Affinity maturation for an optimal balance between long-term immune coverage and short-term resource constraints

Victor Chardès[a], Massimo Vergassola[a], Aleksandra M. Walczak[a,1,2], and Thierry Mora[a,1,2]

[a]Laboratoire de Physique de l'École Normale Supérieure, CNRS, Paris Sciences & Lettres University, Sorbonne Université, Université de Paris, 75005 Paris, France

In order to target threatening pathogens, the adaptive immune system performs a continuous reorganization of its lymphocyte repertoire. Following an immune challenge, the B cell repertoire can evolve cells of increased specificity for the encountered strain. This process of affinity maturation generates a memory pool whose diversity and size remain difficult to predict. We assume that the immune system follows a strategy that maximizes the long-term immune coverage and minimizes the short-term metabolic costs associated with affinity maturation. This strategy is defined as an optimal decision process on a finite dimensional phenotypic space, where preexisting population of cells is sequentially challenged with a neutrally evolving strain. We show that the low specificity and high diversity of memory B cells—a key experimental result—can be explained as a strategy to protect against pathogens that evolve fast enough to escape highly potent but narrow memory. This plasticity of the repertoire drives the emergence of distinct regimes for the size and diversity of the memory pool, depending on the density of de novo responding cells and on the mutation rate of the strain. The model predicts power-law distributions of clonotype sizes observed in data and rationalizes antigenic imprinting as a strategy to minimize metabolic costs while keeping good immune protection against future strains.

B cell repertoire | affinity maturation | optimal decision theory | population dynamics

Adaptive immunity relies on populations of lymphocytes expressing diverse antigen-binding receptors on their surface to defend the organism against a wide variety of pathogens. B lymphocytes rely on a two-step process to produce diversity: First, a diverse naive pool of cells is generated; upon recognition of a pathogen the process of affinity maturation allows B cells to adapt their B cell receptor (BCR) to epitopes of the pathogen through somatic hypermutation (1). This process, which takes place in germinal centers (2), can increase the affinity of naive BCR for the target antigen by up to a thousandfold factor (3). Through affinity maturation, the immune system generates high-affinity, long-lived plasma cells, providing the organism with humoral immunity to pathogens through the secretion of antibodies—the soluble version of the matured BCR—as well as a pool of memory cells with varying affinity to the antigens (4). However, the diversity and coverage of the memory pool, as well as the biological constraints that control its generation, have not yet been fully explored.

Analysis of high-throughput BCR sequencing data has revealed long tails in the distribution of clonotype abundances, identifying some very abundant clonotypes as well as many very rare ones (5, 6). Additionally, many receptors have similar sequences and cluster into phylogenetically related lineages (7–11). These lineages have been used to locally trace the evolution of antibodies in HIV patients (12, 13) and in influenza vaccinees (14, 15). Memory B cell clones are more diverse and less specific

to the infecting antigen than antibody-producing plasma cells (16, 17). This suggests that the immune system is trying to anticipate infections by related pathogens or future escape mutants (18).

Theoretical approaches have attempted to qualitatively describe affinity maturation as a Darwinian coevolutionary process and studied optimal affinity maturation schemes (19–21), as well as optimal immunization schedules to stimulate antibodies with large neutralizing capabilities (22–24). Most of these approaches have been limited to short time scales, often with the goal of understanding the evolution of broadly neutralizing antibodies. Here we propose a mathematical framework to explore the trade-offs that control how the large diversity of memory cells evolves over a lifetime.

Despite long-lasting efforts to describe the coevolution of pathogens and host immune systems (25–29) and recent theoretical work on optimal schemes for using and storing memory in the presence of evolving pathogens (30), few theoretical works have described how the B cell memory repertoire is modified by successive immunization challenges. Early observations in humans (31) have shown that sequential exposure to antigenically drifted

Significance

Humoral immunity relies on the mutation and selection of B cells to better recognize pathogens. This affinity maturation process produces cells with diverse recognition capabilities. Examining optimal immune strategies that maximize the long-term immune coverage at a minimal metabolic cost, we show when the immune system should mount a de novo response rather than rely on existing memory cells. Our theory recapitulates known modes of the B cell response, predicts the empirical form of the distribution of clone sizes, and rationalizes as a trade-off between metabolic and immune costs the antigenic imprinting effects that limit the efficacy of vaccines (original antigenic sin). Our predictions provide a framework to interpret experimental results that could be used to inform vaccination strategies.

PHYSICS

IMMUNOLOGY AND INFLAMMATION

influenza strains was more likely to induce an immune response strongly directed toward the first strain the patients were exposed to (32). This immune imprinting with viral strains encountered early in life was initially called "original antigenic sin" as it can limit the efficiency of vaccination (33). This phenomenon has been observed in a variety of animal models and viral strains (34). Secondary infections with an antigenically diverged influenza strain can reactivate or "backboost" memory cells specific to the primary infecting strain (35). This response is characterized by lower binding affinity but can still have in vivo efficiency thanks to cross-reactive antibodies (36, 37). There is a long-standing debate about how detrimental original antigenic sin is (38, 39). Under what conditions should the immune system invest in keeping an antibody memory of past infections, as opposed to responding de novo to each new infection? When developing memory is preferred to responding de novo, how diverse should that memory be?

We build a theoretical framework of joint virus and repertoire evolution in antigenic space and investigate how acute infections by evolving pathogens have shaped, over evolutionary time scales, the B cell repertoire response and reorganization. Pathogens causing acute infections may be encountered multiple times over time scales of years, especially when they show a seasonal periodicity, while the maturation processes in the B cell repertoire take place over a few weeks. This observation allows us to consider that affinity maturation happens in a negligible time with respect to the reinfection period. Within this approximation, we investigate the optimal immune maturation strategies using a framework of discrete time decision process. We show the emergence of three regimes—monoclonal memory response, polyclonal memory response, and a de novo response—as trade-offs between immune coverage and resource constraint. Additionally, we demonstrate that reactivation of already existing memory clonotypes can lead to self-trapping of the immune repertoire to low reactivity clones, opening the way for original antigenic sin.

Results

Affinity Maturation Strategies for Recurring Infections. B cells recognize pathogens through the binding of their BCR to parts of the pathogen's proteins, called epitopes, which we refer to as "antigens" for simplicity. To model this complex protein–protein interaction problem, we assume that both receptors and antigens may be projected into an effective, d-dimensional antigenic space (Fig. 1), following the generalized shape space idea pioneered by Perelson and Oster (40). Receptor–antigen pairs at close distance in that space bind well, while those that are far away bind poorly. Specifically, we define a cross-reactivity function $0 \le f \le 1$ quantifying the binding affinity between antigen a and receptor x, which we model by a stretched exponential, $f(x, a) = e^{-(\|x-a\|/r_0)^q}$. This choice of function is the simplest that allows for introducing a cross-reactivity radius, r_0, while controlling how sharply recognition is abrogated as the distance between antigen and receptor oversteps that radius, through the stretching exponent q.

For simplicity, we focus on a single pathogen represented by its immunodominant antigen, so that each viral strain is represented by a single point a_t in antigenic space (black square), where $t = 1, 2, \dots$ is a discrete time counting the number of reinfections. It is difficult to estimate the rate of reinfections or exposures to the same pathogen. It can be fairly high in humans, where individuals are exposed to the most common viruses from less than once to several times a year (41). The numbers of lifetime exposures would then range from a few to a few hundred.

The B cell repertoire, on the other hand, is represented by a collection of antigenic coordinates corresponding to each receptor clonotype. We distinguish memory cells (dark blue circles in Fig. 1A), denoted by P_t, which have emerged in response to the presence of the virus, and a dense background of naive and innate
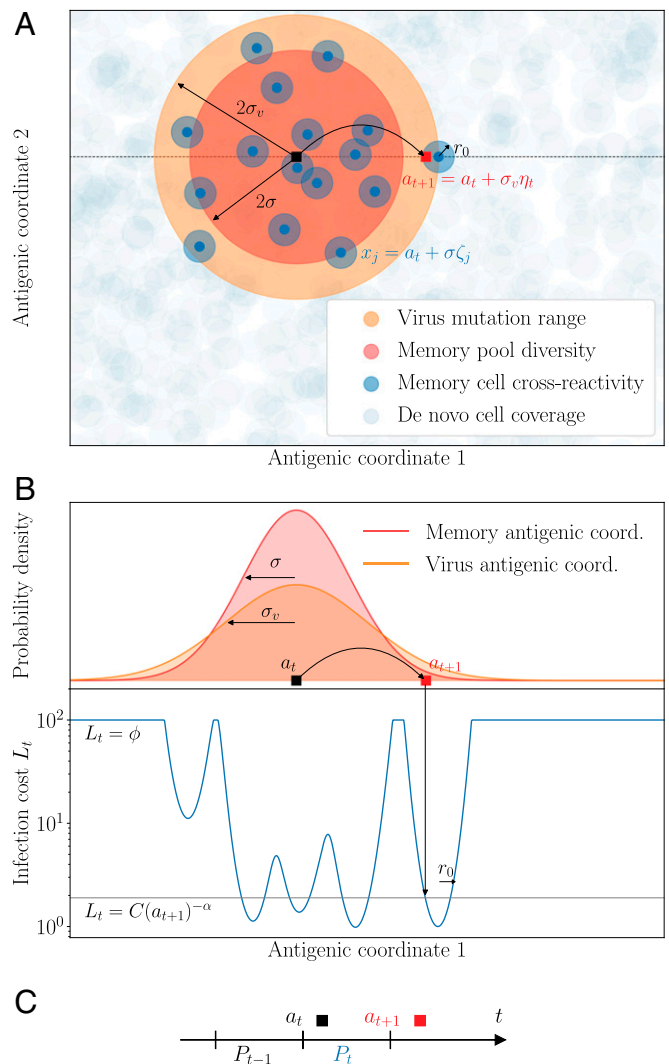

