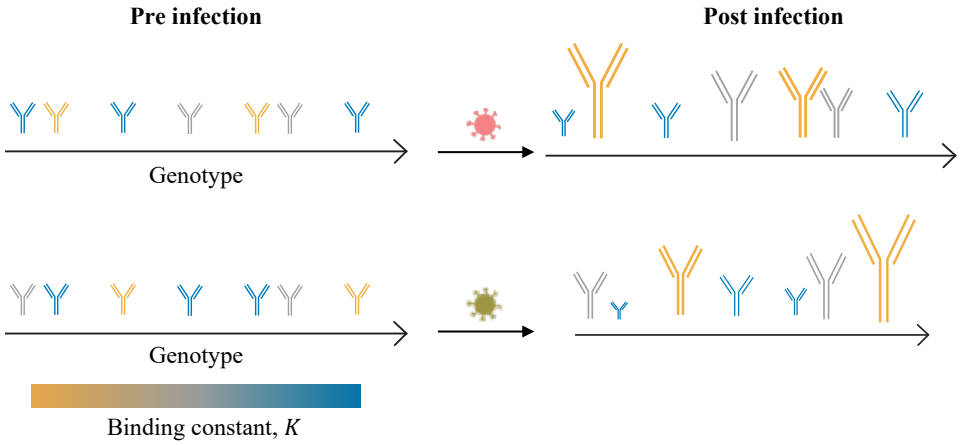


Fig. 5.3. Different levels of disorder in the immune system. The same host (or the same realization of shapes of the B cell repertoire), maps each B cell clone to a different affinity for different viral antigens. Here, exchanging viral antigen (pink for green) changes the binding affinities of the same BCRs in the naive repertoire, and consequently the outcome of the immune response.

same host. We thus assume that the different realizations of the B cell repertoire shown in Figures 5.2A and C, may then represent two different hosts, or the same host at different times during its lifetime. In this context, the host population averages considered in our work (denoted as $\langle \cdot \rangle$) always assume that each host represents a different realization of the B cell repertoire.

As discussed in Chapter 3, the immune response in a high-specificity regime of recognition depends strongly on the disorder of the repertoire. The interplay between the Luria-Delbrück dynamics of infections and the nonequilibrium mechanism of molecular recognition introduces correlations between the large fluctuations in clone size and the quenched affinities of the rare high-affinity B cell clones [62]. This is the proposed mechanism by which these rare high-affinity B cell clones can overcome the large entropy of the naive repertoire.

Another case worth discussing is how the viral antigen under consideration determines the disorder of the repertoire. This point can be a bit involved, so let us break it down carefully. The affinity of a given BCR (or antibody) for a given antigen (pink in Figure 5.3) is determined by its genotype and, of course, by the antigen itself. This means that, for the same realization of quenched genotypes, exchanging the antigen (pink for green in Figure 5.3) would map each genotype to a potential different binding affinity. As we show in Figure 5.3, two hypothetical infections with two different antigens in the same realization of the repertoire disorder will then produce completely different responses. Although this concept sounds trivial in the case of completely different viral antigens, it will be relevant when we consider antigenic evolution and therefore phylogenetically related viral antigens.

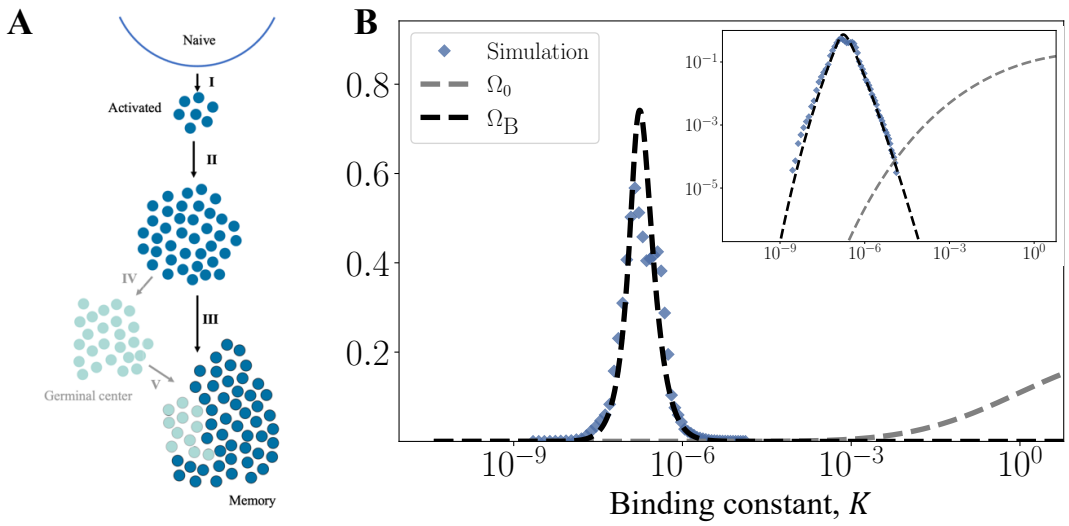


Fig. 5.4. Memory B cell formation A We consider memory formation through the path I-II-II, which includes activation and proliferation of naive B cell and direct differentiation into memory B cells. We do not consider for the moment the complementary path IV-V, which includes germinal center visit and subsequent differentiation into memory B cells. **B** Comparison between naive and activated density of states, Ω_0 and $\Omega_{\mathcal{B}}$. We use the density of activated B cells as the source of memory B cell affinities.

5.2 The B cell memory compartment

Primary response As shown in Chapter 3, during an acute infection, a set of B cell clones $\mathcal{B}$, whose affinities are sampled from the density $\Omega_0(K)$, recognize the pathogen, proliferate at a rate λ_B, and mount a rapid and potent immune response [62]. This process is characterized by the binding affinities of the responding set , $\{K_{\mathbf{b}}\}_{\mathbf{b}\in\mathcal{B}}$, and their corresponding clone sizes, $\{N_{\mathbf{b}}\}_{\mathbf{b}\in\mathcal{B}}$. Figure 5.4B shows a comparison between the affinity spectrum of the naive repertoire, Ω_0, and the affinity spectrum of the responding set, $\Omega_{\mathcal{B}}$[1].

Despite the stochastic nature of this process, we have derived mean-field equations that describe the average affinity, $\langle K\rangle(t', t^*)$, and clone size, $\langle N\rangle(t', t)$, of B cell clones as a function of their activation time, t', given by

$$\langle K\rangle(t', t^*) = \mathcal{K}^* \exp\left[v(t' - t^*)\right] \tag{5.9}$$

$$\langle N\rangle(t, t') = \exp\left[\lambda_B(t - t')\right]. \tag{5.10}$$

with

$$t^* = t_0 + \frac{1}{\lambda_A}\log\left[\left(\frac{\mathcal{K}^*}{K_{\text{step}}}\right)^p\right] \quad \text{and} \quad t_0 = \frac{1}{\lambda_A}\log\left[\left(\frac{\lambda_A}{b_0 k_{\text{on}}\rho_B}\right)^p\right] \tag{5.11}$$

where $\mathcal{K}^*$ is the expected highest available binding affinity in the repertoire, t^* is the expected activation time of a clone with binding affinity $\mathcal{K}^*$[2], and $v = \lambda_A/p$,

[1] In reference [62], we have denoted $\Omega_{\mathcal{B}}$ as Ω_{act}

[2] In references [62, 5], we have denoted the host population average of the highest available

λ_A is the exponential growth rate of the pathogen, λ_B is the exponential growth rate of activated B cells, p is the number of proofreading steps, b_0 is the number of BCRs in the surface of a single B cell, k_{on} is the diffusion-limited association rate and ρ_B the molar concentration of B cell clones [62].

Additionally, we have derived an equation for the expected total number of activated B cell clones, $L_{\text{act}}(t, t^*)$, given by

$$\langle L_{\text{act}} \rangle(t, t^*) = \exp\left[v\beta^*(t - t^*)\right] \tag{5.12}$$

where β^* measures the density of B cell clones at $\mathcal{K}^*$ in affinity space. Activated B cell clones proliferate until they reach a common carrying capacity $\bar{N}$. Here we neglect the nonlinear effects on growth due to saturation closed to $\bar{N}$. Equations 5.9-5.12 describe the response dynamics of the B cell repertoire facing a proliferating antigen. An important consequence of Equations 5.9 and 5.10 is that the expected clone size of activated B cell clones is related to its corresponding expected binding affinity by the relation

$$\langle N \rangle(t) = \langle N^* \rangle(t) \left(\frac{\langle K \rangle}{\mathcal{K}^*} \right)^{-\lambda_B / v} \qquad \text{with} \qquad \langle N^* \rangle(t) = \exp[\lambda_B (t - t^*)] \tag{5.13}$$

where $\langle N^* \rangle(t)$ is the expected clone size of a B cell clone with affinity $\mathcal{K}^*$. We quantify the quality of the response by the potency function $\mathcal{Z}(t)$, a combined quantity of clone size and affinity, given by

$$\mathcal{Z}(t) = \sum_{\mathbf{b} \in \mathcal{B}} z_{\mathbf{b}}(t) \qquad \text{with} \qquad z_{\mathbf{b}}(t) \equiv \frac{N_{\mathbf{b}}(t)}{K_{\mathbf{b}}}. \tag{5.14}$$

Using the potency of a set of B cells, we can estimate the neutralization capacity of antibodies produced by such set after differentiation into plasma cells. For that, we calculate the probability of finding a viral particle bound by an antibody, p_{bound}, as

$$\begin{aligned} p_{\text{bound}} &= 1 - p_{\text{unbound}} \\ &= 1 - \prod_{\mathbf{b}} \left(1 - \frac{1}{1 + \frac{K_{\mathbf{b}}}{\rho_{\mathbf{b}}^{ab}}} \right). \end{aligned} \tag{5.15}$$

where $\rho_{\mathbf{b}}$ is the concentration of antibodies secreted by plasma cells derived from B cell clone $\mathbf{b}$. Assuming $K_{\mathbf{b}}/\rho_{\mathbf{b}}^{ab} \gg 1$ for all $\mathbf{b}$, and neglecting terms of order $\mathcal{O}((\rho_{\mathbf{b}}^{ab}/K_{\mathbf{b}})^2)$ we can approximate Equation 5.15 as

$$p_{\text{bound}} \approx \sum_{\mathbf{b}} \frac{\rho_{\mathbf{b}}^{ab}}{K_{\mathbf{b}}} \tag{5.16}$$

binding affinity in a given repertoire as K^*. In this chapter, we call K^* the random variable that is independently realized in each immune repertoire, distributed according to $P(K^*)$, and we call its host populations average $\langle K^* \rangle \equiv \mathcal{K}^*$.

