CHAPTER 5

Affinity maturation of the antibody response

Introduction

Antibodies play a key role in protecting us from infectious disease-causing pathogens. They act principally to neutralize free pathogens present in blood or extracellular spaces. If an antibody binds to the proteins on the surface of a pathogen (e.g., spike proteins of a virus), it can prevent the pathogen from binding to host cell receptors, thus preventing infection of new cells. As described in Chapter 2, antibody bound pathogens are neutralized in diverse ways by the innate immune system. Antibodies can also target pathogens in other ways. For example, they can bind to pathogens on host cell surfaces, such as virus particles budding out of an infected cell. Various processes that ensue after antibody binding can then eliminate infected cells.

A remarkable aspect of adaptive immunity is that, upon infection by a pathogen, a Darwinian evolutionary process called affinity maturation (AM) ensues. This process results in the production of memory B cells and antibodies that are specifically tailored to combat the invading pathogen. In this chapter, we will study mechanistic models of how the immune system learns to tailor B cell responses to be specific for a particular pathogen. We will first briefly consider the situation where AM occurs in response to a single antigen. The antigen could be a pathogen, some form of a pathogen used in a vaccine, or any foreign substance. In such cases, AM results in the production of antibodies that bind avidly to one specific antigen. As we saw in Chapter 4, mutable pathogens, such as HIV, SARS-CoV-2 and influenza, can mutate to evade antibody responses that target a particular strain. The majority of this chapter will focus on studying how the B cell and antibody response evolves when driven by multiple variants of an antigen. A mechanistic understanding of this process can aid the design of vaccination protocols that aim to direct AM to produce antibodies that can effectively target diverse strains of mutable pathogens. This problem raises many interesting issues that lie at the intersection of non-equilibrium statistical mechanics, evolutionary biology, immunology, and learning theory.

5.1: Affinity maturation in response to a single antigen

Before studying this chapter, readers are encouraged to reacquaint themselves with the biology of AM described in Chapter 2. To briefly recapitulate, germline B cells whose

receptors can bind sufficiently strongly to an antigen can get activated and seed germinal centers (GC) in lymph nodes. Activated GC B cells multiply, and an enzyme called AID that is expressed in these cells acts to introduce mutations into their BCRs at a high rate. These processes occur in the so-called Dark Zone (DZ) of the GC. The B cells then migrate to another region of the GC, called the Light Zone (LZ), and interact with the antigen displayed on the surface of Follicular Dendritic Cells (FDCs). B cells expressing a BCR that bind sufficiently strongly to the antigen can potentially internalize the antigen. The antigenic proteins are chopped into peptides and some of these peptides are displayed on the surface of B cells in complex with MHC (HLA)class II molecules. The B cells that present antigenic peptide-MHC molecules on their surface then compete to interact with T helper cells that were activated by interactions with the same pMHC complexes displayed on other infected APCs. A productive interaction results in a survival signal. A fraction of the positively selected B cells differentiate into memory cells and antibody secreting plasma cells, and exit the GC. The majority of positively selected B cells is recycled for further rounds of mutation and selection.

Mechanistic modeling of affinity maturation played a major role in discovering that most positively selected B cells are recycled in the GC after each round of mutation and selection. Experiments with animal models showed that, as affinity maturation proceeds, the antigen affinity of the generated antibodies can increase by up to a thousand to ten thousand-fold. Affinity increases were observed to occur in stages, accompanied by the accumulation of mutations over time. In 1997, Perelson and co-workers constructed a mathematical model of the GC process, and the results suggested that these observations could only be explained if there were multiple rounds of mutation and selection, not just one. Let us briefly discuss some aspects of this influential study.

In Perelson's model, after about 6 days, proliferation-dependent mutation due to AID was turned on. The mutation probability is estimated to be of the order of 10^{-3} per base pair per replication cycle, and the size of the genes encoding the variable regions of the heavy and light chains of the BCR range from roughly 400 - 600 base pairs. So, mutations occur with a probability of roughly 0.1 per BCR genome. About a third of the mutations are thought to be lethal for reasons that include the inability of the BCR to fold properly. Mutations can also change the affinity of the BCR for the antigen or have no effect on affinity (silent mutations). For affinity changing mutations, deleterious mutations are more likely than beneficial ones because there are only a few BCR sequences that bind well to any given antigen, but there are many that do not. In Perelson's model, B cells could have one of six discrete affinities for the antigen. Two affinity classes (labeled, -1 and -2) had worse antigen affinity compared to germline B cells (class 0) that seeded the GC (with -2 being worse than -1); three affinity classes (labeled, 1, 2, and 3) had

increasingly improved affinities compared to the germline B cells. Mutations that were not silent could change the affinity by one affinity class.

Perelson and co-workers constructed a mean-field model to describe processes that occur in the GC. The abundance of each species in the GC was described by an average concentration, and ordinary differential equations described the dynamics of GC processes. GC B cells multiplied at a certain rate in the DZ. Each type of mutation noted above was described by a probability of occurrence per unit time (or rate). After replication and mutation, GC B cells migrate from the DZ to the LZ, and this process was explicitly modeled. In the LZ, B cells were considered to die rapidly unless they were positively selected (since GC B cells are apoptotic – see Chapter 2). The FDCs had "sites" that contained antigen. The number of sites declined with time because antigen decays and is consumed by B cells, and this was modeled as an exponential decay with a fixed characteristic time. The binding of each B cell to a FDC site was described by a rate of binding and a rate of dissociation. The difference in affinity between two adjacent affinity classes was modeled as a five-fold difference in the rate of binding, with the dissociation rate staying unchanged upon an affinity changing mutation. If the B cell could bind to the FDC site, it could be positively selected. The details of the selection process were not considered, and it was assumed that a fraction of the B cells that could bind to the FDC sites was positively selected. The others could dissociate from the FDC site and become LZ B cells that either died or bound to sites on FDCs again. A fraction of the positively selected B cells differentiated into memory cells and exited the GC, and the rest were recycled to the DZ for further rounds of mutation and selection.

Using values of parameters that were known from experiments or estimated, the differential equations describing the processes noted above were solved numerically. An important prediction of the model was that the extent of experimentally observed increases in antibody affinities during affinity maturation could only be realized if a significant fraction of B cells that were positively selected was recycled for further rounds of mutation and selection (Fig. 5.1). Of course, this also meant that the affinity of the produced antibodies increased with time in steps (or rounds of mutation and selection), as observed in experiments. Definitive proof of the prediction of recycling made by Perelson and co-workers had to wait for more than a dozen years until analyses of results from multiphoton microscopy experiments provided vivid images of the dynamic trajectories of GC B cells, and recycling was observed.

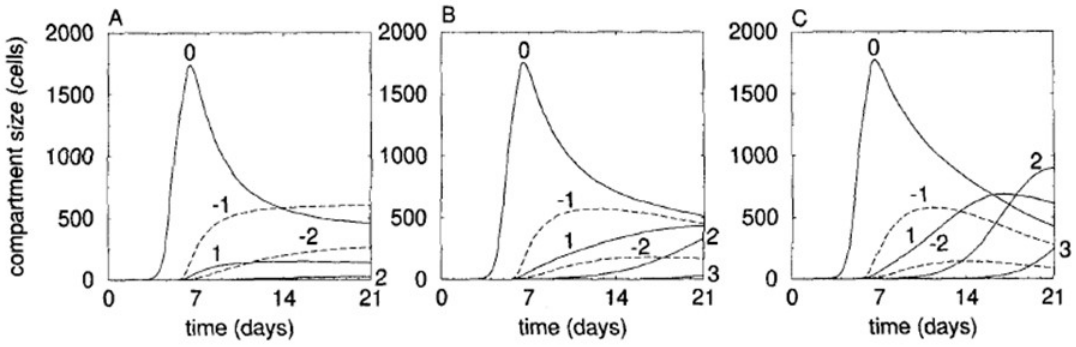


FIG. 5.1: The evolution of B cells in different affinity classes predicted by Perelson's model of GC processes. Affinity class 0 corresponds to germline B cells, cells in classes -1 and -2 have increasingly worse affinity compared to germline B cells, and those in classes 1, 2, and 3 have increasingly better affinities. Panels A, B, and C correspond to results for the fraction of recycled positively selected B cells equal to 0.1, 0.5, and 0.9. High affinity B cells only emerge if the fraction recycled is high.

The principal reason for why recycling is required to achieve the affinity increases observed in experiments is that one round of mutation is unlikely to produce B cells with the large changes in affinity observed in experiments. This conclusion can be reached by making a rough estimate. Let there be M B cells in the DZ that divide N times (and potentially mutate) before migrating to the LZ for selection. M is of the order of a 1000 and N is of the order of 2-4 (i.e., O (1)). Let us estimate the average number of these B cells that can proliferate and mutate in one round in the DZ to produce B cells with an affinity change of ten thousand-fold (i.e., a change in binding free energy equal to about 7 – 8 Kcal/mol at physiological temperature).

The probability, P (x), of observing a mutational trajectory of a B cell that results in a change in free energy of binding equal to x in one cycle of N divisions can be estimated as follows:

$$P\,(x) = \sum_{\{\overrightarrow{X_i}\}}[\prod_{i=1}^{N} \mu * \,(1 - \,p_l)\,p\,(x_i)]\,]\,\delta\,(x - \,\sum_{i=1}^{N} x_i) \qquad (1)$$

where μ is the probability that a mutation occurs in the BCR gene, p_l is the probability that a mutation is lethal, p (x_i) is the probability that in the i^{th} step a change of binding free energy equal to x_i occurred, δ denotes Dirac's delta function, and $\{\overrightarrow{X_i}\}$ denotes summing

over various mutational trajectories with different combinations of values of x_i that sum up to x. In writing Eq. 1, we have ignored mutational trajectories that result in changes in binding free energy greater than x, which can be corrected by replacing the delta function in Eq. 1 with a step function. As we shall see below, the probability of occurrence of trajectories that result in a value of x corresponding to 7 – 8 Kcal/mol is very low. Therefore, this approximation is not concerning.

Given that the probability that there is a mutation in the BCR is roughly 0.1 and p_I is 0.3, the average number of positive outcomes (affinity change equal to x), O (x), is:

$$O\ (x) = \ (0.07)^N\ M\ \sum_{\{\overrightarrow{x_i}\}}[\prod_{i=1}^{N} p\ (x_i)]\ ]\ \delta\ (x - \ \sum_{i=1}^{N} x_i) \qquad (2)$$

The PINT database provides estimates for the probability distribution of the change in binding free energy of a protein-protein interface upon making a single point mutation. This data has been compiled for diverse protein-protein interfaces and is described by a log normal distribution (for x_i), with deleterious mutations being more likely than favorable ones. The probability distribution of binding free energy changes for antibody-antigen interfaces may be different. Nonetheless, using this data, one can estimate p (x_i). By discretizing the changes in x_i that can occur, the value of the quantity in Eq. 2 can be calculated numerically. Let us estimate the value by making a very rough approximation – viz., replacing the log normal distribution for x_i with an exponential distribution. Substituting an exponential distribution for p (x_i) in Eq. 2 obtains:

$$O\ (x) = \ (0.07)^N\ M\ C_1^N\ e^{-cx}\ \sum_{\{\overrightarrow{x_i}\}} 1\ = \ (0.07)^N\ M\ C_1^{N-1}\ p\ (x)\ \alpha \qquad (3)$$

where α is the number of combinations of sequential values of x_i that sum to x, c characterizes the decay of the exponential probability distribution for x_i and C_1 normalizes this probability distribution. As M is of O (1000) and C_1 is of O (1), if we take N to be equal to 4, O (x) is of the order of $2.4 * \ 10^{-2}\ \alpha\ p\ (x)$. The PINT database estimates that p (x), the probability of a random mutation resulting in a beneficial change of 7-8 Kcal/mol, is vanishingly small. For N being about 4, the quantity, α, is not large enough to compensate for the vanishingly small value of $2.4 * \ 10^{-2}$ p (x) because p (x) is tiny. We would reach the same conclusion for an affinity change of a thousand-fold or 2 divisions per cycle because the probability of beneficial mutations that lead to large affinity changes in just a few division/mutation cycles is vanishingly small. Thus, the average number of outcomes that lead to B cells with the large change in affinity observed during affinity maturation in a single round of mutation in the DZ is much less than unity. This extremely low chance of occurrence, the fact that multiple mutations are observed in affinity matured antibodies (compared to germline naïve B cells), and that antibody affinity and

the number of acquired mutations increase with time make it evident that multiple rounds of mutation and selection likely occur during affinity maturation.

Following the seminal work by Perelson and co-workers described above, numerous physics-based models of AM have been studied. These models describe the dynamics of GC processes by employing a variety of methods that include mean-field differential equations, stochastic agent-based computer simulations that reflect the Master equations describing the dynamics of GC processes, and Fokker-Planck equations (or birth-death equations). Until recently, most such studies primarily focused on AM driven by single model antigens, which had also been the focus of most past experimental studies. Rather than study the details of these descriptions of AM induced by a single antigen, let us turn attention to a more complex situation involving AM that is pertinent to a practical problem – viz., developing vaccines that can elicit potent antibodies that protect against infection by highly mutable pathogens. In discussing this problem, we will also describe a few different approaches to studying the pertinent stochastic processes, all of which are also applicable for studying AM initiated by a single antigen.

5.2: The evolution of antibodies that can target diverse mutant strains of pathogens

In Chapter 5, we discussed highly mutable pathogens, with special emphasis on the example of HIV and how knowledge of the mutational fitness landscape of the virus could be obtained. We also discussed how, based on the fitness landscape, immunogens could be designed that may elicit vaccine-induced T cell responses that target the virus' mutational vulnerabilities. Most effective prophylactic vaccines stimulate a strong antibody response, as antibodies can prevent infection of host cells, while T cells kill infected cells. To protect against infection by highly mutable pathogens, a vaccine needs to stimulate the immune system to produce antibodies that can neutralize diverse mutant strains that may infect a person. Such broadly neutralizing antibodies (or bnAbs) against HIV and influenza have been isolated from some patients.

