

# Universal characterization of epitope immunodominance from a multi-scale model of clonal competition in germinal centers

Federica Ferretti<sup>1</sup> and Mehran Kardar<sup>2</sup>

<sup>1</sup>*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

<sup>2</sup>*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

We introduce a novel, multi-scale model for affinity maturation, which aims to capture the intra-clonal, inter-clonal and epitope-specific organization of the B cell population in a germinal center. We describe the evolution of the B cell population via a quasispecies dynamics, with species corresponding to unique B cell receptors (BCRs), where the desired multi-scale structure is reflected on the mutational connectivity of the accessible BCR space, and on the statistical properties of its fitness landscape. Within this mathematical framework, we study the competition among classes of BCRs targeting different antigen epitopes, and construct an effective immunogenic space where epitope immunodominance relations can be universally characterized. We finally study how varying the relative composition of a mixture of antigens with variable and conserved domains allows for a parametric exploration of this space, and identify general principles for the rational design of two-antigen cocktails.

what "indicator" the competition?

## I. INTRODUCTION

The molecular foundation of pathogen recognition and neutralization is the specific, high-affinity binding between antibodies and antigens [1]. Individual antibodies or B cell receptors recognize the antigen at discrete surface accessible regions, known as antigenic determinants or B-cell epitopes. The size of these recognition sites is considerably smaller than the overall size of natural antigens; such as viral proteins or other pathogen-derived molecules. As a result, a population of antibodies interacting with the same antigen can give rise to a multitude of different structural conformations for the bound antigen-antibody complex.

While the accurate identification and prediction of B cell epitopes is still a laborious and expensive experimental task and an outstanding computational challenge, a mixture of experimental and computational methods have been developed to extract a coarse-grained classification of groups of antibodies based on where they bind on the surface of a given antigen [2, 3]. We can then see antigens, from the perspective of immune responders, as displaying a mosaic-like surface, with antibodies that bind on the same tile grouped into distinct classes. The question we are interested in is when and why certain antibody classes can outnumber others, in the antibody repertoire of an individual or of a population.

The composition of each individual's antibody pool is the result of an accelerated evolutionary process called affinity maturation, occurring in the lymph nodes upon the encounter of a foreign agent. In response to this event, the immune system of the host organizes a sophisticated learning machinery in substructures of the lymph nodes known as germinal centers, where low-affinity naive B cells undergo iterated rounds of replication, mutation and selection to acquire the desired specificity. The resulting B cell population produced by this affinity maturation process will compose the memory and antibody repertoires of the infected host [4].

Antigenic drift —i.e. the immune evasion pattern of some viruses— has revealed that the primary antibody response is often focused towards a small subset of epitopes, referred to as immunodominant, rather than being uniformly directed towards the entire set of possible target sites [5]. The physicochemical, structural, and geometrical aspects of the distinct antigen epitopes, as well as potential biases in the naive B cell repertoires, indeed cause the antigen surface to be non homogeneously immunogenic [6, 7]. While in principle these differences in epitope immunogenicity could vary from one individual to another because of the personalized aspects of naive repertoires [8], epitope immunodominance seems to be largely a property of the pathogen, only slightly affected by individual or even organismal particularities (at least for certain viruses) [9].

when exposed also?

The key idea explored in this paper is that B cell epitope immunodominance can be described as a phenomenon emerging from the general statistical features of the affinity landscapes of competing antibody classes. We introduce a new minimal model for affinity maturation, inspired by a long tradition of computational models, such as [10–15], whose central feature is a multi-scale representation of the clonal competition in a germinal center. B cell competition indeed occurs at several different stages:

- (i) inter-class level: following the nomenclature used for SARS-CoV-2 antibodies, B cell receptors (BCRs) are categorized into classes based on the epitopes that they bind to [16].
- (ii) intra-class level: within each class, competition occurs among clonal lineages originating from different B cell ancestors (germlines);
- (iii) intra-clonal level: somatic hypermutations occurring during affinity maturation produce variability in the fitness of B cells within the same clonal family, enabling competition even at this level.

Within this framework of evolving B cell populations, more immunodominant epitopes are associated with higher fixation probabilities of the corresponding BCR classes. We compute the fixation probability as an extreme value problem, and identify, for a specific class of fitness landscape models, a restricted set of universal parameter combinations which modulate immunodominance.

The proposed model can be applied to immunodominance phenomena across different affinity maturation contexts. For example, we investigate how the epitope immunodominance relations can be manipulated by exposing the system to a cocktail of two antigens with varying relative concentrations. This is relevant to the problem of vaccination by a cocktail of antigens, a procedure employed in the development of vaccines against highly mutating pathogens, such as the SARS-CoV-2 bivalent booster [17, 18]. The rationale behind the use of immunogen cocktails is that simultaneous exposure to two or more antigen variants confers a competitive advantage to broadly neutralizing antibodies, i.e. antibodies capable of neutralizing multiple variants of a virus by recognizing conserved regions of the antigen. We employ our model in a toy description of the cocktail context to determine under what general conditions the manipulation of vaccine compositions can alter the immunodominance relations between conserved and variable epitopes.

A more detailed mathematical description of the dynamical model is provided in Sec. II. Results for the construction of a universal immunodominance phase diagram and for the problem of optimizing a bivalent antigen cocktail are presented in Sec. III; implications for future work and data analysis are briefly discussed in Sec. IV.

## II. MODEL

### 1. A schematic description of affinity maturation

Affinity maturation (AM) is the key training process enabling the mammal adaptive immune system to learn upon the encounter of a foreign agent. It encompasses an intricate series of reactions, involving various lymphocytes, cytokines and signaling pathways [4, 19, 20]. At a coarse-grained level, AM is akin to an iterative two-step evolutionary process that takes place in germinal centers (GCs) formed in lymph nodes upon encounter on pathogen: Some (germline) B cells with some affinity to the antigen first undergo replication with a fast accumulation of mutations (in the dark zone of GC), followed by competitive selection (in the light zone of GC). Through repetition of these two steps, the B cell repertoire of an individual is expanded and refined in affinity for the encountered antigen.

We adopt here a simplified model for affinity maturation, which describes the evolution of B cells via parallel mutation, replication and death events. Competition is introduced by looking at the dynamics of the fraction of

each BCR sequence in the germinal center, rather than their absolute number, as in Eigen's well-known quasispecies model [21]. In GCs all B cells compete for the same resources, i.e. interactions with follicular T cells and antigen capture. We consider each productive B cell receptor sequence as a distinct quasispecies, denoted by index  $i = 1, \dots, N$ . Replication, death and mutation events are described as simple first order reactions:

$$B_i \xrightarrow{\lambda_i} B_i + B_i, \quad B_i \xrightarrow{\delta} \emptyset, \quad B_i \xrightarrow{\mu_{ji}} B_j, \quad (1)$$

where  $\lambda_i$  is a replication rate that depends on the affinity of the BCR to the presented antigen,  $\delta$  is a constant death rate, and  $\mu_{ji}$  is the effective mutation rate from sequence  $i$  to sequence  $j$ . For the sake of simplicity, we assume that mutations only occur at a constant rate  $\mu$  between pairs of productive sequences at a Hamming distance of one [22].

### 2. Structure of the B cell receptor space

Once the hopping rate is fixed, only the connectivity of the graph of allowed mutations needs to be specified. In accord with the choice of a homogeneous rate, we decide to focus here only on the *likely* mutations which can accumulate on any BCR sequence during the process of affinity maturation [23]. For each clonal lineage, originating from a distinct germline ancestor, let us encode in a binary string of length  $d$  any of such accessible sequences, so that the resulting mutational graph associated to the lineage is a  $d$ -dimensional hypercube.

Since multiple germ lines are recruited to initiate affinity maturation, we describe the germinal center as a collection of disconnected hypercubes. The mutual distance between two distinct germ lines is indeed typically larger than the mutational distance between somatic B cells and their germline ancestors (usually of the order of 10 residues or less [24, 25]), suggesting that events of convergent evolution can be generally neglected.

To capture the competitive dynamics of B cell subpopulations targeting distinct parts of the antigen, we finally group the clonal lineages into classes [26]. To each class we assign a specific distribution of fitness landscapes, so that the landscapes attributed to all lineages within the same class are treated as independent, quenched realizations derived from these distributions. This model is based on the premise that the fitness of a B cell receptor is largely determined by its binding affinity to the presented antigen, and that the specific geometry and chemical properties of each epitope will sculpt affinity landscapes sharing similar statistical features for groups of B cells within the same class. Our goal is to describe immunodominance in terms of the statistical differences among these class-specific fitness landscape ensembles.

Figure 1 provides an illustration of the resulting structure of the GC space.

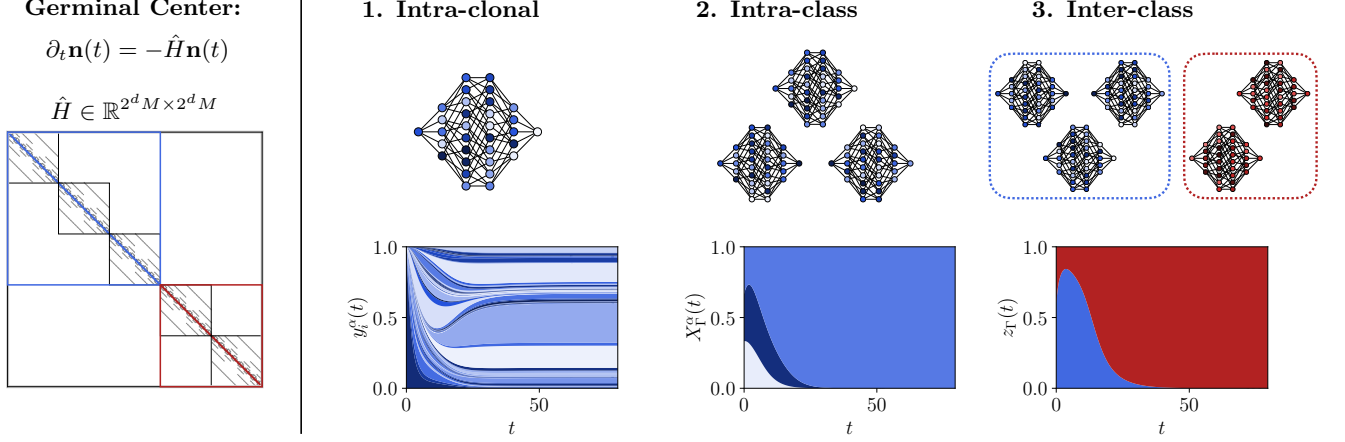


FIG. 1. Structure of the affinity maturation model. B cell evolution in a germinal center is described by a parabolic Anderson model (PAM) on a collection of disconnected  $d$ -dimensional hypercubes. Each clonal lineage evolves on a distinct hypercube, and lineages are grouped into classes, depending on the targeted antigen epitope. The resulting Anderson matrix has a block diagonal structure, with smaller blocks representing individual lineages/hypercubes. The diagonal entries are random variables drawn from the fitness distribution of the class corresponding to the outer, epitope-specific block. B cells are subject to an all-to-all competition, which manifests at three different levels (cf. Eqs. (3)–(5)). On the right panel, we show Muller plots for the evolution of quasi-species at each of these three levels, for a quenched realization of a germinal center seeded by  $M = 5$  germlines belonging to 2 distinct classes ( $m_V = 3, m_C = 2$ ), with Gaussian fitness distributions  $p_V(f) = \mathcal{N}(1.6, 0.5)$ ,  $p_C(f) = \mathcal{N}(1.4, 0.7)$ .