Now, assuming that the concentration of antibodies is proportional to the number of B cells that produce them, we obtain

$$p_{\text{bound}} \approx \sum_{\mathbf{b}} \frac{\alpha N_{\mathbf{b}}}{K_{\mathbf{b}}} = \alpha \mathcal{Z} \tag{5.17}$$

This relationship between $\mathcal{Z}$ and the total binding probability will become relevant to the *in vivo* protection conferred by memory B cells, as we discuss below.

Memory formation As we discussed above, we assume that memory B cell clones are formed directly after the activation-proliferation process during a primary response, neglecting the role of affinity maturation for the moment (See Figure 5.4A). We say that a fraction $\nu = M/\bar{N}$ of activated B cells become memory B cells. Here we assume that the participation of each clone to memory does not depend on its affinity [32]. Here we assume that ν is of the order of 10%. We therefore assume that the expected size of a memory B cell clone, $\hat{N}_{\mathbf{b},0}$, is proportional to its final clone size after a primary infection. We use $\hat{\cdot}$ to denote quantities in the memory compartment. Using Equation 5.13, we say that the expected clone size of a memory B cell clone with binding affinity $\hat{K}$ is given by

$$\langle \hat{N}_0 \rangle (\hat{K}) = \mu \langle \bar{N}^* \rangle \left(\frac{\langle \hat{K} \rangle}{\mathcal{K}^*} \right)^{-\lambda_B/v} . \tag{5.18}$$

where $\langle \bar{N}^* \rangle$ is the expected clone size of a clone with affinity $\mathcal{K}^*$ at carrying capacity at the end of a primary response. Assuming that during a memory response, all members of a given memory clone are activated and start proliferating at the same time, this leads to a modification in Equation 5.10; the initial clone size at activation is now given by $\hat{N}_0(\hat{K})$. The expected clone size of memory B cell clones activated at time t' is given by

$$\begin{aligned} \langle \hat{N} \rangle (t, t') &= \langle \hat{N}_0 \rangle (\hat{K}) \cdot \exp\left[\lambda_B (t - t')\right]. \\ &= \mu \langle \bar{N}^* \rangle \left(\frac{\langle \hat{K} \rangle}{\mathcal{K}^*} \right)^{-\lambda_B/v} \cdot \exp\left[\lambda_B (t - t')\right] \end{aligned} \tag{5.19}$$

5.3 Immune potency dynamics

During a primary infection, the peak viral load and the severity of the infection are not determined solely by the B cell response. In fact, the role of the innate immune system is extremely important, perhaps even more so. Although this process is not fully understood, it is thought that viruses reach a carrying capacity determined by cell death, innate immunity induced inflammation, antibody neutralization from the plasmablast response, and other factors. However, we assume that the severity of a recurrent infection is largely determined by the effective and rapid recall of memory B cells. For that reason, in this section we study the dynamics of the potency function.

5.3.1 Naive vs. Memory dynamics

Primary response We begin to study the dynamics of potency elicited from the naive B cell repertoire during a primary infection. We can approximate $\langle \mathcal{Z}(t) \rangle$ in Equation 5.14 as

$$\langle \mathcal{Z} \rangle(t) = \int_{t^*}^{t} \frac{\langle N \rangle(t,t')}{\langle K \rangle(t',t^*)} \langle \dot{L}_{\text{act}} \rangle(t',t^*)\, \mathrm{d}t' \tag{5.20}$$

where the sum in Equation 5.14 is approximated by an integral using the density of activated clone in the interval $[t', t' + \mathrm{d}t']$. Using Equations 5.9-5.12 we obtain

$$\begin{aligned} \langle \mathcal{Z} \rangle(t) &= v\beta^* \frac{1}{\mathcal{K}^*} e^{\lambda_B t} e^{vt^*} e^{-v\beta^* t^*} \int_{t^*}^{t} e^{-\lambda_B t'} e^{-vt'} e^{v\beta^* t'}\, \mathrm{d}t' \\ &= \frac{v\beta^*}{\sigma} \frac{\langle N^* \rangle(t)}{\mathcal{K}^*} \left[e^{\sigma(t-t^*)} - 1 \right] \end{aligned} \tag{5.21}$$

with $\sigma = v\beta^* - v - \lambda_B$.

An equivalent result is obtained by integrating over affinities K instead of time t as follows

$$\langle \mathcal{Z}(t) \rangle = \int_{\mathcal{K}^*}^{\langle K \rangle(t)} \frac{\langle N \rangle(K)}{K} \Omega_0(K) \mathrm{d} \log K \tag{5.22}$$

Using Equations 5.9 and 5.13, and assuming that the density of B cell clones closed to $\mathcal{K}^*$ is given by $\Omega_0(K) \approx \beta^* (K/\mathcal{K}^*)^{\beta^*}$ [62]

$$\begin{aligned} \langle \mathcal{Z}(t) \rangle &= \frac{\beta^* \langle N^* \rangle(t)}{{\mathcal{K}^*}^{\beta^* - \lambda_B/v}} \int_{\mathcal{K}^*}^{\langle K \rangle(t)} K^{\beta^* - 1 - \lambda_B/v} \mathrm{d} \log K \\ &= \frac{\beta^* \langle N^* \rangle(t)}{\mathcal{K}^* {\mathcal{K}^*}^{\sigma/v}} \int_{\mathcal{K}^*}^{\langle K \rangle(t)} K^{\sigma/v} \mathrm{d} \log K \\ &= \frac{v\beta^*}{\sigma} \frac{\langle N^* \rangle(t)}{\mathcal{K}^*} \left[e^{\sigma(t-t^*)} - 1 \right] \end{aligned} \tag{5.23}$$

recovering the result from Equation 5.21. This equivalence, resulting from the relationship between clone size (or time) and affinity, demonstrates the computational power of this mean-field solution.

From Equation 5.21 we have two distinct regimes for the exponential growth rate of the potency, $\lambda_{\mathcal{Z}}$. If $\sigma < 0$, the first term inside the squared brackets decays rapidly for $t > t^*$ and thus we have that $\langle \mathcal{Z} \rangle(t) \sim \langle N^* \rangle(t) \sim \exp[\lambda_B t]$. If $\sigma > 0$, we have a modified exponent $\langle \mathcal{Z} \rangle(t) \sim \exp[(\lambda_B + \sigma)t] = \exp[(v\beta^* - v)t]$. In summary we can approximate the potency as

$$\langle \mathcal{Z} \rangle(t, t^*) = \langle \mathcal{Z}_0 \rangle \exp[\lambda_{\mathcal{Z}}(t - t^*)] \tag{5.24}$$

with

$$\lambda_{\mathcal{Z}} = \max[\lambda_B, v(\beta^* - 1)] \quad \text{and} \quad \langle \mathcal{Z}_0 \rangle = \frac{v\beta^*}{|\sigma|} \frac{1}{\mathcal{K}^*}. \tag{5.25}$$

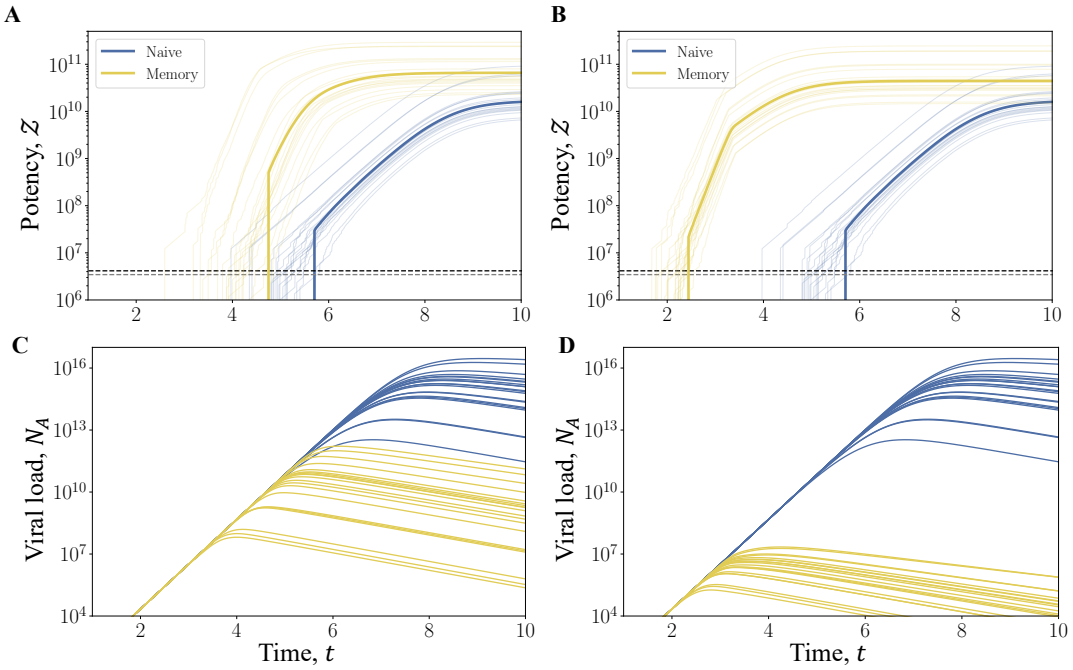


Fig. 5.5. Potency and viral dynamics **A** and **B** show potency $\mathcal{Z}$ as a function of time for primary infections (naive responses) and corresponding memory responses. Transparent thin lines represent single trajectories and thick line average over ensemble of trajectories. Dashed horizontal lines denotes $\mathcal{Z}_0 = 1/K_1$; **C** and **D** show viral load N_A as a function of time for primary infections (naive responses) and corresponding memory responses. Each curve corresponds to a single trajectory shown in **A** and **B**. In **A** and **C** we show memory B cells activated with the same strength of proofreading as the naive B cells ($p_{\mathbf{mem}} = 1$) In **B** and **D** we show memory B cells activated without proofreading . ($p_{\mathbf{mem}} = p$). Other parameters are: growth rates $\lambda_A = 6\mathrm{d}^{-1}$; $\lambda_B = 2\mathrm{d}^{-1}$; kinetic parameters $k_{\mathrm{on}} = 10^6 M^{-1} s^{-1}$ and $k_{\mathrm{step}} = 0.5\mathrm{min}^{-1}$; number of BCRs per cell, $b_0 = 10^5$; repertoire size $L0 = 10^7$; carrying capacity $\bar{N} = 2 \times 10^4$.

The limit $\lambda_{\mathcal{Z}} = \lambda_B$ represents the case where the potency is growth-dominated: few early activated clones take over the activated repertoire and the total potency is dominated by their contribution. The limit $\lambda_{\mathcal{Z}} = v(\beta^* - 1)$ represents the case where the potency is activation-dominated: the activation of B cell clones contributes significantly compared to the growth of individual clones. Interestingly, with physiological values $\lambda_B = 2\,\mathrm{days}^{-1}$, $\lambda_A = 6\,\mathrm{days}^{-1}$, $p = 3$, and $\beta^* = 2.2$, we have an estimate $\sigma \approx 0.4\,\mathrm{days}^{-1}$, which means that a primary response is activation-dominated, but not far from the transition point.