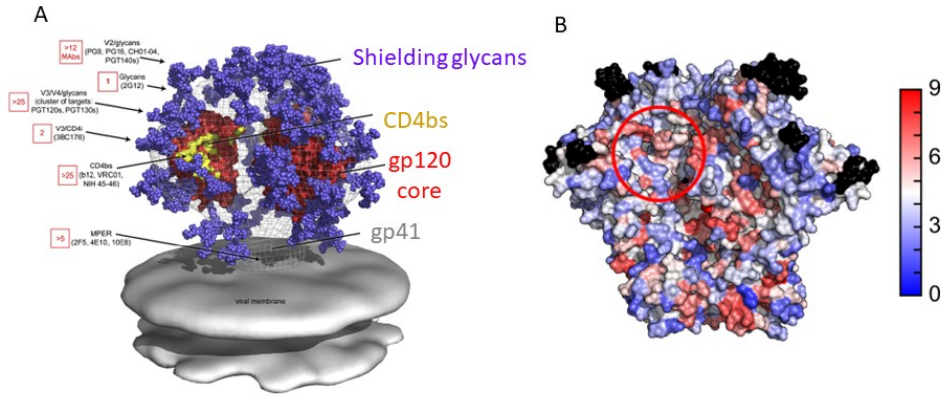
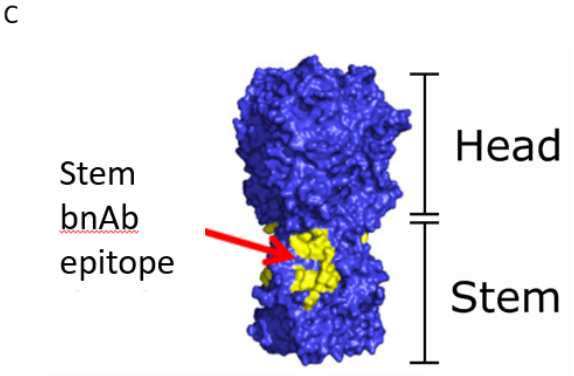


FIG. 5.2: (A) Schematic depiction of the viral spike of HIV. The yellow region corresponds to the relatively conserved CD4 binding site that is targeted by some bnAbs that have been isolated from patients. Some other relatively conserved epitopes that are targeted by known bnAbs are also indicated. The red region is highly variable, and the blue residues represent glycans that can shield the spike protein residues from antibodies. (B) The fitness cost of evolving mutations at a particular residue averaged over for all possible amino acids at that residue in all possible sequence backgrounds (see Chapter 4) is shown superimposed on a representation of the HIV spike trimer. The region surrounding the CD4 binding site is circled in red. Bluer colors indicate lower fitness costs associated with mutations.



(C) Schematic representation of the HA spike on the surface of the influenza virus. The conserved stem epitope is shown in yellow. The conserved residues of the RBS are not shown.

Fig. 5.2 A is a schematic depiction of the spike on the surface of HIV, which is a non-covalently associated trimer composed of two proteins (gp120 and gp41). The viral spike is an important target of antibodies upon natural infection or vaccination. The spike proteins of HIV are highly mutable, thus providing a moving target for the immune system upon natural infection with HIV. However, the HIV spike contains some relatively

conserved sites, and bnAbs bind to such regions. One example is the CD4 binding site, which is relatively conserved because it has to bind to the CD4 co-receptor on human cells in order to propagate infection. However, these relatively conserved receptor binding site (RBS) residues are surrounded by highly variable residues depicted in red in Fig. 5.2 A. Fig. 5.2 B shows this variability of the region surrounding the conserved portion of the CD4 binding site. The fitness cost of evolving mutations at different sites, averaged across all possible sequence backgrounds, is shown superimposed on a representation of the HIV spike trimer. The fitness cost was obtained using the fitness landscape of ENV proteins using methods described in Chapter 4. The typical footprint of an antibody is larger than the size of just the relatively conserved residues of the CD4 binding site. Furthermore, as shown in Fig. 5.2 A, the CD4 binding site is also shielded by glycans (sugars). A number of bnAbs have been isolated from HIV-infected patients that bind to the CD4 binding site. Other relatively conserved regions on the spike that are targeted by bnAbs are also depicted in Fig. 5.2 A. Importantly, upon natural infection, bnAbs are induced in only some people, usually several years after infection, and in relatively low titers (numbers). Most of the antibodies that are produced during natural infection are directed toward other variable epitopes on the HIV spike proteins; i.e., the immunodominant antibody responses are not bnAbs.

The viral spike of influenza is also comprised of a trimer of the hemagglutinin (HA) protein (Fig. 5.2 C). HA binds to sialic acid on human cells to propagate infection. Neuraminidase (NA) is another molecule expressed on the surface of the influenza virus, and it plays roles in virus motion through the respiratory tract and in propagating infection. Both HA and NA are targets of antibodies, but most protective antibodies target HA. A domain of HA, called HA1, comprises the globular “head” of the spike, which is highly variable. The HA2 domain and the termini of the HA1 domain comprise the “stem” of the spike. When HA binds to its target on human cells, conformational changes occur that are predicated on a conserved region on the stem (shown in yellow in Fig. 5.2 C). BnAbs that target this region have been isolated from humans; for example, during the 2009 swine flu epidemic. But the immunodominant antibody response upon natural infection or seasonal vaccination is directed toward the highly variable head epitopes. The density of spikes on the influenza virus surface is very high, thus making it difficult for B cells and antibodies to access the conserved stem epitope. This geometric constraint is likely an important factor contributing to the rare emergence of bnAbs. The sialic acid binding site (RBS) on the head of the HA spike also contains relatively conserved residues, but this region is surrounded by highly variable regions (like HIV’s CD4 binding site).

While bnAbs are subdominant antibody responses upon natural infection by HIV or influenza, the emergence of bnAbs in some infected individuals shows that the human

immune system, through AM, is capable of learning how to evolve bnAbs. This raises the tantalizing possibility that properly designed vaccination strategies could change the immunodominance hierarchy and elicit bnAbs efficiently in the human population. Thus, there are extensive efforts being directed toward developing immunogens and vaccination protocols that can elicit bnAbs targeting HIV and influenza in diverse patients. Success would result in an effective HIV vaccine or a universal influenza vaccine that could protect against seasonal variants and pandemic strains. Moreover, the approaches developed in this regard could help design similar vaccines against other pathogens, such as the coronavirus family of viruses.

Vaccination with a single variant of the viral spike will likely lead to the evolution of strain-specific antibodies because the AM process would be subjected to a selection force that favors the evolution of Abs that bind to the immunodominant epitopes of this particular variant of the viral spike. One would likely have to vaccinate with multiple variants of the molecules comprising the viral spike to induce bnAbs. These variants would share a set of conserved sites (e.g., the RBS), but would have different variable regions. This idea is consistent with data on the temporal evolution of virus sequences in some HIV-infected patients who evolve bnAbs. The emergence of bnAbs is often preceded by a diversification of the viral swarm in an individual. Similarly, depending upon the history of past exposures to influenza strains, some individuals mounted a stem-specific response to the 2009 pandemic strain. This could be because past exposures to different strains led to low levels of antibodies directed against the conserved stem region. When exposed to a completely new strain that shared the conserved stem with past influenza strains but not the previously immunodominant regions, the stem-specific memory B cells expanded. Taken together, these observations suggest that the types of variant antigens to which an individual has been exposed in the past (history of selection forces that the B cell population has been subjected to) influences bnAb evolution.

If multiple variant antigens are used in the vaccine, several practical questions need to be addressed. These include: What should be the variant antigens? How many variant antigens should be used, and how different should their variable parts be from each other? In what temporal order should they be administered (e.g., as a cocktail, sequentially, or permutations thereof)? What should be the concentrations of the variant antigens? Despite significant efforts and advances aimed toward answering these questions, effective universal vaccines that can protect against HIV and influenza infection do not yet exist. An important reason is that the answers to the questions noted above are drawn from a large space of possibilities, which makes it difficult to intuitively design experiments that can search a promising space for answers. A deep mechanistic understanding of how AM is driven by multiple variant antigens could guide the choice of

answers and the design of effective immunogens and vaccination protocols that can change the natural immunodominance hierarchy and elicit bnAbs. Furthermore, such studies would provide fundamental insights into how the immune system works.

The fundamental challenge associated with changing the natural immunodominance hierarchy by vaccination can be stated as follows. AM is a stochastic dynamic process that involves mutation and selection and is driven far from equilibrium. Can we understand how this process can be guided by the choice of selection forces (vaccination strategies) to efficiently evolve bnAbs or any desired sub-dominant response? That is, how can the selection forces be chosen to make normally low probability dynamic trajectories evolve with high probability. This is a problem at the intersection of non-equilibrium statistical physics, evolutionary biology, immunology and learning theory. We will study simplified models that provide some insights into this general problem. Some aspects of these models are chosen motivated by issues pertinent to the evolution of bnAbs that could target the HIV CD4 binding site or the RBS of HA.

We begin by considering a highly simplified model of AM driven by multiple variant antigens. The insights that will emerge will then be tested against computer simulations of more complex models and experimental results.

5.2.1: A simple model of affinity maturation upon immunization with multiple variant antigens

The viral spike of HIV or influenza contains far fewer BCR epitopes that contain conserved sites than those that do not. So, unless the frequency of naïve (or germline) B cells that target the conserved regions is reasonably high or their affinities for these regions is very high, vaccination with a form of the spike as the immunogen would likely lead to GCs seeded overwhelmingly by germline B cells whose receptors bind to epitopes that do not contain any conserved sites. The precursor frequency of germline B cells that can target the RBS on influenza's HA spike is reasonable, and so may not get outcompeted if properly designed immunogens can be developed for a vaccine. But, in other circumstances, such as the CD4 RBS of the HIV spike and the conserved stem region of the influenza spike, only a rather restricted set of germline B cells can evolve into bnAbs. So, for example, as a first step in efforts to elicit bnAbs against HIV's RBS, immunization with a simpler antigen (not the whole spike) has been shown to be effective in selectively activating and expanding germline B cells with receptors that can bind to epitopes containing the pertinent conserved residues. These activated germline B cells and corresponding antibodies are not bnAbs, but could potentially evolve to become bnAbs. Our focus will be on immunization with variant antigens that mimic the entire viral spike after such

germline B cells have been activated. That is, we will assume that only these B cells seed GCs upon immunization with spike-like antigens. In order to develop basic concepts and intuition, we will, for now, ignore the fact that upon immunization with the spike proteins germline B cells that target other regions could also seed GCs. At the end of this chapter, we will comment on such issues and their importance.

The breadth of coverage that an antibody provides to a panel of variant antigens is defined by the fraction of these antigens that it can neutralize. We will use the affinity (free energy of binding) of the antibody for an antigen as a proxy for neutralization ability; i.e., neutralization corresponds to the affinity exceeding a threshold. The affinity of a BCR or antibody for an antigen is determined by its amino acid sequence and the 3-dimensional conformations of the regions of the antibody and antigen that interact with each other. Thus, the affinity depends on many variables coupled together in complicated ways. Since the affinity of an antibody depends on many variables, its breadth of coverage for a panel of variant antigens also depends upon these variables. In the spirit of Occam's razor, let us assume that the breadth of an antibody (or parent BCR) is characterized by a single dimension. This dimension may be thought of as a projection of the high dimensional space spanned by the many variables noted above on to one that defines an appropriate collective property that determines breadth. We will later consider more detailed models where affinity, or breadth, is defined by multiple variables.

Each point on the breadth dimension described above defines a breadth state – i.e., the fraction of variant antigens that a BCR corresponding to that state can bind to with an affinity exceeding a threshold. Because the pool of B cells that seed GCs are not bnAbs, we expect the probability distribution describing their breadth states to be as shown in Fig. 5.3 A (black line).

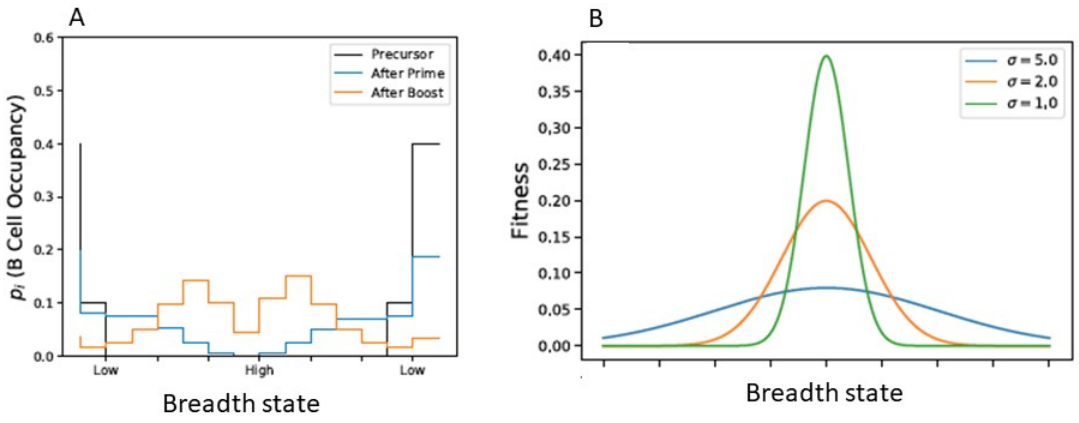


Fig. 5.3: A one-dimensional projection of BCR/antibody sequence defined in a higher-dimensional space is assumed to define the breadth of a B cell. The state at the center of the axis representing breadth corresponds to the highest breadth, and edges correspond to the lowest. (A) Schematic depiction of the probability distribution of B cells in different breadth states: B cells that seed GCs (black line), after a first immunization or prime (blue line), and after a second immunization or boost (orange line). (B) Schematic depictions of the probability of B cells of different breadth being positively selected per unit time (or "fitness"). The probability distributions depicted correspond to Gaussian distributions with different values of the variance (σ)

The immunogens used in a vaccine impose selection forces on this population of B cells as it evolves during AM. We need to characterize how the selection force depends upon the choice of a particular variant antigen administered as the immunogen. A B cell that binds more strongly to an antigen will likely internalize more antigen and display more pMHC molecules on its surface, and thus has a higher probability of being positively selected compared to its peers. Since the immunogen in a shot of the vaccine is different from the one that activated the germline B cells, GC B cells with a higher breadth will have a higher probability of internalizing more antigen, and thus being positively selected. So, we can represent the selection force imposed on B cells by a shot of the vaccine by a breadth-dependent probability of being positively selected per unit time (Fig. 5.3 B). This probability distribution will be peaked at the highest breadth state if the vaccine immunogen is sufficiently different from the previously administered immunogen. It will be more sharply peaked around the highest breadth state if the difference between the first administered immunogen and the one that activated the right germline B cells is greater. Biochemically, this is because, if the difference is larger, the positive selection of B cells that exhibit stronger interactions with the shared conserved sites will be more strongly favored. Stronger binding to the conserved sites is likely to result in higher breadth. For brevity, we will refer to the probability of a B cell being positively selected per unit time (Fig. 5.3 B) as it's "fitness".

If a B cell mutates during AM, it's breadth state will change. Mutations occur via discrete changes to the codons that represent BCR amino acids. Furthermore, it is known that bnAbs usually acquire many mutations compared to their ancestor germline B cells. These considerations suggest that the axis describing states of breadth of B cells can be discretized. Mutations can result in transitions between these states, but can also result in non-functional BCRs (e.g., misfolded ones). To represent these lethal mutations, as well as the fact that GC B cells are inherently apoptotic and die if not positively selected, mutations can also lead to transitions to a "dead" state of B cells. Let us discretize the breadth space into K states (Fig. 5.4). States 1 through K-1 represent different breadth states, with the $K/2^{th}$ state corresponding to the highest breadth (a bnAb) and states 1 and K-1 correspond to the lowest breadths. State 0 corresponds to dead B cells.