Fig. 1. Model of sequential affinity maturation. (*A*) An infecting strain is defined by its position a_n in antigenic space (dark square). In response, the immune system creates m new memory clonotypes x_j (blue points) from a Gaussian distribution of width σ centered in a_t (red area). These new clonotypes create a cost landscape (blue areas) for the next infection, complemented by a uniform background of naive, innate, and T cells (basal coverage, light blue). The next infecting strain (red square) is drawn from a Gaussian distribution of width σ_v centered in a_t (orange area). The position of this strain on the infection landscape is shown with the arrow. Antigenic space is shown in two dimensions for illustration purposes but can have more dimensions in the model. (*B*) Cross-section of the distributions of memories and of the next strain, along with the infection cost landscape L_t (in blue). Memories create valleys in the landscape, on a baseline protection ϕ. (*C*) Sequential immunization. Strain a_t modifies the memory repertoire into P_t, which is used to fight the next infection a_{t+1}. P_t is made of all newly created clonotypes (blue points in A) as well as some previously existing ones (not shown). Clonotype abundances are boosted following each infection as a function of the cross-reactivity, and each individual cell survives from one challenge to the other with a probability γ.

cells N (light blue circles), which together provide a uniform but weakly protective coverage of any viral strain (subsumed into the parameter ϕ defined later in this section).

The viral strain evolves randomly in antigenic space, sequentially challenging the existing immune repertoire. This assumption is justified by the fact that for acute infections with a drifting viral strain, such as influenza, the immune pressure exerted on the strain does not happen in hosts but rather at the population level (25). Viral evolution is not neutral, but it is unpredictable

from the point of view of individual immune systems. Specifically, we assume that upon reinfection, the virus is represented by a new strain, which has moved from the previous antigenic position a_t to the new one a_{t+1} according to a Gaussian distribution with typical jump size σ_v, called "divergence" (*Materials and Methods*).

Upon infection by a viral strain at a_t, available cross-reactive memory or naive cells will produce antibodies whose affinities determine the severity of the disease. We quantify the efficiency of this early response to the strain a_t with an infection cost I_t:

$$I_t = \min \left[\phi, \left(\sum_{x \in P_{t-1}} n_{x,t} f(x, a_t) \right)^{-\alpha} \right], \qquad [1]$$

where $\phi > 0$ is a maximal cost corresponding to using a de novo response and where $n_{x,t}$ denotes the size of clonotype x at time t. This infection cost is a decreasing function of the coverage of the virus by the preexisting memory repertoire, P_{t-1}, $C(a_t) = \sum_{x \in P_{t-1}} n_{x,t} f(x, a_t)$, with a power α governing how sharp that dependence is. Intuitively, the lower the coverage, the longer it will take for memory cells to mount an efficient immune response and clear the virus, incurring larger harm on the organism (42, 43).

When memory coverage is too low, the response of the naive B cell repertoire and the rest of the immune system, including its innate branch as well as T cell cytotoxic activity, offer protection through a de novo response, incurring a maximal cost fixed to ϕ. Memory cells respond more rapidly than naive cells, which is indirectly encoded in our model by the de novo cost ϕ being larger than the cost when specific memory cells are present (of order 1 or less). In *SI Appendix*, we show how this de novo cutoff may be derived in a model where the immune system activates its memory and de novo compartments in response to a new infection, when, for example, naive clonotypes are very numerous but offer weak protection. In that interpretation, ϕ scales like the inverse density of naive cells. We will refer to ϕ^{-1} as "de novo density," keeping in mind that this basal protection levels also includes other arms of the immune system. In Fig. 1*B*, we plot an example of the infection cost along a cross-section of the antigenic space.

After this early response, activated memory cells proliferate and undergo affinity maturation to create new plasma and memory cells targeting the infecting strain. To model this immune repertoire reorganization in response to a new infection a_t, we postulate that its strategy has been adapted over evolutionary time scales to maximize the speed of immune response to subsequent challenges, given the resource constraints imposed by affinity maturation (42). This strategy dictates the stochastic rules according to which the BCR repertoire evolves from P_{t-1} to P_t as a result of affinity maturation (Fig. 1*C*).

We consider the following rules inspired by known mechanisms available to the immune system (2). After the infection by a_t has been tackled by existing receptors and the infection cost has been paid, new receptors are matured to target future versions of the virus. Their number m_t is distributed according to a Poisson law, whose mean is controlled by the cost of infection, $\bar{m}(I_t)$. This dependence accounts for the feedback of the early immune response on the outcome of affinity maturation, consistent with extensive experimental evidence of the history dependence of the immune response (44). Each new receptor is roughly located around a_t in antigenic space with some added noise $\sigma(I_t)$ and starts with clonotype size $n_{x,t} = 1$ by convention. The diversification parameter σ can be tuned by the immune system through the permissiveness of selection in germinal centers, through specific regulation factors induced at the early stage of affinity-based selection (45): $\sigma = 0$ means that affinity maturation only keeps the best binders to the antigens, while $\sigma > 0$ means that selection is weaker.

At the same time, each clonotype $x \in P_{t-1}$ from the previous repertoire may be reactivated and be subsequently duplicated through cell divisions (18), with probability $\mu f(x, a_t)$ (*Materials and Methods*), proportional to the cross-reactivity, where $0 \leq \mu \leq 1$ is a proliferation parameter. These previously existing cells and their offspring may then die before the next infection. We denote by γ their survival probability, so that the average lifetime of each cell is $(1 - \gamma)^{-1}$. The proliferation and death parameters μ and γ are assumed to be constrained and fixed. The net mean growth is thus given by $\langle n_{x,t} \rangle = (1 + \mu f(x, a_t)) \gamma n_{x,t-1}$. $\Gamma \equiv (1 + \mu)\gamma$ is defined as the maximum growth factor. At the end of the process, the updated repertoire P_t combines the result of this proliferation and death process applied to P_{t-1} with the new receptors obtained from affinity maturation.

To assess the performance of a given strategy, we define an overall cost function at each time step:

$$L_t = I_t + \kappa m_t. \qquad [2]$$

The second term κm_t corresponds to a plasticity cost encoding the resources necessary to generate and maintain new memory clonotypes with affinity maturation. This cost enforces a minimal homeostatic constraint on the memory repertoire. We neglect any dependence of the cost on the diversification σ, which is secondary and would require adding additional parameters without affecting the qualitative picture. We assume that over evolutionary time scales, the immune system has minimized the average cumulative cost $\langle L_t \rangle$ over a large number of infections (*Materials and Methods*). This optimization yields the optimal parameters of the strategy, namely, the best functions $m^*(I)$ and $\sigma^*(I)$ describing the extent and diversity of affinity maturation and how they should depend on the strength of the infection I. For the sake of simplicity, in *Phase Diagram of Optimal Affinity Maturation Strategies*, *Population Dynamics of Optimized Immune Systems*, and *Comparison to Experimental Clone Size Distributions*, we specialize to the case of constant functions $m(I) \equiv m$ and $\sigma(I) \equiv \sigma$. We come back to the general case in *Inhibition of Affinity Maturation and Antigenic Imprinting*.

Phase Diagram of Optimal Affinity Maturation Strategies. We obtain optimal constant strategies $\bar{m}(I) = m^*, \sigma(I) = \sigma^*$, by minimizing the simulated long-term cost $\mathcal{L}(\mathcal{S})$ (Eq. 6) in a two-dimensional antigenic space (see *Materials and Methods* for details of the simulation, optimization procedures, and phase determination). By varying two key parameters, the cost ϕ associated to the use of the de novo response, and the virus divergence σ_v, we see a phase diagram emerge with three distinct phases: the de novo response, monoclonal response, and polyclonal response phases (Fig. 2*A*). In Fig. 2 *B* and *C*, we show examples of the stochastic evolution of memory repertoires with optimal rules in the two phases (monoclonal and polyclonal responses). Fig. 2 *D–F* show the behavior of the optimal parameters, as well as the fraction of infections for which the de novo response is used (when the maximal infection cost ϕ is paid). The general shape and behavior of this phase diagram depends only weakly on the parameter choices (*SI Appendix*, Fig. S1).