Memory response During a recurrent infection, activation of memory B cell clones follows a similar dynamics as naive B cell clones. We can estimate the potency due to memory as

$$\hat{\mathcal{Z}}(t) = \sum_{\mathbf{b}} \frac{\hat{N}_{\mathbf{b}}}{\hat{K}_{\mathbf{b}}} \tag{5.26}$$

Simlarly, we can approximate the host population average as

$$\langle \hat{\mathcal{Z}} \rangle(t) = \int_{t^*}^{t} \frac{\langle \hat{N} \rangle(t,t')}{\langle K \rangle(t',t^*)} \langle \dot{L}_{\text{act}} \rangle(t') \, \mathrm{d}t' \tag{5.27}$$

Note that the difference between Equation 5.20 and Equation 5.27 is the initial clone size of the activated B cell clones. Using Equations 5.9,5.12, 5.19 we obtain

$$\begin{aligned} \langle \hat{\mathcal{Z}} \rangle(t) &= \langle \bar{N}^* \rangle \nu v \beta^* \frac{1}{\hat{\mathcal{K}}^*} e^{\lambda_B t^*} e^{\lambda_B t} e^{v t^*} e^{-v\beta^* t^*} \int_{t^*}^{t} e^{-2\lambda_B t'} e^{-v t'} e^{v\beta^* t'} \, \mathrm{d}t' \\ &= \frac{\nu v \beta^*}{\hat{\sigma}} \frac{\langle \hat{N}^* \rangle(t)}{\hat{\mathcal{K}}^*} \left[e^{\hat{\sigma}(t-t^*)} - 1 \right] \end{aligned} \tag{5.28}$$

with $\hat{\sigma} = v\beta^* - v - 2\lambda_B = \sigma - \lambda_B$ and where $\langle \hat{N}^* \rangle(t) = \langle \bar{N}^* \rangle \exp[\lambda_B(t - t^*)]$. Note that $\hat{\mathcal{K}}^*$ in Equation 5.28 is not always the same as the naive equivalent $\mathcal{K}^*$ in Equation 5.21. Initially they correspond to the same values but, as we discuss below, this might change due to antigenic viral evolution. Similar to a primary response, we can approximate the potency as

$$\langle \hat{\mathcal{Z}} \rangle(t, \hat{t}^*) = \langle \hat{\mathcal{Z}}_0 \rangle \exp[\hat{\lambda}_{\mathcal{Z}}(t - t^*)] \tag{5.29}$$

with

$$\hat{\lambda}_{\mathcal{Z}} = \max[\lambda_B, v(\beta^* - 1) - \lambda_B] \quad \text{and} \quad \langle \hat{\mathcal{Z}}_0 \rangle = \frac{\nu v \beta^*}{|\hat{\sigma}|} \frac{\langle \bar{N}^* \rangle}{\hat{\mathcal{K}}^*} \tag{5.30}$$

This last expression suggests that after the first activation, the potency has an initial value a factor $\nu \langle \bar{N}^* \rangle$ larger than in the primary response. Under our assumptions, this is not possible because B cells clones do not change their affinity when they become memory. Simulations show that the increase in initial clone size is translated into an earlier activation time for memory responses. Based on that, we define a memory activation time

$$\hat{t}^* = t^* - \frac{1}{\hat{\lambda}_{\mathcal{Z}}} \log \left[\nu \langle \bar{N}^* \rangle \right]. \tag{5.31}$$

The final mean-field approximation for the potency is given by

$$\langle \hat{\mathcal{Z}} \rangle(t, \hat{t}^*) = \langle \hat{\mathcal{Z}}_0 \rangle \exp[\hat{\lambda}_{\mathcal{Z}}(t - \hat{t}^*)] \tag{5.32}$$

with

$$\langle \hat{\mathcal{Z}}_0 \rangle = \frac{v\beta^*}{|\hat{\sigma}|} \frac{1}{\hat{\mathcal{K}}^*} \tag{5.33}$$

In this case, for physiological relevant parameters, $\hat{\sigma} < 0$. This means that a memory response is dominated by growth with negligible contribution from activation. This represents a change compared to a primary response. Figure 5.5A shows individual potency trajectories for primary and memory responses, as well as the population average over an ensemble of different infections with the same antigen.

Memory depth We can now estimate the expected average potency difference between a recall memory response and a naive response as follows

$$\begin{aligned}\langle \Delta\hat{\mathcal{Z}} \rangle &\equiv \log\left[\frac{\langle \hat{\mathcal{Z}} \rangle(t,\hat{t}^*)}{\langle \mathcal{Z} \rangle(t,t^*)}\right] \\ &= \log\left[\frac{\langle \hat{\mathcal{Z}}_0 \rangle}{\langle \mathcal{Z}_0 \rangle}\right] + \hat{\lambda}_{\mathcal{Z}}(t-\hat{t}^*) - \lambda_{\mathcal{Z}}(t-t^*) \\ &= \log\left[\frac{\sigma}{\hat{\sigma}}\right] + t\left(\hat{\lambda}_{\mathcal{Z}} - \lambda_{\mathcal{Z}}\right) + \lambda_{\mathcal{Z}} t^* - \hat{\lambda}_{\mathcal{Z}}\hat{t}^* \end{aligned} \tag{5.34}$$

Note that this is a definition. We call $\langle \Delta\hat{\mathcal{Z}} \rangle$ the *depth* of the B cell memory protection. It can be understood as the time-dependent expected gain in neutralization capacity of a *in vivo* memory response compared to a naive primary response. Assuming $\hat{\lambda}_{\mathcal{Z}} \approx \lambda_{\mathcal{Z}}$ and $\hat{\sigma} \approx \sigma$, we get that

$$\langle \Delta\hat{\mathcal{Z}} \rangle \approx \lambda_{\mathcal{Z}}(t^* - \hat{t}^*) \approx \log(\nu\langle \bar{N}^* \rangle) \tag{5.35}$$

In the following sections, we will study how memory potency is reduced due to viral antigenic evolution. In this context, the depth of the response can be interpreted as the necessary decrease of memory potency at which a memory response is basically indistinguishable from a typical naive response. This can be used to estimate the transition from a memory protective response to a new naive response triggered by a breakthrough infection.

5.3.2 Do memory B cells proofread?

An important quality of memory B cells is that they are stimulated by the presence of the cognate antigen easier than naive B cells. We ask the question whether it makes sense for memory B cells to reduce the strength of proofreading that characterizes the naive response. Let us define the number of proofreading steps that memory B cells perform as p_{mem}. Figure 5.5B shows individual potency trajectories for primary and memory responses, as well as the population average over an ensemble of different infections with the same antigen. In this case, memory B cells perform a single activation step in the recognition process ($p_{\text{mem}} = 1$, no proofreading). Compared to Figure 5.5A in which memory B cells perform the same amount of proofreading steps as naive B cells ($p_{\text{mem}} = p$), it is clear that the non-proofreading case produces a more rapid response than the proofreading case.

We interpret this result as follow. As it was shown in Chapter 3, a tuned strength of proofreading ($p \approx \beta^*$) is used by the naive B cells in order to overcome the high entropy of the naive repertoire. In this case, a high-specificity response is achieved at expense of a delay in response time. However, proofreading is justified because the average affinity of the B cell clones that would dominate a low-specificity response is significantly lower than that of a high-specificity response. In this case, proofreading solves the problem of selecting for activation the rare high-affinity B cell clones in the vast B cell repertoire. In the case of a

memory response, by removing the need for proofreading, we eliminate the delay in response of the previously selected B cell clones.

Is there another reason why memory B cell would prefer to avoid kinetic proofreading? As mention above, with a minimum strength of kinetic proofreading it is possible to generate immune responses in the high-specificity regime [62]. In this regime, we showed that affinity and clone size are correlated as given by Equation 5.13. We can rewrite the potency of an individual activated B cell clone from Equation 5.14 as

$$z_{\mathbf{b}} = N_{\mathbf{b}} \exp\left[-\Delta E_{\mathbf{b}}\right] \quad \text{with} \quad \Delta E_{\mathbf{b}} \equiv \log K_{\mathbf{b}} \tag{5.36}$$

and define a probability distribution over B cell clones as

$$\mathcal{P}_{\mathbf{b}} = \frac{z_{\mathbf{b}}}{\mathcal{Z}}. \tag{5.37}$$

The distribution $\mathcal{P}_{\mathbf{b}}$ becomes the Boltzmann distribution if all clones have clone size $N_{\mathbf{b}} = 1$. When this is not the case and B cell clones have in general clone size different than 1, we can rewrite $\mathcal{P}_{\mathbf{b}}$ as

$$\mathcal{P}_{\mathbf{b}} = \frac{\exp\left[-\Delta E_{\mathbf{b}} + \log N_{\mathbf{b}}\right]}{\mathcal{Z}}. \tag{5.38}$$

Now, using the mean-field Equation 5.13 we can rewrite

$$\begin{aligned} \mathcal{P}_{\mathbf{b}} &= \frac{\exp\left[-\Delta E_{\mathbf{b}} - \frac{\lambda_B}{v}\Delta E_b + const\right]}{\mathcal{Z}} \\ &= \frac{\exp\left[-(1+\frac{\lambda_B}{v})\Delta E_{\mathbf{b}} + const\right]}{\mathcal{Z}} \end{aligned} \tag{5.39}$$

From this last expression, we can define a new probability distribution over B cells clones with a different effective inverse temperature as

$$\mathcal{P}_{\mathbf{b}}(\theta) = \frac{\exp\left[-\theta\Delta E_{\mathbf{b}}\right]}{\mathcal{Z}_\theta} \quad \text{with} \quad \theta = 1 + \frac{\lambda_B}{v} \tag{5.40}$$

and with $\mathcal{Z}_\theta = \sum_{\mathbf{b}} \exp\left[-\theta\Delta E_{\mathbf{b}}\right]$. The new effective inverse temperature θ measures how much concentrated are cells in the high-affinity B cell clones. In other words, how strong are affinity and clone size correlated. In order to read θ from the simulations, we calculate the entropy of $\mathcal{P}_{\mathbf{b}}(\theta)$ as

$$\begin{aligned} S(\mathcal{P}_{\mathbf{b}}(\theta)) &= -\sum_{\mathbf{b}} \mathcal{P}_{\mathbf{b}}(\theta) \log \mathcal{P}_{\mathbf{b}}(\theta) \\ &= -\sum_{\mathbf{b}} \frac{\exp\left[-\theta\Delta E_{\mathbf{b}}\right]}{\mathcal{Z}_\theta} \log \frac{\exp\left[-\theta\Delta E_{\mathbf{b}}\right]}{\mathcal{Z}_\theta} \\ &= \theta\, \mathbb{E}[\Delta E_{\mathbf{b}}] + \log Z_\theta \end{aligned} \tag{5.41}$$

where the expected value $\mathbb{E}[\cdot]$ is calculated over B cells clones.