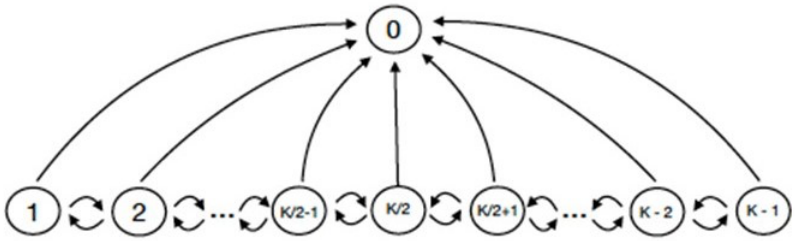


Fig. 5.4: Discrete representation of breadth states. The $K/2^{th}$ bin (middle bin) is the highest breadth state, and bin 1 and bin K-1 represent the lowest breadth states. Bin 0 corresponds to dead B cells (absorbing state). B cells can transition between breadth states by mutation, and transitions to the dead state occur either due to lethal mutations or a basal death rate.

Once an immunogen is administered, the existing B cell population will seed GCs and evolve by mutation and selection to learn and adapt to the new environment. If adaptation is perfect, the probability distribution of the evolved B cells would mirror the fitness distribution. Cast in the language of information theory, perfect adaptation implies that the B cell population has acquired the maximum amount of information available about the environment. This maximal amount of information can be quantified as the Kullback-Leibler divergence (KLD) between the probability distribution of the breadth states characterizing the initial GC B cell population (p) and the breadth-dependent probability distribution for a B cell to be positively selected per unit time (the fitness function, f) imposed by administering the immunogen. The KLD, D_{KL} (p||f), is a quantitative measure of the selection force imposed by the immunogen and corresponds to the distance between the two probability distributions. It is defined as follows:

$$\boldsymbol{D_{KL}}\ (\boldsymbol{p}||\boldsymbol{f}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{K-1}} \boldsymbol{p_i} \log\left(\frac{\boldsymbol{p_i}}{\boldsymbol{f_i}}\right) \qquad (4)$$

The KLD can also be interpreted as a thermodynamic force that acts on the initial GC B cell population upon immunization with a particular variant antigen. This force results in a flux of B cells through breadth space due to the non-equilibrium process of AM, potentially leading to adaptation.

There are three possible outcomes of AM: 1] If all the antigen is consumed by B cells that internalize antigen, AM will come to an end. B cells that internalize antigen and are then positively selected by T helper cells multiply. Therefore, a proxy for antigen being consumed is that the GC B cell population reaches a maximal size. 2] All the B cells die, and the population becomes extinct. 3] Antigen also decays because of various degradation processes, thus ending AM. For this simple model, we will consider only the first two possibilities as conditions that bring AM to an end.

If the population of B cells does not go extinct during AM induced by the first administered immunogens (termed the prime), then a pool of memory B cells is created. The memory B cells are actually produced after every round of mutation and selection during AM, but in the spirit of Occam's razor, let us assume that the final population of B cells in the GC is the memory pool. Now, if we immunize with another shot of the vaccine containing new variant immunogens (termed a boost), both these memory cells and B cells from the naïve population can potentially seed the new GCs that can form. Here we assume that only the memory cells seed new GCs in response to the boost. The immunogen (s) therein imposes a selection force on these B cells, which can again be represented as a breadth-dependent probability of being positively selected per unit time (f^2). A larger difference between the immunogens in the boost and prime would result in f^2 being more sharply peaked around the K/2th bin. The thermodynamic force acting on the population of existing memory B cells that results in AM is again given by the KLD. In general, we can define the KLD for sequential rounds of immunization as follows:

$$\boldsymbol{D_{KL}}\ \left(\boldsymbol{p^l}||\boldsymbol{f^{l+1}}\right) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{K-1}} \boldsymbol{p_i^l} \log\left(\frac{\boldsymbol{p_i^l}}{\boldsymbol{f_i^{l+1}}}\right) \qquad (5)$$

where $\boldsymbol{p^l}$is the probability distribution characterizing the memory B cell population after the $\boldsymbol{l^{th}}$ immunization and $\boldsymbol{f^{l+1}}$ is the fitness distribution imposed on these B cells by the immunogens in the subsequent immunization. The logistics associated with vaccinating a population at a reasonable cost imply that only a small number of immunizations is practical. Here, we will consider either two shots of different immunogens administered sequentially or a single shot containing a cocktail of multiple variant antigens.

If the vaccine is comprised of a single shot of a cocktail of multiple variant antigens, then the AM process will occur only once. For AM induced by a cocktail of variant antigens, we have to consider whether the variant antigens in the cocktail are displayed homogeneously on FDCs or whether the distribution is heterogeneous. If homogeneously distributed, every time a B cell interacts with the FDC, it encounters multiple types of antigens. If the distribution is highly heterogeneous, a B cell would interact with only one type of antigen. Experimental data does not inform us about which of these situations is true. The answer likely depends upon the concentration of antigens displayed on the FDC surface, with higher concentrations resulting in more homogeneous distributions. Let us consider the case where the distribution is heterogeneous first. Suppose that the variable regions surrounding the shared conserved region in these antigens are separated from each other by large mutational distances. In successive rounds of mutation and selection, the B cells will likely interact with different variant antigens. Since the variable regions are very different, after a few cycles of mutation and selection, only the ones with high breadth are likely to have survived. So, for immunization with such a cocktail, if we coarse grain time in our simple model such that each time step corresponds to AM dynamics over a few rounds of mutation and selection, only the B cells with high breadth will have a significant probability of being positively selected per unit time. This can be represented by a fitness distribution that is sharply peaked around the $K/2^{th}$ bin. If the variable regions of the antigens in the cocktail are more similar, this probability distribution will be less sharply peaked. If the variant antigens are distributed homogeneously on the FDCs, then since a B cell can see the same antigen in successive rounds, the probability of being positively selected per unit time for B cells of low breadth will be quite high. So, the fitness distribution will not be sharply peaked. These arguments lead to the conclusion that, for immunization with a cocktail, we can still use the KLD between the probability distribution of being positively selected and the initial distribution of breadths of the GC B cell population as a measure of the thermodynamic force that drives AM. We will study the outcome of AM induced by a cocktail of immunogens or sequentially administered variant immunogens as the character of the immunogens are varied; i.e, as the KLD defined earlier is varied.

We now need a description of how the probability of observing B cells in different breadth states changes with time as AM ensues. This can be accomplished using Master equations that account for the probabilities of occurrence of various processes by which B cells can be added or removed from each breadth state, i, and the state 0. The equation that describes the temporal evolution of the probability that there are n B cells in any breadth bin, i, $\boldsymbol{p_i}\ (\boldsymbol{n}, \boldsymbol{t})$, is:

$$\frac{dp_i(n,t)}{dt} = -p_i(n,t)\left[n\left(f_i + \sum_{i\neq j}\mu_{ij}\right) + \sum_{j\neq i}\mu_{ji}\sum_m m\,p_j(m,t)\right] + \sum_{i\neq j}\mu_{ij}\,(n+1)\,p_i(n+1,t) + p_i(n-1,t)\left[(n-1)f_i + \sum_{j\neq i}\mu_{ji}\sum_m m\,p_j(m,t)\right] \quad (6)$$

The terms proportional to $p_i(n,t)$ describe positive selection of B cells resulting in replication, the mutation of B cells in bin i to another bin j (at rate μ_{ij}), and mutations of B cells from any other bin, j, to the breadth state, i (at rate μ_{ji}). These stochastic processes reduce the probability of there being n cells in the bin corresponding to breadth state, i. The term proportional to $p_i(n+1,t)$ describes mutations away from bin i to any other bin j when there are (n+1) B cells in bin, i. This process adds to the probability of there being n B cells in bin, i. The terms proportional to $p_i(n-1,t)$ describe positive selection and replication of B cells in bin, i, and mutations of a B cell in any bin, j, to a breadth state corresponding to bin, i. These stochastic processes increase $p_i(n,t)$ because a B cell is added to bin, i, when it has (n – 1) B cells. The combined effects of lethal mutation and basal death rate of GC B cells is denoted by a rate of transition to the state 0, μ_{i0}. Transitions from state 0 to other states are disallowed.

Eq. 6 is not analytically solvable. One way to solve it numerically is to obtain the corresponding mean-field equations that describe the temporal evolution of the average number of B cells in each bin, i, and then express the processes they describe in terms of a set of "chemical reactions". Standard Monte-Carlo approaches, such as the Gillespie algorithm, can then be used to obtain stochastic evolutionary trajectories as these "chemical reactions" occur. To obtain the mean-field equations corresponding to Eq. 6, multiply both sides of the equation by n and sum over n. This obtains:

$$\frac{d<N_i>}{dt} = -\left(f_i + \sum_{i\neq j}\mu_{ij}\right)<N_i^2> - <N_i>\sum_{j\neq i}\mu_{ji}\sum_m m\,p_j(m,t) + <N_i^2>\sum_{i\neq j}\mu_{ij} - <N_i>\sum_{i\neq j}\mu_{ij} + f_i<N_i^2> + f_i<N_i> + N_i\sum_{j\neq i}\mu_{ji}\sum_m m\,p_j(m,t) + \sum_{j\neq i}\mu_{ji}\sum_m m\,p_j(m,t) \quad (7)$$

To obtain the equation above, one must observe that n (n + 1) = $[(n + 1)^2 – (n + 1)]$, and that n (n -1) = $[(n-1)^2 + (n – 1)]$. Cancelling terms proportional to $< N_i >$ and $< N_i^2 >$

with opposite signs in Eq. 7, and noting that $\sum_m \boldsymbol{m}\, \boldsymbol{p_j}\,(\boldsymbol{m}, \boldsymbol{t})$ = <N_j>, we obtain the following mean-field equations:

$$\frac{d<N_i>}{dt} = \left(f_i - \sum_{i \neq j} \mu_{ij}\right) < N_i > + \sum_{j \neq i} \mu_{ji} < N_j > \qquad (8)$$

Eq. 8 can be interpreted as follows. B cells in any bin, i, are positively selected per unit time according to the fitness corresponding to that bin and its occupancy. B cells in bin, i, are depleted according to the rate of mutation to other bins and the occupancy of the bin, and B cells are added to bin, i, by mutations from other bins, j. If i = 0, the first term in Eq. 8 is zero. The "chemical reactions" that correspond to the processes described by Eq. 8 are shown in Table 5.1.

For the edge states $i = 1$ and $i = 15$,

$$\mathrm{N}_1 \xrightarrow{f_1} 2\,\mathrm{N}_1$$
$$\mathrm{N}_1 \xrightarrow{\mu_{1,2}} \mathrm{N}_2$$
$$\mathrm{N}_1 \xrightarrow{\mu_{1,0}} \mathrm{N}_0$$

$$\mathrm{N}_{15} \xrightarrow{f_{15}} 2\,\mathrm{N}_{15}$$
$$\mathrm{N}_{15} \xrightarrow{\mu_{15,14}} \mathrm{N}_{14}$$
$$\mathrm{N}_{15} \xrightarrow{\mu_{15,0}} \mathrm{N}_0$$

For all states between $i = 1$ and $i = 15$,

$$\mathrm{N}_i \xrightarrow{f_i} 2\,\mathrm{N}_i$$
$$\mathrm{N}_i \xrightarrow{\mu_{i,i-1}} \mathrm{N}_{i-1}$$
$$\mathrm{N}_i \xrightarrow{\mu_{i,i+1}} \mathrm{N}_{i+1}$$
$$\mathrm{N}_i \xrightarrow{\mu_{i,0}} \mathrm{N}_0$$

Table 5.1: List of reactions corresponding to Eq. 8 for stochastic simulation.

Given a set of parameters, we can now use the Gillespie algorithm to study the effects of different immunization protocols by varying $D_{KL}\ \left(p_i^l || f_i^{l+1}\right)$. It is not possible to choose the precise values of parameters in our simple model to equal experimentally measured numbers because the model is highly coarse-grained. However, the choice of parameters can be guided by consistency with known phenomena. The expectation that the population of B cells should not be extinguished if the imposed probability of being positively selected per unit time (fitness) is distributed uniformly allows us to set bounds

on μ_{i0}. As there are K-1 discrete bins characterizing breadth space and the fitness distribution is normalized as it corresponds to a probability, μ_{i0} must be less than 1/(K-1); otherwise, the basal death rate, rather than the fitness landscape, would principally dominate the outcome of GC processes. Experimental data show that bnAbs usually evolve several mutations compared to the germline B cells. Thus, in our coarse-grained representation the mutation rates, μ_{ij}, must be relatively low as each mutation correspond to a change in breadth, which likely requires multiple actual BCR mutations. For the same reason, mutations between distal bins are unlikely, and we will only allow mutations between adjacent breadth states (bins). The probability of making breadth enhancing mutations should be lower than that for evolving mutations that reduce breadth. This is because there are fewer BCR sequences that can bind with high affinity to multiple variant strains compared to those that bind to just a few variant strains. For the results we will discuss, the values of the parameters are: K = 16; μ_{i0} = 0.02; μ = 0.05; if i < K/2 $\mu_{i\ i+1}$ = 0.125 μ, $\mu_{i+1\ i}$ = 0.875 μ; if i > K/2, $\mu_{i\ i+1}$ = 0.875 μ, $\mu_{i+1\ i}$ = 0.125 μ. For mutations from bin, K/2, $\mu_{i\ i+1}$ and $\mu_{i\ i-1}$ = 0.5 μ.

A population of 50 B cells is chosen from the distribution, p_i^l, a fitness landscape corresponding to a value of $D_{KL}\ \left(p_i^l||f_i^{l+1}\right)$ is chosen and the stochastic version of Eq. 8 is then simulated 100 times until one of the stop conditions is reached. Each calculation corresponds to a particular stochastic realization of a GC process under the imposed conditions. For sequential immunization, 50 new B cells are chosen from the population of B cells in each GC that is not extinguished, and then the calculation is repeated with a new $D_{KL}\ \left(p_i^l||f_i^{l+1}\right)$ corresponding to the next immunization. The total number of bnAbs (B cells in bin K/2) in the GCs that survive after completion of the two immunizations divided by the number of simulated GCs is a measure of the numbers of resulting bnAbs. This number can be regarded to be the number of bnAbs produced on average in a vaccinated person. This is because, in any person, a number of GCs form in the lymph node upon vaccination or natural infection.