### 3. Evolution dynamics

Let  $n_i^\alpha$  be the expected size of quasispecies  $i$  belonging to lineage  $\alpha$ ; from (1), we can derive a simple linear ODE for  $\mathbf{n}(t) \in \mathbb{R}^{N \times M}$ :

$$\partial_t n_i^\alpha = [f_i^\alpha \delta_{ij} + \mu \Lambda_{ij}] n_j^\alpha, \quad i = 1, \dots, N, \quad \alpha = 1, \dots, M, \quad (2)$$

where  $N = 2^d$  is the number of nodes of each hypercube [27], and  $M$  is the number of clonal lineages in a single germinal center, typically of the order of  $10 - 10^2$  [4]. The  $M$  clonal lineages are organized into a small number of classes, indexed by  $\Gamma$ , each of which contains  $m_\Gamma$  elements. The  $\Lambda_{ij}$  matrix indicates the Laplacian of the  $d$ -dimensional hypercube; the fitness is the net growth rate  $f_i^\alpha = \lambda_i^\alpha - \delta^\alpha$ . Initially the population is localized on the precursors' sequences: hence, without loss of generality,  $n_i^\alpha(0) = \delta_{i,0} \forall \alpha$ .

In the case of independent identically distributed (IID) fitness variables, the model in Eq. (2) is known as parabolic Anderson model (PAM) [28]; analytical results for the PAM dynamics in the thermodynamic limit are known for several types of graphs, including hypercubes [29].

We are interested here in the evolution of B cell species frequencies. This is a common way to introduce competition in evolutionary models with only first order reactions, which has been demonstrated to capture the behavior of classical fixed population models (such as Wright-Fisher's or Moran's) in the limit of infinite population size [21]. Given the structure of the model, we can focus on competition at different levels:

(i) intra-clonal level:

$$\partial_t y_i^\alpha = [f_i^\alpha - \bar{f}^\alpha(t)] y_i^\alpha + \mu \sum_{j=1}^N \Lambda_{ij} y_j^\alpha, \quad (3)$$

where  $y_i^\alpha(t) = n_i^\alpha(t) / \sum_{i=1}^N n_i^\alpha(t)$  is the fraction of identical clones  $i$  within the lineage, and  $\bar{f}^\alpha(t) = \sum_{i=1}^N f_i^\alpha y_i^\alpha(t)$  is the average fitness of the clonal family;

(ii) intra-class level:

$$\partial_t X_\Gamma^\alpha = [\bar{f}^\alpha(t) - \Phi_\Gamma(t)] X_\Gamma^\alpha, \quad (4)$$

where  $X_\Gamma^\alpha(t) = \sum_i n_i^\alpha(t) / \sum_{\alpha \in \Gamma} \sum_i n_i^\alpha(t)$  is the fraction of lineages in the class, and  $\Phi_\Gamma(t) = \sum_{\alpha \in \Gamma} \bar{f}^\alpha(t) X_\Gamma^\alpha(t)$  is the population-averaged fitness of class  $\Gamma$ ;

(iii) inter-class level:

$$\partial_t z_\Gamma = [\Phi_\Gamma(t) - \bar{F}(t)] z_\Gamma, \quad (5)$$

where  $z_\Gamma(t) = \sum_{\alpha \in \Gamma} \sum_i n_i^\alpha(t) / \sum_\alpha \sum_i n_i^\alpha(t)$  is the relative size of class  $\Gamma$ , and  $\bar{F}(t) = \sum_\Gamma \Phi_\Gamma(t) z_\Gamma(t)$  is the total population-averaged fitness.

At the intra-clonal level, the system reaches at long times a state of mutation-selection balance. At the intra-class and inter-class level, due to the disconnected structure of the global graph, the stable fixed points of Eqs. (4)–(5) correspond to fixation of the asymptotically

fittest quasi-species and extinction of the rest. The eventual dominance of a single clonal lineage in the GC population is consistent with experimental observations, even though affinity maturation typically terminates before the GC becomes completely monoclonal [30, 31]. In this work, we will use the quenched average of  $z_\Gamma(t)$  at long times to study epitope immunodominance from the GC

response.

In the asymptotic-time limit, for any class  $\Gamma$ ,  $z_\Gamma$  can only take values 0 or 1, depending on the realized collection of random fitness landscapes that describes the germinal center. The quenched average  $\mathbb{E}[\lim_{t \rightarrow \infty} z_{\Gamma^*}(t)]$  is then equal to the fixation probability of class  $\Gamma^*$ , which can be computed as an extreme value problem:

$$\mathbb{E}[\lim_{t \rightarrow \infty} z_{\Gamma^*}(t)] = P_{fix, \Gamma^*} = m_{\Gamma^*} \int dx \rho_{\Gamma^*}(x) P_{\Gamma^*}(x)^{m_{\Gamma^*}-1} \prod_{\Gamma \neq \Gamma^*} P_\Gamma(x)^{m_\Gamma}, \quad (6)$$

where  $\rho_\Gamma(x)$  is the probability density function (p.d.f.) of the asymptotic growth rate of the hypercube “mass,” and  $P_\Gamma(x)$  is its cumulative density function (c.d.f.). The mass of hypercube  $\alpha$  is defined as  $N^\alpha(t) = \sum_i n_i^\alpha(t)$  and its asymptotic growth rate is

$$x^\alpha = \lim_{t \rightarrow \infty} \partial_t \log N^\alpha(t). \quad (7)$$

Given the linear nature of Eq. (2), the asymptotic growth rate of any hypercube mass is given by the ground state eigenvalue of the matrix  $H_{ij}^\alpha = -f_i^\alpha \delta_{ij} - \mu \Lambda_{ij}$  (corresponding to an Anderson Hamiltonian when the  $f_i$ ’s are independent identically distributed, or IID, variables). There is no general expression for the p.d.f. of the ground state eigenvalue of this random matrix, but we can derive approximate expressions using first order perturbation theory in two limiting regimes.

Without loss of generality, let us fix the mutation rate to  $\mu = 1/d$ : all growth rates  $f_i^\alpha$  are then measured in rescaled units, such that one mutation per unit time is expected. Let  $\sigma_\Gamma$  denote the spread of the fitness values for any class  $\Gamma$ : the two limiting regimes are obtained when  $d\sigma_\Gamma \ll 1$  (delocalized limit) or  $d\sigma_\Gamma \gg 1$  (localized limit). The corresponding p.d.f.s for the ground state eigenvalues are (see App. A):

$$\rho_\Gamma^{del}(x) = \int d\mathbf{f} \pi_\Gamma(\mathbf{f}) \delta\left(x - \frac{1}{N} \sum_{i=1}^N f_i\right); \quad (8)$$

$$\rho_\Gamma^{loc}(x) = \int d\mathbf{f} \pi_\Gamma(\mathbf{f}) \delta\left(x - \max_{i=1 \dots N} \{f_i\} + 1\right); \quad (9)$$

where  $\pi_\Gamma(\mathbf{f})$  is the distribution of the realized disordered fitness landscape  $\mathbf{f}$ . When  $d\sigma_\Gamma \ll 1$ , the asymptotic growth rate can be identified with the average fitness on the hypercube, thanks to the delocalized nature of the ground state eigenvector. This regime corresponds to a scenario in which competition within the same lineage is not strong. By contrast, a strong selection-weak mutation regime corresponds to the localized limit,  $d\sigma_\Gamma \gg 1$ , where, neglecting the constant offset, the asymptotic growth rate can be identified with the extremum of the hypercube fitness values, where the ground state eigenvector is localized.

#### 4. Antigen cocktails

To model affinity maturation in the context of vaccination or infection by an antigen cocktail, we must specify the fitness landscape imposed by individual antigens on any class of BCRs, as well as the rules by which these landscapes are combined when the antigens are mixed.

Let us focus on a simplified scenario, where only two antigens are included in the vaccine, in relative proportions  $c$  and  $1 - c$ . We further assume that the antigens are two variants of the same protein where we can distinguish, at a coarse-grained level, a mutated dominant epitope and a conserved subdominant epitope. The mutants are significantly distant in antigenic space, so that most of the antibodies generated after a primary immunization with one antigen may not neutralize the unseen antigen, as observed in the case of the SARS-CoV2 wild type strain and omicron variant [32, 33]. Hence, at a first approximation, cross-reactivity can only be achieved by targeting the conserved subdominant epitope.

In this scenario, at least three classes of antibodies must be introduced: the classes targeting the mutant epitopes on antigen 1 and antigen 2, respectively named  $A$  and  $B$ , and the class targeting the shared epitope ( $C$ ). Let  $f_i^{\Gamma,a}$  indicate the fitness value of a BCR belonging to class  $\Gamma \in \{A, B, C\}$  under exposure to antigen  $a \in \{1, 2\}$ . We can interpret  $f_i^{\Gamma,a} dt$  as the probability to have a replication event for the BCR  $i$  in a time interval  $dt$ , conditioned to the encounter of antigen  $a$ . Then the marginal probability of a replication event per unit time, in the presence of multiple antigens, reads:

$$f_i^\Gamma = c f_i^{\Gamma,1} + (1 - c) f_i^{\Gamma,2}. \quad (10)$$

This weighted average describes the effective fitness of the BCRs in the cocktail when all antigen types are presented homogeneously and abundantly on the follicular dendritic cells, so that each B cell, during its residency in the light zone, effectively samples their relative concentration. In the rest of this paper, we will work in this condition.

Similarly, the frequencies of activated precursors will be impacted by the composition of the cocktail. Let us assume that, for a fixed total amount of antigen, the

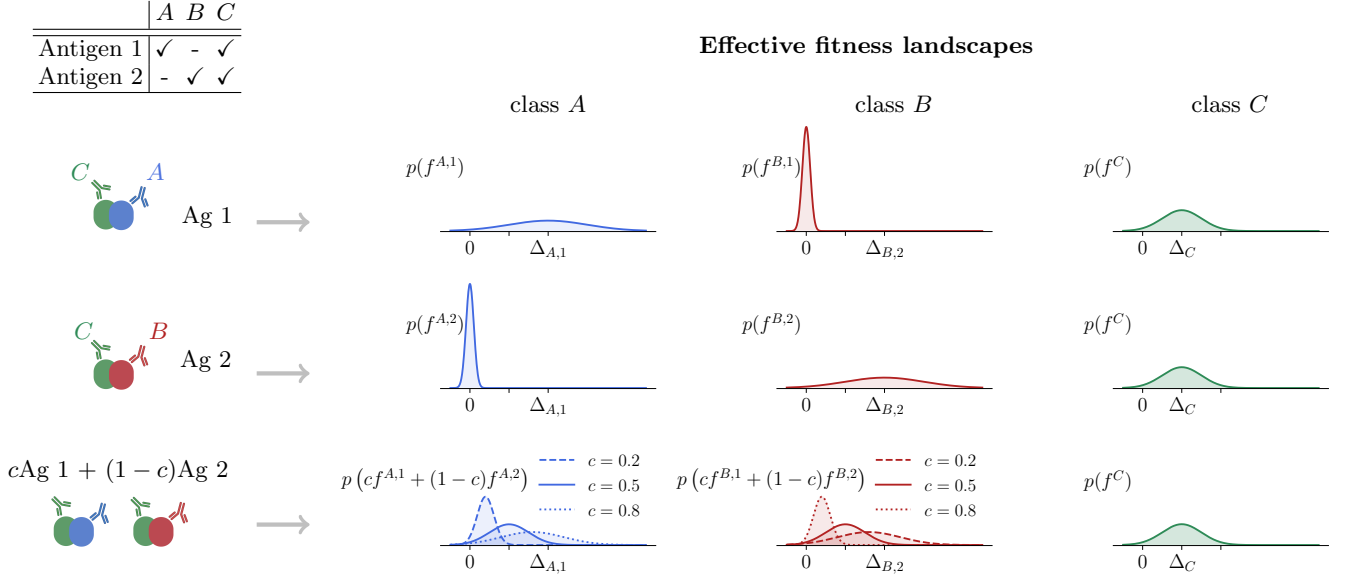


FIG. 2. The table summarizes the reactivity pattern of antigens and BCR classes. More precisely, reactivity is modeled by a Gaussian fitness distribution assigned to each antigen-BCR class pair. Non-reactive pairs have a narrow fitness distribution centered around zero, while reactive pairs have mostly positive fitness values. The effective “cocktail fitness landscape” is the weighted average of the fitness landscapes imposed by individual antigens, with weights equal to the relative antigen concentrations. The distribution of class  $C$  remains unaffected by  $c$  since the targeted epitope (in green) is the same on both antigens.

expected number of precursors entering the GC is fixed to  $M$ , but the number of precursors of each class  $m_\Gamma$  depends on the original abundance  $\nu_\Gamma$  of reactive B cells of class  $\Gamma$  in the repertoire and on the concentration of the targeted epitope. Imposing the condition  $\sum_\Gamma m_\Gamma = M$ , we have:

$$\frac{m_A}{M} = \frac{c\nu_A}{Z(c)}; \quad \frac{m_B}{M} = \frac{(1-c)\nu_B}{Z(c)}; \quad \frac{m_C}{M} = \frac{\nu_C}{Z(c)}; \quad (11)$$

where  $Z(c) = c\nu_A + (1-c)\nu_B + \nu_C$ .