We perform a linear fit on the $S(\mathcal{P}) - \log Z$ vs. $\mathbb{E}[\Delta E]$ plot as shown in Figure 5.6A. Figure 5.6B shows the effective temperature θ^{-1} for different populations of B cell clones after primary and memory responses. Let us break

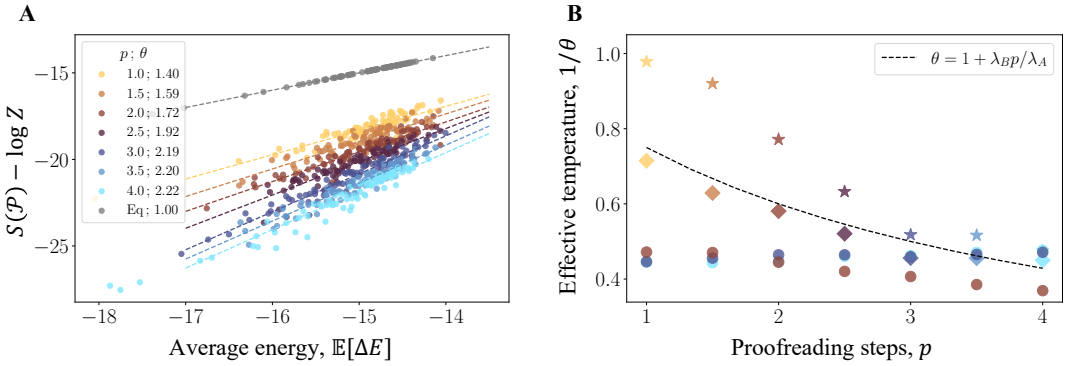


Fig. 5.6. Effective temperature of an immune response A Linear fit of Equation 5.41. We show the entropy/average energy relation for an ensemble of 80 primary infections for different values of p. Inferred inverse temperatures are shown in the caption. **B** Effective temperature for the set of 10 largest B cell clones in a primary response (stars), 10 highest-affinity B cell clones in a primary response (diamonds) and 10 highest-affinity clones in B cell a memory response (circles). In stars and diamonds, each symbol corresponds to a value of p given by its x-coordinate. In this case, each color is associated with a value of p. For circles, the color corresponds to the value of p used in the corresponding primary infection and p_{mem} is given by its x-coordinate. Other parameters as in Figure 5.5

down the content of this Figure 5.6B. First, we consider the population of the 10 largest clones after a primary response (see stars). As expected from the results of Chapter 3, the mean-field solution for the largest B cell clones is only valid in the high-specificity given by $p \gtrsim \beta^*$. If we consider the population of the B cell clones with the highest affinity after a primary response (see diamonds), they follow the mean field solution. Interestingly, for $p = 1$ corresponding to no kinetic proofreading, we observe an effective temperature smaller than 1. In this case, we define a modified inverse temperature

$$\theta_0 = 1 + \frac{\lambda_B}{\lambda_A} \tag{5.42}$$

as the strength of *exponential proofreading*, a new mechanism by which a system can harvest exponential growth to generate greater specificity beyond the limits of equilibrium.

We generate memory responses in which memory B cell clones are produced during primary responses with $p = 4$ (see circles). In this case, we vary p_{mem}. Preliminary results show that although the effective temperature decrease compared to the one achieved during the primary response, the relation with p_{mem} is now opposite. Smaller values of p_{mem} generate smaller effective temperatures, although the effect is not large.

5.3.3 Viral collapse under *in vivo* memory response

Now we want to investigate the effect of the rapid increase of $\hat{\mathcal{Z}}$ on the viral load. For that, we start by considering the relation between $\mathcal{Z}$ and the binding

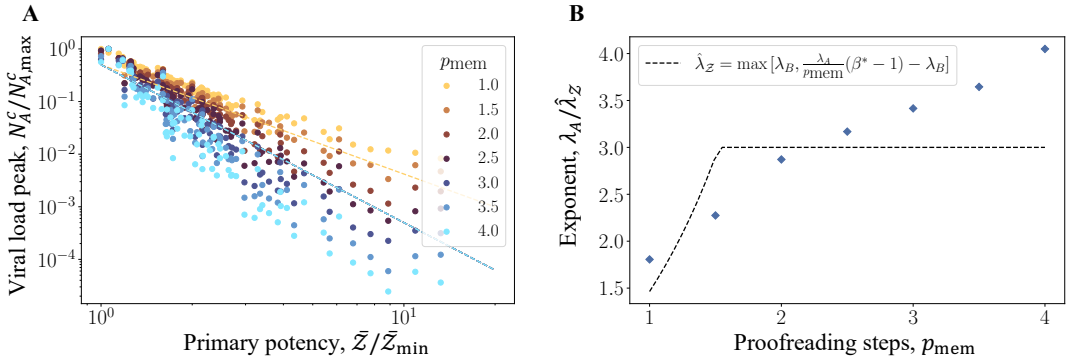


Fig. 5.7. Predicting memory response from potency of primary response A Linear fit of log-log plot of Equation 5.48. We show the viral load peak in memory responses as a function of the potency at carrying capacity of the corresponding primary infection. We show results for an ensemble of 80 primary/memory infections for different values of p_{mem} and $p = 4$ **B** Inferred exponent $\lambda_A/\hat{\lambda}\mathcal{Z}$. Diamonds show the values from linear fits and dashed line the expectation from theoretical model. Other parameters as in Figure 5.5

probability of the virus and all circulating antibodies. Let assume that the dynamics of the number of viral particles is given by

$$\dot{N}_A = \lambda_A(\hat{\mathcal{Z}})N_A + \text{decay terms}, \tag{5.43}$$

We consider a potency-dependent viral growth of the virus as follows

$$\lambda_A(\hat{\mathcal{Z}}) = \lambda_A\left[1 - p_{\text{bound}}(\hat{\mathcal{Z}})\right] \tag{5.44}$$

where we have assumed that only the unbound fraction of viruses are able to replicate and with $p_{\text{bound}}(\hat{\mathcal{Z}})$ given by Equation 5.17. In the limit low binding probability regime given by $\alpha\hat{\mathcal{Z}} \ll 1$, the modified viral growth rate is given by

$$\lambda_A(\hat{\mathcal{Z}}) \approx \lambda_A \cdot (1 - \alpha\hat{\mathcal{Z}}(t)). \tag{5.45}$$

We define the *viral collapse* as the time point when the viral growth is approximately zero. This condition defines a viral collapse time t^c. Using our mean-field approximation of Equation 5.32, the condition of viral collapse is given by

$$\alpha\langle\hat{\mathcal{Z}}_0\rangle \exp\left[\hat{\lambda}_{\mathcal{Z}}(t^c - \hat{t}^*)\right] = 1. \tag{5.46}$$

From here we get that the mean-field critical collapse time is given by

$$t^c = \hat{t}^* - \frac{1}{\hat{\lambda}_{\mathcal{Z}}} \log\left[\alpha\langle\hat{\mathcal{Z}}_0\rangle\right]. \tag{5.47}$$

Equation 5.44 implies that the neutralization produced by antibodies shifts rapidly from very weak to almost complete as the potency grows exponentially. Assuming that the number of antigen particles grows as $N_A(t) = \exp(\lambda_A t)$ (neglecting saturation at carrying capacity), the collapse viral load is given by

$$N_A^c \equiv N_A(t^c) = \left[\alpha \exp[-\hat{\lambda}_Z \hat{t}^*]\hat{\mathcal{Z}}_0\right]^{-\lambda_A/\hat{\lambda}_{\mathcal{Z}}} \tag{5.48}$$

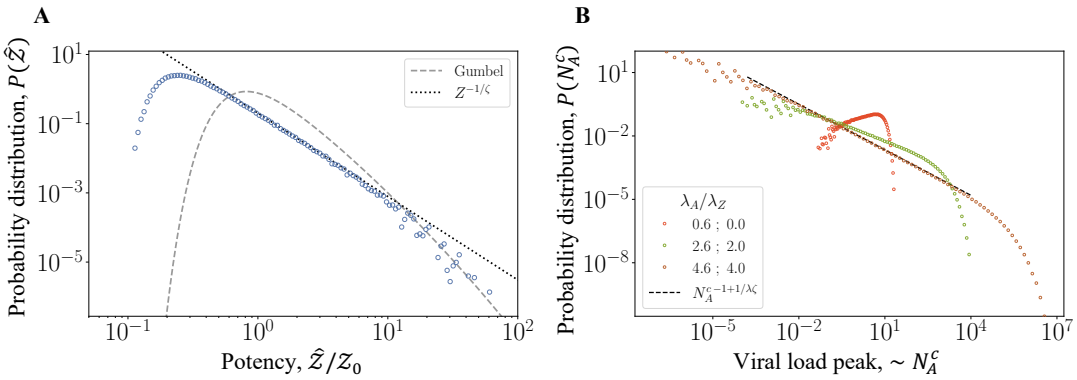


Fig. 5.8. Fluctuations of potency and critical viral load A Fluctuations in the potency at carrying capacity after a primary response. Circles show the probability distribution over an ensemble of 10^5 infections. Black dashed line shows the expected fluctuations from the clone-size fluctuations deriving from the Luria-Delbrück dynamics. Grey dashed curve shows the Gumbel distribution that describes the probability of the highest affinity B cell clones in the repertoire $P(K^*)$. Potency axis is scaled by $\mathcal{Z} = 1/\mathcal{K}^*$ **B** Fluctuations in the critical viral load of hypothetical memory responses for each primary response in **A**. The critical viral load is calculated using Equation 5.48. All parameters as in Figure 5.5

As discussed above, the term inside the squared brackets in Equation 5.48 is fully determined by the primary response. This suggests that we can estimate the collapse viral load from the potency of the primary response. We generated an ensemble of primary and memory responses and show the outcome in Figure 5.7 Here we show the collapse viral load of a memory response, N_A^c, as a function of the carrying capacity potency, $\bar{\mathcal{Z}}$ of the corresponding primary response. We show the results for different values of p_{mem}. In this case, the potency exponente is given by

$$\hat{\lambda}_{\mathcal{Z}} = \max\left[\lambda_B, \frac{\lambda_A}{p_{\text{mem}}}(\beta^* - 1) - \lambda_B\right] \tag{5.49}$$

We found a strong correlation between N_A^c and $\bar{\mathcal{Z}}$, following a power-law behavior. However, the exponent differs from the expected $\lambda_A/\hat{\lambda}_{\mathcal{Z}}$. The accurate calculation of this exponent remains an open problem in the project.