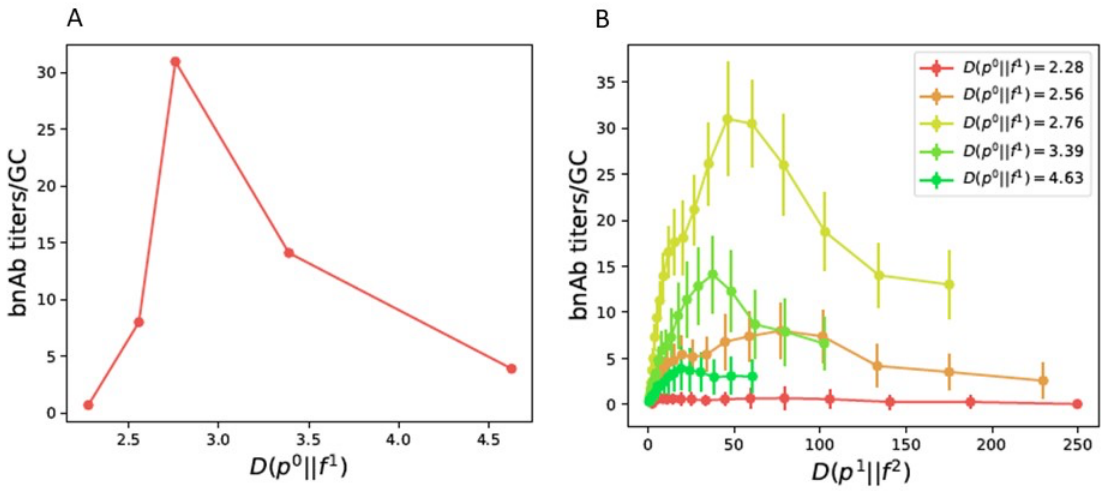


Fig. 5.5: (A) The number (proxy for titers) of bnAbs/GC after the first immunization is shown as a function of the KLD (D ($p^0||f^1$)) between the initial B cell population and the fitness distribution imposed by the immunogens in the first immunization. (B) For each value of D ($p^0||f^1$), the number (proxy for titers) of bnAbs/GC is shown as a function of the KLD (D ($p^1||f^2$)) between the B cell distribution produced after the first immunization and the fitness distribution imposed by the immunogens in the second immunization. The titers are actually related to the binding association constant of antibodies to the viral spike epitope multiplied by their number.

Fig. 5.5 A shows that the number of bnAbs per GC is maximized for a particular value of D ($p^0||f^1$) corresponding to the prime immunization. Fig. 5.5 B shows that, for each value of D ($p^0||f^1$), there is a value of D ($p^1||f^2$) for the boost immunization that maximizes bnAb production. Taken together, these results show that there is an optimal prime-boost protocol that maximizes bnAb production. Let us parse the mechanistic origins of this result.

If the value of D ($p^0||f^1$) is not too large, the B cell population begins to learn about its environment by trying to adapt to the imposed selection force. Thus, after a certain time, trajectories collected from each GC show that the KLD between the evolving B cell population and f^1 decreases with time (Fig. 5.6 A). When the B cell population reaches its maximal size, the GC reaction comes to an end and adaptation stops at a finite value of this KLD. A KLD value of zero would correspond to perfect adaptation. Perfect adaptation is not possible for small values of D ($p^0||f^1$). There is a flux toward B cells in low breadth states because mutations are more likely to reduce the breadth of B cells, but when D ($p^0||f^1$) is small, the resulting low breadth B cells still can be positively selected (Fig. 5.3 B, curve with highest variance). So, the GCs quickly fill up with B cells, and AM ends before there is enough time for B cells to acquire the mutations required to increase their breadth. Taken together, these effects result is a state with the B cell population largely occupying states of low breadth (Fig. 5.7).

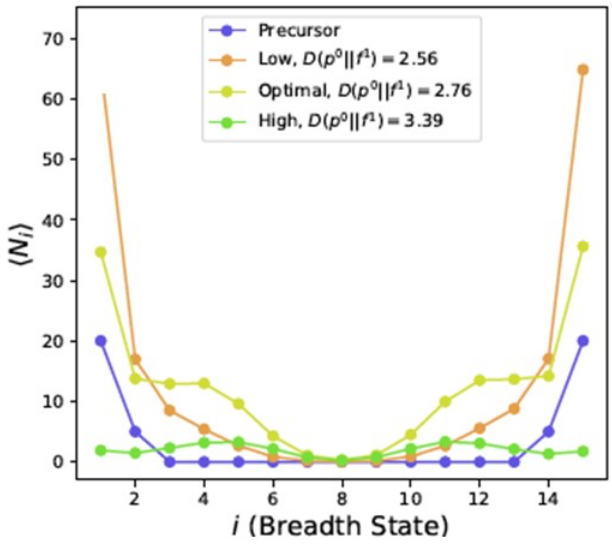


Fig. 5.7: The average number of B cells, $\langle N_i \rangle$, in each breadth state, i, of the precursor B cells that seed GCs, and B cells obtained after the first immunization for different values of the KLD, D ($p^0||f^1$). For low values of the KLD, after immunization, most B cells occupy states of low breadth, and for large values of the KLD, most GC B cells die and the B cell population is very small. For the optimal value of the KLD, after the first immunization, B cells occupy states with relatively high breadth with reasonable probability.

Large values of D ($p^0||f^1$) corresponds to a large thermodynamic driving force. Typically, this should lead to fast adaptation. Fig. 5.6 B shows a few trajectories of the change in KLD with time when D ($p^0||f^1$) is large, and indeed adaptation is faster than when the imposed selective pressure to evolve breadth is small (compare Figs.5.6 A and 5.6 B). However, the trajectories shown in Fig. 5.6 B are very rare. When D ($p^0||f^1$) is large, the

B cell population in most GCs goes extinct. This is because f^1 is sharply peaked (Fig. 5.3 B, curve with the smallest variance). So, the flux of B cells to low breadth states (because mutations are more likely to reduce breadth) results in B cells that have a very low probability of being positively selected. Thus, with high probability, the B cell population quickly becomes extinct. So, on average, the B cell population is very small after the prime (Fig. 5.7). In the few evolutionary trajectories where extinction does not occur, B cells rapidly acquire mutations that enhance breadth substantially. The mutated B cells then multiply as the probability of being positively selected is high for B cells that acquire high breadth. Thus, the rare evolutionary trajectories that can avoid extinction rapidly adapt to become bnAbs and the KLD declines substantially with time quickly (Fig. 5.6 B).

For the optimal value of D ($p^0||f^1$), the probability of being positively selected at the edge states of lowest breadth is slightly below the basal probability of transitioning to the dead state, bin 0 (Fig. 5.8 A). But the B cells in higher breadth states have a reasonable probability of being positively selected. Some of these B cells can mutate to even higher breadth states as they replicate. This allows the flux of B cells toward higher breadth states and the flux toward lower breadth states due to biased mutational probabilities to be better balanced than when D ($p^0||f^1$) is too high. So, the B cell population does not go extinct with high probability. However, the replication probability of low breadth B cells does not compare as well with high breadth B cells as when D ($p^0||f^1$) is too low. So, the B cells do not quickly multiply and fill up the GCs before B cells have time to acquire several breadth-enhancing mutations. For optimal D ($p^0||f^1$), AM continues for a relatively long time, which allows B cells to acquire the mutations necessary to enhance breadth. Adaptation is slower than if D ($p^0||f^1$) is higher than the optimum, but B cells in a far larger fraction of GCs can adapt to the new environment. Thus, after the prime, a significant fraction of the B cell population occupies states adjacent to the bnAb state (Fig. 5.7). Note that there is a very sharp transition in the fraction of GCs that go extinct (Fig. 5.8 B) if one exceeds the optimal value of D ($p^0||f^1$). This is because in this circumstance bins in states with breadth higher than the edge states also have a probability of being positively selected that is lower than the probability of transitioning to state 0 (Fig. 5.8 A).

Fig. 5.9 summarizes the change in the KLD from D ($p^0||f^1$) after prime for conditions corresponding to low, optimal, and high values of D ($p^0||f^1$). For values of D ($p^0||f^1$) that exceed the optimal value, except for the rare trajectories that adapt considerably, in most of the GCs, the B cell population becomes extinct as signaled by an increase in KLD after prime. The KLD increases because the B cell population loses information about the environment upon extinction. For low values of D ($p^0||f^1$), the decrease in KLD, or extent of adaptation, is less than for the optimal D ($p^0||f^1$). For the optimal value of D ($p^0||f^1$), the KLD decreases significantly as the B cell population adapts. A small fraction of GCs become extinct under optimal conditions, as evidenced by a small mode of the ΔD distribution with an increase in KLD.

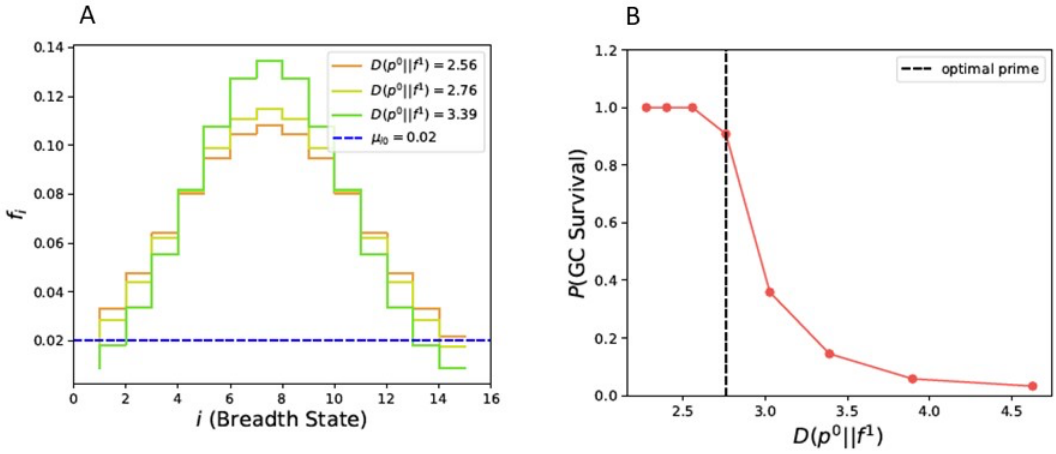


Fig. 5.8: (A) The probability of being positively selected per unit time (f_i) corresponding to various values of D ($p^0||f^1$). The curve for the optimal value of D ($p^0||f^1$) = 2.76 goes below the basal death rate at the edge states. If D ($p^0||f^1$) is higher (e.g., 3.39), the value of f_i goes below the basal death rate for states with higher breadth. (B)The fraction of GCs that do not go extinct, P (GC Survival) is shown as a function of D ($p^0||f^1$). The value of D ($p^0||f^1$) corresponding to the optimal prime is shown as the vertical dashed line.

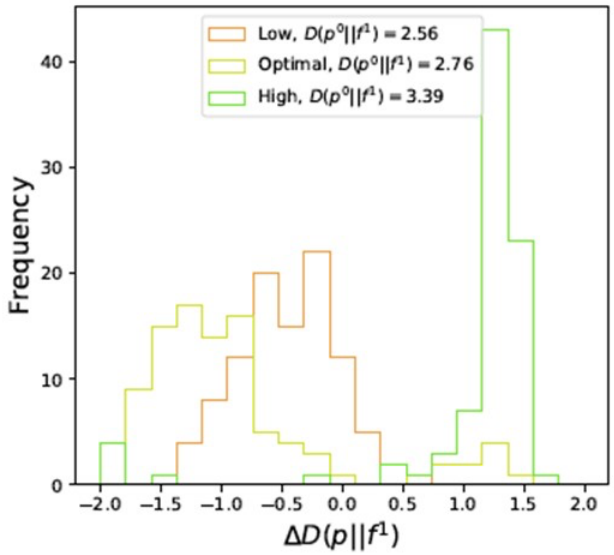


Fig. 5.9: For low (brown), optimal (yellow), and high (green) values of D ($p^0||f^1$), the change in D after GC reactions, $\Delta D\ (p||f^1)$, is recorded for each simulated GC and the results are shown as a histogram. More negative values of $\Delta D\ (p||f^1)$ correspond to better adaptation of the B cell population to the immunogen, which is the desired response to vaccination.

For each value of D ($p^0||f^1$), we find that there is an optimal value of D ($p^1||f^2$) that maximizes bnAb production (Fig. 5.5 B). The determinants of the optimal value of D ($p^1||f^2$) are the same as for D ($p^0||f^1$). For distributions of f^2 corresponding to lower than optimal D ($p^1||f^2$), B cells that are relatively far from the bnAb state have a probability of being positively selected that is larger than the basal death rate. They can replicate and the GC processes end before the mutations necessary to become bnAbs can occur with a sufficiently high probability. For distributions of f^2 corresponding to higher than optimal D ($p^1||f^2$), only B cells that are very close to the bnAb state have a significant probability of being positively selected. As the occupancy of pre-existing B cells in these states is low, a small number of bnAbs evolve. For the distribution of f^2 corresponding to the optimal D ($p^1||f^2$), B cells further away from the bnAb state have a significant chance of being positively selected. But the probability of being positively selected for B cells in these states is lower than for distributions of f^2 corresponding to sub-optimal values of D ($p^1||f^2$). So, for the optimal D ($p^1||f^2$), these B cells can replicate and acquire mutations that allow them to evolve to the bnAb state, while at the same time not replicating too quickly and filling up the GC with B cells that are not bnAbs.

The optimal prime-boost combination maximizes bnAb production because the optimal value of D ($p^0||f^1$) produces the right diversity of the B cell population prior to the boost (p^1) . Fig. 5.10 shows the number of trajectories originating from different breadth states, i, that evolve into bnAbs for low, optimal, and high values of D ($p^0||f^1$); for each case, D ($p^1||f^2$) equals the corresponding optimal value. When D ($p^0||f^1$) is lower than the optimal value, the probability of being positively selected during the boost has to be sharply peaked to promote bnAb evolution. This is because most B cells occupy low breadth states after the prime and evolving the many mutations needed by repeated replication from these states is not likely. Only the few B cells that evolved relatively high breadth during the prime have a chance of evolving into bnAbs, and so the number of bnAbs after the boost is low. If D ($p^0||f^1$) is larger than the optimal, the optimal value of D ($p^1||f^2$) can be smaller than for sub-optimal values of D ($p^0||f^1$). This provides B cells in states adjacent to the bnAb state to also have some probability of evolving into bnAbs. But since only a few B cells survive after a supra-optimal prime (Fig. 5.7), the number of bnAbs produced is small. The optimal prime-boost combination enables the evolution of bnAbs from a wider range of B cell states following prime than for sub-optimal or supra-optimal values of D ($p^0||f^1$). That is, the optimal prime generates the right diversity of B cells, which can subsequently be boosted to generate many bnAb producing evolutionary trajectories.

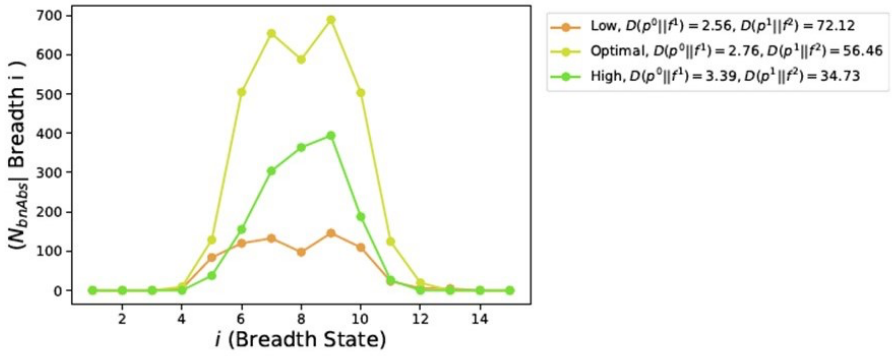


Fig. 5.10: The number of evolutionary trajectories that result in bnAbs originating from B cells in different breadth states, i, (($N_{bnAbs}|$ Breadth i), after prime. Examples of low and high values of D ($p^0||f^1$), and the optimal value are shown. For each value of D ($p^0||f^1$), the boost immunogen corresponds to the optimal value of D ($p^1||f^2$).