When the de novo response is sufficiently protective (small ϕ) or when the virus mutates too much between infections (large σ_v), the optimal strategy is to produce no memory cells at all ($\bar{m}^* = 0$) and rely entirely on the de novo response, always paying a fixed cost $L_t = \mathcal{L} = \phi$ (de novo phase).

When the virus divergence σ_v is small relative to the cross-reactivity range r_0, it is beneficial to create memory clonotypes ($\bar{m}^* > 0$) but with no diversity, $\sigma^* = 0$ (monoclonal response). In this case, all newly created clonotypes are invested into a single antigenic position a_t that perfectly recognizes the virus. This strategy is optimal because subsequent infections, typically caused by similar viral strains of the virus, are well recognized by these memory clonotypes.
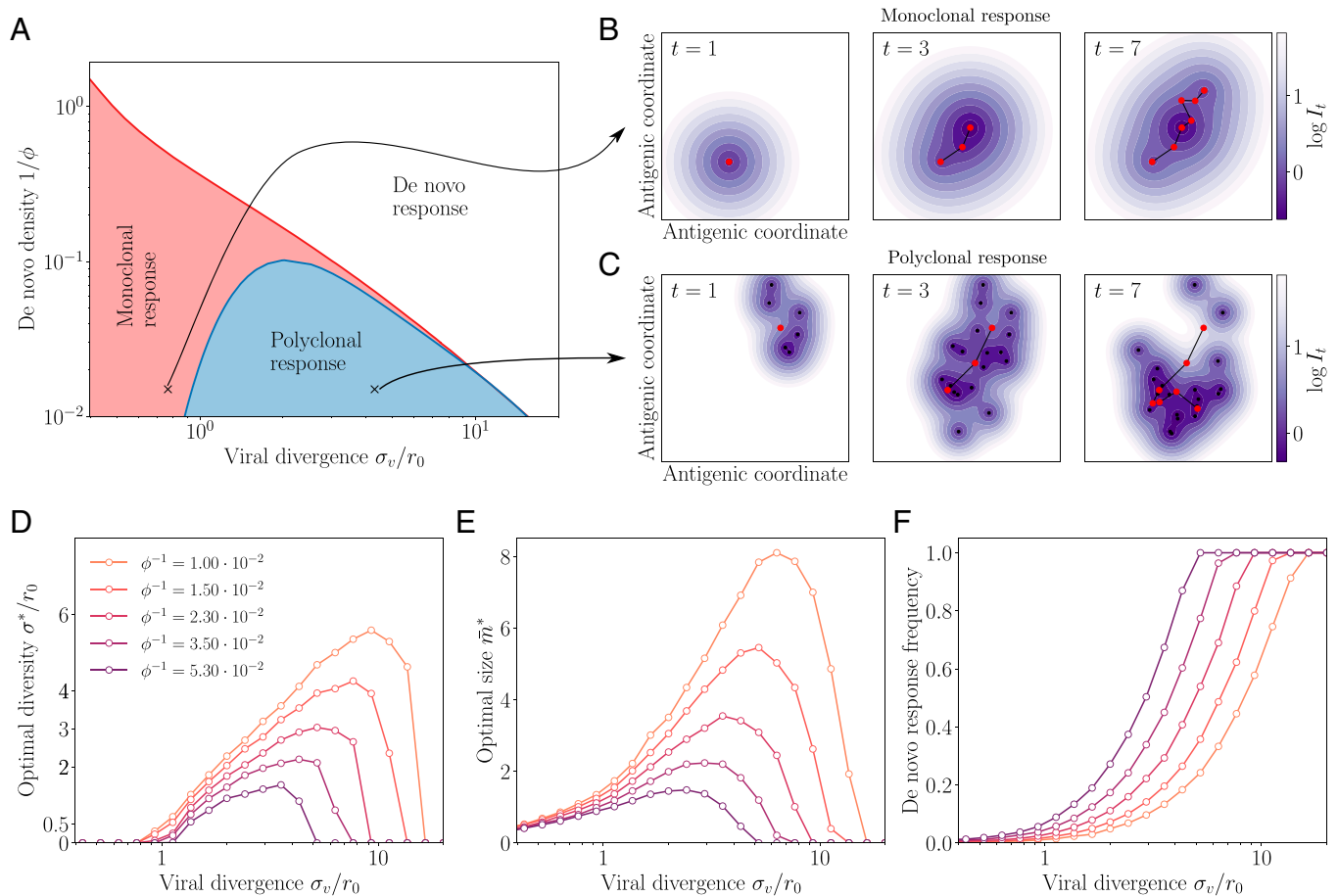
Chardès et al.
Affinity maturation for an optimal balance between long-term immune coverage and short-term resource constraints

PNAS | 3 of 10
https://doi.org/10.1073/pnas.2113512119

Fig. 2. Regimes of affinity maturation. (*A*) Phase diagram of the model as a function of the de novo density $1/\phi$ and viral divergence σ_v, in a two-dimensional antigenic map. Three phases emerge: monoclonal memory (red), polyclonal memory (blue), and de novo response (white). Snapshots in antigenic space of the sequential immunization by a viral strain in the (*B*) monoclonal and (*C*) polyclonal phases. We show the viral position (red dots), memory clonotypes (black dots), and viral trajectory (black line). The color map shows the log infection cost. Parameters σ_v and ϕ correspond to the crosses on the phase diagram in *A*, with their respective optimal σ^*, $\bar{m}^*$ (see arrows). (*D*) Diversity σ^*, (*E*) optimal size $\bar{m}^*$, and (*F*) frequency of de novo response usage to an immunization challenge for different de novo densities $1/\phi$. Parameters values: $\kappa = 3.3$, $\alpha = 1$, $q = 2$, $d = 2$, $\gamma = 0.85$, and $\mu = 0.5$.

For larger but still moderate virus divergences σ_v, this perfectly adapted memory is not sufficient to protect from mutated strains: the optimal strategy is rather to generate a polyclonal memory response, with $\bar{m}^* > 0$, $\sigma^* > 0$. In this strategy, the immune system hedges its bet against future infections by creating a large diversity of clonotypes that cover the vicinity of the encountered strain. The created memories are thus less efficient against the current infection, which they never will have to deal with. The advantage of this strategy is to anticipate future antigenic mutations of the virus. This diversified pool of cells with moderate affinity is in agreement with recent experimental observations (2, 18, 46, 47). The diversity of the memory pool is supported by a large number of clonotypes $\bar{m}^*$ (Fig. 2*D*). As the virus divergence σ_v is increased, the optimal strategy is to scatter memory cells farther away from the encountered strain (increasing σ^*; Fig. 2*E*). However, when σ_v is too large, both drop to zero as the de novo response takes over (Fig. 2*F*). Increasing the de novo density ϕ^{-1} also favors the de novo phase. When there is no proliferation on average, i.e., $\Gamma = (1 + \mu)\gamma < 1$, there even exists a threshold ϕ_c^{-1} above which the de novo response strategy is always best (*SI Appendix,* Fig. S1 and *Appendix B: Transition from Monoclonal to De Novo Phase at $\sigma_v = 0$* for estimates of that threshold).

We derived analytical results and scaling laws in a simplified version of the model, where cross-reactivity is ideally sharp and where memory cells survive only until the next infection.

The results are derived in *SI Appendix* and illustrated in *SI Appendix,* Fig. S2. In this simplified setting, the monoclonal to polyclonal transition occurs at $\sigma_v \sim r_0$, consistent with the intuition that diversification occurs when the virus is expected to escape out of the cross-reactivity radius. The polyclonal to de novo response transition occurs at $(\sigma_v/r_0)^d \sim \phi/\kappa$, when basal coverage by the naive repertoire and the rest of the immune system outcompetes the protection afforded by memory cells relative to their cost κ.

In summary, the model predicts the two expected regimes of de novo response and memory use depending on the parameters that set the costs of infections and memory formation. However, in addition, it shows a third phase of polyclonal response, where affinity maturation acts as an anticipation mechanism whose role is to generate a large diversity of cells able to respond to future challenges. The prediction of a less focused and thus weaker memory pool observed experimentally is thus rationalized as a result of a bet-hedging strategy.