Memory B cell protection In this section, we would begin to construct a *in vivo* protection function for the memory B cell response. First, we can ask how likely is that a host is protected against a reinfection with the same virus. To do this, we want to consider how much the final potency can fluctuate after a primary infection. As we have shown in Chapter 3, potency has large fluctuation characteristic of a generalized Luria-Delbrück dynamics. Therefore, we would expect that protection also has large fluctuations at the population level. This is also shown in Figure 5.5C and D, where the peak of the different viral load curves fluctuates over several order of magnitude for different memory responses against the same viral strain.

5.3.4 Antigen viral evolution

As discussed in the introduction of the present chapter, when a virus changes by an antigenic mutation, we can model such change by an effective increase in the binding energy, $\Delta\Delta E_{\mathbf{b}}$, with the different memory B cell clones targeting such epitope

$$\hat{K}_{\mathbf{b}} \to \hat{K}_{\mathbf{b}} \cdot \exp\left(\Delta\Delta E_{\mathbf{b}}\right) \tag{5.50}$$

We call $\Delta\Delta E$ the average of $\Delta\Delta E_{\mathbf{b}}$ over all memory B cell clones.

We are interested to understand how antigenic changes affect the memory potency $\hat{\mathcal{Z}}(t, \hat{t}^*)$. We model the collective shift in binding energy at the level of the memory repertoire as a modification on $\hat{\mathcal{K}}^*$ given by

$$\hat{\mathcal{K}}^* \to \hat{\mathcal{K}}^* \exp\left(\Delta\Delta E\right). \tag{5.51}$$

This shift in $\hat{K}^*$ has an additional effect on the response; the expected activation time of the memory B cell clones targeting changes by

$$\hat{t}^* \to \hat{t}^* + \Delta\Delta E/v. \tag{5.52}$$

Given these two changes, we can express the memory potency from Equation 5.32 as

$$\begin{aligned}
\langle\hat{\mathcal{Z}}\rangle(t, \hat{t}^*, \Delta\Delta E) &= \frac{v\beta^*}{|\hat{\sigma}|} \frac{1}{\hat{\mathcal{K}}^* \exp\left(\Delta\Delta E\right)} \exp\left[-\hat{\lambda}_{\mathcal{Z}}\left(t - \hat{t}^* - \Delta\Delta E/v\right)\right] \\
&= \langle\hat{\mathcal{Z}}\rangle(t, \hat{t}^*) \exp\left[-(1 + \hat{\lambda}_{\mathcal{Z}}/v)\Delta\Delta E\right] \\
&= \langle\hat{\mathcal{Z}}\rangle(t, \hat{t}^*) \exp\left[-\hat{\phi}\Delta\Delta E\right]
\end{aligned} \tag{5.53}$$

with $\hat{\phi} = 1 + \hat{\lambda}_{\mathcal{Z}}/v$

Antiserum vs. memory antigen effects As mentioned above, antiserum antibodies produced by long-lived plasma cells have undergone substantial affinity maturation, increasing their affinity by up to 10^3 times compared to their germline counterparts. As shown in Figure 5.9A, this means that these new B cell clones that are successfully produced during affinity maturation are located closer to the master sequences (in K space) than any other naive B cell clone . As a result of the different mutations acquired by mature antibodies, the expected effect of antigenic changes in viral epitopes on these mature antibodies is different from the effect on the unmatured and clonally related memory B cell clones.

5.4 Multi-epitope system

Among the available epitopes, some of them are preferentially targeted by an immune response, a phenomenon known as *immunodominance*. For example, in the case of Influenza, 5 different epitopes have been characterized, three of

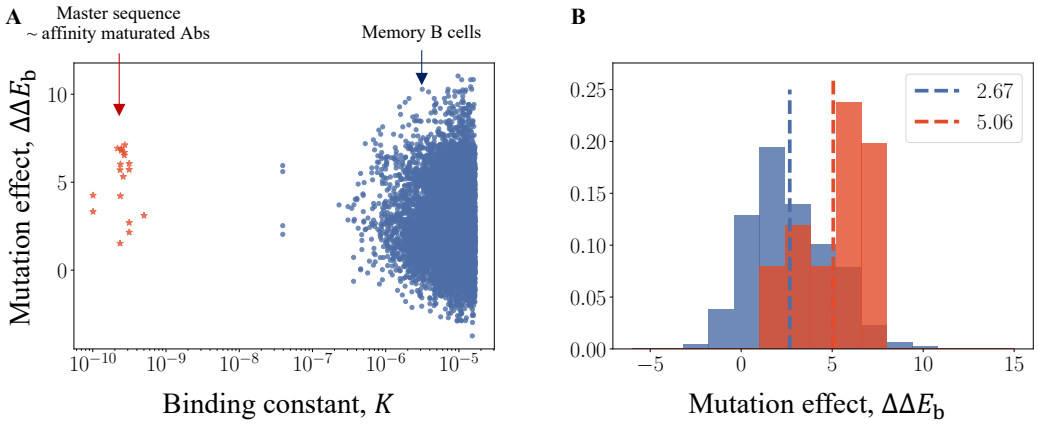


Fig. 5.9. Global epistasis in the underlying energy model A For three random mutations of the initial epitope, we show an scatter plot of the initial binding constant K and the effect of the mutation on the binding energy $\Delta\Delta E\mathbf{b}$ of different populations of B cell clones. Red dots represent the *master sequence*: best binder for an epitope. Blue dots represent B cell clones drawn from $\Omega_{\mathcal{B}}$ closed to $\mathcal{K}^*$ in affinity space. **B** Histogram of the mutational effects $\Delta\Delta E\mathbf{b}$ from **A**. Dashed lines show average mutational effect. Binding energies are calculated using the same linear model as in Chapter 3 and we used the TCRen matrix for the energy effects [66, 62].

which tend to be the ones more frequently targeted by B cell immunity, or more immunodominant [26, 27].

The theory presented in Chapters 3 and 4 and in this chapter has so far assumed that B cells target a single epitope in the viral antigen. In this section we extend the activation-proliferation theory, and in particular Equation 5.24, to the case of a multi-epitope viral antigen. First, we assume that a viral antigen has g independent, equally accessible, and statistically equivalent epitopes [5]. By this we mean that the immune response to all epitopes is characterized by the different densities of state $\Omega^e(K)$, and therefore different distributions for the best available binding affinity in each realization of the repertoire, $P^e(K^{*e})$. We say that the average best available binding affinity of epitope e is $\mathcal{K}^{*e}$.

The g distinct epitopes are characterized by g independent realizations

$$\{K^{*1}, K^{*2}, \ldots, K^{*g}\} \tag{5.54}$$

in each realization of a B cell repertoire. The ranking of affinities in this set defines the immunodominance of the viral epitopes. For a given realization of a B cell repertoire and an ordered set of g independent realizations $\{K^{*e}\}$, let us express

$$K^{*e} = K^{*1} \cdot \exp\left[\Delta\Delta\mathcal{E}^e\right] \quad \text{with} \quad \begin{cases} \Delta\Delta\mathcal{E}^e = 0 & (e = 1) \\ \\ \Delta\Delta\mathcal{E}^e \geqslant 0 & (e = 2, 3, \ldots, g) \end{cases} . \tag{5.55}$$

Note that energy differences $\Delta\Delta E$ are different from intrinsic energy differences $\Delta\Delta\mathcal{E}$. The former are intrinsic epitope energy differences with respect to the immunodominant epitope. The later are energy differences due to changes in the binding epitope with respect to the WT one.

5.4.1 Multi-epitope potency dynamics

Primary response Let us first calculate the potency dynamics against a single epitope e assuming that it does not interact with any other epitopes. In this case, we are assuming that the mean-field solution applies for each single realization of a B cell repertoire. Thus, in this section we do not use the average brackets $\langle\cdot\rangle$. We have for the potency of epitope e the following expression

$$\mathcal{Z}^e(t,t^*,K^{*e}) = \frac{v\beta^*}{\sigma}\frac{1}{K^{*e}}\exp[\lambda_{\mathcal{Z}}(t-t^{*e})] \tag{5.56}$$

where t^{*e} is the associated expected activation time for epitope e. We know that this time is determined by K^{*e} [62]. Using our formalism we can express this time in terms of the energy difference as $\Delta t^{*e} = \Delta\Delta\mathcal{E}^e/v$. This means that, from Equation 5.56 the expected behavior of the $e-$specific potency is given by

$$\begin{aligned}
\mathcal{Z}^e(t,t^*,\Delta\mathcal{E}^e) &= \frac{v\beta^*}{\sigma}\frac{1}{K^{*e}}\exp[\lambda_{\mathcal{Z}}(t-t^{*e})] \\
&= \frac{v\beta^*}{\sigma}\frac{1}{K^{*1}e^{\Delta\Delta\mathcal{E}^e}}\exp[\lambda_{\mathcal{Z}}(t-t^*-\Delta\Delta\mathcal{E}^e/v)] \\
&= \frac{v\beta^*}{\sigma}\frac{1}{K^{*1}}\exp[\lambda_{\mathcal{Z}}(t-t^*)]\exp[-\Delta\Delta\mathcal{E}^e(1+\lambda_{\mathcal{Z}}/v)] \\
&= \mathcal{Z}^{(1)}(t,t^*)\exp[-\Delta\Delta\mathcal{E}^e(1+\lambda_{\mathcal{Z}}/v)]
\end{aligned} \tag{5.57}$$

with $\mathcal{Z}^{(1)}(t,t^*)$ as given by Equation 5.24. From here we can see that the intrinsic drop in potency due to immunodominance is independent of time and given by

$$\frac{\mathcal{Z}^e(t,t^*,\Delta\mathcal{E}^e)}{\mathcal{Z}^{(1)}(t,t^*)} = \exp[-\Delta\Delta\mathcal{E}^e(1+\lambda_{\mathcal{Z}}/v)] \tag{5.58}$$

Until here, we have calculated the potency dynamics specific for epitope e assuming that it is independent of the other epitopes. However, we would like to assume that multi epitopes do not generate a greater response, but that the hypothetical response of a dominant epitope is distributed among all epitopes according to the energy differences $\{\Delta\Delta\mathcal{E}^e\}$, this suggest the following construction for the potency dynamics specific for epitope e as

$$\mathcal{Z}^e(t,t^*,\theta,\Delta\Delta\mathcal{E}^e) \equiv \mathcal{Z}^{(1)}(t,t^*)\mathcal{I}^e(\phi) \tag{5.59}$$

with the immunodominance weight given by

$$\mathcal{I}^e(\phi) = \frac{\exp[-\phi\Delta\Delta\mathcal{E}^e]}{\sum_e \exp[-\phi\Delta\Delta\mathcal{E}^e]} \quad \text{and} \quad \phi = 1+\lambda_{\mathcal{Z}}/v. \tag{5.60}$$

and with $\mathcal{Z}^{(1)}(t,t^*)$ given by Equation 5.24. Thus, the total potency is given by

$$\mathcal{Z}(t,t^*,\phi,\{\Delta\Delta\mathcal{E}^e\}) = \mathcal{Z}^{(1)}(t,t^*)\sum_e \mathcal{I}^e(\phi,\Delta\Delta\mathcal{E}^e) \tag{5.61}$$

This last expression implies that in our construction, the potency that the most immunodominant epitope alone would have, is distributed among the remaining epitopes.