Upon natural infection, bnAbs that target diverse strains of highly mutable pathogens do not evolve with high probability. The results of our simple model lead us to the conclusion that a prime-boost sequential immunization strategy can potentially be optimized by choosing the right variant antigens and other conditions such that rare evolutionary trajectories that lead to bnAbs become more likely. That is, the variant antigens employed

for sequential immunization can be chosen such that the immunodominance hierarchy is modulated to maximize bnAb production for reasons elaborated above.

The results we have described above for the prime apply for immunization with cocktails comprised of immunogens with varying extents of differences between their variable regions. As we have noted earlier, if the B cells interact with only one type of variant antigen in the cocktail during each round of selection, greater differences between the variable regions of the variant antigens will result in more sharply peaked distributions of the probability of getting positively selected per unit time (larger values of KLD). Our simple model predicts that there is an optimal cocktail of antigens separated by a specific average mutational distance corresponding to the optimal value of KLD (Fig. 5.5A). Finally, we note that, if B cells can see all types of variant antigens in a cocktail during each round of selection, the KLD is small, and bnAbs are unlikely to evolve. Biologically, this is because, if a B cell can interact with one of the variants in every round of mutation and selection, there is no driving force to evolve breadth. A diverse set of B cells each with a different specificity is likely to evolve.

In the next sections, we will explore whether computer simulations with more complex models support these findings, and what experiments have to say.

5.2.2: Computer simulations of more complex models

The affinity, and therefore breadth, of a BCR/antibody for a panel of variant antigens is defined by multiple features of its sequence and structure. It is not clear how to determine an appropriate one-dimensional projection that defines breadth. Therefore, the first step toward testing the veracity of the findings described in the previous section is to examine whether a higher-dimensional model for BCR-antigen affinity yields similar results. In this section and the next, we will consider two such models. It is important to note that these models are also highly simplified, and many details of processes that occur in the GC are ignored. Hopefully, these simple models still provide insights into the general mechanistic principles that we seek.

Simulations with a "string model"

In one model, both the sites of the antigen's epitope and the BCR sites that interact with them (called paratope) are described as strings of amino acids (Fig. 5.11).

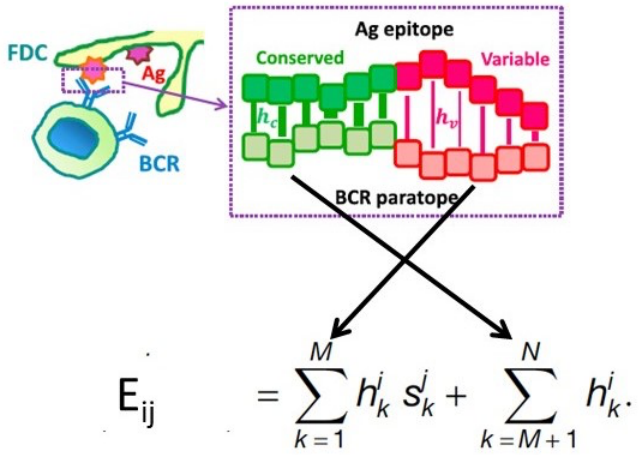


Fig. 5.11: A schematic depiction of the string model described in the text in detail. The interaction free energy is that described in Eq. 9.

The CD4 binding site epitope on the HIV spike or the receptor binding site of the influenza virus is comprised of some relatively conserved sites and variable ones that surround it. A simple caricature of such an epitope is to model the antigen as a string of N "spins" characterized by a variable s_k for each site, k, which can take values equal to ± 1. There are M variable sites with s_k = -1 corresponding to a "mutated" variable site, and the rest of the are the conserved sites (s_k is always equal to +1). We can write the interaction free energy, E_{ij}, between strings of amino acids representing a BCR paratope, i, and an antigen's epitope, j, as:

$$E_{ij} = \sum_{k=1}^{M} h_k^i s_k^j + \sum_{k=M+1}^{N} h_k^i \qquad (9)$$

The set, $\{h_k^i\}$, is a representation of the interaction characteristics of BCR amino acids at the sites that comprise its paratope. For example, if hydrophobicity was the only metric of the strength of interaction of an amino acid with other amino acids, h_k^i, would be a measure of the hydrophobicity of the amino acid at residue, k, of the BCR, i. We will use a convention wherein larger values of E correspond to higher affinity. As a concrete example, we will study variant antigens with 28 variable sites (M = 28) and 18 conserved sites (N = 46 in Eq. 9).

The values of h_k^i can change as the BCR sequence mutates during AM. One could choose the values of h_k^i to also be ± 1, but then the interaction free energy would vary rather abruptly upon BCR mutation. As large beneficial single point mutations are rare (see section 5.1), we choose the values of h_k^i from a continuous distribution. Unlike TCR-

peptide interactions that we studied in Chapter 3, BCR (or antibody) interactions with antigenic epitopes depends on the 3-dimensional conformations of the interacting pair. A string model is clearly not a good representation of how conformations of molecules might influence interaction free energies, and Eq. 9 will be embellished below to take some conformational effects into account.

We can simulate the stochastic dynamics of AM for the model described above on a computer, by numerically solving the Master equations corresponding to B cell dynamics during AM. One way to do this is to develop a stochastic simulation method that executes a set of rules that describe AM. The following rules are a simple representation of the key aspects of AM.

The GC reaction is seeded with a few B cells that bind to the immunogen with a free energy that exceeds a threshold value (E_a). The value of E_a was chosen to be $9k_BT$ for the results we will discuss, but its specific value does not matter for this simple model as it just serves as a reference or scale for all other values of free energies. As before, we will study AM following immunization with complex antigens after a simpler antigen has activated the germline B cells that can bind to epitopes that contain the conserved sites. So for the B cells that seed a GC, the values of the fields, h_k^i, corresponding to interactions with the conserved sites of the antigen's epitope are drawn from a distribution that reflects favorable binding to them; viz., a uniform distribution that spans positive values (between 0.3 and 0.6). The values h_k^i for the BCR amino acids that interact with the variable sites are drawn from a broader uniform distribution, ranging from -0.18 to 0.9. The specific ranges of these values are chosen for numerical convenience.

The GC is seeded with 10 germline B cells, which expand for a few divisions before replication-dependent mutations due to the action of AID turns on. Seeding GCs with larger numbers of B cells does not affect the qualitative results that we discuss. Consistent with experimental data, B cells divide twice per GC cycle. Mutations occur at all sites of the BCR with a uniform probability chosen in accord with experiments (~ 0.1 per BCR sequence per division). This is an approximation as AID acts more effectively on certain sequence motifs, and so the mutation probabilities are not uniform across the BCR sites. It is estimated that the probability of occurrence of lethal mutations is 0.5 (the B cell is removed from the simulation), silent mutations is 0.3 (no change in the values of h_k^i upon mutation), and affinity affecting mutations is 0.2.

Affinity affecting mutations are more likely to decrease, rather than increase, affinity. The reason is similar to that which we noted in the preceding sub-section for mutations that affect breadth. A catalog of measurements of changes in the the affinity of protein-protein interfaces due to mutations is available (the PINT database), and this data can be

fit to a lognormal distribution with advantageous mutations being less likely. In the simulation results we discuss, for a B cell with affinity affecting mutations, the change in its binding free energy to the antigen's epitope (ΔE) is chosen from the following distribution:

$$\Delta E = \varepsilon - \exp(\mu + \sigma r) \qquad (10)$$

In Eq. 10, r is a normal distribution with mean zero and standard deviation equal to one. The values of the parameters, ε, μ, and σ, are chosen such that the tail of the log normal distribution fits the experimental data reported in the PINT database, and only 5 % of the mutations lead to an increase in ΔE. The specific values of the parameters thus obtained are: μ = 1.9, σ = 0.5, and ε = 3 $k_B T$. The values of ΔE are also bounded to lie within a range ($\pm$ 1 $k_B T$) for reasons noted below.

Once a value of ΔE is chosen, the value of h_k^i corresponding to a randomly picked BCR site, k, is changed to $h_k^i + \Delta h_k^i$, with $\Delta E = \Delta h_k^i \, s_k^j$. Because mutations can change the range of BCR affinities beyond that characterizing the B cells that seed a GC, the allowed range of values of h_k^i is expanded to span – 1.0 to +1.5. The specific ranges of these values are chosen for numerical convenience. For example, if the range is too wide, the evolution of a single mutation with a large value of h can dominate the results. If residue k on the BCR that was randomly chosen to undergo an affinity change contacts a conserved site, no other affinity changes occur.

Changes in the variable regions surrounding the conserved sites of the CD4 binding site on the HIV spike can lead to loops, which serve as steric barriers for a BCR (or antibody) to bind to the conserved sites. BCR mutations that can avoid interactions with such loops would be able to access and bind to the conserved sites more easily. Inspired by this example, let us assume that if a mutation occurs in a BCR site, x, that reduces its interaction with a variable site on the antigen (smaller value of h_x), then the value of h_y for a randomly chosen BCR site, y, that interacts with a conserved site is increased. This mimics the effect of enhanced access of the BCR to the conserved sites if the BCR is better able to avoid interactions with loops that present a steric barrier. Conversely, if a mutation occurs in a BCR site that increases its interaction with a variable site on the antigen, then the value of h_y corresponding to a randomly chosen BCR site that interacts with a conserved antigen site is decreased. The change in interaction free energy with the conserved residue is taken to be -α ΔE, where α is a factor that reflects the degree of conformational steric screening and ΔE is the change in interaction free energy with the variable site due to BCR mutation (Eq. 10). In the results we describe here, α was chosen to be 0.25.

B cells must internalize antigen in order to be positively selected. When the B cell membrane imposes a pulling force, if the antigen-FDC bond breaks rather than the BCR-antigen bond, antigen is internalized (see Chapter 2). This process can be explicitly modeled to calculate the number of antigens internalized. In a coarse-grained treatment, we can model affinity-dependent antigen internalization in a simpler way by assuming that the probability with which antigen is internalized grows with the affinity of the BCR for the antigen and then saturates as per a Langmuir-like function:

$$P_j^i = \frac{c_j\, e^{\rho(E_{ij}-E_a)}}{1+ c_j\, e^{\rho(E_{ij}-E_a)}} \qquad (11)$$

where P_j^i is the probability that B cell i is successful in internalizing antigen, j, c_j is the antigen concentration on the FDC, E_{ij} is the interaction free energy for B cell, i, interacting with antigen, j (Eq. 9), and ρ is a parameter with units of $(k_BT)^{-1}$ whose meaning is discussed below. There is experimental evidence that GC B cells in the LZ can make a few attempts, separated by refractory periods, to internalize antigen. If the B cell is not positively selected in a few attempts, apoptosis occurs. The few attempts to be positively selected have been coarse-grained into one set of probabilities in Eq. 11.

B cells that succeed in internalizing antigen display antigen-derived pMHC molecules on their surface. These B cells compete with each other to interact with helper T cells, and receive a survival signal if a productive interaction occurs. This requires that a B cell encounters a T helper cell whose TCR is specific for a peptide-MHC molecule displayed on its surface. Depending upon the question one seeks to answer, these and other complex effects may need to be accounted for at different levels of detail. For our search for qualitative mechanistic insights here, let us employ the simplest model. The B cells that successfully internalize antigen are ordered according to their affinities for the antigen, and B cells in the top fraction, F_s, are positively selected. This simple model is justified by noting that B cells with BCR that have a higher affinity for the antigen are more likely to internalize more antigen, and so display more peptide-MHC molecules on their surface and will be more likely to interact productively with T helper cells.

As discussed earlier, a fraction of the B cells that are positively selected in the LZ (F_r) are recycled for further rounds of mutation and selection. The rest exit the GC as antibody producing plasma cells and memory cells. GC processes end when one of three conditions discussed earlier is satisfied: 1] all GC B cells die; 2] a threshold number of B cells is reached (chosen to be ~ 5000 B cells); 3] a maximum time has elapsed (chosen to be 200 GC cycles), which reflects antigen decay from FDC surfaces.

Computer simulations of the agent-based model described above provide stochastic trajectories, and the B cells remaining in the GC in the end are collected. For each immunization protocol studied, the typical properties of these B cells are obtained by simulating many GCs (1000 for the results we describe below) and averaging the value of a property across the B cells in all these GCs. As discussed in section 6.2.1, such averages can be considered representative of the outcome for a typical person. The principal property of interest to us is whether bnAbs evolve, and in what numbers. When experimentalists measure the breadth of an antibody, they determine the ability of an antibody to prevent (neutralize) infection of target cells by a virus, *in vitro*. Different variant strains of the virus are used, and the breadth of an antibody is the fraction of these strains it can neutralize. The relationship between the ability of an antibody to "neutralize" a virus and its affinity for an antigenic epitope is complex. Our simulations only allow us to compute affinities. So, we calculate the fraction of virus variants to which an antibody can bind with an affinity above a threshold and use this quantity as a proxy for breadth.

The affinity of the B cells that remain in the GC at the end of a simulation to a set of variant antigens different from the ones used in the immunogens is calculated using Eq. 9. The variant antigens share the conserved sites, but the identities of their variable sites are chosen to be ± 1 with equal probability. The threshold affinity that serves as a proxy for neutralization was chosen to be 12 k_BT. Given the other parameters, choosing this value of the threshold led to B cells with low breadth upon immunization with a single antigen. This is consistent with the observation that upon natural infection bnAbs do not evolve quickly. Using 100 or 1000 variant antigens produced the same results for breadth. We can calculate the average number of B cells in a GC (across 1000 simulations) with a breadth of coverage above 0.8 and use this quantity as a proxy for bnAb titers/GC. Using values different from 0.8 do not alter qualitative results if the number is sufficiently high.

In the simulation results discussed below, the values of F_s, F_r, ρ, and δ were chosen to be 0.7, 0.7, 0.08 $(k_BT)^{-1}$, and 1, respectively, to ensure that the simulation results are consistent with observations of GC dynamics upon administering a single antigen (without prior germline targeting). Specifically, these parameter values made the simulation results consistent with the following results: 1] About 10 mutations resulting in a roughly 1000-fold increase in affinity occurs after AM is complete; 2] After the initial period of multiplication, the GC population declines, then plateaus, and then rises again relatively rapidly.

We are now ready to discuss the results of the computer simulations and compare them to the simple one-dimensional model considered earlier. Let us first consider what happens upon immunization with variant antigens that differ from the germline targeting antigen by different mutational distances. Fig. 5.12 shows that bnAb titers are found to be maximal for an intermediate mutational distance, a result analogous to that shown in Fig. 5.5 A with respect to the KLD, D ($p^0||f^1$). Simulations using the string model also show that the reasons underlying the optimum are the same as that revealed by the simple model. For small mutational distances, the GCs fill up quickly with B cells that have undergone very few mutations, and hence exhibit very low breadth (Fig. 5.13). For too high a mutational distance, most GCs are extinguished, and so the bnAb titers are low. Indeed, Fig. 5.14 shows that, just as for the simple model, the optimal value of the mutational distance corresponds to the condition when GC extinction begins to rise sharply (see Fig. 5.8 B for comparison).