For the sake of simplicity, let us assume that the fitness associated to each non-reactive antigen-BCR pair is exactly zero, while fitness values associated to reactive antigen-BCR pairs are mostly positive. In that case,

$$f_i^A = cf_i^{A,1}, \quad f_i^B = (1-c)f_i^{B,2}, \quad f_i^C = f_i^C, \quad (12)$$

where we assume that the fitness of the BCRs that target the unchanged epitope (class  $C$ ), is unaffected by the antigen background. The same prescription as Eq. (12) also applies to the effective asymptotic growth rates of the hypercubes’ mass. Therefore:

$$P_{fix,C} = \int dx \frac{\partial}{\partial x} P_C(x)^{m_C(c)} P_A\left(\frac{x}{c}\right)^{m_A(c)} P_B\left(\frac{x}{1-c}\right)^{m_B(c)}. \quad (13)$$

### III. THE RANDOM ENERGY MODEL

#### 1. Universality of immunodominance for REM landscapes

So far we described a flexible paradigm which can be adapted to any fitness landscape model. We must now specify what type of fitness landscape is associated to the hypercubes where the clonal lineages evolve. Several theoretical models have been proposed over the decades to effectively model the topography of real fitness land-

scapes [34–38] and, since the advent of high-throughput sequencing techniques, an increasing number of them has been empirically reconstructed [39].

In the case of affinity maturation, the fitness landscape is shaped by the presented antigen. In order to achieve reproducible antibody evolution, at least at the phenotypic level, the fitness of BCRs must be largely determined by their binding affinity to the antigen; a phenotype which has been long believed to have an almost-linear relation to the genotype [20, 40]. However, recent experimental



results have demonstrated that other random effects can contribute to determining the effective fitness landscape at our level of description, where stochastic sub-processes—like antigen capture from follicular dendritic cells, antigen presentation on the B cell surface, and the encounter and interaction with T-helper cells—are not resolved [41].

Motivated by these findings, by some evidence of epistasis in the binding affinity of antibodies to antigens [42–45], and for the sake of tractability, we study here a maximally epistatic model known as *House of Cards* or Random Energy Model (REM) [35, 36, 46, 47], where the fitness values associated to each node in the graph are independent random variables drawn from an identical distribution:  $\pi_\Gamma(\mathbf{f}) = \prod_i p_\Gamma(f_i)$ . Under this assumption, the asymptotic growth rate distributions in Eqs. (8)–(9) are guaranteed to converge to universal laws as  $N \rightarrow \infty$ .

Thanks to the generalized central limit theorem, the desired distribution in the delocalized regime reads

$$\rho_\Gamma^{\text{del}}(x) \approx h_{\alpha_\Gamma, \beta_\Gamma} \left( \frac{x - \mu_{\Gamma, N}}{\sigma_{\Gamma, N}} \right), \quad \text{for } N \gg 1, \quad (14)$$

where  $h_{\alpha, \beta}(z)$  is a stable distribution (for  $0 < \alpha \leq 2$ ,  $-1 \leq \beta \leq 1$ ), whose canonical representation is given in terms of its characteristic function [48]. When the parent fitness variable has a finite variance, Eq. (14) reduces to the standard central limit theorem, where  $\alpha = 2$  and  $\beta$  is irrelevant, and  $h_{2, \beta}(z)$  corresponds to the Gaussian distribution.

Similarly, it is known from the theory of extreme value

statistics that the c.d.f. of the maximum of a set of IID variables converges (for most distributions) to one of three types of functions, rewritten in compact form as:

$$\int_{-\infty}^x du \rho_\Gamma^{\text{loc}}(u) \approx G_{\gamma_\Gamma} \left( \frac{x - b_{\Gamma, N}}{a_{\Gamma, N}} \right), \quad \text{for } N \gg 1, \quad (15)$$

where

$$G_\gamma(z) = \begin{cases} e^{-(1+\gamma z)^{-1/\gamma}} & \gamma \neq 0, \ 1 + \gamma z \geq 0 \\ e^{-e^{-z}} & \gamma = 0 \end{cases}, \quad (16)$$

and  $b_{N, \Gamma}$  and  $a_{N, \Gamma}$  are, respectively, average and standard deviation of  $\max_{i=1 \dots N} \{f_i\}$  [49].

In summary, in both the localized and delocalized cases, the c.d.f. of the mass growth rate is of the form

$$P_\Gamma(x) \approx \Phi_{K_\Gamma} \left( \frac{x - \Delta_{\Gamma, N}}{\Sigma_{\Gamma, N}} \right), \quad (17)$$

where  $K_\Gamma$  is a set of parameters indicating the shape and skewness of the asymptotic growth rate distribution in the two considered regimes. Specifically,  $K_\Gamma = (\alpha, \beta)$  in the delocalized limit, for the stable distribution  $h_{\alpha, \beta}(x)$ ;  $K_\Gamma = \gamma$  in the localized limit, for the extreme value distribution  $G'_\gamma(x)$ . The two sets of parameters  $\Delta_{\Gamma, N}$  and  $\Sigma_{\Gamma, N}$  represent respectively the shift and scale parameters of the variable of interest (average or maximum); they will depend on  $N$  and on the parameters of the parent fitness distribution.

Using the functional form of Eq. (17) in Eq. (6), we obtain

$$P_{fix, \Gamma^*} \approx \int dx \frac{\partial}{\partial x} \left( \Phi_{K_{\Gamma^*}} \left( \frac{x - \Delta_{\Gamma^*, N}}{\Sigma_{\Gamma^*, N}} \right)^{m_{\Gamma^*}} \right) \prod_{\Gamma \neq \Gamma^*} \Phi_{K_\Gamma} \left( \frac{x - \Delta_{\Gamma, N}}{\Sigma_{\Gamma, N}} \right)^{m_\Gamma}, \quad N \gg 1. \quad (18)$$

Equation (18) shows that, thanks to universality, the fixation probability of each class depends only on a handful of parameters derived from the fitness distributions  $p_\Gamma(f)$ , and on the germline abundance of the different BCR classes. For any combination of these model parameters, we can use Eq. (18) to study the inter-class competition in the asymptotic time limit, and identify the class with the highest fixation probability as the dominant one. We can then immediately construct universal immunodominance phase diagrams in the joint parameter space of the limit distributions associated to the asymptotic growth rates of all the considered classes.

Let us first focus on the simplest, two-class problem, where only two distinct epitopes are considered, targeted by two classes of BCRs named  $V$  (variable) and  $C$  (conserved). We are interested in determining when each one of the epitopes is dominant or recessive. Plots of  $P_{fix, C}$

for some example combinations of universal distributions are shown in Fig. 3 as a function of the following dimensionless combinations of the distribution parameters [50]:

$$\tilde{\Delta}_N = \frac{\Delta_{V, N} - \Delta_{C, N}}{\Sigma_{C, N}} \quad \text{and} \quad \tilde{\Sigma}_N = \frac{\Sigma_{C, N}}{\Sigma_{V, N}}. \quad (19)$$

Assuming that the GC capacity  $M$  is constrained by the total antigen amount and resource availability, the variables of our problem are only  $\tilde{\Delta}_N$ ,  $\tilde{\Sigma}_N$ , and  $m_C/M$ . These parameters represent the axes of a three-dimensional immunogenic space, where different *phases* can be identified, as the immunodominant epitope switches from  $C$  to  $V$  (cf. Fig. 3).

Note that the parameters appearing in Eq. (18), or their dimensionless combinations in Eq. (19), all refer to the distributions of the asymptotic growth rates. The practical question is how to relate these parameters of

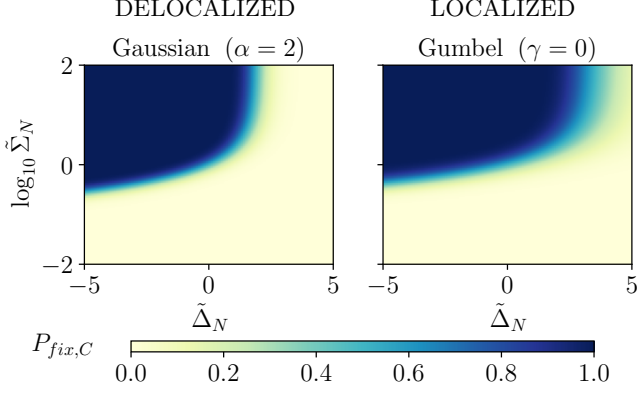


FIG. 3. Fixation probability for the two-class competition problem ( $C$  vs.  $V$ ), in the case of REM fitness landscapes, with asymptotic mass growth rate distributions of the two classes  $C$  and  $V$  belonging to the same universality class ( $K_C = K_V$  in Eq. (17)). Specifically, we take Gaussian limit stable laws ( $\alpha_C = \alpha_V = 2$ ) in the delocalized regime, and Gumbel extreme value distributions ( $\gamma_C = \gamma_V = 0$ ) in the localized regime. In this plot,  $m_V = m_C = 0.5M$ ,  $M = 50$ .

the universal distributions to the parameters of the parent fitness distributions  $p_\Gamma(f)$ ? To answer this question, it is useful to resort to derivations of the stable limit laws and extreme value distributions based on the renormalization group (RG) method. This idea has been repeatedly explored in the literature [51–57] and used to derive finite-size corrections to the limit distributions, as well as the asymptotic scaling of shift and scale parameters—corresponding to  $\Delta_{\Gamma,N}$  and  $\Sigma_{\Gamma,N}$  in our notation. A concise derivation of the flow equations in the space of probability density functions in both the localized and delocalized case is reported in Appendix B.