Memory response We can extend Equation 5.29 to the case of multiple epitopes in the same way as the construction in Equation 5.61. We have then

$$\hat{\mathcal{Z}}^e(t,\hat{t}^*,\hat{\phi},\Delta\Delta\mathcal{E}^e) \equiv \hat{\mathcal{Z}}(t,\hat{t}^*)\mathcal{I}^e(\hat{\phi}) \tag{5.62}$$

with $\hat{\mathcal{Z}}(t,\hat{t}^*)$ as given by Equation 5.29 and with

$$\hat{\phi} = 1 + \hat{\lambda}_{\mathcal{Z}}/v \tag{5.63}$$

and thus the total memory potency given by

$$\hat{\mathcal{Z}}(t,\hat{t}^*,\hat{\phi},\{\Delta\Delta\mathcal{E}^e\}) = \hat{\mathcal{Z}}(t,\hat{t}^*)\sum_e \mathcal{I}^e(\hat{\phi},\Delta\Delta\mathcal{E}^e) \tag{5.64}$$

Antigenic evolution From Equation 5.53 we see that an antigen effect on epitope e, $\Delta\Delta E^e$, modifies the epitope-specific potency by a factor $\exp(-\phi\Delta\Delta E^e)$. Therefore, using Equations 5.62 and 5.53, we construct the potency dynamics specific for epitope e as

$$\hat{\mathcal{Z}}^e(t,\hat{t}^*,\hat{\phi},\Delta\Delta\mathcal{E}^e,\Delta\Delta E^e) = \hat{\mathcal{Z}}(t,\hat{t}^*)\hat{\chi}^e(\hat{\phi},\Delta\Delta\mathcal{E}^e,\Delta\Delta E^e) \tag{5.65}$$

with

$$\begin{aligned}\hat{\chi}^e(\hat{\phi},\Delta\Delta\mathcal{E}^e,\Delta\Delta E^e) &= \mathcal{I}^e(\hat{\phi},\Delta\Delta\mathcal{E}^e)\cdot\exp\left[-\hat{\phi}\Delta\Delta E^e\right] \\ &= \frac{\exp\left[-\hat{\phi}\left(\Delta\mathcal{E}^e+\Delta E^e\right)\right]}{\sum_e \exp\left[-\hat{\phi}\Delta\Delta\mathcal{E}^e\right]}\end{aligned} \tag{5.66}$$

and thus the total memory potency given by

$$\hat{\mathcal{Z}}(t,\hat{t}^*,\hat{\phi},\{\Delta\Delta\mathcal{E}^e\},\{\Delta\Delta E^e\}) = \hat{\mathcal{Z}}(t,\hat{t}^*)\sum_e \chi^e(\hat{\phi},\Delta\Delta\mathcal{E}^e,\Delta\Delta E^e) \tag{5.67}$$

We note that while the immunodominance weights sum up to one

$$\sum_e \mathcal{I}^e = 1$$

this is not the case in general for the weights χ^e

$$\sum_e \chi^e \neq 1.$$

We now evaluate the drop in potency due to antigen evolution as

$$\begin{aligned}\Delta\hat{Z} &= \log\left[\frac{\hat{\mathcal{Z}}(t,\hat{t}^*,\hat{\phi},\{\Delta\Delta\mathcal{E}^e\},\{\Delta\Delta E^e\})}{\hat{\mathcal{Z}}(t,\hat{t}^*)}\right] \\ &= \log\sum_e \chi^e(\hat{\phi},\Delta\Delta\mathcal{E}^e,\Delta\Delta E^e) \\ &= \log\sum_e \mathcal{I}^e(\hat{\phi},\Delta\Delta\mathcal{E}^e)\cdot\exp\left[-\phi\Delta\Delta E^e\right]\end{aligned} \tag{5.68}$$

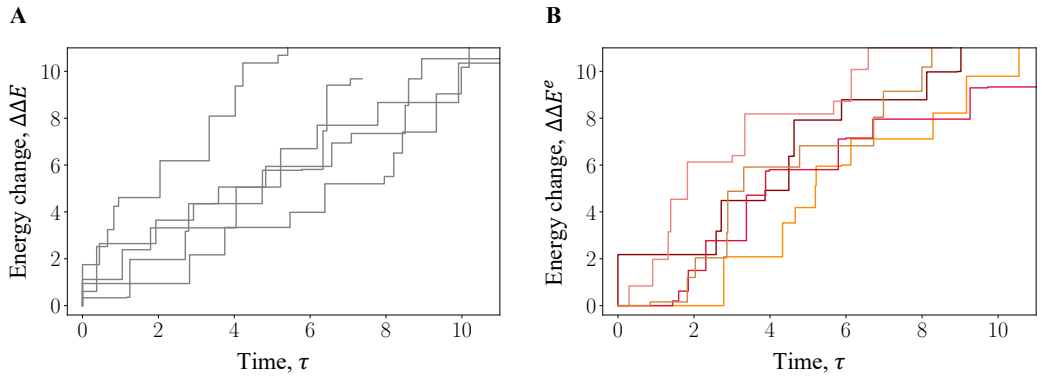


Fig. 5.10. Trajectories of multi-epitope antigen evolution A Five independent trajectories $\Delta\Delta E(\tau)$ for a single epitope system. **B** For a single trajectory $\mathbf{\Delta\Delta E}(\tau)$, the five independent trajectories $\Delta\Delta E^e(\tau)$ for a multi-epitope system.

Antigen evolution rate We understand the antigenic evolution of g epitopes as the independent, parallel, and monotonic increase of all $\Delta\Delta E^e$ values due to the accumulation of mutations in all epitopes. Assuming a constant mutation rate per epitope μ, increasing the number of epitopes also increases the number of mutations that are accumulated in the antigen per unit of time. This means that the antigenic evolution rate of the whole antigen is the same as the rate of a single epitope. However, this is not the case for the expected decrease in potency due to antigenic evolution.

The drop in memory potency from Equation 5.68 is non linear in the quantities $\Delta\Delta E^e$. This means that the total antigenic change of the different epitopes affects the potency drop in a different way as is the case of a single epitope. In this case, the sum is dominated by the epitope with the less antigen evolution (or smaller value of $\Delta\Delta E^e$). This means that the drop in potency requires all epitopes to evolve antigenically.

Let us break up this concept more in detail. Let us consider a vector

$$\mathbf{\Delta\Delta E}(\tau) = \{\Delta\Delta E^1(\tau), \dots \Delta\Delta E^g(\tau)\} \tag{5.69}$$

that has as components the total energy change in each epitope due to antigen mutations up to time τ. We measure the time in units of inverse total mutation rate $(g\mu)^{-1}$. Figure 5.10 shows an example of $\Delta\Delta E^1(\tau)$ trajectories. Figure 5.10A shows different trajectories of a single epitope in the case $g = 1$. Figure 5.10B shows the trajectory of 5 different epitopes in the same antigen. We see that each epitope in a multi-epitope antigen evolves identically as a single epitope. Now we compute the potency drop, $\Delta\hat{\mathcal{Z}}(\tau)$ due to antigenic evolution given by Equation 5.68 and parametrized by τ.

We simulate an ensemble of trajectories $\{\mathbf{\Delta\Delta E}(\tau)\}$ for $g = 5$ and $\Delta\Delta\mathcal{E}^e = 0$ for all e (potency is distributed equally). To reduce the complexity of the problem for the moment, we use $\hat{\phi} = 1$. Figure 5.11A shows the drop of potency as a function of τ. Interestingly, the drop is slower than that for $g = 1$. Antigen evolution is slower for greater number of epitopes [5].

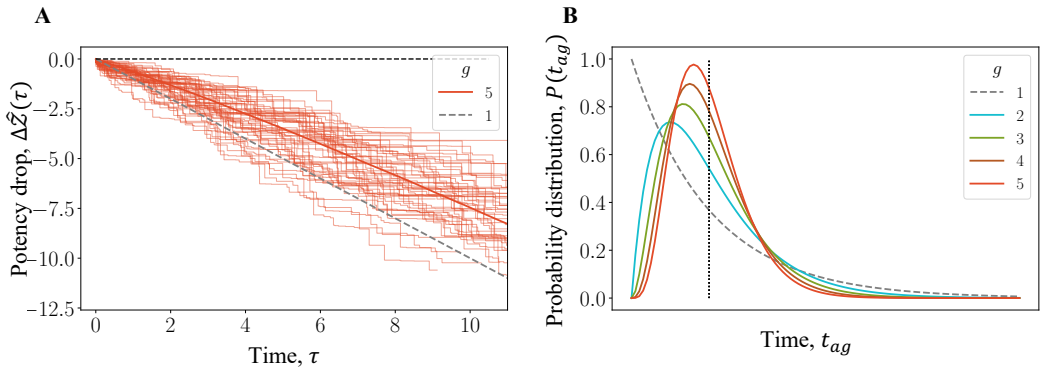


Fig. 5.11. Speed of multi-epitope antigen evolution A For an ensemble of 10^5 trajectories $\{\mathbf{\Delta\Delta E}(\tau)\}$ as in Figure 5.10B, we show the time evolution of $\chi(\tau)$ for $\hat{\phi} = 1$ and $\Delta\Delta\mathcal{E}^e = 0$ for all e. Lines show the expected behavior for $g = 1$ (dashed grey) and $p = 5$ (solid red). **B** Distribution of the time for the first g events in a Poisson process with average μ^{-1}. The distribution for $g = 1$ is Exponential with rate parameter μ and for $g > 1$ is Gamma with shape parameter g and rate parameter $g\mu$. Vertical dashed line shows the shared average value for all distributions.

The explanation for this decrease in speed for an increase in the number of epitopes is the following. For $g = 1$, $\Delta\hat{\mathcal{Z}}$ changes proportional to $\Delta\Delta E^1(\tau)$.

$$\Delta\hat{\mathcal{Z}} = -\Delta\Delta E^1(\tau) \tag{5.70}$$

However, for $g > 1$, it is given by

$$\Delta\hat{\mathcal{Z}} = \log \sum_e \exp\left[-\Delta\Delta E^e(\tau)\right] \tag{5.71}$$

which is dominated by the smallest $\Delta\Delta E^e(\tau)$.

In this case, we can say that, approximately, a drop in potency occurs whenever all g epitopes have acquired at least one mutation. Let us call this time t_{ag}. Mutations in each epitope can be described as a Poisson process with rate μ. Therefore, for the single epitope case, the waiting time for the next mutation, which in this case corresponds to t_{ag}, is exponentially distributed with average μ^{-1} and variance μ^{-2}.