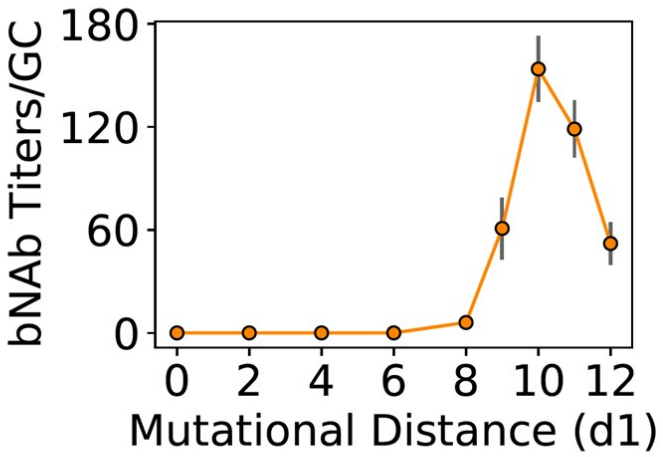


Fig. 5.12: The number (proxy for titers) of bnAbs/GC is shown as a function of the mutational distance between the first immunogen and the GL targeting antigen, d1 (see text) at a fixed value of the immunogen concentration. The maximum is analogous to the maximum shown in Fig. 5.5 A.

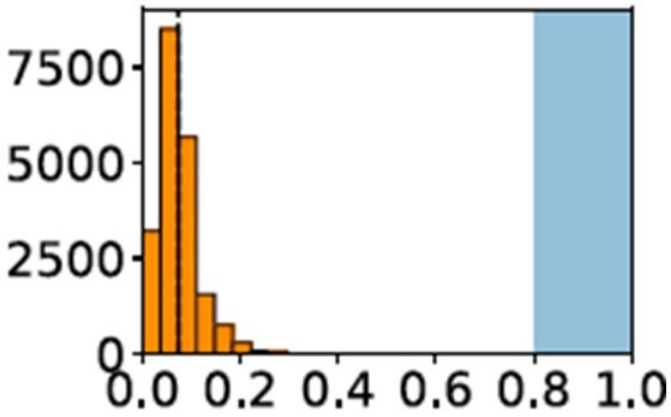


Fig. 5.13: A histogram showing the distribution of breadths of the antibodies produced upon the first immunization when the mutational distance separating the first immunogen and the GL targeting antigen is small (d = 4 in Fig. 5.12). The blue shaded region corresponds to bnAbs.

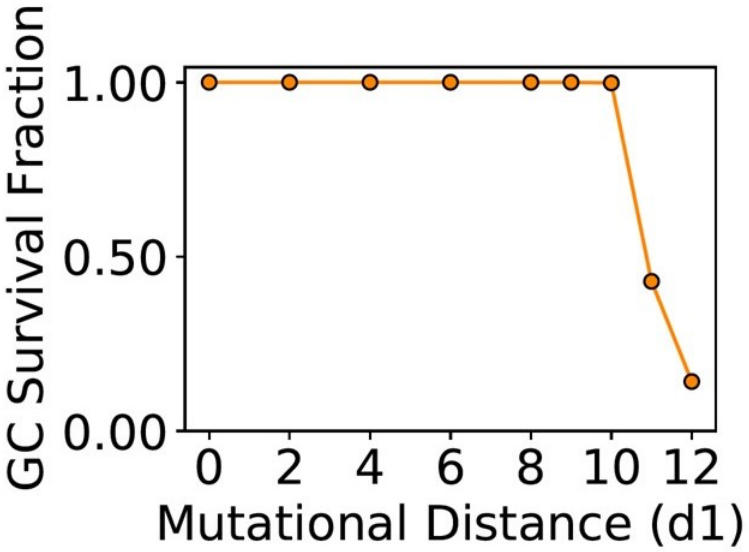


Fig. 5.14: The GC survival rate goes down dramatically around the optimal value of the mutational distance for the first immunogen. This behavior is analogous to that shown in Fig. 5.8 B using the simple one-dimensional model. All simulation conditions are identical to that corresponding to the results shown in Fig. 5.12.

Next, let us ask what happens when a second immunogen is administered. As in the calculation with the simple model, the new GCs are seeded by cells taken from the pool of memory cells that evolved during affinity maturation induced by the first immunogen. In reality, naïve B cells can also seed GCs, and some may not even target an epitope containing the conserved sites. Here, we ignore this potentially significant complexity and

seed the second GC with 10 memory cells from the pool that was produced during affinity maturation.

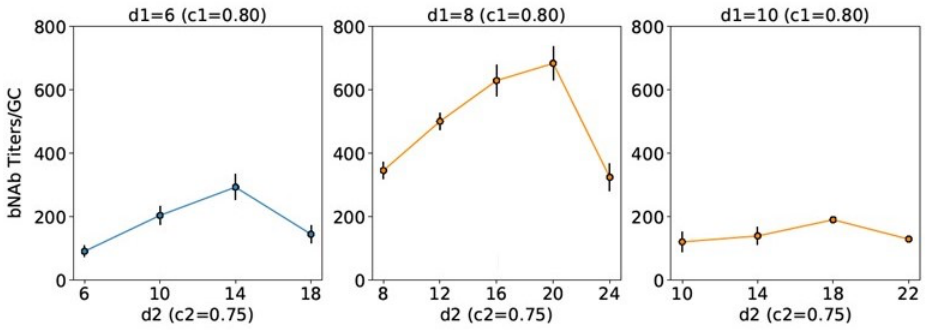


Fig. 5.15: The bnAb titers/GC after the second immunization graphed against the mutational distance (d_2) between the second and first immunogens for three different values of the mutational difference between the first immunogen and the GL targeting antigen (d_1). The antigen concentrations, c1 and c2, for the two immunizations, respectively are also noted.

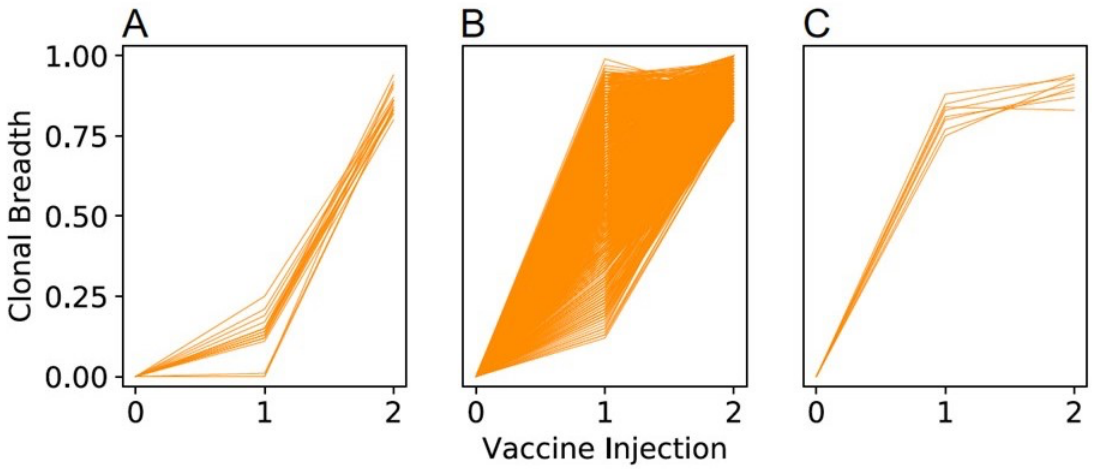


Fig. 5.16: The number of evolutionary trajectories that lead to bnAbs. The abscissa labels 1 and 2 correspond to the first and second immunizations, respectively. Results for high (A), optimal (B), and low (C) antigen concentration during the first immunization are shown. The mutational distance, d_1 ($d_{Ag1\text{-}GL}$) is equal to 4, and d_2 ($d_{Ag1\text{-}Ag2}$+$d_{Ag2\text{-}GL}$) is equal to 8.

Fig. 5.15 shows the bnAb titers/GC that are produced as a function of the mutational distance (d_2) separating the variable regions of the second and first immunogens. The quantity, d_2, is analogous to D ($p^1||f^2$) in the simple model we considered earlier. As in Fig. 5.5 B, we see that for each value of d_1 (analog of D ($p^0||f^1$)), there is an optimal value of d_2. So, consistent with the simplified model, the more detailed simulations also show that there is an optimal combination of prime and boost that maximizes the probability of realizing evolutionary trajectories that produce bnAbs. As we saw with the simplified

model, the optimal prime produces the right kind of diversity, resulting in many possible evolutionary trajectories that produce bnAbs upon boosting (Fig. 5.16). The results in Fig. 5.16 are shown as a function of three different antigen concentrations during the prime, for a fixed mutational distance between the first immunogen and the germline targeting immunogen. As Eq. 11 shows, the selection force can be modulated by changes in the mutational distance or the concentration. A high concentration of antigen during the prime allows most B cells to be positively selected and the GC reaction quickly comes to an end, while a sufficiently low concentration leads to GC extinction. So, the sharpness of the fitness distribution (Fig. 5.3 B) and D_{KL} can be modulated by either mutational distance or antigen concentration. Like the optimal mutational distance, the optimal concentration leads to the desired outcome.

For sequential immunization with variant antigens, the results obtained using the simple model that we developed to search for general principles are consistent with those obtained using a more detailed model. Let us next consider what happens when we immunize with a cocktail of immunogens in one shot, a situation also described by the simple model. For this purpose, we will illustrate the use of another representation for the free energies describing BCR-antigen interactions that has proven to be useful in different contexts.

Simulations with a "shape space" model

The model for BCR-antigen interactions we will use to study immunization with an antigen cocktail was introduced by Perelson and Oster and is called the shape space model. The basic idea is that interaction free energy between antibodies/BCR and antigens depends upon a few coarse-grained degrees of freedom, such as charge, polarity, hydrophobicity and shapes of the interacting regions. The values that each of these features could acquire can be represented on different axes (Fig. 5.17).

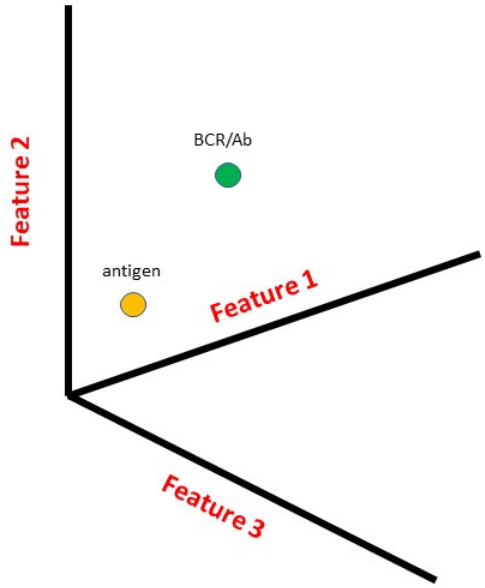


Fig. 5.17: A schematic depiction of "shape space" for the free energy of interaction between a BCR/antibody and an antigen. Each axis corresponds to a particular feature (e.g., charge, hydrophobicity, conformation, etc) of the BCR/antibody and the complementary feature of the antigen. A particular BCR/antibody or an antigen is thus represented as a point in this shape space. For simplicity, a 3-dimensional shape space is shown.

The space spanned by the axes is labeled "shape space". A particular BCR or antigen is defined by specific values of each of the features important for interactions, and thus is a point in this abstract shape space. In this model, the shorter the distance between a BCR and an antigen, the larger (more favorable) their interaction free energy. Note that, in some instances, this implies that a particular axis may represent complementary features of the BCR and the antigen. For example, if a feature represents charge, an axis in shape space will define how positively charged the interaction region of a BCR is. The same axis will represent how negatively charged the corresponding region on the antigen is. As long as such mappings defining complementary axes for BCR-antigen interactions exist, shape space is a useful abstract representation for the corresponding free energies (E), which can be written as:

$$E = \frac{1}{N_d} |\vec{r} - \vec{a}|^2 \qquad (12)$$

where N_d is the number of dimensions, or axes, that defines shape space, and $\vec{r}$ and $\vec{a}$ are the positions of the BCR/antibody and antigen in shape space, respectively. The equilibrium constant of association (K_a) can be defined as:

$$K_a = \exp(-(E - E_a)/k_B T) \qquad (13)$$

E_a is a reference free energy and E = 0, corresponds to the strongest possible interaction between a BCR and an antigen. The value of K_a can increase by a 1000-fold or more during AM driven by a single antigen. If we choose the value of E_a to be $8k_BT$ and consider it to be the threshold value of interaction free energy required for B cells to seed a GC, a roughly 3000-fold increase in K_a becomes possible.

We wish to study a cocktail of variant antigens that share conserved residues but have different variable regions. We can consider shape space to be comprised of N_v axes that define the features of the variable regions of the antigens, and N_c axes that define the features of the common conserved regions ($N_d = N_v + N_c$). It is convenient to choose the shared conserved residues to be at the origin, and the positions of the variable regions of different variant antigens on each of the N_v axes to be drawn from independent Gaussian distributions. That is, the probability of a variant antigen being located at a position, $\vec{a}$, in shape space is:

$$P\left(\vec{a}\right) = \prod_{i=1}^{N_v} \phi\left(a_i, 0, \sigma_v\right) \prod_{j=1}^{N_c} \delta\left(a_j\right) \qquad (14)$$

$\phi\left(a_i, 0, \sigma_v\right)$ is a Gaussian function with zero mean and variance, σ_v, that represents the positions of each coordinate, a_i, describing the features of the variable regions, $\delta\left(a_j\right)$ is the Kroenecker delta function denoting that all antigen coordinates, a_j, corresponding to the conserved residues are located at the origins of their respective axes. Using a Gaussian function to represent the positions of the variable regions of the variant antigens implies that, if a particular BCR is perfectly matched with a particular antigen's variable residues (E = 0), then its interaction free energy with another antigen's variable regions will differ by a typical value, E_v, which equals:

$$\frac{E_v}{K_B T} = \frac{1}{N_d} \langle |\vec{a}|^2 \rangle = \frac{N_v}{N_d} \sigma_v^2 \qquad (15)$$

We expect that the difference between the interaction free energies of a BCR with two variant antigens will be larger if the mutational distance between their variable regions is larger. So, given Eq. 15, the value of σ_v^2 is a measure of the mutational distances between the variable regions of variant antigens. We will vary σ_v^2 to study the effects of immunization with a cocktail of variant antigens that differ in their variable regions to different extents.

During affinity maturation, the BCRs mutate, and this can be represented by a random displacement of a particular BCR in shape space.

$$\vec{r}^{t+1} = \vec{r}^{t} + \Delta\vec{r} \qquad (16)$$

where $\vec{r}^{t+1}$ is the position of the BCR in consideration after mutation, $\vec{r}^{t}$ is its position before mutation, and $\Delta\vec{r}$ is the random displacement, which can be chosen (for simplicity) to be drawn from a Gaussian distribution with zero mean and variance, σ_m; i.e.,

$$P\ (\Delta\vec{r}) = \prod_{i=1}^{N_d} \phi\ (\Delta r_i, 0, \sigma_m) \qquad (17)$$

where ϕ is a Gaussian distribution.