In the delocalized case we restrict to the Gaussian case; other cases refer to parent fitness distributions with diverging first or second moments, which are not biologically relevant. Identifying  $\log N = s$  (treated as a continuous variable for large  $N$ ) and rescaling the sums in such a way that first and second moments do not change with  $N$  (see Appendix B), we find the following flow equations for the parameters of the asymptotic mass growth rate:

$$\partial_s \Delta_\Gamma(s) = 0, \quad \partial_s \Sigma_\Gamma(s) = -\frac{1}{2} \Sigma_\Gamma(s). \quad (20)$$

Recalling the definition in Eq. (19) for the two-class problem,

$$\partial_s \tilde{\Delta}(s) = \frac{1}{2} \tilde{\Delta}(s), \quad \partial_s \tilde{\Sigma}(s) = 0. \quad (21)$$

In the localized regime,  $\Delta_{\Gamma,N}$  and  $\Sigma_{\Gamma,N}$  can be computed if we know the cumulative density function

$$F_\Gamma(z) = \int_{-\infty}^z df p_\Gamma(f) \equiv e^{-e^{-\varphi_\Gamma(z)}}, \quad (22)$$

via the set of equations (see Appendix B, [49]):

$$\Delta_\Gamma(s) = \varphi_\Gamma^{-1}(s); \quad \Sigma_\Gamma = \Delta'_\Gamma(s). \quad (23)$$

From Eqs. (23), the flow equations for  $\tilde{\Delta}(s)$  and  $\tilde{\Sigma}(s)$  read

$$\partial_s \tilde{\Delta}(s) = \frac{1}{\tilde{\Sigma}(s)} - 1 - \gamma_C(s) \tilde{\Delta}(s), \quad (24)$$

$$\partial_s \tilde{\Sigma}(s) = [\gamma_C(s) - \gamma_V(s)] \tilde{\Sigma}(s), \quad (25)$$

where  $\gamma_\Gamma(s) = \Sigma'_\Gamma(s)/\Sigma_\Gamma(s) \rightarrow \gamma_\Gamma$  as  $s \rightarrow \infty$ , for  $\Gamma \in \{C, V\}$ . Asymptotically,

$$\tilde{\Sigma}_N \approx \tilde{\Sigma}_0 N^{\gamma_C - \gamma_V}, \quad (26)$$

$$\tilde{\Delta}_N \approx \tilde{\Delta}_0 N^{-\gamma_C} - \frac{1 - N^{-\gamma_C}}{\gamma_C} + \frac{N^{-\gamma_C}}{\gamma_V \tilde{\Sigma}_0} (N^{\gamma_V} - 1), \quad (27)$$

where the initial conditions  $\tilde{\Sigma}_0 = \Sigma_C(0)/\Sigma_S(0)$  and  $\tilde{\Delta}_0 = (\Delta_S(0) - \Delta_C(0))/\Sigma_C(0)$  are defined after the following relations:

$$F_\Gamma(-\Sigma_\Gamma(0)/\Delta_\Gamma(0)) = F'_\Gamma(-\Sigma_\Gamma(0)/\Delta_\Gamma(0)) = e^{-1}. \quad (28)$$

The flow Eqs. (21), (24) and (25) can be exploited to construct an immunodominance phase diagram in the space of ‘bare’ control parameters  $\tilde{\Delta}_0, \tilde{\Sigma}_0$ —which can be directly reconstructed from the parent distribution  $p_\Gamma(f)$ —rather than  $\tilde{\Delta}_N, \tilde{\Sigma}_N$ . The idea is to identify from these equations the stable fixed points and their basins of attraction, and to associate to each basin the asymptotic value of  $P_{fix,C}$  computed at the corresponding fixed point.

Figure 4 depicts the flow equations and of the resulting immunodominance phase diagrams in some illustrative special cases. If we work in the delocalized regime and assume, as in Fig. 3, that the tails of the two parent fitness distributions are of the same type, so that the shape parameters associated to  $V$  and  $C$  converge to the same value at the same rate, then  $\tilde{\Sigma}_\infty$  is a constant and  $\tilde{\Delta}_\infty$  diverges, keeping the same sign as  $\tilde{\Delta}_0$ . Then we obtain only two asymptotic values for  $P_{fix,C}$ , respectively equal to 0 (for  $\tilde{\Delta}_0 > 0$ ) or 1 (for  $\tilde{\Delta}_0 < 0$ ). In the localized case, when  $\gamma_V(s) = \gamma_C(s) \rightarrow \gamma$ ,  $\tilde{\Sigma}$  is a constant and

$$\tilde{\Delta} = \frac{1}{\gamma} \left( 1/\tilde{\Sigma} - 1 \right) \quad (29)$$

identifies a line of fixed points, for  $\gamma \neq 0$ , whose stability is determined by the sign of  $\gamma$ . When  $\gamma < 0$  the points are unstable [58]: since  $\lim_{\tilde{\Delta} \rightarrow \infty} P_{fix,C} = 0$  and  $\lim_{\tilde{\Delta} \rightarrow -\infty} P_{fix,C} = 1$ , the line of fixed points becomes a transition line between a phase where the  $C$  epitope is dominant and a phase where it is subdominant.

Let us finally note that in the  $N \rightarrow \infty$  limit the parameter  $m_C/M$ , indicating the fraction of class  $C$  germ lines in the GC, becomes irrelevant: the phase diagram is thus the same for all values of  $m_C/M \in (0, 1)$ , as long as  $M \ll N$ . Finite size corrections may however be important: it is known that the slow convergence to the asymptotic distribution in the size  $N$  of the data set is

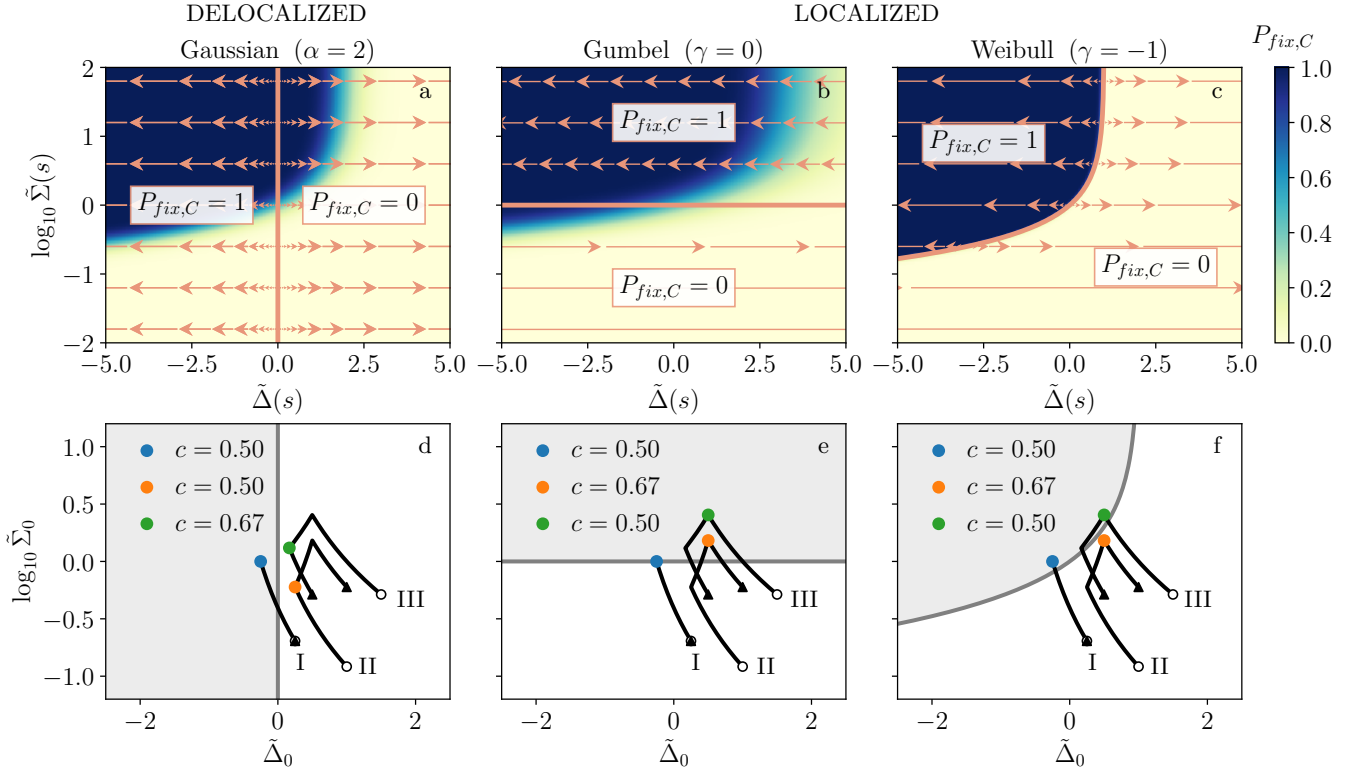


FIG. 4. **First row:** Identification of the basins of attraction of the RG flow equations for the fixed points  $\tilde{\Delta} = -\infty$  ( $P_{fix,C} = 1$ ) and  $\tilde{\Delta} = \infty$  ( $P_{fix,C} = 0$ ) in three example cases: **a** - Delocalized Gaussian case; **b** - Localized case, with  $\gamma_C = \gamma_V = 0$ ; **c** - Localized case, with  $\gamma_C = \gamma_V = -1$ . On the background, we show the fixation probability of class  $C$  in the space of asymptotic parameters  $\tilde{\Delta}_N, \tilde{\Sigma}_N$ , for  $M = 50$ ,  $m_C = m_V = 0.5M$ . In the foreground, solutions of the flow equations (21) (a) and (24)–(25) (b–c) are shown in orange. **Second row:** Parametric exploration of the immunogenic space along the lower bound in Eq. (30). The shaded region indicates the  $C$  immunodominant phase in the two-class setting. The black parametric curves are described by Eqs. (34)–(35), with parameters: (I) (symmetric case)  $\Delta_{A,1} = \Delta_{B,2} = \Sigma_C$ ;  $\Delta_C = 0.75\Sigma_C$ ;  $\Sigma_{A,1} = \Sigma_{B,2} = 2\Sigma_C$ ; (II) (asymmetric  $\Sigma$ )  $\Delta_{A,1} = \Delta_{B,2} = 1.5\Sigma_C$ ;  $\Delta_C = \Sigma_{A,1} = 2.5\Sigma_C$ ;  $\Sigma_{B,2} = 1.25\Sigma_C$ ; (III) (asymmetric  $\Delta$ )  $\Delta_{A,1} = \Sigma_C$ ;  $\Delta_{B,2} = 2\Sigma_C$ ;  $\Delta_C = 0.5\Sigma_C$ ;  $\Sigma_{A,1} = \Sigma_{B,2} = 1.33\Sigma_C$ . For each curve, the empty circle denotes the  $c = 0$  point, the filled triangle the  $c = 1$  point. **d** - Delocalized regime (Gaussian): here the transition line is vertical and the colored circles indicate the leftmost points along the parametric curves. These points correspond to  $c^* = 1/2$  when  $\Delta_{A,1} = \Delta_{B,2}$  and to  $c^* \neq 1/2$  when  $\Delta_{A,1} \neq \Delta_{B,2}$ . **e** - Localized regime,  $\gamma = 0$  (Gumbel): here the transition line is horizontal, and thus sensitive to imbalances of the kind  $\Sigma_{A,1} \neq \Sigma_{B,2}$ . **f** - Localized regime,  $\gamma = -1/2$  (Weibull): colored points represent here the farthest points from the transition line, in the immunodominant phase of epitope  $C$ .

problematic for the quantitative application of extreme value theory to real situations [59]. In our problem, finite size effects may become especially non-negligible if the effective size of the uncorrelated sequence space explored by each clonal lineage is not too big, as it can happen in the case of short-lasting affinity maturation or if we allow for strongly correlated fitness landscapes.

This analysis, which we illustrated for the two-class problem, can be extended to multiple classes of BCRs targeting an increasing number of distinct epitopes.

## 2. General principles for optimal antigen cocktail design

The immunogenic space construction introduced in the previous Section can be used to investigate under what conditions antigen cocktails can be used to invert the natural immunodominance relations between conserved and

highly mutable epitopes. Under the assumptions outlined in Sec. II 4, manipulating the relative antigen concentration  $c$  in a two-antigen cocktail describes a parametric exploration of the immunogenic space of the three-class problem, as described by the combination rules for the fitness landscapes of Eq. (12) and precursor frequencies of the various BCR classes in Eq. (11).