Population Dynamics of Optimized Immune Systems. We now turn to the population dynamics of the memory repertoire. When the virus drifts slowly in antigenic space (small σ_v), the same clonotypes get reactivated multiple times, causing their proliferation, provided that $\Gamma = \gamma(1 + \mu) > 0$. This reactivation continues until the virus leaves the cross-reactivity range of the original clonotype, at which point the memory clone decays and eventually
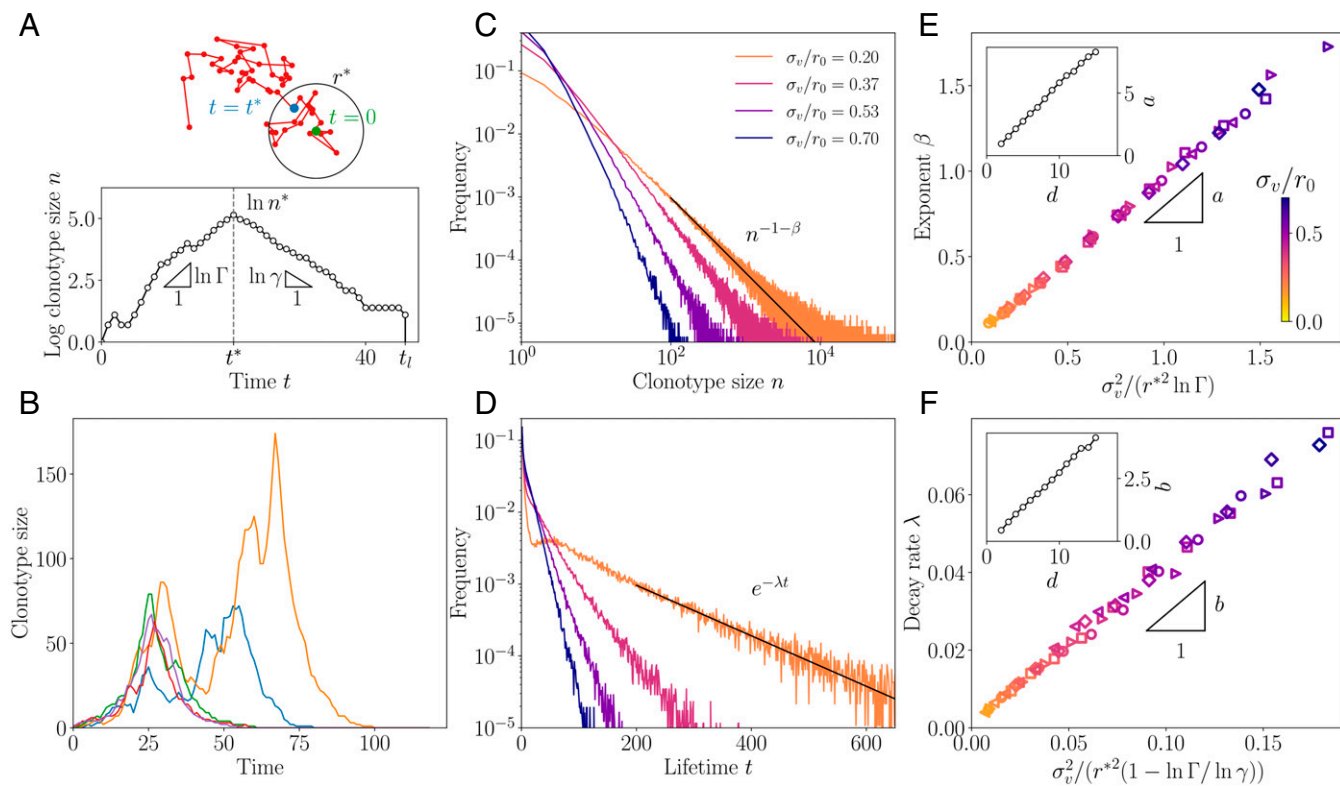
Fig. 3. Clonotype dynamics and distribution. (*A*) Sketch of a recall response generated by sequential immunization with a drifting strain. Clonotypes first grow with multiplicative rate $\Gamma = \gamma(1 + \mu)$, until they reach the effective cross-reactivity radius r^*, culminating at n^*, after which they decay with rate γ until extinction at time t_l. (*B*) Sample trajectories of clonotypes generated by sequential immunization with a strain of mutability $\sigma_v/r_0 = 0.53$. (*C*) Distribution of clonotype size for varying virus mutability σ_v/r_0. (*D*) Distribution of the lifetime of a clonotype for varying virus mutability σ_v/r_0. In *B–D*, the proliferation parameters are set to $\gamma = 0.85$, $\gamma = 0.5$, i.e., $\Gamma = 1.275$. (*E*) Scaling relation of the power law exponent for varying values of the parameters. (*Inset*) Dependence of the proportionality factor a on dimension. (*F*) Scaling relation of the decay rate λ for varying σ_v/r_0, with scaling of the proportionality factor b. In both *E* and *F*, the different parameters used are ($\gamma = 0.82$, $\mu = 0.65$), i.e., $\Gamma = 1.353$ (diamonds); ($\gamma = 0.8$, $\mu = 0.62$), i.e., $\Gamma = 1.296$ (squares); ($\gamma = 0.85$, $\mu = 0.5$), i.e., $\Gamma = 1.275$ (circles); ($\gamma = 0.87$, $\mu = 0.4$), i.e., $\Gamma = 1.21$ (triangles pointing to the right); and ($\gamma = 0.9$, $\mu = 0.35$), i.e., $\Gamma = 1.21$ (triangles pointing to the left). In *B–F*, the strategy was optimized for $\phi = 100$ and $\kappa = 0.5/(1 - \gamma)$. The color code for σ_v/r_0 is consistent across *C–F*. In *B–F*, the other parameters used are $\alpha = 1$, $q = 2$, and $d = 2$.

goes extinct (Fig. 3*A*). Typical clonotype size trajectories from the model are shown in Fig. 3*B*. They show large variations in both their maximal size and lifetime. The distribution of clonotype abundances, obtained from a large number of simulations, is indeed very broad, with a power law tail (Fig. 3*C*). The lifetime of clonotypes, defined as the time from emergence to extinction, is distributed according to an exponential distribution (Fig. 3*D*). The exponents governing the tails of these distributions, β and γ, depend on the model parameters, in particular the divergence σ_v.

We can understand the emergence of these distributions using a simple scaling argument, detailed in *SI Appendix*. The peak size of a clonotype depends on the number of successive infections by viral strains remaining within a distance $r^* = r_0 \ln[\gamma\mu/(1 - \gamma)]^{1/q}$ from the clonotype, under which it continues expanding. This number has a long exponential tail with characteristic time $t_s \sim (r^*/\sigma_v)^2$. One can show that this translates into a power law tail for the distribution of clonotype sizes,

$$p(n^*) \sim \frac{1}{n^{*1+\beta}}, \quad \text{with } \beta \sim \frac{\sigma_v^2}{r^{*2}\ln\Gamma}, \qquad [3]$$

and an exponential tail for the lifetime of clonotypes,

$$p(t_l) \sim e^{-\lambda t_l}, \quad \lambda \sim \frac{\sigma_v^2}{r^{*2}}\left(1 + \frac{\ln\Gamma}{\ln(1/\gamma)}\right)^{-1}. \qquad [4]$$

This simple scaling argument predicts the exponents β and λ fairly well: Fig. 3 *E* and *F* confirm the validity of the scaling relations Eqs. **3** and **4** against direct evaluation from simulations, for

$d = 2$ and $q = 2$. These scalings still hold for different parameter choices (*SI Appendix*, Fig. S3).

These scaling relations are valid up to a geometry-dependent prefactor, which is governed by dimensionality and the shape of the cross-reactivity kernel. In *SI Appendix*, we calculate this prefactor in the special case of an all-or-nothing cross-reactivity function, $q = \infty$. Generally, β increases with d, as shown in Fig. 3 *E* and *F*, *Insets*, for $q = 2$. In higher dimensions, there are more routes to escape the cross-reactivity range and thus a faster decaying tail of large clonotypes. This effect cannot be explained by having more dimensions in which to mutate, since the antigenic variance is distributed across each dimension, according to σ_v^2/d. Rather, it results from the absence of antigenic back-mutations: in high dimensions, each mutation drifts away from the original strain with a low probability of return, making it easier for the virus to escape and rarer for memory clonotypes to be recalled upon infections by mutant strains.