The representation of mutations and free energies of interaction described above has the pleasing feature that the average change in a BCR's interaction free energy with a given antigen upon BCR mutation is positive – i.e., changes in free energy of binding are more likely to be deleterious, rather than beneficial, consistent with expectations and experiments. This is because the average change in free energy of binding with a particular antigen upon BCR mutation, $< \Delta E_m >$, is given by:

$$\frac{<\Delta E_m>}{k_B T} = \frac{1}{N_d}\langle|\overrightarrow{r}^{t+1} - \vec{a}|^2 - |\overrightarrow{r}^{t} - \vec{a}|^2\rangle = \frac{1}{N_d}\langle|\Delta\vec{r}|^2\rangle \sim \sigma_m^2 > 0 \qquad (18)$$

With this model in place, we can now carry out affinity maturation simulations following the steps described earlier for simulations using the string model, but now with cocktails of different variant antigens separated by varying mutational distances (σ_v). The results described below are for $N_c = N_v = 4$, the number of antigens in the cocktail can vary, and so can the antigen concentration. The antigens are distributed in shape space according to Eq. 14. The seeding B cells are chosen to lie on a hypersphere centered around the origin that has a radius of $\sqrt{8N_d}$; i.e., the founder B cells bind to an antigen at the origin with a binding free energy of $E_a = 8\ k_BT$. As we have noted earlier, when considering a cocktail of antigens, two extreme situations can be envisaged: the B cells interact with only one type of variant antigen during each round of mutation and selection, or they interact with all types of variants. To account for both these possibilities, Eq. 11 for the probability of a B cell internalizing antigen can be modified to read as follows:

$$P_a^i = \frac{\sum_j C_j\, e^{\rho(E_{ij}-E_a)}}{1+\sum_j C_j\, e^{\rho(E_{ij}-E_a)}} \qquad (19)$$

The sum runs over all the variant antigens in the cocktail only if all types of variant antigens can interact with B cells during each round of mutation and selection, otherwise only the antigen, j, that interacts with the B cell is considered.

After every round of selection, 5 % of the positively selected B cells with an affinity of at least 4 k_BT for one of the variant antigens is removed and considered to have become an antibody-secreting plasma cell. A threshold is applied because there is some evidence (although not certain) that the B cells with higher affinity BCRs differentiate into plasma

cells. This condition was not applied for the simulations carried out using the string model. The qualitative mechanistic insights are not different if we change this threshold, and as we shall see the results are consistent with those described using our simple model (section 6.2.1). Multiple rounds of mutation and selection ensue until one of the stop conditions noted in the preceding section is met.

To determine the breadth of the antibodies generated for each simulated condition, a panel of variant antigens is generated using Eqs. 14. The mutational distance between the variant antigens used in this panel can be different from that used in the immunization cocktail. For the results shown, the mutational distance between variant antigens in the panel was chosen such that E_v (Eq. 15) was equal to 2 k_BT, and 10^4 variants were used to assess breadth. Choosing a different level of diversity characterizing the variant antigens in the panel used to assess breadth does not change qualitative results. By running many simulations for each simulated condition, the average breadth of the antibodies was determined using 10^5 antibodies generated during the GC reaction.

Let us first consider the situation wherein, during any one round of selection, a B cell interacts with only one type of randomly chosen variant antigen. As we have argued earlier, a cocktail with variant antigens separated by larger (smaller) mutational distances corresponds to a larger (smaller) KLD in our simple model (section 5.2.1). Consistent with the results of this model, the results obtained using the shape space model show that there is an optimal value of σ_v^2 (Fig. 5.18 A).

The mechanistic reason for this optimum is as we described earlier. Consider a B cell that is positively selected upon interacting with a particular variant antigen in one round of selection. During the early stages of affinity maturation, the affinity of the B cells for the variant antigens is not high, and so it is unlikely that strong interactions with the conserved residues shared by the variant antigens have evolved. Therefore, if the average mutational distance separating the antigens in the cocktail is large (large σ_v^2), the affinity of this B cell for another variant antigen is likely to be weak. Consequently, if in the next round of selection this B cell interacts with a different variant antigen, the probability of being positively selected is small. Indeed, simulations using the shape space model show that GC extinction occurs with high probability in this situation. If the average mutational distance separating the variant antigens is small (small σ_v^2), then a B cell's probability of being positively selected in successive rounds of mutation and selection is large, and the GC quickly fills up with B cells that have not acquired the mutations required to achieve high breadth. The optimal value of σ_v^2 provides the right balance for the GC processes to continue for a sufficiently long time to acquire several mutations without a high probability of GC extinction.

If the mutational distance between the variant antigens is fixed and the number of variant antigens changes, an optimal number of variant antigens maximizes breadth (Fig. 5.18 B). If there is only one antigen, then strain specific antibodies will be produced by affinity maturation. If there are many variant antigens, the probability of encountering the same antigen in successive rounds is low and GC extinction occurs with high probability. For the optimal number of variant antigens, the selection force for binding to the shared conserved residues of the antigens can act on the B cell population without a high probability of GC extinction (i.e., it corresponds to an optimal KLD). Finally, as we noted for the results obtained using the string model, antigen concentration can also change the KLD. Thus, one can also optimize the performance of a cocktail immunogen with a fixed number of antigens separated by a fixed average mutational distance by changing antigen concentration.

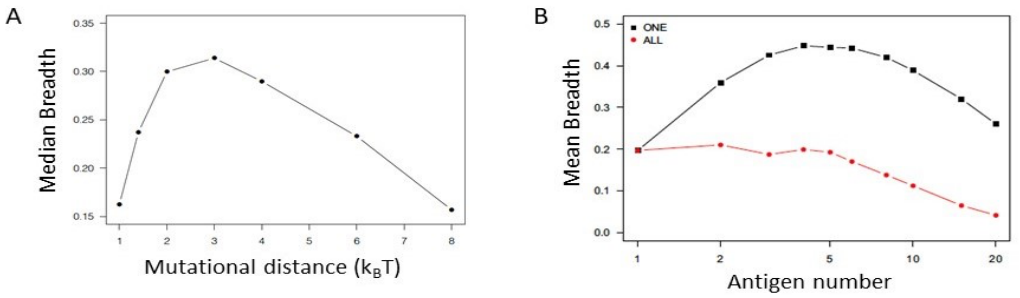


Fig. 5.18: Optimal cocktails of variant antigens maximize the breadth of antibodies. (A) The median breadth of antibodies produced upon immunization with a cocktail of variant antigens whose variable regions are separated by different mutational distances. The mutational distances are specified in units of k_BT as per Eq. 15. The number of antigens is 5, and the antigen concentration is 20 (units as per Eq. 19). (B) The mean breadth of antibodies produced as a function of the number of variant antigens. The mutational distance between the antigens is 2 k_BT, and c = 0.2. A comparison between the cases when only one variant antigen is encountered in every round of selection, and all are encountered is also shown.

Now, let us consider the situation where all types of variant antigens are encountered at the B cell-FDC synapse in every round of selection. In this case, B cells can be positively selected by the same variant antigen during every round of mutation and selection, and so a population of strain-specific B cells exhibiting low breadth will emerge. Indeed, this is what the simulation results show (Fig. 5.18 B). As noted earlier, GC B cells attempt to be positively selected a few times with refractory periods separating these attempts; our coarse-grained models for the probability of internalizing antigen (Eqs. 11 or 19) represent an average over these attempts. If there are very few variant antigens in a cocktail, it is possible that, even if in any one attempt a B cell interacts with only one

variant antigen, overall all types of variant antigens are available during every round of selection. In such a scenario, bnAbs will be unlikely to evolve.

The results discussed for cocktail immunization suggest that evolutionary trajectories that result in bnAbs can be made more likely by optimizing the number of variant antigens, their concentrations, or the mutational distances that separate them. However, how exactly to tune these variables finely in the design of a vaccine is difficult to say. Sequential immunization seems to offer an easier path to optimizing bnAb production. Also, results obtained using the string model show that, using the same variant antigens, sequential immunization is more likely to result in bnAbs than immunization with a cocktail containing the same antigens.

5.3: Experiments and outlook

Many studies have been carried out to trace the evolutionary paths that led to bnAbs upon HIV infection in humans. Antibodies obtained from temporally ordered clinical samples were sequenced and thus the evolution of the lineage of B cells that evolved to bnAbs as well as the corresponding germline precursor were inferred. Immunogens have been designed that can activate germline B cells that can potentially evolve to bnAbs. Experiments aimed toward eliciting bnAbs by vaccination in animal models have also been reported for HIV and influenza antigens. Here, we briefly discuss some studies that are directly pertinent to the topics and mechanisms discussed in this chapter, and those that point out some of the complexities that we have not considered.

GP120 is one of the two proteins that constitutes the trimeric viral spike of HIV and contains the CD4 receptor binding site. Mutations were introduced into the surface residues of a modified form of the GP120 monomer to obtain 4 variant immunogens which shared the conserved CD4 binding site. Note that these immunogens are not variants of an intact trimeric spike. Therefore, immunization with these immunogens cannot result in antibodies that can neutralize the actual virus. But one can ask whether cross-reactive antibodies that focus their binding on the shared conserved residues could be elicited upon immunization with these variant antigens. One group of mice was immunized every two weeks sequentially with each of the four immunogens, while another received a cocktail of the same immunogens four times. A third group of mice was immunized with one of the variant immunogens, which served as a control as this protocol is not expected to elicit cross-reactive antibodies.

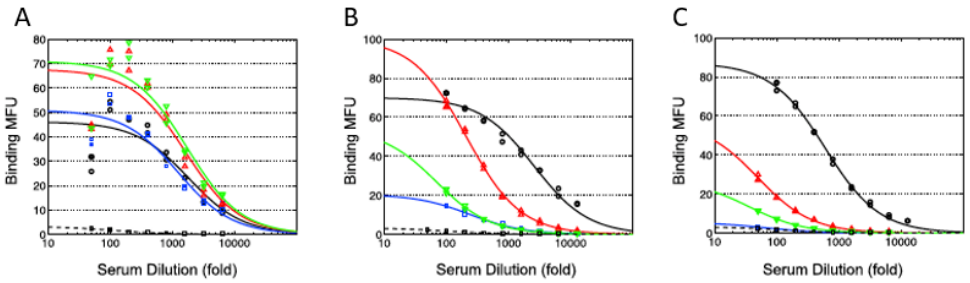


Fig. 5.19: Two groups of mice were immunized with four variant immunogens composed of a modified form of the monomer of GP120. The ability of the generated antibodies to bind to each of the four variant antigens (black, red, green and blue curves) is shown. The abscissa corresponds to the extent to which serum taken from mice is diluted; smaller values correspond to a higher concentration of antibodies. The ordinate is a measure of the amount of bound antibodies measured using a fluorescence assay; MFU corresponds to mean fluorescent units. Data from representative immunized mice are shown. The black dashed line corresponds to serum taken from a mouse that was not immunized. (A) Sequential immunization. (B) Immunization with a cocktail of the same variant immunogens. (C) Immunization with one of the variant immunogens.

Fig. 5.19 shows data from representative mice from each group. Serum (blood) from a mouse is progressively diluted (so that it contains fewer antibodies), and the antibodies that bind to an antigen are measured at each dilution level using a fluorescence assay. The antibodies generated upon sequential immunization bind to the four variant antigens with affinities that are more similar to each other compared to antibodies generated upon immunization with a cocktail of antigens (compare Figs. 5.19 A and B). Moreover, comparing Figs. 5.19 B and C shows that the distribution of antibody binding affinities for the four variant antigens for mice immunized with a single immunogen and a cocktail of four variants is similar. These data suggest that mice immunized sequentially with variant antigens are more likely to result in cross-reactive antibodies than those immunized with a cocktail. Our theoretical/computational predictions are consistent with this result. In the experiments with the cocktail, we do not see a dearth of antibodies, which would be true if there was massive GC extinction. This is because if the antigen concentration is high enough, the FDCs likely display the variant antigens homogeneously. In this circumstance, the theoretical prediction is that strain-specific antibodies would evolve (see earlier discussion), and indeed that is what the data shows.

Do the cross-reactive antibodies produced by sequential immunization focus their binding on the shared conserved residues of the variant antigens? To address this question, the following experiment was performed. VRCO1 is a potent bnAb isolated from humans that binds to the conserved CD4 binding site. This antibody was displayed on yeast, and GP120 was then added. GP120 is expected to bind to VRCO1 via its CD4 binding site, which would

thus be occupied. If antibodies added now can bind to GP120, it implies that they do not bind to the conserved CD4 binding site. As Fig. 5.20 shows, the antibodies in serum from mice immunized sequentially with the four variant antigens do not bind much, while antibodies obtained from mice immunized with a cocktail or a single variant do bind. These data show that, consistent with theoretical expectations, mice immunized sequentially with variant antigens are more likely to evolve cross-reactive antibodies that focus their binding footprint to the shared conserved residues.

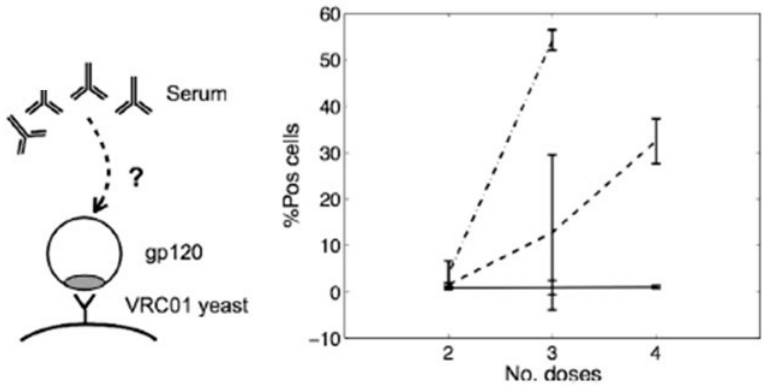


Fig. 5.20: VRCO1, a HIV bnAb that binds to the CD4 binding site of gp 120 (shown as shaded area in left panel) was displayed on the surface of yeast. The gp 120 antigen was then added, followed by serum antibodies. The percentage of serum antibodies that bound is shown in the right panel for sera taken from mice immunized sequentially with variant antigens (solid line), antigen cocktail (dashed line), and a single antigen (dot dashed line). The data show results obtained by following the same mice across multiple immunizations. The error bars reflect the standard deviation across mice that were immunized using the same immunization protocol. Due to technical details of the assay, 60 % positive cells is the maximum signal that can be observed.