This space is higher-dimensional than the examples of Sec. III 1, but if we are only interested in the immunodominance of one epitope (e.g. the conserved epitope targeted by B cells of class  $C$ ), we can exploit a lower bound on  $P_{fix,C}^{3cl}(c)$  that casts the parametric three-class problem into an effective 2-class problem:

$$P_{fix,C}^{3cl}(c) \geq P_{fix,C}^{2cl}(\tilde{\Delta}^{eff}(c), \tilde{\Sigma}^{eff}(c), m_C(c)), \quad (30)$$

where



$$P_{fix,C}(c) \approx \int dx \frac{\partial}{\partial x} \left( \Phi_{K_C} \left( \frac{x - \Delta_{C,N}}{\Sigma_{C,N}} \right)^{m_C(c)} \right) \Phi_{K_A} \left( \frac{x - \Delta_{A,N}(c)}{\Sigma_{A,N}(c)} \right)^{m_A(c)} \Phi_{K_B} \left( \frac{x - \Delta_{B,N}(c)}{\Sigma_{B,N}(c)} \right)^{m_B(c)}, \quad (31)$$

with

$$\Delta_{A,N}(c) = c\Delta_N^{A,1}; \quad \Delta_{B,N}(c) = (1-c)\Delta_N^{B,2}; \quad (32)$$

$$\Sigma_{A,N}(c) = c\Sigma_N^{A,1}; \quad \Sigma_{B,N}(c) = (1-c)\Sigma_N^{B,2}. \quad (33)$$

In Eq. (30)  $P_{fix,C}^{2cl}$  is the fixation probability of  $C$  against a single effective class  $V$ , representing the most successful of  $A$  and  $B$ , with effective parameters

$$\tilde{\Delta}_N^{eff}(c) = \frac{\max\{c\Delta_N^{A,1}, (1-c)\Delta_N^{B,2}\} - \Delta_N^C}{\Sigma_N^C}; \quad (34)$$

$$\tilde{\Sigma}_N^{eff}(c) = \frac{\Sigma_N^C}{\max\{c\Sigma_N^{A,1}, (1-c)\Sigma_N^{B,2}\}}. \quad (35)$$

Let us note that Eqs. (30)–(35) are valid for any  $N$ , even though the bound may not be equally tight.

For the sake of simplicity, we work in the space of ‘bare’ parameters  $(\tilde{\Delta}_0, \tilde{\Sigma}_0)$ , where the immunodominance phase diagram exhibits a sharp transition. From a comparison between the shape of the parametric curves described by Eqs. (34)–(35) and that of the transition line in the  $N \rightarrow \infty$  limit, we can gain insight on several questions of interest, i.e.: given a pair of antigens, what is the optimal cocktail formulation that maximizes the production of class  $C$  B cells? And what are the conditions under which the optimized cocktail can make an epitope which is naturally immunorecessive de facto immunodominant?

As regards the optimization of the cocktail composition, an obvious solution exists if the system is symmetric under exchange of  $c$  and  $1-c$ , i.e. if  $p_{A,1}(f) = p_{B,2}(f)$  and  $\nu_A = \nu_B$ . In that case, the optimum corresponds to a balanced mixture of antigens ( $c^* = 1/2$ ) for any finite but large  $N$ , in both evolutionary regimes (see Appendix C). When the symmetry is broken, the optimal composition will deviate from the even mixture, in order to balance the competition exerted by classes  $A$  and  $B$  on class  $C$ . The specific value of  $c^*$  is only implicitly determined for finite  $N$  in this asymmetric case, but a qualitative analysis can be extracted from the study of the immunodominance phase diagrams in the  $N \rightarrow \infty$  limit. When  $\Delta_{A,1} \neq \Delta_{B,2}$  (curve III in Fig. 4 d–f), the asymmetry stretches the parametric lower bound curve in the horizontal direction: as a result, this asymmetry is best sensed in the delocalized case, where the transition line is perpendicular to that direction. Similarly, the localized Gumbel case exhibits a transition line in the horizontal direction, making the system most sensitive to asymmetries in the  $\Sigma$  parameters of classes  $A$  and  $B$ , which stretch the parametric lower bound curve in the vertical direction (curve II in Fig. 4 d–f).

A sufficient condition for the inversion of the immunodominance hierarchy is that the parametric curve of

Eqs. (34)–(35) crosses the manifold  $P_{fix,C}^{2cl} = 1/2$  for some  $c \in (0, 1)$ . Clearly there is a restricted range of combinations of the original parameter values for the three BCR classes such that this crossing can be achieved. In the delocalized limit, this condition is tied to the  $\Delta_\Gamma$  parameters of the parent single-antigen fitness distributions, while in the localized limit, for  $\gamma = 0$ , it is constrained by the  $\Sigma_\Gamma$  parameters. For values of the shape parameter  $\gamma < 0$ , the conditions for inversion look generally more intricate, unless classes  $A$  and  $B$  have symmetric distributions for antigens 1 and 2.

These general guidelines are robust to the specific details of the problem, as they only arise from the identification of asymmetries of epitope-specific fitness distributions and of the universality class of the problem. The bulk of the presented results is based on the assumption  $N \gg 1$  and on long-time asymptotics: we did not investigate in depth finite size or time corrections. However, we remark that finite time effects may be important in the strongly localized limit, where the effective size  $N$  of the explored sequence space is dramatically reduced. In such case, since finite size corrections also extend to the third axis of the immunogenic space —i.e. germline abundance—, cocktail optimization becomes sensitive to possible imbalances in the relative germline abundances, with greater sensitivity exhibited for smaller  $N$ .

#### IV. CONCLUSION

Understanding epitope immunodominance hierarchies is of paramount importance for developing universal vaccines against highly mutable pathogens and for studying pathogen coevolution in immunized hosts or populations. Using a coarse-grained model for affinity maturation and an asymptotic definition of immunodominance, we proposed here a simple framework to characterize immunodominance hierarchies from the statistical features (or rather *differences* in them) of epitope-specific fitness landscapes of germinal-center B cells.

We analyzed the paradigmatic case of Random Energy Model landscapes, showing that in this case, as the dimension of the genotypic space explored during affinity maturation increases, the details of the problem become less and less important and a form of universality emerges—at least in evolutionary regimes that are far from the localization transition. At long times, the impact of the precursor cell abundance vanishes (provided that the number of precursors per class remains sufficiently large for our deterministic approximation) and the average GC population becomes insensitive to the details of the fitness distributions of the various antibody classes. The fixation probability of any class, serving as a

proxy for the immunogenic advantage of the targeted epitope, ultimately depends on a small, universally defined set of parameters.

This set of parameters defines what we refer to as *immunogenic space*, where a phase transition occurs between immuno-dominant and immuno-recessive states for a given epitope. We believe that such construction of an immunodominance phase space, which can be extended to other fitness landscape models, can be a useful tool to visualize and characterize pathogen evolution or to determine under what general conditions the manipulation of vaccine compositions may invert the immunodominance relations between epitopes. While, to the best of our knowledge, comprehensive datasets enabling a systematic comparison of the binding affinity landscapes of BCRs across various classes are currently lacking, recent advances in epitope mapping and deep mutational scanning techniques indicate the potential for their acquisition and analysis [42, 60, 61].

In conclusion, we believe that, despite its extreme simplicity, the framework we have proposed presents a promising avenue for a deeper understanding, prediction, and manipulation of immunodominance relations within the context of affinity maturation. To fully realize its potential, an invaluable input would be high-throughput experimental data that capture without bias the coarse-grained statistical properties of epitope binding affinities.

## ACKNOWLEDGMENTS

We warmly thank Arup Chakraborty for suggesting the research problem and engaging in insightful discussions. MK acknowledges support from NSF grant DMR-2218849.

## Appendix A: First order perturbation theory for the ground state eigenvalue of the Anderson model

For each clonal lineage  $\alpha$ , we have from Eq. (2)

$$H_{ij}^\alpha = -f_i^\alpha \delta_{ij} - \frac{1}{d} \Lambda_{ij}, \quad \Lambda_{ij} = -d\delta_{ij} + A_{ij}, \quad (\text{A1})$$

where  $A_{ij}$  is the adjacency matrix of the  $d$ -dimensional hypercube. Without loss of generality, let us take a zero-mean random diagonal and designate by  $\sigma$  the scale parameter of the I.I.D. variables, however defined (e.g. standard deviation, when not diverging). Let us define  $\tilde{f}_i^\alpha = f_i^\alpha / \sigma$  a new random fitness variable with the same type of distribution but unit scale. The parameter that controls the Anderson localization transition is  $d\sigma = \epsilon$ , with  $\epsilon_c \sim O(1)$ . Two limiting regimes can be identified:

- $d\sigma = \epsilon \ll 1$  — delocalized regime. We can rewrite

$$H_{ij}^\alpha = \frac{1}{d} (\epsilon H_{ij}^1 + H_{ij}^0), \quad (\text{A2})$$

where  $H_{ij}^0 = -\Lambda_{ij}$  and  $H_{ij}^1 = -\tilde{f}_i^\alpha \delta_{ij}$ . Spectrum and eigenvectors of the unperturbed Hamiltonian are exactly known for the  $d$ -dimensional hypercube. However, at first order in perturbation theory we are only interested in the unperturbed ground state eigenvalue,  $\lambda_0^0 = 0$ , and the associated eigenvector,  $\mathbf{v}_0^0 = \frac{1}{\sqrt{N}}(1, 1, \dots, 1)$ , which are common to any graph Laplacian. Given the structure of  $\mathbf{v}_0^0$ , the first order correction for the asymptotic growth rate is,

$$x = -\lambda_0 \simeq -\frac{1}{d} (\lambda_0^0 + \epsilon \lambda_0^1) = \frac{1}{N} \sum_{i=1}^N f_i^\alpha, \quad (\text{A3})$$

from which we obtain the distribution  $\rho^{del}(x)$  in Eq. (8).

- $d\sigma = \epsilon \gg 1$  — localized regime. Let us rewrite:

$$H_{ij}^\alpha = \sigma (H_{ij}^0 + \epsilon^{-1} H_{ij}^1) \quad (\text{A4})$$

where the unperturbed Hamiltonian  $H_{ij}^0 = (-\tilde{f}_i^\alpha + \sigma^{-1}) \delta_{ij}$  is already diagonal, with ground state localized on the site with the maximum fitness. The perturbation is  $H_{ij}^1 = -A_{ij}$ . Since the adjacency matrix has null diagonal entries, at first order we have no correction to the ground state eigenvalue, leading to

$$x \simeq \max\{f_i^\alpha, i = 1, \dots, N\} - 1. \quad (\text{A5})$$

Both results hold true even when the random energy landscape is correlated; what changes, in such case, is only how the distributions of the empirical mean and of the maximum are computed.

## Appendix B: Renormalization group equations for sum and extreme value statistics of IID variables

For completeness, we present in this appendix a concise derivation of the renormalization group (RG) equations for the shift and scale parameters of the stable laws describing the sum and extreme value statistics of a large set of IID variables. The presentation is largely based on Refs. [49, 51].