Comparison to Experimental Clone Size Distributions. The power law behavior of the clone size distribution predicted from the model (Fig. 3*E*) can be directly compared to existing data on bulk repertoires. While the model makes a prediction for subsets of the repertoire specific to a particular family of pathogens, the same power law prediction is still valid for the entire repertoire, which is a mixture of such subrepertoires. Power laws have been widely observed in immune repertoires: from early studies of repertoire sequencing data of BCR in zebrafish (5, 48), to the distribution of sizes of clonal families of human immunoglobulin
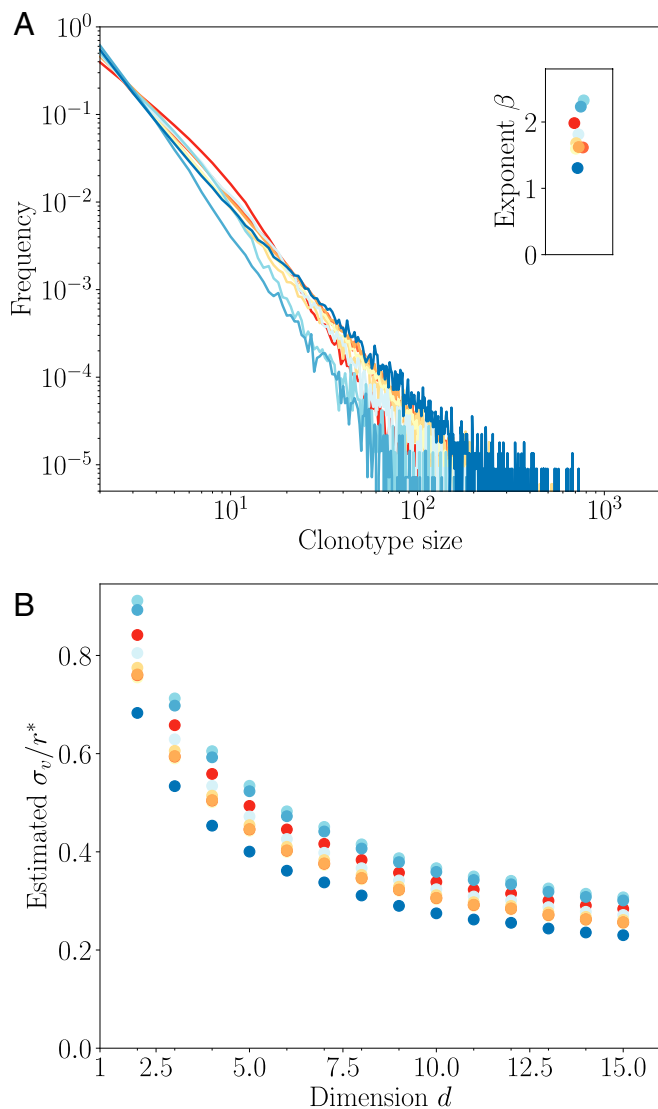
Fig. 4. Comparison to repertoire data. (*A*) Clonotype abundance distribution of IgG repertoires of healthy donors from ref. 50. (*B*) Estimated mutatibility σ_v in units of the rescaled cross-reactivity r^*, defined as the antigenic distance at which clonotypes stop growing. σ_v is obtained as a function of d by inverting the linear relationship estimated in Fig. 3*E, Inset*, assuming $q = 2$ and $\Gamma = 1.4$ (estimated from ref. 18).

G (IgG) BCR (number of unique sequences in a lineage stemming from a common naive ancestor) (14, 49), as well as in T cell receptor repertoires (6). However, these power laws have not yet been reported in the clonotype abundance distribution of human BCR (number of copies of unique BCR sequences).

To fill this gap, we used publicly available IgG repertoire data of nine human donors from a recent ultradeep repertoire profiling study of immunoglobulin heavy chains (IGH) (50). The data were downloaded from Sequence Read Archive and processed as in ref. 49. Repertoires were obtained from the sequencing of IGH messenger RNA molecules to which unique molecular identifiers (UMI) were appended. For each IGH nucleotide clonotype found in the dataset, we counted the number of distinct UMI attached to it, as a proxy for the overall abundance of that clonotype in the expressed repertoire. The distributions of these abundances are shown for all nine donors in Fig. 4*A*. In agreement with the theory, they display a clear power law decay $p(n) \sim n^{-1-\beta}$, with $\beta = 1.2$ to 2.4.

Since the experimental distribution is derived from small subsamples of the blood repertoire, the absolute abundances cannot be directly compared to those of the model. In particular, subsampling means that the experimental distribution focuses on the very largest clonotypes. Thus, comparisons between model and data should be restricted to the tail behavior of the distribution, namely, on its power law exponent β. The bulk repertoire is a mixture of antigen-specific subrepertoires, each predicted to be a power law with a potentially different exponent. The resulting distribution is still a power law dominated by the largest exponent.

We used this comparison to predict from the exponent β the virus divergence between infections. To do so, we fit a linear relationship to Fig. 3*E, Inset*, and invert it for various values of the dimension d to obtain σ_v/r^*. We fixed $\Gamma = 1.4$, which corresponds to a 40% boost of memory B cells upon secondary infection, inferred from a fourfold boost following four sequential immunizations reported in mice (18). The result is robust to the choice of donor but decreases substantially with dimension because higher dimensions mean faster escape and thus a lower divergence for a given measured exponent (Fig. 4*B*). The inferred divergence σ_v is always lower than, but of the same order as, the effective cross-reactivity range r^*, suggesting that the operating point of the immune system falls in the transition region between the monoclonal and polyclonal response phases (Fig. 2*A*).

Inhibition of Affinity Maturation and Antigenic Imprinting. In this section we come back to general strategies where the process of affinity maturation depends on the immune history through the infection cost I experienced by the system during the early immune response, which controls the number and diversity of newly created memories following that response: $\bar{m}(I)$ and $\sigma(I)$. The optimization of the loss function in Eq. **2** is now carried out with respect to two functions of I. To achieve this task, we optimize with respect to discretized functions $(\bar{m}_1, \ldots, \bar{m}_n)$ and $(\sigma_1, \ldots, \sigma_n)$ taken at n values of the infection cost I between 0 and ϕ. From this optimization, a clear transition emerges between a regime of complete inhibition of affinity maturation $(\bar{m}^*(I) = 0)$ at small infection costs and a regime of affinity maturation $(\bar{m}^*(I) > 0)$ at larger infection costs (Fig. 5*A*). In the phase where affinity maturation occurs, the optimal diversity $\sigma^*(I)$ is roughly constant (Fig. 5*B*).

This transition means that when preexisting protection is good enough, the optimal strategy is not to initiate affinity maturation at all, to save the metabolic cost κm_t. This inhibition of affinity maturation is called antigenic imprinting and is linked to the notion of original antigenic sin, whereby the history of past infections determines the process of memory formation, usually by suppressing it. This phenomenon leads to the paradoxical prediction that a better experienced immune system is less likely to form efficient memory upon new infections. Importantly, in our model this behavior does not stem from a mechanistic explanation, such as competition for antigen or T cell help between the early memory response and germinal centers, but rather as a long-term optimal strategy maximizing immune coverage while minimizing the costs of repertoire rearrangement.

To simplify the investigation of antigenic imprinting, we approximate the optimal strategies in Fig. 5 *A* and *B* by step functions, with a suppressed phase, $\bar{m}(I) = 0$, for $I < \xi$ and an active phase, $\bar{m}(I) \equiv \bar{m} > 0$ and $\sigma(I) \equiv \sigma$, for $I > \xi$. The threshold ξ is left as an optimization parameter, in addition to σ and $\bar{m}$. Optimizing with respect to these three parameters, we observe that the frequency of affinity maturation events mostly depends on σ_v (Fig. 5*C*). While this threshold remains approximately constant, the frequency of affinity maturation events increases as σ_v increases. At small σ_v, the optimal strategy is to extensively backboost existing memory cells; for large σ_v, the growing unpredictability of the next viral move makes it more likely
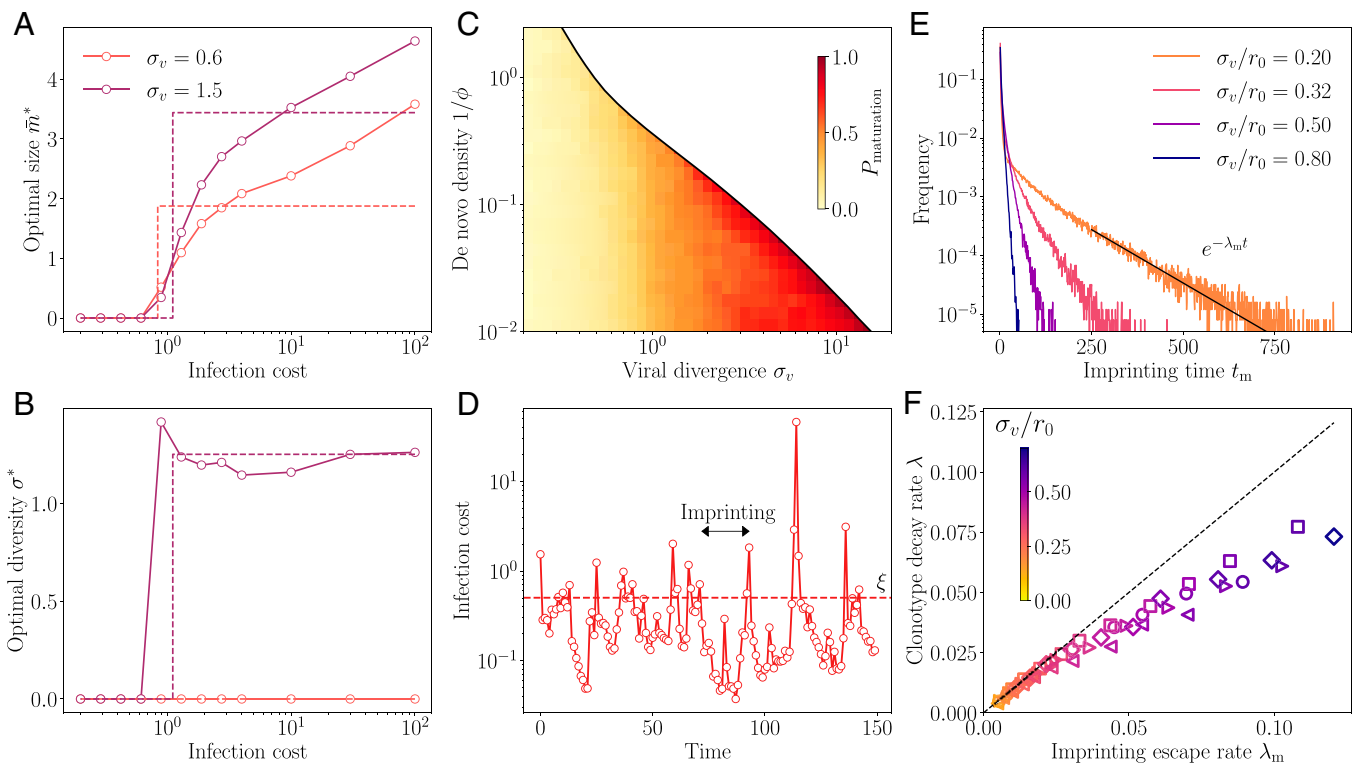
Fig. 5. Imprinting and backboosting. Optimal regulatory functions for (*A*) the number $m^*(I)$ and (*B*) the diversity $\sigma^*(I)$ of new memories as a function of the infection cost I, for two values for the viral divergence. These functions show a sharp transition from no memory formation to some memory formation, suggesting that they be replaced by simpler step functions (dashed lines). This step function approximation is used in *C–F*. (*C*) Frequency of infections leading to affinity maturation in the optimal strategy. The frequency increases with the virus divergence σ_v, up to the point of the transitions to the de novo phase where memory is not used at all. (*D*) Typical trajectory of infection cost in sequential infections at $\sigma_v/r_0 = 0.5$. When the cost goes beyond the threshold ξ, affinity maturation is activated, leading to a drop in infection cost. These periods of suboptimal memory describe an original antigenic sin, whereby the immune system is frozen in the state imprinted by the last maturation event. (*E*) Distribution of imprinting times, i.e., the number of infections between affinity maturation events, decays exponentially with rate λ_m. The proliferation parameters in *A–E* are set to $\gamma = 0.85$ and $\mu = 0.5$. (*F*) Predicted scaling of λ_m with the clonotype decay rate λ from Fig. 3 *D* and *F*. In *F*, the different parameters used are ($\gamma = 0.82$, $\mu = 0.65$), i.e., $\Gamma = 1.353$ (diamonds); ($\gamma = 0.8$, $\mu = 0.62$), i.e., $\Gamma = 1.296$ (squares); ($\gamma = 0.85$, $\mu = 0.5$), i.e., $\Gamma = 1.275$ (circles); ($\gamma = 0.87$, $\mu = 0.4$), i.e., $\Gamma = 1.21$ (triangles pointing to the right); and ($\gamma = 0.9$, $\mu = 0.35$), i.e., $\Gamma = 1.21$ (triangles pointing to the left). The color code for σ_v/r_0 is consistent across *E* and *F*. In *A–F*, the strategy was optimized for $\phi = 100$ and $\kappa = 0.5/(1-\gamma)$. In *D–F*, the other parameters used are $\alpha = 1$, $q = 2$, and $d = 2$.