Other studies using more realistic antigens have also shown that sequential immunization can lead to cross-reactive antibodies in animal models. We briefly summarize one such study here. PGT121 is a potent bnAb that neutralizes diverse strains of HIV, and its germline precursor has been inferred. Engineered mice were generated in which a form of this germline precursor was expressed in the majority of its B cells. The heavy chain of the precursor was that of the fully mature PGT121 bnAb but the light chain has to evolve multiple mutations that are important for PGT121's activity. Like the germline precursor of PGT121, the germline B cells in the engineered mice did not bind to a mimic of the HIV spike. So, as noted earlier, a germline targeting immunogen has to be designed to activate these B cells. A germline targeting immunogen was designed that achieved this goal. The activated cells and associated antibodies did not exhibit high breadth (like in our simple model, Fig. 5.3A). The engineered mice were sequentially immunized with HIV spike antigens that represent intermediates between the germline targeting antigen and the

real HIV spike. Finally, they were immunized with a cocktail of variants of the HIV spike. The sera from most immunized mice were cross-reactive to various HIV strains. However, the generated antibodies exhibited lower breadth than PGT121. These antibodies evolved many mutations compared to the germline sequences, and monoclonal antibodies isolated from the sera exhibited a broad range of breadth of coverage. Overall, the results showed that sequential immunization with variant antigens led to progress toward evolution of bnAbs in engineered mice.

Note that in the models that we studied in this chapter the antigens that we considered after the germline B cells were activated were variants of the full HIV spike. This is why the fitness landscape was chosen to be peaked at the highest breadth state. In the experiments with the engineered mice that we just described the sequentially administered antigens were intermediates between the germline targeting antigen and the full HIV spike. Thus, the peak of the fitness landscape shifts from lower breadth states to the bnAb state upon sequential immunization with these antigens. The models we studied can be generalized to consider this case.

There is currently very limited data from experiments that systematically probe whether sequential vaccination protocols can be optimized by the choice of variant antigens to maximize evolution of cross-reactive antibodies or bnAbs. However, there is some evidence that an optimal cocktail of antigens may promote the evolution of bnAbs upon natural infection. Typically, upon natural infection bnAbs have been observed to evolve upon sequential interactions with different strains of the virus. However, some studies suggest that bnAbs emerge during natural infection when the immune system is exposed to a swarm (or cocktail) of viral strains that are neither too different nor too similar. Such a situation can arise upon infection with more than one viral strain either at the same time or spaced by a short time interval. Usually only one strain of the HIV virus establishes infection and such an infection with multiple strains is rare. We briefly describe one study where such a situation was realized. In this study an individual evolved bnAbs after being infected with two viruses. Recombination between these two strains led to the evolution of two types of virus lineages. The resulting viral strains had closely related spike sequences, but not as closely related as the viral strains that evolve after typical infection with a single strain. The latter situation typically does not lead to bnAb evolution. These data suggest that perhaps an optimal difference between variant antigens administered as a cocktail may promote bnAb evolution, as predicted by the results from our models (Figs. 5.5 and 5.18).

In the models we have discussed in this chapter, we assumed that only the memory cells produced after one immunization enter GCs upon exposure to another variant antigen.

While memory cells can enter newly formed GCs, recent mouse experiments suggest that mostly naïve cells enter such secondary GCs upon re-exposure to antigen. Understanding how the balance of naïve cells versus memory cells entering GCs influences the humoral immune response is an important problem. This is especially true in light of findings on the way in which previous exposure to an antigen influences the recall response upon re-exposure to the same antigen. Studies in mice as well as humans sequentially vaccinated with multiple shots of the identical sequence of the SARS-CoV-2 spike proteins have made vivid the importance of feedback loops that influence the recall response. Upon re-exposure to the same antigen, memory cells generated upon previous exposure to antigen are rapidly expanded outside GCs. Higher affinity memory cells are selectively expanded by processes very similar to those that occur in GCs except that there is little mutation. The majority of these expanded memory cells then differentiate into short-lived plasma cells that rapidly secrete antibodies. This leads to the first wave of antibodies produced during the recall response to confer protection.

Upon first exposure to a new antigen, circulating generic IgM antibodies bind to the antigen and form immune complexes that are deposited on FDCs. These antigens displayed on FDCs drive GC processes. Soluble antigen decays quite rapidly (a few days in monkeys), and so the amount of antigen that these weakly binding antibodies can bind and deposit on FDCs before antigen decays is small. Thus, germline B cells that bind to their epitopes with higher affinity or larger precursor frequencies are more likely to enter GCs and outcompete those germlines that bind to epitopes with lower affinity or are less prevalent. That is, immunodominant responses are more likely to emerge after the first exposure to antigen and sub-dominant responses are muted.

Upon re-exposure to the same antigen secondary GCs form. The antibodies produced during the first response and those that are rapidly produced outside GCs upon re-exposure are specific for the antigen. Thus, they can bind to and deposit far more antigen on FDCs. The availability of more antigen can promote the entry of naïve cells into secondary GCs that target epitopes with lower affinity, and they can also compete better in GCs with B cells that target the immunodominant epitopes. Thus, sub-dominant responses are promoted in secondary GCs. Sub-dominant responses are further amplified by the fact that antibodies targeting the immunodominant epitopes produced during the first antigen exposure or from expansion of memory cells outside GCs upon re-exposure to antigen can enter secondary GCs. These antibodies bind to their epitopes thereby lowering their effective availability in secondary GCs. This effect, called epitope masking, further promotes the evolution of sub-dominant responses as their epitopes are not masked. Thus, repeated boosting with the same antigen can promote a more diverse

antibody response due to the feedback loops described above. The effects of these feedback loops on the humoral immune response needs far more study. For example, their effects in the responses that emerge after sequential immunization with variant antigens that share a conserved immunodominant epitope needs further exploration.

Many other variables influence the immunodominance hierarchy of the humoral response. For example, consider sequential immunization in a context pertinent to eliciting bnAbs that target the conserved stem epitope of the HA spike on the influenza virus. As mentioned earlier, this epitope is normally not targeted significantly upon natural infection and targeting this region effectively by vaccination requires that the natural immunodominance hierarchy be altered. One reason that this epitope is sub-dominant is that the geometry of the HA spike and its high density on the influenza virus sterically shields this epitope, in contrast to the exposed variable epitopes on the HA head that are immunodominant. BnAbs that target the conserved stem epitope also evolve from germline B cells that use a specific heavy chain gene. The frequency of occurrence of such germline B cells in humans is 0.14 %. Therefore, to investigate how vaccination protocols can be designed to maximize the probability of evolutionary trajectories that result in stem-targeting influenza bnAbs, we have to consider the role of antigen geometry, and germline precursor frequency and affinity on GC reactions.

Other variables that can affect immunodominance hierarchies during GC reactions are also being explored. The effects of the relative stringency of selection by T helper cells and delivering the vaccine dose in a temporal pattern rather than just a bolus of antigen are just two examples.

The duration over which memory is maintained for effective protection varies greatly for infection with different pathogens and immunization with different vaccines. The factors and mechanisms that control the period over which memory of a past exposure to a pathogen or vaccine immunogen is maintained are not well-understood. This fundamental question is of obvious practical import.

In this chapter, we have also not considered how antibody responses change over time upon infection with a highly mutable pathogen that evolves to evade immune responses. This is an interesting problem as, for a persistent infection like HIV, the antibody response and the virus population co-evolve in an arms race within an individual. A similar arms race plays out at the population level for acute infections as exemplified by the emergence of new influenza and SARS-CoV-2 strains that evade population-level immune responses generated during past exposures.

We have also not considered how stochastic models of affinity maturation can be combined with atomistically detailed representations of the BCR-antigen binding free energies. Large data sets on antibody-antigen binding affinities are beginning to be generated and machine learning approaches are being applied with the goal of generating predictive models of binding free energies. Studies that combine these models with stochastic models of processes inside and outside GCs may allow accurate predictions of the humoral immune response to specific infections or vaccine antigens. These, and many other issues concerning the evolution of antibodies upon natural infection or vaccination, and how these processes can be modulated, are exciting opportunities to explore at the intersection of statistical physics, immunology, and evolutionary biology. Close collaborations between physical and life scientists in exploring these issues are likely to yield fundamentally new insights into immunology, which will also help enhance human health and well-being.

Suggested Readings

1] G.D. Victora, M.C. Nussenzweig, "Germinal Centers", *Annual Reviews in Immunology*, **40**, 413 (2022).

2] T.B. Kepler, A.S. Perelson, "Somatic hypermutation in B cells: an optimal control treatment", *J. Theoretical Biology*, **164**, 37 (1993).

3] M. Oprea, A.S. Perelson, "Somatic mutation leads to efficient affinity maturation when centrocytes recycle back to centroblasts", *J. Immunology*, **158**, 5135 (1997).

4] H.N. Eisen, G.W. Siskund, "Variations in affinities of antibodies during immune response", *Biochemistry*, **3**, 996 (1964).

5] L.A. Steiner, H.N. Eisen, "Sequential changes in the relative affinity of antibodies synthesized during the immune response", *J. Experimental Medicine*, **126**, 1143 (1967).
6] J. Zhang, E.I. Shakhnovich, "Optimal mutation and selection in germinal centers", *PLOS Comp. Biol*., 6, e1000800 (2010).

7] M.D.S. Kumar, M.M. Gromiha, "PINT: protein-protein interactions", *Nucleic Acids Research*, **34**, 195 (2006).

8] F. Batista, M.S. Neuberger, "Affinity dependence of the B cell response to antigen: a threshold, a ceiling and the importance of off-rate", *Immunity*, **8**, 751 (1998).

9] M. Meyer-Hermann et al., "Germinal centres seen through the mathematical eye: B cell models on the catwalk", *Trends in Immunology*, **30**, 157 (2009).

10] R.J. De Boer, A.S. Perelson, "How Germinal Centers Evolve Broadly Neutralizing Antibodies: the Breadth of the Follicular Helper T Cell Response", *J. Virology*, **91**: e00983-17. https://doi.org/10.1128/JVI.00983-17 (2017).

11] A. Nourmohammad et al., "Host-Pathogen Coevolution and the Emergence of Broadly Neutralizing Antibodies in Chronic Infections", *PLOS Genetics*, **12(7)**: e1006171. https://doi.org/10.1371/journal.pgen.1006171 (2016).

12] T.B. Kepler, A.S. Pereslson, "Cyclic re-entry of germinal center B cells and the efficiency of affinity maturation", *Trends in Immunology*, **14**, 412 (1993).

13] B. Haynes et al., "B-cell–lineage immunogen design in vaccine development with HIV-1 as a case study", *Nature Biotechnology*, **30**, 423 (2012).

14] D.R. Burton, "Advancing an HIV vaccine; advancing vaccinology", Nature Reviews Immunology, **19**, 77 (2019).

15] M. Caskey et al., "Broadly Neutralizing Antibodies for HIV-1 Prevention or Immunotherapy", *New England Journal of Medicine*, **375**, 2019 (2016).

16] M. Sangesland, D. Lingwood**, "**Antibody Focusing to Conserved Sites of Vulnerability: The Immunological Pathways for 'Universal' Influenza Vaccines", *Vaccines*, **9**, 125 (2021) *https://doi.org/10.3390/vaccines9020125*.

17] Y. Chen et al., "Broadly neutralizing antibodies to SARS-CoV-2 and other human coronaviruses", Nature Reviews Immunology, **23**, 189 (2023).

18] R. Abbott et al., "Precursor frequency and affinity determine B cell competitive fitness in germinal centers, tested with germline-targeting HIV vaccine immunogens", *Immunity*, **48**, 133 (2018).

19] A. Amitai et al., "Defining and manipulating B cell immunodominance hierarchies to elicit broadly neutralizing antibody responses against influenza virus", *Cell Systems*, **11**, 573 (2020).

20] J. G. Jardine et al., "HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen", *Science*, *351*, 1458 (2016).

21] R. S. Ganti, A.K. Chakraborty, "Mechanisms underlying vaccination protocols that may optimally elicit broadly neutralizing antibodies against highly mutable pathogens", *Physical Review E*, **103**, 052408 (2021).

22] C. Gardiner, "*Stochastic methods - A handbook for the natural and social sciences*", Fourth Edition, Springer-Verlag, Berlin (2009).

23] D.T. Gillespie, “Exact Stochastic Simulation of Coupled Chemical Reactions”, *Journal of Physical Chemistry*, **81**, 2340 (1977).

24] T.M. Cover, J.A. Thomas, “*Elements of information theory*”, Second Edition, John Wiley and Sons, Hoboken, New Jersey (2006).

23] S. Wang et al., “Manipulating the selection forces during affinity maturation to generate cross-reactive HIV antibodies”, *Cell*, **160**, 785 (2015).

24] K.G. Sprenger et al., “Optimizing immunization protocols to elicit broadly neutralizing antibodies”, *Proc. Natl. Acad. Sci*., **117**, 20077 (2020).

25] J.S. Shaffer et al., “Optimal immunization cocktails can promote induction of broadly neutralizing Abs against highly mutable pathogens”, *Proc. Natl. Acad. Sci*., **113**, E7039 (2016).

26] A.S. Perelson, G.F. Oster, “Theoretical studies of clonal selection: Minimal antibodyrepertoire size and reliability of self-non-self discrimination”, *J. Theoretical Biology*, **81**, 645 (1979).

27] A. Escolano et al., “Sequential immunization elicits broadly neutralizing anti-HIV-1 antibodies in Ig knockin mice”, *Cell*, **166**, 1445 (2016).

28] T. Mohan et al., “Sequential immunizations with a panelof HIV-1 Env virus-like particles coach immune system to make broadly neutralizingantibodies”, *Scientific Reports*, **8,** 7807 (2018).

29] A. Escolano et al., “Sequential immunization of macaques elicits heterologous neutralizing antibodies targeting the V3-glycan patch of HIV-1 Env”, *Science Translational Medicine*, **13**, eabk1533 (2021).

30] L. Mesin et al., “Restricted clonality and limited germinal center reentry characterize memory B cell reactivation by boosting”, Cell, 180, 92 (2020).

31] F. Muecksch et al., “Increased memory B cell potency and breadth after a SARS-CoV-2 mRNA boost”, *Nature,* **607**, 128 (2022).

32] L. Yang et al., “Antigen presentation dynamics shape the antibody response to variants like SARS-CoV-2 Omicron after multiple vaccinations with the original strain”, *Cell Reports*, 42, 112256 (2023).

33] J.M.J. Tas et al., “Antibodies from primary humoral responses modulate the recruitment of naïve B cells during secondary responses”, *Immunity*, **55**, 1856 (2022).

34] D. Schaefer-Babajew, "Antibody feedback regulates immune memory after SARS-CoV-2 mRNA vaccination", *Nature*, **613**, 735 (2023).

35] H.H. Tam et al., "Sustained antigen availability during germinal center initiation enhances antibody responses to vaccination", *Proc. Natl. Acad., Sci*., **113**, E6639 (2016).

36] K.M. Cirelli, "Slow delivery immunization enhances HIV neutralizing antibody and germinal center responses via modulation of immunodominance", *Cell*, **177**, 1153 (2019).

37] C. Joyce et al., "Antigen pressure from two founder viruses induces multiple insertions at a single antibody position to generate broadly neutralizing HIV antibodies", PLOS Pathogens, **19(6)**: e1011416 (2023).

38] T. Mora, A.M. Walczak, "Quantitative Theory of Viral-Immune Coevolution May Be within Reach", *PRX Life*, **1**, 011001 (2023).

CHAPTER 6

Intracellular signaling in lymphocytes

6.1: Introduction