Let us consider a set of  $N$  IID variables  $x_1, \dots, x_N$  with distribution  $p_x$ , characteristic function  $\phi_x$ , and cumulative  $F_x$ . The assumption underlying the RG construction is that, upon an  $N$ -dependent linear transformation of the variable of interest (sum or maximum), the p.d.f. of the transformed variable converges to a well-defined limit. Let us define:

$$\zeta_N = \frac{\sum_{i=1}^N x_i - b_N}{a_N}; \quad \xi_N = \frac{\max_{i=1 \dots N} x_i - b_N}{a_N}; \quad (\text{B1})$$

and correspondingly their characteristic and cumulative functions:

$$\phi_{\xi_N}(t) = \int d\zeta p_{\xi_N}(\zeta) e^{-it\zeta} = \phi_x \left( \frac{t}{a_N} \right)^N e^{ib_N \frac{t}{a_N}}; \quad (\text{B2})$$

$$F_{\xi_N}(z) = \int^z d\xi p_{\xi_N}(\xi) = F_x(a_N z + b_N)^N. \quad (\text{B3})$$

Since the maximum and the sum of a set of IID variables can be obtained iteratively, we can divide the set of  $N$  variables into  $p$  groups of equal size  $N' = N/p$  and compute the overall sum as the sum over the sums of each group, and the total maximum as the maximum of the maxima of the  $p$  groups. The same procedure can be iterated to compute sums and maxima within groups until we reach groups of size 1. Let us then introduce the flowing functions  $\phi(t, p)$  and  $F(z, p)$ , with  $p = 1 + \epsilon$  parametrizing the flow ( $\epsilon$  small):

$$\phi(t, p) = \phi_x \left( \frac{t}{a_p} \right)^p e^{ib_p \frac{t}{a_p}}; \quad (\text{B4})$$

$$F(z, p) = F_x(a_p z + b_p)^p. \quad (\text{B5})$$

For convenience, redefine  $s = \log p \approx \epsilon$ ,  $g(t, s) = \log \phi(t, e^s)$  and  $f(z, s) = -\log[-\log F(z, e^s)]$ , so that Eqs. (B4)–(B5) become

$$g(t, s) = e^s \left[ g_x \left( \frac{t}{a(s)} \right) + i \frac{t}{a(s)} e^{-s} b(s) \right]; \quad (\text{B6})$$

$$f(z, s) = f_x(a(s)z + b(s)) - s. \quad (\text{B7})$$

The RG equations for the sum and extreme value statistics are obtained by rewriting, respectively, Eqs. (B6) or (B7) as PDEs for  $g(t, s)$  or  $f(z, s)$  where  $g_x$  and  $f_x$  do not explicitly appear any more. This can be achieved by an appropriate manipulation of the partial derivatives of  $g(t, s)$  and  $f(z, s)$  with respect to their arguments. The resulting PDEs read

$$\partial_s g(t, s) = g(t, s) - t \frac{a'(s)}{a(s)} \partial_t g(t, s) - it \left[ \frac{b(s)}{a(s)} - \frac{b'(s)}{a(s)} \right]; \quad (\text{B8})$$

$$\partial_s f(z, s) = \left[ \frac{a'(s)}{a(s)} z + \frac{b'(s)}{a(s)} \right] \partial_z f(z, s) - 1. \quad (\text{B9})$$

Let us notice that, while  $f(z, s) \in \mathbb{R}$ ,  $g(t, s) \in \mathbb{C}$ , and Eq. (B8) must be read as a pair of equations for the real and imaginary part of the flowing function  $g(t, s) = u(t, s) + iv(t, s)$ . The shift and scale parameters,  $b(s)$  and  $a(s)$ , on the contrary, must be real.

Recall that the assumption behind the RG construction is the existence of an asymptotically stable fixed point for Eqs. (B8) and (B9), requiring that the  $s$ -dependent coefficients converge to constants as  $s \rightarrow \infty$ . Precisely, let us denote these constant limits of the coefficients of Eq. (B8) as

$$\frac{a'(s)}{a(s)} \rightarrow \alpha^{-1}, \quad \frac{b(s)}{a(s)} - \frac{b'(s)}{a(s)} \rightarrow \beta, \quad (\text{B10})$$

so that the invariant solution reads

$$g(t) = \begin{cases} C_1 t + i C_2 t + i \beta t \log|t| & \text{if } \alpha = 1, \\ i \frac{\alpha \beta}{\alpha - 1} t + |t|^\alpha \left( C_1 + i C_2 \frac{t}{|t|} \right) & \text{if } \alpha \neq 1, \end{cases} \quad (\text{B11})$$

where  $\alpha > 0$  and  $C_1, C_2 \in \mathbb{R}$  are arbitrary constants which will be fixed by the choice of suitable boundary conditions for the PDE (B8) —interpreted as physically invariant conditions that relate the parent distribution to the asymptotic stable law, as in traditional RG procedure— and from constraints coming from the support of the parent distribution  $p_x(x)$  [51].

From the specification of these boundary conditions, the flow equations for the scale and shift parameters  $a(s)$  and  $b(s)$  are also derived. Let us focus here on the case where the parent distribution has finite first and second moments (considerations about the scaling with  $s$  of  $a(s)$  and  $b(s)$  in the general case can be found in [51]). Let us firstly impose that, for any  $s$ , the first moment of the distribution reconstructed from  $\phi(t, s) = e^{g(t, s)}$  is equal to zero, i.e. that  $\partial_t u(t, s)|_{t=0} = \partial_t v(t, s)|_{t=0} = 0$ . Using this condition in Eq. (B8), we deduce

$$\beta(s) = \frac{b(s)}{a(s)} - \frac{b'(s)}{a(s)} = 0 \quad \forall s \iff b(s) = b(0)e^s, \quad (\text{B12})$$

where  $e^s$  can be identified with the number of variables of which we are computing the sum, and  $b(0)$  is the initial shift for the parent distribution, chosen in such a way that  $\partial_t v(t, s)|_{t=0} = 0$ . Trying to impose the same condition on Eq. (B11), it is clear that this is only possible for  $\alpha > 1$ . Let us now impose the boundary conditions  $\partial_{tt}^2 u(t, s)|_{t=0} = 1$ ,  $\partial_{tt}^2 v(t, s)|_{t=0} = 0$ , corresponding to fixing the second moment to be equal to 1. From Eq. (B8), this condition implies

$$\alpha(s) = 2 \quad \forall s \iff a(s) = a(0)(e^s)^{1/2}, \quad (\text{B13})$$

and fixes the values of the arbitrary constants to  $C_1 = 1$ ,  $C_2 = 0$ .

For the coefficients of Eq. (B9), let us denote the limits

$$\frac{a'(s)}{a(s)} \rightarrow \gamma, \quad \frac{b'(s)}{a(s)} \rightarrow \delta \quad (\text{B14})$$

and the corresponding invariant solution

$$f(z) = \frac{1}{\gamma} \log(\gamma z + \delta) + C, \quad (\text{B15})$$

for  $\gamma z + \delta > 0$ , where again  $C$  is an arbitrary constant fixed by the boundary conditions of the PDE (B9). Let us impose in this case that  $f(0, s) = 0 \quad \forall s$  and  $\partial_z f(z, s)|_{z=0} = 1 \quad \forall s$ , corresponding to the conditions  $F(0, s) = \partial_z F(z, s)|_{z=0} = e^{-1} \quad \forall s$ . From these conditions (using Eqs. (B9) and (B7)) we obtain

$$a(s) = b'(s), \quad b(s) = f_x^{-1}(s), \quad (\text{B16})$$

also implying  $\delta = 1$  and  $C = 0$ .

To obtain the flow equations for the dimensionless parameters defining the axes of the immunogenic phase space, it is sufficient to identify  $\Delta_{\Gamma,N}, \Sigma_{\Gamma,N}$  with  $b(s), a(s)$  in the localized/extreme value regime, and with  $b(s)e^{-s}, a(s)e^{-s}$  in the localized/average regime ( $s = \log N$ ).

### Appendix C: Symmetric optimization

When  $P_{fix,C}(c) = P_{fix,C}(1-c)$ , the symmetry of the problem imposes that  $c = 1/2$  be a stationary point.

That this stationary point is also a maximum and no spontaneous symmetry breaking occurs must be proved by the concavity of this function.

Let us rewrite Eq. (6) for the symmetric case:

$$P_{fix,C}(c) = m_C M \int dx \rho_C(x) P_C(x)^{m_C M - 1} P_A(x; c)^{m_A(c)M} P_B(x; c)^{m_B(c)M}, \quad (C1)$$

where  $m_C$  is independent of  $c$ , and  $m_A = \nu m_C$ ,  $m_B = (1-c)\nu m_C$ , with  $\nu = \nu_A/\nu_C = \nu_B/\nu_C$ . Because of the symmetry of the effective cocktail fitness distributions,

$$p_A(f; c) = p_B(f; 1-c) \implies P_A(x; c) = P_B(x; 1-c), \quad (C2)$$

independently of the regime in which affinity maturation occurs. Working in the same setting as in the main text, where  $f^A = cf^{A,1}$  and  $f^B = (1-c)f^{B,2}$ , meaning that replication of a BCR is conditioned to the encounter of the reactive epitope, we can deduce that  $x^A/c$  and  $x^B/(1-c)$  are equal in probability. Here  $x^A$  and  $x^B$  represent the asymptotic mass growth rates of hypercubes in class  $A$  and  $B$ . For all the considered limiting regimes, the transformations that map  $f$  to  $x$  are indeed linear—extremum for localized regime at asymptotic times; identity for localized regime at early times; average for delocalized regime—, allowing us to rewrite the cumulative distribution of growth rates for classes  $A$  and  $B$  in terms of a reference distribution for the variable  $\hat{x} = x^A/c$  or  $\hat{x} = x^B/(1-c)$ , independent of  $c$ :

$$P_A(x; c) = \hat{P}\left(\frac{x}{c}\right); \quad P_B(x; c) = \hat{P}\left(\frac{x}{1-c}\right). \quad (C3)$$

We can now recognize that  $\hat{P}(x)^{\nu m_C M} \equiv P(x)$  is the

cumulative distribution of the maximum of a sequence of  $\nu m_C M$  I.I.D. variables distributed as  $\hat{x}$ . Analogously,  $m_C M \rho_C(x) P_C(x)^{m_C M - 1} \equiv \tilde{\rho}(x)$  is the p.d.f. of the maximum of  $m_C M$  independent variables distributed as  $x^C$ . Therefore

$$P_{fix,C}(c) = \int dx \tilde{\rho}(x) P\left(\frac{x}{c}\right)^c P\left(\frac{x}{1-c}\right)^{1-c}. \quad (C4)$$

It is not guaranteed, for any p.d.f.  $\tilde{\rho}$  and any c.d.f.  $P$ , that (C4) is a concave function at  $c^* = 1/2$ . When the fixation probability is twice differentiable w.r.t.  $c$ , the general condition reads  $\frac{\partial^2 P_{fix,C}}{\partial c^2} \Big|_{c^*=1/2} \leq 0$ , with

$$\frac{\partial^2 P_{fix,C}}{\partial c^2} \Big|_{c^*=1/2} = \int dx \tilde{\rho}(x) 16x^2 P(2x) \frac{\partial^2 \log P(z)}{\partial z^2} \Big|_{z=2x}. \quad (C5)$$

Nonetheless, if we identify  $\hat{P}(a_N z + b_N)$  with  $G_\gamma(z)$  in Eq. (16), the concavity condition can be easily proved. In any of the three cases,  $\gamma = 0$ ,  $\gamma < 0$  or  $\gamma > 0$ , the former identification implies that  $P(a_N \nu m_C M z + b_N \nu m_C M)$  can also be identified with  $G_\gamma(z)$ . Hence we obtain, with a simple change of variables:

$$\frac{\partial^2 P_{fix,C}}{\partial c^2} \Big|_{c^*=1/2} = \int dx \tilde{\rho}_N(x) 16 \left(x + \frac{b_N}{2}\right)^2 \left(G_\gamma(z) \frac{\partial^2 \log G_\gamma(z)}{\partial z^2}\right) \Big|_{z=2x}, \quad \text{with } \tilde{\rho}_N(x) = \int ds \tilde{\rho}(s) \delta\left(x - \frac{s - b_N}{a_N}\right). \quad (C6)$$