to have recourse to affinity maturation. In other words, when the virus is stable (low σ_v), the immune system is more likely to capitalize on existing clonotypes and not implement affinity maturation because savings made on the plasticity cost outweigh the higher infection cost. As the virus drifts away with time, this infection cost also increases, until it reaches the point where affinity maturation becomes worthwhile again.

Trajectories of the infection cost show the typical dynamics induced by backboosting, with long episodes where existing memory remains sufficient to keep the cost below ξ (Fig. 5*D*), interrupted by infections that fall too far away from existing memory, triggering a new episode of affinity maturation and concomitant drop in the infection cost.

We call the time between affinity maturation events t_m. Its mean $\langle t_m \rangle$ is equal to the inverse of the frequency of maturation events and thus decreases with σ_v. Its distribution, shown Fig. 5*E*, has an exponential tail with exponent λ_m. The exponential tail of the distribution of t_m is dominated by episodes where the viral strain drifted less than expected. In that case, the originally matured clonotype grows to a large size, offering protection for a long time, even after it has stopped growing and only decays. We therefore expect that in the case of a slowly evolving virus $\sigma_v \ll r_0$, the escape rate from the suppressed phase is given by the clonotype decay rate: $\lambda_m \sim \lambda$. We verify this prediction in Fig. 5*F*. Interestingly, for slowly evolving viruses, the typical clonotype lifetime diverges, leading to a lifelong imprinting by the

primary immune challenge. Conversely, as the viral divergence σ_v grows, the imprinting time decays faster than the typical clonotype lifetime, and the extent of the imprinting phenomenon is limited.

Discussion

Adaptive immunity coordinates multiple components and cell types across entire organisms over many space and time scales. Because it is difficult to empirically characterize the immune system at the system scale, normative theories have been useful to generate hypotheses and fill the gap between observations at the molecular, cellular, and organismal scales (51, 52). Such approaches include clonal selection theory (53), or early arguments about the optimal size and organization of immune repertoires (40, 54, 55) and of affinity maturation (20, 21). While these theories do not rely on describing particular mechanisms of immune function, they may still offer quantitative insights and help predict new mechanisms or global rules of operation.

Previous work developed models of repertoire organization as a constrained optimization problem where the expected future harm of infection, or an ad hoc utility function, is minimized (30, 42, 43, 56). In ref. 43, it was assumed clonotypes specific to all antigens are present at all times in the repertoire; the mechanism of immune memory then merely consists of expanding specific clonotypes at the expense of others. This assumption describes T cell repertoires well, where there are naive cells with good affinity

to essentially any antigen (57). For B cells the situation is more complex because of affinity maturation. In addition, reorganizing the repertoire through mutation and selection has a cost and is subject to metabolic and physical constraints.

Our work addresses these challenges by proposing a framework of discrete time decision process to describe the optimal remodeling of the B cell repertoire following immunization, through a combination of affinity maturation and backboosting. While similar to ref. 30, our approach retains the minimal amount of mechanistic details and focuses on questions of repertoire remodeling, dynamics, and structure. The specific choices of the cost functions were driven by simplicity, while still retaining the ability to display emergent behavior. Generalizing the metabolic cost function to include, e.g., costs of diversification (through a dependence on σ) or of cell proliferation is not expected to affect our results qualitatively, although it may shift the exact positions of the transition boundaries.

We investigated strategies that maximize long-term protection against subsequent challenges and minimize short-term resource costs due to the affinity maturation processes. Using this model, we observed that optimal strategies may be organized into three main phases as the pathogen divergence and de novo coverage are varied. We expect these distinct phases to coexist in the same immune system as there exists a wide range of pathogen divergences, depending on their evolutionary speed and typical frequency of recurrence.

For fast recurring or slowly evolving pathogens, the monoclonal response ensures a very specific and targeted memory. This role could be played by long-lived plasma cells. These cells are selected through the last rounds of affinity maturation, meaning that they are very specific to the infecting strain (58). Yet, despite not being called memory cells, they survive for long times in the bone marrow, providing long-term immunity.

For slow recurring or fast evolving pathogens, the polyclonal response provides a diverse memory to preempt possible antigenic drift of the pathogen. The function could be fulfilled by memory B cells, which are released earlier in the affinity maturation process, implying that they are less specific to the infecting strain, but collectively can cover more immune escape mutations. While affinity maturation may start from both memory and naive B cells during sequential challenges, the relative importance of each is still debated (59–61). Our model does not commit on this question since we assume that the main benefit of memory is on the infection cost, rather than its reuse in subsequent rounds of affinity maturation.

For simplicity our model assumed random evolution of the virus. However, there is evidence, backed by theoretical arguments, that successive viral strains move in a predominant direction in antigenic space, as a result of immune pressure by the host population (29, 62, 63). While it is unlikely that the immune system has evolved to learn how to exploit this persistence of antigenic motion in a specific manner, such a bias in the random walk is expected to affect the optimal strategy, as we checked in simulations (*SI Appendix*, Fig. S4). The bias of the motion effectively increases the effective divergence of the virus, favoring the need for more numerous and more diverse memory cells. However, it does not seem to affect the location of the polyclonal to de novo response transition.

The model is focused on acute infections, motivated by the assumption that recurring infections and antigenic drift are the main drivers of affinity maturation evolution. However, much of the model and its results can be reinterpreted for chronic infections. In that context, the sequential challenges of our model would correspond to selective sweeps in the viral population giving rise to new dominant variants. While the separation of time scales between the immune response and the rate of reinfections would no longer hold, we expect some predictions, such as the distribution of clonotype sizes and the emergence of imprinting, to hold true. Chronic infections also imply that the virus evolves

as a function of how the immune system responds. Including this feedback would require a game-theoretic treatment. We speculate that it would drive antigenic motion in a persistent direction, as argued earlier and evaluated in *SI Appendix*, Fig. S4.

We investigated strategies where the outcome of affinity maturation is impacted by the efficiency of the early immune response. It is known that the extrafollicular response can drastically limit antigen availability and T cell help, decreasing the extent of affinity maturation and the production of new plasma and memory cells (64). Our general framework allows for but does not presume the existence of such negative feedbacks. Instead, they naturally emerge from our optimization principle. We further predict a sharp transition from no affinity maturation to some affinity maturation as a function of the infection cost. This prediction can be interpreted as the phenomenon of antigenic imprinting widely described in sequential immunization assays (35), or original antigenic sin (34). It implies that having been exposed to previous strains of the virus is detrimental to mounting the best possible immune response. Importantly, while antigenic imprinting has been widely described in the literature, no evolutionary justification was ever provided for its existence. Our model explains it as a long-term optimal strategy for the immune system, maximizing immune coverage while minimizing repertoire rearrangements (encoded in the cost κm).