In the multiple-epitope case, t_{ag} is the time at which all g epitopes have had a mutation. In this case, t_{ag} is Gamma distributed with average μ^{-1} and variance $(g\mu)^{-2}$. Figure 5.11B shows the ditribution $P(t_{ag})$ for different values of g.

Interestingly, although all share the same average value for t_{ag}, their variance are different. A single-epitope antigen is more likely to have antigen mutations at times $t_{ag} < \mu^{-1}$. Multiple-epitopes antigens evolve statistically slower than single-epitope antigen.

Viral collapse revisited We can now recalculate the dynamics of viral collapse for a multi-epitope antigen and including antigenic evolution. Using Equation 5.67,

the condition of viral collapse is given by

$$\alpha\hat{\mathcal{Z}}_0 \exp\left[\hat{\lambda}_{\mathcal{Z}}(t^c - \hat{t}^*)\right] \sum_e \chi^e(\hat{\phi}, \Delta\Delta\mathcal{E}^e, \Delta\Delta E^e) = 1. \tag{5.72}$$

From here we get that the critical collapse time is given by

$$t^c = \hat{t}^* - \frac{1}{\lambda_{\mathcal{Z}}} \log\left[\alpha\hat{\mathcal{Z}}_0 \sum_e \chi^e(\hat{\phi}, \Delta\Delta\mathcal{E}^e, \Delta\Delta E^e)\right]. \tag{5.73}$$

and the collapse viral load given by

$$N_A^c = \exp[\lambda_A \hat{t}^*] \left[\alpha\hat{\mathcal{Z}}_0 \sum_e \chi^e(\hat{\phi}, \Delta\Delta\mathcal{E}^e, \Delta\Delta E^e)\right]^{-\lambda_A/\hat{\lambda}_{\mathcal{Z}}} \tag{5.74}$$

This expression is the most complete form for the mean-field critical viral load developed in our mathematical model so far. It incorporates the information of different epitopes in the form of immunodominance energies $\{\Delta\Delta\mathcal{E}^e\}$ and the effect of antigenic viral evolution $\{\Delta\Delta E^e\}$. For future work, we expect to properly understand the fluctuations around the mean-field solutions and used the complete theory to construct a memory B cell *in vivo* protective-function that could inform epidemiological measures and viral surveillance protocols.

Part III

Epidemiology of infectious outbreaks

Chapter 6

Infectious outbreaks: Establishment size and surveillance

The content of this chapter has appeared as,

- Morán-Tovar, R, Gruell, H., Klein, F. & Lässig, M, *Stochasticity of infectious outbreaks and consequences for optimal interventions*. J. Phys. A: Math. Theor. **55**, 384008 (2022). [67].

On this work, I am first author and I have performed the research and analysis under supervision of the last author. All authors formulated the original research problem. The paper is co-written with all co-authors.

During a global pandemic, when strict social distancing and confinement measures are taken, some essential institutions must remain open and functioning. This is the case of hospitals, schools, daycare facilities or food production plants. The prevention of local outbreaks in those institutions becomes very relevant and it is important to formulate appropriate surveillance protocols. In this chapter we show one way to optimize preventive testing protocols using knowledge about the underlying stochasticity of the process.

We developed stochastic epidemiological models for small populations. We decided to work with an extension of the classical SIR model that consider exposed individuals that carry the infections but are not yet able to infect other individuals: The SEIR model. Moreover, we consider populations that interact through a contact network that recreates human interactions. We focus on the initial phase of an infectious outbreak, in which stochasticity is crucial due to the low number of infected individuals. In this case, we analyze the transition between a small outbreak, characterized by few infections and a high probability of extinction of the infections, and a breakthrough epidemic, in which the infection has taken over a large fraction of the population. We study the probability of this transition under different conditions.

In particular, we analyze the probability of a breakthrough epidemic as a

function of the connectivity of the patience zero within the network of contacts. We used this analytical result to inform a random testing protocol. Using numerical simulations of epidemics in scale-free contact networks, we showed that random testing in which the frequency at which individuals are tested is proportional to the connectivity of individuals in the population generates overall earlier detection compared to uniform frequencies.

Part IV

Closing

Chapter 7

Conclusions and Perspectives

In this dissertation we have presented a biophysically based theory of immune activation by dynamic antigens: the recognition of proliferating signals. Our approach, which uses tools from statistical physics, has a novel insight compared to the state of the art. Namely, we have mapped the diversity of BCRs to their function: we based our work on statistical properties of the distribution of binding affinities of the BCR repertoire. This biophysically motivated map allowed us to reduce the complexity of the problem and characterize the BCR repertoire with few macroscopic variables. In particular, we proposed the density of high-affinity BCRs, summarized by the *repertoire exponent* β^*, as a new relevant quantity that characterizes adaptive immune responses. In addition, we propose that the exponent β^*, an emerging quantity of the genotype-phenotype map in the BCR-antigen binding problem, may be under different evolutionary pressures. For example, the exponent influences the clone size statistics of responding B cell clones during immune responses, and therefore it also affects its potency. In this case, the repertoire exponent can be modulated by the binding motifs of the BCR or by the size of the repertoire.

Under this framework, we have shown that the immune response against different immunization protocols resembles a Luria-Delbrück dynamics. In the case of immune responses, we have extended the initial Luria-Delbrück setup to a new class of models that we call *generalized Luria-Delbrück* models. In this case, exponential growth, which is the main driving force in the problem, can be harvested to discriminate between energetic states beyond the limits of thermodynamics equilibrium. We have not yet coined a name for this mechanism and we continue investigating in what other systems is relevant. For the moment, we use the term *exponential proofreading*.

In the case of a primary immune response, we have shown (retrospectively), that exponential proofreading is not sufficient to overcome the large entropy contained in the BCR repertoire. In this context, we proposed a combination with kinetic proofreading as a likely mechanism to generate rapid, potent and highly specific primary immune responses. We found that the strength of proofreading required for an optimal response is determined by the repertoire exponent and, in general, by the repertoire complexity. In the case of memory immune responses,

where reducing kinetic proofreading to a minimum appears to be optimal, we continue to investigate the potential relevance of exponential proofreading as a major source of high specificity in the system. What is the role of these different forms of proofreading in the process of affinity maturation? We expect that this new framework will provide the right tools to address this question as well.

Our framework also has important clinical consequences. First, as a hallmark of generalized Luria-Delbrück processes, we showed that immune responses experience giant fluctuations. For example, we predict the existence of primary elite neutralizers. Jackpot individuals in which a single high-affinity B cell clone dominates the immune response, providing protection orders of magnitude higher than the population average. Second, our model explains the observed age-dependent decline in the potency of primary immune responses as a result of the decrease in the density of high affinity B cell clones. Finally, we are working to construct an *in vivo* immune cross-protective function. Leveraging our understanding of the primary immune response, we aim to understand how memory B cell clones respond rapidly to recurrent infections and are able to neutralize viral proliferation in real time. Combining such a model with antigenic evolution of the virus could provide a first answer to what was the initial question of this PhD project. Why do we get sick again? I believe that this new understanding of the real-time memory protection will be useful to inform public health measures and preventive protocols against seasonal and endemic pathogens, as well as potential new pandemics.

The adaptive immune system in vertebrates has proven to be an excellent system to study general properties of living matter. In recent decades, it has challenged many paradigms in biology and medicine. For example, it has shown that mutation rates can be manipulated by cells in order to rapidly generate diversity in a target gene and selected fitter variants in the course of a few weeks. Another example, perhaps more important for this dissertation, is that the immune system has challenged our assumptions about how much information we can store in the genome. With limited space, the immune system is able to create such a large diversity of proteins, that for a long time it was not easy to believe it was true. In this sense, I have learned from this project that it is important to go back to the origin of the scientific paradigms on which we stand. This process, we would not only provide us with the right context and the best questions for our research. It will also show us the intricacies our nowadays assumptions have gone through. In some cases, it might also show us that many of our current assumptions are still open problems.

I was surprised to learn how strongly the theory of clonal selection, one of the major paradigms in immunology, is supported by experimental evidence. I was also surprised to learn that the theory lacks a complete explanation for one of its main assumptions. Namely, that antigens choose the right immune receptors from a diverse set of pre-existing ones. Little progress has been made in mathematical and mechanistic models of this crucial step of the theory. I hope I did not miss this in the sea of literature.

Finally, I was surprised to learn how similar the debate about acquired immunity is to the debate in evolution about the source of mutations. And it is with

this thought that I would like to close this dissertation. Some time ago, I heard a prominent immunologist say that with the invention of affinity maturation in GCs, evolution has reinvented itself[1]. I would say that evolution has actually done this twice with the adaptive immune system. Before GCs, it also reinvented the neutral evolution of natural variation in expanding populations.

[1]The real quote is ≪[Affinity maturation is] "Nature's most accomplished self-portrait" by Vasco Barreto. It was shown by Gabriel Victora during a presentation as part of the *Delbrück Lecture series on evolution* in April 2022.≫

List of Figures

2.1 Luria-Delbrück fluctuation experiment I 7
2.2 Luria-Delbrück fluctuation experiment II 9
2.3 B cell immune response 18

5.1 Layers of Immune protection 42
5.2 Different levels of disorder in the immune system I 43
5.3 Different levels of disorder in the immune system II 44
5.4 Memory B cell formation 45
5.5 Potency and viral dynamics 49
5.6 Effective temperature of an immune response 53
5.7 Predicting memory response from potency of primary response . . 54
5.8 Fluctuations of potency and critical viral load 55
5.9 Global epistasis in the underlying energy model 57
5.10 Trajectories of multi-epitope antigen evolution 60
5.11 Speed of multi-epitope antigen evolution 61

Bibliography

1. *Phage and the Origins of Molecular Biology: The Centennial Edition* Centennial Edition (eds Cairns, J., Stent, G. S. & Watson, J. D.) ISBN: 9780879698003 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2007).
2. Jenkins, D. J. & Stekel, D. J. De novo evolution of complex, global and hierarchical gene regulatory mechanisms. *Journal of molecular evolution* **71,** 128–140 (2010).
3. Higgs, P. G. & Lehman, N. The RNA World: molecular cooperation at the origins of life. *Nature Reviews Genetics* **16,** 7–17 (2015).
4. Tkachenko, A. V. & Maslov, S. Onset of natural selection in populations of autocatalytic heteropolymers. *The Journal of chemical physics* **149** (2018).
5. Röschinger, T., Morán-Tovar, R., Pompei, S. & Lässig, M. *Adaptive ratchets and the evolution of molecular complexity* 2025. arXiv: `2111.09981 [q-bio.PE]`. `https://arxiv.org/abs/2111.09981`.
6. Fang, X., Kruse, K., Lu, T. & Wang, J. Nonequilibrium physics in biology. *Rev. Mod. Phys.* **91,** 045004 (4 2019).
7. *Nobel Prize in Physiology or Medicine 1969* `https://www.nobelprize.org/prizes/medicine/1969`. Accessed: 2025-03-14.
8. Luria, S. E. & Delbrück, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28,** 491–511. ISSN: 0016-6731 (6 1943).
9. Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* **270,** 313–321 (2003).
10. Glanville, J. *et al.* Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 20216–20221. ISSN: 00278424 (48 2009).
11. Weinstein, J. A., Jiang, N., White, R. A., Fisher, D. S. & Quake, S. R. High-throughput sequencing of the zebrafish antibody repertoire. *Science* **324,** 807–810. ISSN: 00368075 (5928 2009).
12. Briney, B., Inderbitzin, A., Joyce, C. & Burton, D. R. Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature* **566,** 393–397. ISSN: 14764687 (7744 2019).