We see from Eq. (16) that  $G_\gamma(z)$  is not differentiable in  $\mathbb{R}$  when  $\gamma \neq 0$ ; however we can rewrite in this case

$$G_\gamma(z) = \begin{cases} e^{-(1+\gamma z)^{-1/\gamma}} \Theta(1+\gamma z), & \gamma > 0 \\ e^{-(1+\gamma z)^{-1/\gamma}} \Theta(1+\gamma z) + \Theta(-1-\gamma z), & \gamma < 0 \end{cases} \implies \log G_\gamma(z) = -(1+\gamma z)^{-1/\gamma} \Theta(1+\gamma z), \quad \text{for } \gamma \neq 0, \quad (C7)$$

and approximate the Heaviside functions with smooth sigmoidal functions, e.g.  $\Theta_l(s) = \frac{1}{\pi} \tan^{-1}(s/l) + 1/2$ , such that  $\Theta(s) = \lim_{l \rightarrow 0} \Theta_l(s)$ . As a result,

$$\frac{\partial^2 P_{fix,C}}{\partial c^2} \Big|_{c^* = \frac{1}{2}} = \lim_{l \rightarrow 0} \begin{cases} - \int dx \tilde{\rho}_N(x) 16 \left(x + \frac{b_N}{2}\right)^2 e^{-2x - e^{-2x}}, & \gamma = 0; \\ - \int dx \tilde{\rho}_N(x) 16 \left(x + \frac{b_N}{2}\right)^2 G_\gamma(2x) (1 + 2\gamma x)^{-\frac{1}{\gamma}} \left[ (1 + \gamma) (1 + 2\gamma x)^{-2} \Theta_l(1 + 2\gamma x) - \right. \\ \left. 2\gamma (1 + 2\gamma x)^{-1} \Theta'_l(1 + 2\gamma x) + \gamma^2 \Theta''_l(1 + 2\gamma x) \right], & \gamma \neq 0. \end{cases} \quad (C8)$$

When  $\gamma = 0$   $P_{fix,C}$  has a negative well-defined second derivative, which guarantees that  $c^* = 1/2$  is a smooth maximum. On the contrary, when  $\gamma \neq 0$ ,  $P_{fix,C}(c)$  can have a kink at  $c^* = 1/2$ . By definition,  $\lim_{l \rightarrow 0} \Theta_l(s) = \Theta(s)$ , and  $\lim_{l \rightarrow 0} \Theta'_l(s) = \delta(s)$ . Hence the first term in the bracket has the sign of  $1 + \gamma$ . This term is finite when  $\gamma > 0$ , while it can diverge for  $\gamma < 0$ —depending on whether and how fast  $\tilde{\rho}_N(s)$  converges to 0 for  $s \rightarrow -1/2\gamma$ . The second term is null for any  $\gamma > -1$ . In the third term,  $\Theta''_l$  formally converges to the derivative of a Dirac's delta function, which yields again a null integral for any  $\gamma > -1$ . In conclusion,  $\frac{\partial^2 P_{fix,C}}{\partial c^2} \Big|_{c^* = \frac{1}{2}} \leq 0$  for any  $\gamma > -1$ , and it possibly diverges for  $-1 < \gamma < 0$ .

One can also make a hand-wavy argument for the concavity of  $P_{fix,C}(c)$  in the delocalized regime: in the limit

$N \gg 1$ ,  $\hat{P}(x)$  is a monotonically increasing function that varies very steeply between 0 and 1. Thus we can roughly approximate  $\hat{P}(x) \simeq \theta(x - x^*)$ : if  $\nu m_C M \ll N$ , this implies  $P(x) \simeq \theta(x - x^*)$ , at the same order of approximation. Therefore

$$P_{fix,C}(c) \simeq \int_{\max\{cx^*, (1-c)x^*\}} \tilde{\rho}(x), \quad (C9)$$

implying  $\arg \max_c P_{fix,C}(c) = \arg \min_c \max\{cx^*, (1-c)x^*\} = 1/2 \forall x^* \in \mathbb{R}_0$ . The same line of reasoning can be applied to cases where we have more than two BCR classes competing with the cross-reactive one, which target equally immunogenic variants of the variable epitope. In this scenario, the optimum  $c^* = \arg \min_{c: \sum_i c_i = 1} \max\{c_i x^*\}_{i=1, \dots, E}$  still corresponds to the balanced cocktail.

- 
- [1] A. J. Charles, P. Travers, M. Walport, and M. J. Shlomchik, *Immunobiology: The Immune System in Health and Disease. 5th edition* (Garland Science, New York, 2001).
  - [2] L. Potocnakova, M. Bhide, and L. B. Pulzova, An introduction to b-cell epitope mapping and in silico epitope prediction, *Journal of Immunology Research* 10.1155/2016/6760830 (2016).
  - [3] K. M. Hastie, H. Li, D. Bedinger, S. L. Schendel, S. M. Dennison, K. Li, V. Rayaprolu, X. Yu, C. Mann, M. Zandonatti, R. D. Avalos, D. Zyla, T. Buck, S. Hui, K. Shaffer, C. Hariharan, J. Yin, E. Olmedillas, A. Enriquez, D. Parekh, M. Abrahama, E. Feeney, G. Q. Horn, C.-D. team1, Y. Aldon, H. Ali, S. Aracic, R. R. Cobb, R. S. Federman, J. M. Fernandez, J. Glanville, R. Green, G. Grigoryan, A. G. L. Hernandez, D. D. Ho, K.-Y. A. Huang, J. Ingraham, W. Jiang, P. Kellam, C. Kim, M. Kim, H. M. Kim, C. Kong, S. J. Krebs, F. Lan, G. Lang, S. Lee, C. L. Leung, J. Liu, Y. Lu, A. MacCamy, A. T. McGuire, A. L. Palser, T. H. Rabbitts, Z. R. Tehrani, M. M. Sajadi, R. W. Sanders, A. K. Sato, L. Schweizer, J. Seo, B. Shen, J. L. Snitselaar, L. Stamatatos, Y. Tan, M. T. Tomic, M. J. van Gils, S. Youssef, J. Yu, T. Z. Yuan, Q. Zhang, B. Peters, G. D. Tomaras, T. Germann, and E. O. Saphire, Defining variant-resistant epitopes targeted by sars-cov-2 antibodies: A global consortium study, *Science* **374**, 472 (2021), <https://www.science.org/doi/pdf/10.1126/science.abh2315>.
  - [4] G. D. Vitoria and M. C. Nussenzweig, Germinal centers, *Annual Review of Immunology* **40**, 413 (2022), pMID: 35113731.
  - [5] M. O. Altman, D. Angeletti, and J. W. Yewdell, Antibody immunodominance: The key to understanding influenza virus antigenic drift., *Viral immunology* **31**, 10.1089/vim.2017.0129 (2018).
  - [6] M. Yuan, D. Huang, C.-C. D. Lee, N. C. Wu, A. M. Jackson, X. Zhu, H. Liu, L. Peng, M. J. van Gils, R. W. Sanders, D. R. Burton, S. M. Reincke, H. Prüss, J. Kreye, D. Nemazee, A. B. Ward, and I. A. Wilson, Structural and functional ramifications of antigenic drift in recent sars-cov-2 variants, *Science* **373**, 818 (2021), <https://www.science.org/doi/pdf/10.1126/science.abh1139>.
  - [7] A. Amitai, L. Mesin, G. D. Vitoria, M. Kardar, and A. K. Chakraborty, A population dynamics model for clonal diversity in a germinal center, *Frontiers in Microbiology* **8**, 10.3389/fmicb.2017.01693 (2017).
  - [8] G. Altan-Bonnet, T. Mora, and A. M. Walczak, Quantitative immunology for physicists, *Physics Reports* **849**, 1 (2020), quantitative immunology for physicists.
  - [9] M. O. Altman, J. R. Bennink, J. W. Yewdell, and B. R. Herrin, Lamprey vlrp response to influenza virus supports universal rules of immunogenicity and antigenicity, *eLife* **4**, e07467 (2015).
  - [10] M. Oprea and A. S. Perelson, Somatic mutation leads to efficient affinity maturation when centrocytes recycle back to centroblasts., *Journal of immunology* **158** **11**, 5155 (1997).
  - [11] S. Wang, J. Mata-Fink, B. Kriegsman, M. Hanson, D. J. Irvine, H. N. Eisen, D. R. Burton, K. D. Wittrup, M. Kardar, and A. K. Chakraborty, Manipulating the selection forces during affinity maturation to generate cross-reactive hiv antibodies, *Cell* **160**, 10.1016/j.cell.2015.01.027 (2015).