We believe this framework can be generalized to investigate interactions between slow and fast varying epitopes, which are known to be at the core of the low effectiveness of influenza vaccines (32). When during sequential challenges only one of multiple epitopes changes at a time, it may be optimal for the immune system to rely on its protections against the invariant epitopes. Only after all epitopes have escaped immunity does affinity maturation get reactivated concomitantly to a spike of infection harm, similar to our result for a single antigen.

Our model can explain previously reported power laws in the distributions of abundances of BCR clonotypes. However, there exist alternative explanations to such power laws (65, 66) that do not require antigenically drifting antigens. Our model predictions could be further tested in a mouse model, by measuring the B cell recall response to successive challenges (35), but with epitopes carefully designed to drift in a controlled manner, to check the transition predicted in Fig. 5*A*. While not directly included in our model, our results also suggest that the size of the inoculum, which would affect the infection cost, should also affect backboosting. This effect could also be tested in mouse experiments. The predicted relationships between viral divergence and the exponents of the power law and clonotype lifetimes (Fig. 3 *E* and *F*) could be tested in longitudinal human samples, by sequencing subrepertoires specific to pathogens with different rates of antigenic evolution. This would require computational prediction of what BCRs are specific to what pathogen, which in general is difficult.

We only considered a single pathogenic species at a time, with the assumption that pathogens are antigenically independent, so that the costs relative to different pathogens are additive. Possible generalizations could include secondary infections, as well as antigenically related pathogens showing patterns of cross-immunity (such as cowpox and smallpox or different groups of influenza), which could help us shed light on complex immune interactions between diseases and serotypes, such as negative interference between different serotypes of the Dengue fever leading to hemorrhagic fever or of the human Bocavirus affecting different body sites (34).

Materials and Methods

Mathematical Model. The viral strain is modeled by its antigenic position, which follows a discrete random walk:

$$a_{t+1} = a_t + \sigma_v \eta_{t+1},$$ [5]

where η_t is a normally distributed d-dimensional variable with $\langle \eta_t \rangle = 0$ and $\langle \eta_t \cdot \eta_{t'} \rangle = \delta_{tt'}$.

The positions of newly created memory receptors are drawn at random according to $x_j = a_t + \sigma \xi_j$, $j = 1, \ldots, m_t$, where ξ_j is normally distributed with $\langle \xi_j \rangle = 0$ and $\langle \xi_j^2 \rangle = 1$. Their initial sizes are set to $n_{x_j,t} = 1$.

Upon further stimulation, the new size $n'_{x,t}$ of a preexisting clonotype right after proliferation is given by $n'_{x,t} - n_{x,t-1} \sim \text{Binom}(n_{x,t-1}, \mu f(x, a_t))$, where $f(x, a) = e^{-(\|x-a\|/r_0)^q}$ is the cross-reactivity kernel. After proliferation, each memory may die with probability γ, so that the final clonotype size after an infection cycle is given by $n_{x,t} \sim \text{Binom}(n'_{x,t}, \gamma)$.

The objective to be minimized is formally defined as a long-term average:

$$\mathcal{L}(m, \sigma) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} L_t. \qquad [6]$$

The optimal strategy is defined as

$$(\bar{m}^*, \sigma^*) = \underset{(m,\sigma)}{\arg \min} \, \mathcal{L}(m, \sigma), \qquad [7]$$

where m^* and σ^* are, in the general case, full functions of the infection cost I, $m^*(I)$ and $\sigma^*(I)$. In all results except in *Inhibition of Affinity Maturation and Antigenic Imprinting*, we use the Ansatz of constant functions, $m(I) \equiv m$, $\sigma(I) \equiv \sigma$. In *Inhibition of Affinity Maturation and Antigenic Imprinting*, we first perform optimization over discretized functions $m = (m_1, \ldots, m_n)$, $\sigma = (\sigma_1, \ldots, \sigma_n)$, defined over n chosen values of $I = (I_1, \ldots, I_n)$. Then, we parametrize the functions as step functions, $\sigma(I) = 0$ and $m(I) = 0$ for $I < \xi$ and $\sigma(I) = \sigma$ and $m(I) = m$ for $I > \xi$, and optimize over the three parameters σ, m, ξ.

Monte Carlo Estimation of the Optimal Strategies. The average cumulative cost $\mathcal{L}$ in Eq. 6 is approximated by a Monte Carlo method. To ensure the simulated repertoire reaches stationarity, we start from a naive repertoire and discard an arbitrary number of initial viral encounters. Because the process is ergodic, simulating a viral-immune trajectory over a long time horizon is equivalent to simulating M independent trajectories of smaller length T. To ensure the independence of our random realizations across our M parallel streams we use the random number generators method *split* provided in Tina's Random Number Generator library (67). The cumulative cost function $\mathcal{L}$ is convex for the range of parameters tested. To optimize $\mathcal{L}$ under positivity constraints for the objective variables σ, $\bar{m}$, and ξ, we use Py-BOBYQA (68), a coordinate descent algorithm supporting noisy objective functions.

The polyclonal to monoclonal (red curve) and memory to de novo response (blue curve) boundaries of the phase diagrams in *SI Appendix*, Figs. S2 and S3, are obtained by solving $\partial \mathcal{L}/\partial \sigma = 0$ in the monoclonal phase and $\partial \mathcal{L}/\partial \bar{m} = 0$ in the de novo phase. Both these derivatives can be approximated by finite differences with arbitrary tolerances on σ and $\bar{m}$. We fix the tolerance on σ to 0.2 and the tolerance on $\bar{m}$ to 0.01. To obtain the root of these difference functions, we use a bisection algorithm. In order to further decrease the noise level, we compute the difference functions across pairs of simulations, each pair using an independent sequence of pathogens a_t of length $L = 400$. The number of independent pairs of simulations used for each value of σ_v and ϕ is $M \sim 10^5$.

1. P. Nieuwenhuis, D. Opstelten, Functional anatomy of germinal centers. *Am. J. Anat.* **170**, 421–435 (1984).
2. G. D. Victora, M. C. Nussenzweig, Germinal centers. *Annu. Rev. Immunol.* **30**, 429–457 (2012).
3. H. N. Eisen, G. W. Siskind, Variations in affinities of antibodies during the immune response. *Biochemistry* **3**, 996–1008 (1964).
4. F. Weisel, M. Shlomchik, Memory B cells of mice and humans. *Annu. Rev. Immunol.* **35**, 255–284 (2017).
5. J. A. Weinstein, N. Jiang, R. A. White, 3rd, D. S. Fisher, S. R. Quake, High-throughput sequencing of the zebrafish antibody repertoire. *Science* **324**, 807–810 (2009).
6. T. Mora, A. M. Walczak, *Quantifying Lymphocyte Receptor Diversity in Systems Immunology* (CRC Press, 2019).
7. C. Kocks, K. Rajewsky, Stepwise intraclonal maturation of antibody affinity through somatic hypermutation. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8206–8210 (1988).
8. S. H. Kleinstein, Y. Louzoun, M. J. Shlomchik, Estimating hypermutation rates from clonal tree data. *J. Immunol.* **171**, 4639–4649 (2003).
9. T. B. Kepler, Reconstructing a B-cell clonal lineage. I. Statistical inference of unobserved ancestors. *F1000 Res.* **2**, 103 (2013).
10. G. Yaari, S. H. Kleinstein, Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med.* **7**, 121 (2015).
11. D. K. Ralph, F. A. Matsen, 4th, Likelihood-based inference of B cell clonal families. *PLOS Comput. Biol.* **12**, e1005086 (2016).
12. H. X. Liao et al.; NISC Comparative Sequencing Program, Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496**, 469–476 (2013).
13. A. Nourmohammad, J. Otwinowski, M. Łuksza, T. Mora, A. M. Walczak, Fierce selection and interference in B-cell repertoire response to Chronic HIV-1. *Mol. Biol. Evol.* **36**, 2184–2194 (2019).
14. N. Jiang et al., Lineage structure of the human antibody repertoire in response to influenza vaccination. *Sci. Transl. Med.* **5**, 171ra19 (2013).
15. F. Horns, C. Vollmers, C. L. Dekker, S. R. Quake, Signatures of selection in the human antibody repertoire: Selective sweeps, competing subclones, and neutral drift. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1261–1266 (2019).
16. K. G. C. Smith, A. Light, G. J. V. Nossal, D. M. Tarlinton, The extent of affinity maturation differs between the memory and antibody-forming cell compartments in the primary immune response. *EMBO J.* **16**, 2996–3006 (1997).
17. F. J. Weisel, G. V. Zuccarino-Catania, M. Chikina, M. J. Shlomchik, A temporal switch in the germinal center determines differential output of memory B and plasma cells. *Immunity* **44**, 116–130 (2016).
18. C. Viant et al., Antibody affinity shapes the choice between memory and germinal center B cell fates. *Cell* **183**, 1298–1311.e11 (2020).
19. M. Oprea, A. S. Perelson, Somatic mutation leads to efficient affinity maturation when centrocytes recycle back to centroblasts. *J. Immunol.* **158**, 5155–5162 (1997).
20. M. Oprea, E. van Nimwegen, A. S. Perelson, Dynamics of one-pass germinal center models: Implications for affinity maturation. *Bull. Math. Biol.* **62**, 121–153 (2000).
21. T. B. Kepler, A. S. Perelson, Somatic hypermutation in B cells: An optimal control treatment. *J. Theor. Biol.* **164**, 37–64 (1993).
22. S. Wang et al., Manipulating the selection forces during affinity maturation to generate cross-reactive HIV antibodies. *Cell* **160**, 785–797 (2015).