13. Desponds, J., Mora, T. & Walczak, A. M. Fluctuating fitness shapes the clone-size distribution of immune repertoires. *Proceedings of the National Academy of Sciences of the United States of America* **113,** 274–279. ISSN: 10916490 (2 2016).

14. Desponds, J., Mayer, A., Mora, T. & Walczak, A. M. in *Mathematical, computational and experimental T cell immunology* 203–221 (Springer, 2021).

15. Chardès, V., Vergassola, M., Walczak, A. M. & Mora, T. Affinity maturation for an optimal balance between long-term immune coverage and short-term resource constraints. *Proceedings of the National Academy of Sciences of the United States of America* **119,** 1–10. ISSN: 10916490 (8 2022).

16. Pauling, L. in *Festschrift Arthur Stoll zum siebzigsten Geburtstag* 597–602 (Birkhäuser, 1958).

17. Boeger, H. Kinetic proofreading. *Annual Review of Biochemistry* **91,** 423–447 (2022).

18. Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences of the United States of America* **71,** 4135–4139. ISSN: 00278424 (10 1974).

19. Ninio, J. Kinetic amplification of enzyme discrimination. *Biochimie* **57,** 587–595. ISSN: 61831638 (5 1975).

20. Gromadski, K. B. & Rodnina, M. V. Kinetic Determinants of High-Fidelity tRNA Discrimination on the Ribosome. *Molecular Cell* **13,** 191–200. ISSN: 1097-2765 (2004).

21. Johansson, M., Zhang, J. & Ehrenberg, M. Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection. *Proceedings of the National Academy of Sciences* **109,** 131–136 (2012).

22. Murugan, A., Huse, D. A. & Leibler, S. Speed, dissipation, and error in kinetic proofreading. *Proceedings of the National Academy of Sciences* **109,** 12034–12039 (2012).

23. Mckeithan, T. W. Kinetic proofreading in T-cell receptor signal transduction. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 5042–5046. ISSN: 00278424 (11 1995).

24. François, P., Voisinne, G., Siggia, E. D., Altan-Bonnet, G. & Vergassola, M. Phenotypic model for early T-cell activation displaying sensitivity, specificity, and antagonism. *Proceedings of the National Academy of Sciences of the United States of America* **110,** E888–E897. ISSN: 00278424 (10 2013).

25. Pettmann, J. *et al.* The discriminatory power of the t cell receptor. *eLife* **10,** 1–42. ISSN: 2050084X (2021).

26. Gerhard, W., Yewdell, J., Frankel, M. E. & Webster, R. Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. *Nature* **290,** 713–717 (1981).

27. Frank, S. A. in *Immunology and Evolution of Infectious Disease* Chapter 5 (Princeton University Press, 2002).

28. Morbach, H., Eichhorn, E. M., Liese, J. G. & Girschick, H. J. Reference values for B cell subpopulations from infancy to adulthood. *Clinical and Experimental Immunology* **162,** 271–279. ISSN: 13652249 (2 2010).

29. Elhanati, Y. *et al.* Inferring processes underlying B-cell repertoire diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370.** ISSN: 14712970 (1676 2015).

30. Altan-Bonnet, G., Mora, T. & Walczak, A. M. Quantitative immunology for physicists. *Physics Reports* **849,** 1–83. ISSN: 03701573 (2020).

31. Tas, J. M. *et al.* Visualizing antibody affinity maturation in germinal centers. *Science* **351,** 1048–1054. ISSN: 10959203 (6277 2016).

32. Mesin, L. *et al.* Restricted Clonality and Limited Germinal Center Reentry Characterize Memory B Cell Reactivation by Boosting. *Cell* **180.** Publisher: Elsevier, 92–106.e11. ISSN: 0092-8674, 1097-4172. (2024) (Jan. 2020).

33. Glaros, V. *et al.* Limited access to antigen drives generation of early B cell memory while restraining the plasmablast response. *Immunity* **54.** early memory formation. ISSN: 10974180 (9 2021).

34. Viant, C. *et al.* Antibody Affinity Shapes the Choice between Memory and Germinal Center B Cell Fates. *Cell* **183,** 1298–1311. ISSN: 10974172 (5 2020).

35. Hirst, G. K. The Agglutination of Red Cells by Allantoic Fluid of Chick Embryos Infected with Influenza Virus. *Science* **94,** 22–23 (1941).

36. Engvall, E. & Perlmann, P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* **8,** 871–874 (1971).

37. Heidelberger, M. & Pedersen, K. O. The Molecular Weight Of Antibodies. *Journal of Experimental Medicine* **65,** 393–414. ISSN: 0022-1007 (Mar. 1937).

38. Tiselius, A. & Kabat, E. A. An Electrophoretic Study Of Immune Sera And Purified Antibody Preparations. *Journal of Experimental Medicine* **69,** 119–131. ISSN: 0022-1007 (Jan. 1939).

39. Porter, R. R. The hydrolysis of rabbit γ-globulin and antibodies with crystalline papain. *Biochemical Journal* **73,** 119–127. ISSN: 0006-2936 (Sept. 1959).

40. Edelman, G. M. Dissociation Of γ-Globulin. *Journal of the American Chemical Society* **81,** 3155–3156 (1959).

41. Pauling, L., Campbell, D. H. & Pressman, D. The Nature Of The Forces Between Antigen And Antibody And Of The Precipitation Reaction. *Physiological Reviews* **23,** 203–219 (1943).

42. Breinl, F. & Haurowitz, F. Chemische Untersuchung des Präzipitates aus Hämoglobin und Anti-Hämoglobin-Serum und Bemerkungen über die Natur der Antikörper. *Biological Chemistry* **192,** 45–57 (1930).

43. Hodgkin, P. D., Heath, W. R. & Baxter, A. G. The clonal selection theory: 50 years since the revolution. *Nature immunology* **8,** 1019–1026 (2007).

44. Landsteiner, K. & Chase, M. W. Studies On The Sensitization Of Animals With Simple Chemical Compounds : Ix. Skin Sensitization Induced By Injection Of Conjugates. *Journal of Experimental Medicine* **73,** 431–438. ISSN: 0022-1007 (Mar. 1941).

45. Kuhn, T. S. *The Structure of Scientific Revolutions* 1st. ISBN: 9780226458120 (University of Chicago Press, Chicago, 1962).

46. Ehrlich, P. Croonian lecture.—On immunity with special reference to cell life. *Proceedings of the Royal Society of London* **66,** 424–448 (1900).

47. Jerne, N. K. The Natural-Selection Theory Of Antibody Formation. *Proceedings of the National Academy of Sciences* **41,** 849–857 (1955).

48. Talmage, D. W. Allergy and Immunology. *Annual Review of Medicine* **8,** 239–256. ISSN: 1545-326X (1957).

49. Burnet, F. M. A Modification of Jerne's Theory of Antibody Production using the Concept of Clonal Selection. English. **20,** 67–9. ISSN: 0365-3668 (3 1957).

50. Forsdyke, D. R. The Origins of the Clonal Selection Theory of Immunity as a Case Study for Evaluation in Science. *The FASEB Journal* **9,** 164–166 (1995).

51. Burnet, F. M. *The Clonal Selection Theory of Acquired Immunity* (Cambridge University Press, Cambridge, UK, 1959).

52. Hozumi, N. & Tonegawa, S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proceedings of the National Academy of Sciences* **73,** 3628–3632 (1976).

53. Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* **280,** 288–294 (1979).

54. Bell, G. I. Mathematical Model of Clonal Selection and Antibody Production. *Nature* **228,** 739–744 (1970).

55. Perelson, A. S. & Weisbuch, G. Immunology for physicists. *Reviews of modern physics* **69,** 1219 (1997).

56. Perelson, A. S., Mirmirani, M. & Oster, G. F. Optimal strategies in immunology: I. B-cell differentiation and proliferation. *Journal of mathematical biology* **3,** 325–367 (1976).

57. Perelson, A. S. & DeLisi, C. Receptor clustering on a cell surface. I. Theory of receptor cross-linking by ligands bearing two chemically identical functional groups. *Mathematical Biosciences* **48,** 71–110 (1980).

58. Eisen, H. N. & Siskind, G. W. Variations in affinities of antibodies during the immune response. *Biochemistry* **3,** 996–1008 (1964).

59. Wang, A., Wilson, S., Hopper, J., Fudenberg, H. & Nisonoff, A. Evidence for control of synthesis of the variable regions of the heavy chains of immunoglobulins G and M by the same gene. *Proceedings of the National Academy of Sciences* **66,** 337–343 (1970).

60. Kepler, T. B. & Perelson, A. S. Cyclic re-entry of germinal center B cells and the efficiency of affinity maturation. *Immunology today* **14,** 412–415 (1993).

61. Kepler, T. B. & Perelson, A. S. Somatic hypermutation in B cells: an optimal control treatment. *Journal of theoretical biology* **164,** 37–64 (1993).

62. Morán-Tovar, R. & Lässig, M. Nonequilibrium Antigen Recognition during Infections and Vaccinations. *Phys. Rev. X* **14,** 031026 (3 2024).

63. Lam, J. H. & Baumgarth, N. The Multifaceted B Cell Response to Influenza Virus. *The Journal of Immunology* **202,** 352–359. ISSN: 0022-1767 (2 2019).

64. Hobson, D., Curry, R., Beare, A. & Ward-Gardner, A. The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses. *Epidemiology & Infection* **70,** 767–777 (1972).

65. Khoury, D. S. *et al.* Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nature medicine* **27,** 1205–1211 (2021).

66. Karnaukhov, V. *et al.* Structure-based prediction of T cell receptor recognition of unseen epitopes using TCRen. *Nature Computational Science,* 1–12 (July 2024).

67. Morán-Tovar, R., Gruell, H., Klein, F. & Lässig, M. Stochasticity of infectious outbreaks and consequences for optimal interventions. *Journal of Physics A: Mathematical and Theoretical* **55,** 384008 (2022).