- [12] M. Molari, K. Eyer, J. Baudry, S. Cocco, and R. Monas-son, Quantitative modeling of the effect of antigen dosage on b-cell affinity distributions in maturing germinal centers, *eLife* **9**, e55678 (2020).
- [13] L. Yang, M. Van Beek, Z. Wang, F. Muecksch, M. Canis, T. Hatzioannou, P. D. Bieniasz, M. C. Nussenzweig, and A. K. Chakraborty, Antigen presentation dynamics shape the antibody response to variants like sars-cov-2 omicron after multiple vaccinations with the original strain, *Cell Reports* **42**, 10.1016/j.celrep.2023.112256 (2023).
- [14] K. G. Sprenger, J. E. Louveau, P. M. Murugan, and A. K. Chakraborty, Optimizing immunization protocols to elicit broadly neutralizing antibodies, *Proceedings of the National Academy of Sciences* **117**, 20077 (2020), <https://www.pnas.org/doi/pdf/10.1073/pnas.1919329117>.
- [15] L. Yang, T. M. Caradonna, A. G. Schmidt, and A. K. Chakraborty, Mechanisms that promote the evolution of cross-reactive antibodies upon vaccination with designed influenza immunogens, *Cell Reports* **42**, 10.1016/j.celrep.2023.112160 (2023).
- [16] C. O. Barnes, C. A. Jette, M. E. Abernathy, K.-M. A. Dam, S. R. Esswein, H. B. Gristick, A. G. Malyshev, N. G. Sharaf, K. E. Huey-Tubman, Y. E. Lee, D. F. Robbiani, M. C. Nussenzweig, A. P. West, and P. J. Bjorkman, Sars-cov-2 neutralizing antibody structures inform therapeutic strategies, *Nature* **588**, 682 (2020).
- [17] S. Chalkias, C. Harper, K. Vrbicky, S. R. Walsh, B. Essink, A. Brosz, N. McGhee, J. E. Tomassini, X. Chen, Y. Chang, A. Sutherland, D. C. Montefiori, B. Girard, D. K. Edwards, J. Feng, H. Zhou, L. R. Baden, J. M. Miller, and R. Das, A bivalent omicron-containing booster vaccine against covid-19, *New England Journal of Medicine* **387**, 1279 (2022).
- [18] S. M. Scheaffer, D. Lee, B. Whitener, B. Ying, K. Wu, C.-Y. Liang, H. Jani, P. Martin, N. J. Amato, L. E. Avena, D. M. Berrueta, S. D. Schmidt, S. O'Dell, A. Nasir, G.-Y. Chuang, G. Stewart-Jones, R. A. Koup, N. A. Doria-Rose, A. Carfi, S. M. Elbashir, L. B. Thackray, D. K. Edwards, and M. S. Diamond, Bivalent sars-cov-2 mrna vaccines increase breadth of neutralization and protect against the ba.5 omicron variant in mice, *Nature Medicine* **29**, 247 (2023).
- [19] G. D. Victora and M. C. Nussenzweig, Germinal centers, *Annual Review of Immunology* **30**, 429 (2012), PMID: 22224772.
- [20] L. Mesin, J. Ersching, and G. D. Victora, Germinal center b cell dynamics, *Immunity* **45**, 471 (2016).
- [21] R. Cerf and J. Dalmau, *The Quasispecies Equation and Classical Population Models*, Probability Theory and Stochastic Modelling (Springer Cham, 2022).
- [22] This simplifying assumption overlooks the current understanding of the statistics of hypermutations, which have been observed to be far from uniform and to result into enrichment of specific nucleotide motifs [8, 23, 62, 63]. Nonetheless, we believe that a more realistic depiction of the mutational network will not significantly change our qualitative results and is not considered in this work.
- [23] N. Spisak, A. M. Walczak, and T. Mora, Learning the heterogeneous hypermutation landscape of immunoglobulins from high-throughput repertoire data, *Nucleic Acids Research* **48**, 10702 (2020).
- [24] K. G. C. Smith, A. Light, G. J. V. Nossal, and D. M. Tarlinton, The extent of affinity maturation differs between the memory and antibody-forming cell compartments in the primary immune response, *The EMBO Journal* **16**, 2996 (1997), <https://www.embopress.org/doi/pdf/10.1093/emboj/16.11.2996>.
- [25] W. Dejnirattisai, D. Zhou, H. M. Ginn, H. M. Duyvesteyn, P. Supasa, J. B. Case, Y. Zhao, T. S. Walter, A. J. Mentzer, C. Liu, B. Wang, G. C. Paesen, J. Slon-Campos, C. López-Camacho, N. M. Kafai, A. L. Bailey, R. E. Chen, B. Ying, C. Thompson, J. Bolton, A. Fyfe, S. Gupta, T. K. Tan, J. Gilbert-Jaramillo, W. James, M. Knight, M. W. Carroll, D. Skelly, C. Dold, Y. Peng, R. Levin, T. Dong, A. J. Pollard, J. C. Knight, P. Klennerman, N. Temperton, D. R. Hall, M. A. Williams, N. G. Paterson, F. K. Bertram, C. A. Siebert, D. K. Clare, A. Howe, J. Radecke, Y. Song, A. R. Townsend, K.-Y. A. Huang, E. E. Fry, J. Mongkolsapaya, M. S. Diamond, J. Ren, D. I. Stuart, and G. R. Screaton, The antigenic anatomy of sars-cov-2 receptor binding domain, *Cell* **184**, 2183 (2021).
- [26] We note that the class definition is somewhat arbitrary: for an example of experimental methods used to identify antibody epitope classes, see [16] or [3].
- [27] We make here the simplifying assumption that  $d$  is constant.
- [28] W. König, *The Parabolic Anderson Model: Random Walk in Random Potential*, Pathways in Mathematics (Birkhäuser Cham, 2016).
- [29] L. Avena, O. Gün, and M. Hesse, The parabolic anderson model on the hypercube, *Stochastic Processes and their Applications* **130**, 3369 (2020).
- [30] J. M. J. Tas, L. Mesin, G. Pasqual, S. Targ, J. T. Jacobsen, Y. M. Mano, C. S. Chen, J.-C. Weill, C.-A. Reynaud, E. P. Browne, M. Meyer-Hermann, and G. D. Victora, Visualizing antibody affinity maturation in germinal centers, *Science* **351**, 1048 (2016), <https://www.science.org/doi/pdf/10.1126/science.aad3439>.
- [31] J. Lee, D. R. Boutz, V. Chromikova, M. G. Joyce, C. Vollmers, K. Leung, A. P. Horton, B. J. DeKosky, C.-H. Lee, J. J. Lavinder, E. M. Murrin, C. Chrysostomou, K. H. Hoi, Y. Tsybovsky, P. V. Thomas, A. Druz, B. Zhang, Y. Zhang, L. Wang, W.-P. Kong, D. Park, L. I. Popova, C. L. Dekker, M. M. Davis, C. E. Carter, T. M. Ross, A. D. Ellington, P. C. Wilson, E. M. Marcotte, J. R. Mascola, G. C. Ippolito, F. Krammer, S. R. Quake, P. D. Kwong, and G. Georgiou, Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination, *Nature Medicine* **22**, 1456 (2016).
- [32] Y.-F. Hu, J.-C. Hu, H.-R. Gong, A. Danchin, R. Sun, H. Chu, I. F.-N. Hung, K. Y. Yuen, K. K.-W. To, B.-Z. Zhang, T. Yau, and J.-D. Huang, Computation of antigenicity predicts sars-cov-2 vaccine breakthrough variants, *Frontiers in Immunology* **13**, 10.3389/fimmu.2022.861050 (2022).
- [33] A. J. Greaney, T. N. Starr, P. Gilchuk, S. J. Zost, E. Binshtein, A. N. Loes, S. K. Hilton, J. Huddleston, R. Eguia, K. H. Crawford, A. S. Dingens, R. S. Nargi, R. E. Sutton, N. Suryadevara, P. W. Rothlauf, Z. Liu, S. P. Whelan, R. H. Carnahan, J. Crowe, James E., and J. D. Bloom, Complete mapping of mutations to the sars-cov-2 spike receptor-binding domain that escape antibody recognition, *Cell Host & Microbe* **29**, 44 (2021).
- [34] J. A. G. de Visser and J. Krug, Empirical fitness landscapes and the predictability of evolution, *Nature Reviews Genetics* **15**, 480 (2014).

- [35] J. F. C. Kingman, A simple model for the balance between selection and mutation, *Journal of Applied Probability* **15**, 1 (1978).
- [36] S. Franz, L. Peliti, and M. Sellitto, An evolutionary version of the random energy model, *Journal of Physics A: Mathematical and General* **26**, L1195 (1993).
- [37] S. A. Kauffman and E. D. Weinberger, The nk model of rugged fitness landscapes and its application to maturation of the immune response, *Journal of Theoretical Biology* **141**, 211 (1989).
- [38] T. Aita, H. Uchiyama, T. Inaoka, M. Nakajima, T. Kokubo, and Y. Husimi, Analysis of a local fitness landscape with a model of the rough mt. fuji-type landscape: Application to prolyl endopeptidase and thermolysin, *Biopolymers* **54**, 64 (2000).
- [39] C. Blanco, E. Janzen, A. Pressman, R. Saha, and I. A. Chen, Molecular fitness landscapes from high-coverage sequence profiling, *Annual Review of Biophysics* **48**, 1 (2019), pMID: 30601678, <https://doi.org/10.1146/annurev-biophys-052118-115333>.
- [40] C. Berek and C. Milstein, Mutation drift and repertoire shift in the maturation of the immune response, *Immunological Reviews* **96**, 23 (1987), <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-065X.1987.tb00507.x>.
- [41] G. D. Victora, T. A. Schwickert, D. R. Fooksman, A. O. Kamphorst, M. Meyer-Hermann, M. L. Dustin, and M. C. Nussenzweig, Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter, *Cell* **143**, 10.1016/j.cell.2010.10.032 (2010/11/12).
- [42] A. M. Phillips, K. R. Lawrence, A. Moulana, T. Dupic, J. Chang, M. S. Johnson, I. Cvijovic, T. Mora, A. M. Walczak, and M. M. Desai, Binding affinity landscapes constrain the evolution of broadly neutralizing anti-influenza antibodies, *eLife* **10**, e71393 (2021).
- [43] A. M. Phillips, D. P. Maurer, C. Brooks, T. Dupic, A. G. Schmidt, and M. M. Desai, Hierarchical sequence-affinity landscapes shape the evolution of breadth in an anti-influenza receptor binding site antibody, *eLife* **12**, e83628 (2023).
- [44] R. M. Adams, J. B. Kinney, A. M. Walczak, and T. Mora, Epistasis in a fitness landscape defined by antibody-antigen binding free energy, *Cell Systems* **8**, 10.1016/j.cels.2018.12.004 (2019).
- [45] M. C. Julian, L. Li, S. Garde, R. Wilen, and P. M. Tessier, Efficient affinity maturation of antibody variable domains requires co-selection of compensatory mutations to maintain thermodynamic stability, *Scientific Reports* **7**, 10.1038/srep45259 (2017).
- [46] B. Derrida, Random-energy model: Limit of a family of disordered models, *Phys. Rev. Lett.* **45**, 79 (1980).
- [47] B. Derrida, Random-energy model: An exactly solvable model of disordered systems, *Phys. Rev. B* **24**, 2613 (1981).
- [48] B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sums of independent random variables*, Addison-Wesley Mathematics Series (Addison-Wesley, 1954).
- [49] J.-Y. Fortin and M. Clusel, Applications of extreme value statistics in physics, *Journal of Physics A: Mathematical and Theoretical* **48**, 183001 (2015).
- [50] We choose to define  $\tilde{\Delta}_N$  in a non-symmetric way for reasons that will be clear in the study of the parametric 3-class problem, Sec. III 2.
- [51] A. Amir, An elementary renormalization-group approach to the generalized central limit theorem and extreme value distributions, *Journal of Statistical Mechanics: Theory and Experiment* **2020**, 013214 (2020).
- [52] G. Jona-Lasinio, The renormalization group: A probabilistic view, *Il Nuovo Cimento* **26 B**, 99 (1975).
- [53] G. Jona-Lasinio, Renormalization group and probability theory, *Physics Reports* **352**, 439 (2001), renormalization group theory in the new millennium. III.
- [54] D. Li and Y. G. Sinai, An application of the renormalization group method to stable limit laws, *Journal of Statistical Physics* **157**, 10.1007/s10955-014-1098-4 (2014).
- [55] G. Györgyi, N. R. Moloney, K. Ozogány, and Z. Rácz, Finite-size scaling in extreme statistics, *Phys. Rev. Lett.* **100**, 210601 (2008).
- [56] G. Györgyi, N. R. Moloney, K. Ozogány, Z. Rácz, and M. Droz, Renormalization-group theory for finite-size scaling in extreme statistics, *Phys. Rev. E* **81**, 041135 (2010).
- [57] I. Calvo, J. C. Cuchí, J. G. Esteve, and F. Falceto, Extreme-value distributions and renormalization group, *Phys. Rev. E* **86**, 041109 (2012).
- [58] One can reconstruct the phase diagram  $P_{fix,C}$  in the plane of control parameters  $(\tilde{\Delta}_0, \tilde{\Sigma}_0)$  when  $\gamma_V(s) = \gamma_C(s) \rightarrow \gamma > 0$ , but it will not exhibit a sharp phase transition. Even in the  $N \rightarrow \infty$  limit,  $P_{fix,C}$  will be a smooth function of  $\tilde{\Sigma}_0$ , which must be reconstructed by calculating Eq. (18) along the fixed point line.
- [59] R. A. Fisher and L. H. C. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Mathematical Proceedings of the Cambridge Philosophical Society* **24**, 180–190 (1928).
- [60] E. Engelhart, R. Emerson, L. Shing, C. Lennartz, D. Guion, M. Kelley, C. Lin, R. Lopez, D. Younger, and M. E. Walsh, A dataset comprised of binding interactions for 104,972 antibodies against a sars-cov-2 peptide, *Scientific Data* **9**, 10.1038/s41597-022-01779-4 (2022).
- [61] Q. He, L. Wu, Z. Xu, X. Wang, Y. Xie, Y. Chai, A. Zheng, J. Zhou, S. Qiao, M. Huang, G. Shang, X. Zhao, Y. Feng, J. Qi, G. F. Gao, and Q. Wang, An updated atlas of antibody evasion by sars-cov-2 omicron sub-variants including bq.1.1 and xbb, *Cell Reports Medicine* **4**, 100991 (2023).
- [62] J. K. Hwang, F. W. Alt, and L.-S. Yeap, Related mechanisms of antibody somatic hypermutation and class switch recombination, *Microbiology Spectrum* **3**, 10.1128 (2015).
- [63] J. M. Di Noia and M. S. Neuberger, Molecular mechanisms of antibody somatic hypermutation, *Annual Review of Biochemistry* **76**, 1 (2007), pMID: 17328676.