23. V. Sachdeva, K. Husain, J. Sheng, S. Wang, A. Murugan, Tuning environmental timescales to evolve and maintain generalists. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12693–12699 (2020).
24. M. Molari, K. Eyer, J. Baudry, S. Cocco, R. Monasson, Quantitative modeling of the effect of antigen dosage on B-cell affinity distributions in maturating germinal centers. *eLife* **9**, e55678 (2020).
25. B. T. Grenfell et al., Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332 (2004).
26. F. Blanquart, S. Gandon, Time-shift experiments and patterns of adaptation across time and space. *Ecol. Lett.* **16**, 31–38 (2013).
27. S. Cobey, P. Wilson, F. A. Matsen, 4th, The evolution within us. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140235 (2015).
28. K. Koelle, S. Cobey, B. Grenfell, M. Pascual, Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **314**, 1898–1903 (2006).
29. J. Marchi, M. Lässig, A. M. Walczak, T. Mora, Antigenic waves of virus-immune coevolution. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2103398118 (2021).
30. O. H. Schnaack, A. Nourmohammad, Optimal evolutionary decision-making to store immune memory. *eLife* **10**, e61346 (2021).
31. T. Francis, On the doctrine of original antigenic sin. *Proc. Am. Philos. Soc.* **104**, 572–578 (1960).
32. S. Cobey, S. E. Hensley, Immune history and influenza virus susceptibility. *Curr. Opin. Virol.* **22**, 105–111 (2017).
33. T. W. Hoskins, J. R. Davies, A. J. Smith, C. L. Miller, A. Allchin, Assessment of inactivated influenza-A vaccine after three outbreaks of influenza A at Christ's Hospital. *Lancet* **1**, 33–35 (1979).
34. A. Vatti et al., Original antigenic sin: A comprehensive review. *J. Autoimmun.* **83**, 12–21 (2017).
35. J. H. Kim, I. Skountzou, R. Compans, J. Jacob, Original antigenic sin responses to influenza viruses. *J. Immunol.* **183**, 3294–3301 (2009).
36. D. J. Smith, S. Forrest, D. H. Ackley, A. S. Perelson, Variable efficacy of repeated annual influenza vaccination. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14001–14006 (1999).
37. S. L. Linderman, S. E. Hensley, Antibodies with 'original antigenic sin' properties are valuable components of secondary immune responses to influenza viruses. *PLoS Pathog.* **12**, e1005806 (2016).
38. J. W. Yewdell, J. J. S. Santos, Original antigenic sin: How original? How sinful? *Cold Spring Harb. Perspect. Med.* **11**, 038786 (2021).
39. M. Worobey, S. Plotkin, S. E. Hensley, Influenza vaccines delivered in early childhood could turn antigenic sin into antigenic blessings. *Cold Spring Harb. Perspect. Med.* **10**, 038471 (2020).
40. A. S. Perelson, G. F. Oster, Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-non-self discrimination. *J. Theor. Biol.* **81**, 645–670 (1979).
41. C. Cohen et al.; PHIRST group, Asymptomatic transmission and high community burden of seasonal influenza in an urban and a rural community in South Africa, 2017-18 (PHIRST): A population cohort study. *Lancet Glob. Health* **9**, e863–e874 (2021).
42. A. Mayer, V. Balasubramanian, T. Mora, A. M. Walczak, How a well-adapted immune system is organized. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5950–5955 (2015).
43. A. Mayer, V. Balasubramanian, A. M. Walczak, T. Mora, How a well-adapting immune system remembers. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8815–8823 (2019).

Chardès et al.
Affinity maturation for an optimal balance between long-term immune coverage and short-term resource constraints

PNAS | 9 of 10
https://doi.org/10.1073/pnas.2113512119

44. R. J. Oidtman et al., Influenza immune escape under heterogeneous host immune histories. Trends Microbiol. 29, 1072–1082 (2021).

45. R. Nakagawa et al., Permissive selection followed by affinity-based proliferation of GC light zone B cells dictates cell fate and ensures clonal breadth. Proc. Natl. Acad. Sci. U.S.A. 118, e2016425118 (2021).

46. J. M. J. Tas et al., Visualizing antibody affinity maturation in germinal centers. Science 351, 1048–1054 (2016).

47. M. Kuraoka et al., Complex antigens drive permissive clonal selection in germinal centers. Immunity 44, 542–552 (2016).

48. T. Mora, A. M. Walczak, W. Bialek, C. G. Callan, Jr., Maximum entropy models for antibody diversity. Proc. Natl. Acad. Sci. U.S.A. 107, 5405–5410 (2010).

49. N. Spisak, A. M. Walczak, T. Mora, Learning the heterogeneous hypermutation landscape of immunoglobulins from high-throughput repertoire data. Nucleic Acids Res. 48, 10702–10712 (2020).

50. B. Briney, A. Inderbitzin, C. Joyce, D. R. Burton, Commonality despite exceptional diversity in the baseline human antibody repertoire. Nature 566, 393–397 (2019).

51. A. K. Chakraborty, A perspective on the role of computational models in immunology. Annu. Rev. Immunol. 35, 403–439 (2017).

52. G. Altan-Bonnet, T. Mora, A. M. Walczak, Quantitative immunology for physicists. Phys. Rep. 849, 1–83 (2020).

53. F. M. Burnet, A modification of Jerne's theory of antibody production using the concept of clonal selection. Aust. J. Sci. 20, 67–69 (1957).

54. A. S. Perelson, M. Mirmirani, G. F. Oster, Optimal strategies in immunology. I. B-cell differentiation and proliferation. J. Math. Biol. 3, 325–367 (1976).

55. A. S. Perelson, M. Mirmirani, G. F. Oster, Optimal strategies in immunology. II. B memory cell production. J. Math. Biol. 5, 213–256 (1978).

56. R. Marsland, 3rd, O. Howell, A. Mayer, P. Mehta, Tregs self-organize into a computing ecosystem and implement a sophisticated optimization algorithm for mediating immune response. Proc. Natl. Acad. Sci. U.S.A. 118, e2011709118 (2021).

57. J. J. Moon et al., Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. Immunity 27, 203–213 (2007).

58. M. Akkaya, K. Kwak, S. K. Pierce, B cell memory: Building two walls of protection against pathogens. Nat. Rev. Immunol. 20, 229–238 (2020).

59. J. S. Turner et al., Human germinal centres engage memory and naive B cells after influenza vaccination. Nature 586, 127–132 (2020).

60. L. Mesin et al., Restricted clonality and limited germinal center reentry characterize memory B cell reactivation by boosting. Cell 180, 92–106.e11 (2020).

61. R. Wong et al., Affinity-restricted memory B cells dominate recall responses to heterologous flaviviruses. Immunity 53, 1078–1094.e7 (2020).

62. T. Bedford, A. Rambaut, M. Pascual, Canalization of the evolutionary trajectory of the human influenza virus. BMC Biol. 10, 38 (2012).

63. T. Bedford et al., Integrating influenza antigenic dynamics with molecular evolution. eLife 3, e01914 (2014).

64. T. Arulraj, S. C. Binder, P. A. Robert, M. Meyer-Hermann, Germinal centre shutdown. Front. Immunol. 12, 2730 (2021).

65. J. Desponds, T. Mora, A. M. Walczak, Fluctuating fitness shapes the clone-size distribution of immune repertoires. Proc. Natl. Acad. Sci. U.S.A. 113, 274–279 (2016).

66. M. U. Gaimann, M. Nguyen, J. Desponds, A. Mayer, Early life imprints the hierarchy of T cell clone sizes. eLife 9, e61639 (2020).

67. H. Bauke, S. Mertens, Random numbers for large-scale distributed Monte Carlo simulations. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 75, 066701 (2007).

68. C. Cartis, J. Fiala, B. Marteau, L. Roberts, Improving the flexibility and robustness of model-based derivative-free optimization solvers. ACM Trans. Math. Softw. 45, 32:1–32:41 (2019).

10 of 10 | **PNAS**
https://doi.org/10.1073/pnas.2113512119

Chardès et al.
Affinity maturation for an optimal balance between long-term immune coverage and short-term resource constraints