Non-equilibrium antigen recognition in acute infections

Roberto Morán-Tovar[1] and Michael Lässig[1]

[1]Institute for Biological Physics, University of Cologne, Zülpicherstr. 77, 50937 Köln, Germany

Abstract

The immune response to an acute primary infection is a coupled process of antigen proliferation, molecular recognition by naive B-cells, and their subsequent proliferation and antibody shedding. Here we show B-cells can efficiently recognise new antigens by a tuned kinetic proofreading mechanism, where the number of proofreading steps and the characteristic rate of each step are set by the complexity of the immune repertoire. This process produces potent, specific and fast recognition of antigens, maintaining a spectrum of genetically distinct B-cell lineages as input for affinity maturation. We show that the proliferation-recognition dynamics of a primary infection can me mapped onto a generalised Luria-Delbrück process, akin to the dynamics of the classic fluctuation experiment. We derive the resulting statistics of the activated immune repertoire: antigen binding affinity, expected size, and frequency of active B-cell clones are related by power laws. Their exponents depend on the antigen and B-cell proliferation rate, the number of proofreading steps, and the lineage density of the naive repertoire. Empirical data of mouse immune repertoires are found to be consistent with activation involving at least three proofreading steps. Our model predicts key clinical characteristics of acute infections. The primary immune response to a given antigen is strongly heterogeneous across individuals; few elite responders are distinguished by early activation of high-affinity clones. Conversely, ageing of the immune system, by reducing the density of naive clones, degrades potency and speed of pathogen recognition.

Introduction

B-cells are a central part of the human adaptive immune system. These cells recognise antigens by specific binding: B-cell receptors (BCRs) located in the cellular membrane bind to antigenic epitopes, which are cognate binding loci on the surface of pathogens. To capture a wide range of a priori unknown pathogens, humans produce a large and diverse naive B-cell repertoire, estimated to contain about $L_0 \sim 10^9$ lineages with distinct BCR genotypes [1, 2, 3] and a comparable number of circulating naive cells [4]. An acute infection is characterised by rapid, initially exponential growth of an antigen population, often starting with few particles and reaching peak densities of order $10^8 \mathrm{ml}^{-1}$ within a few days [5, 6, 7, 8]. At some stage of this process, free antigens start to bind to B-cells in lineages of sufficiently high binding affinity. Antigen binding can activate B-cells, inducing rapid proliferation and shedding of free antibodies that eventually clear the antigen. Activated B-cells also form germinal centres and create immunological memory [9, 10]. Primary infections are estimated to generate multiple activated B-cell lineages, $L_{\mathrm{act}} \gtrsim 10^2$ [11].

The exponential growth of antigens, together with a large number of B-cell lineages, generates a formidable, real-time specificity problem for recognition. Consider an antigen that activates a high-affinity B-cell lineage at a given point of time. One day later, at a > 100fold higher population density, the antigen can potentially activate a large number of low-affinity lineages, generating a poor overall response of the immune repertoire. The actual process activates only a tiny fraction of the pre-infection repertoire. Previous work has established upper bounds of order 10^{-5} [12, 13], and a lower bound follows from recent data, $L_{\mathrm{act}}/L_0 \gtrsim 10^{-7}$ [11]. How is such highly specific immune response possible?

To address this problem, we develop a model for the recognition dynamics that consists of three steps: antigen proliferation, molecular activation, and subsequent proliferation of activated B-cells (Fig. 1). The B-cell activation process contains kinetic proofreading: a series of multiple, thermodynamically irreversible steps [14, 15]. In an appropriate regime of rate parameters, processes with kinetic proofreading are known to increase the affinity discrimination of their output compared to near-equilibrium processes. For the activation process of Fig. 1, we show that the activation rate of weak binding B-cells depends on the antigen-BCR binding constant and on the number of activation steps, $u_{\mathrm{act}} \sim K^{-p}$.

Kinetic proofreading has been recognised as a key step in the activation of T-cell immunity [16, 17, 18, 19]. For B-cells, evidence for proofreading comes from experimental observations of characteristic time lags in activation, but little is known about the underlying molecular mechanism [20]. Mechanisms observed in immune cell activation by membrane-bound antigens include BCR clustering, quo-

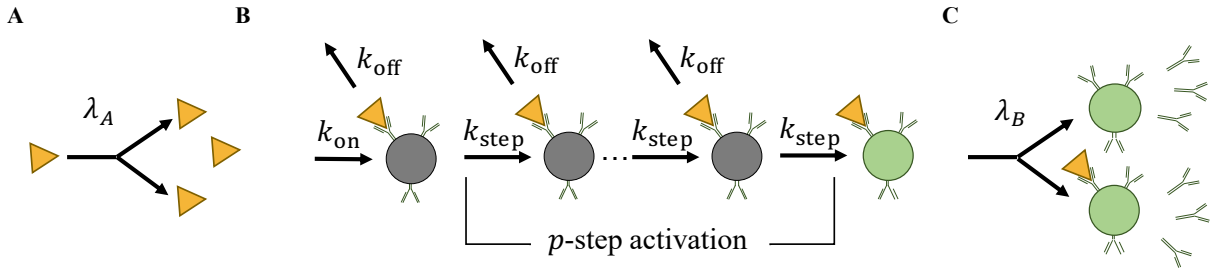

Fig. 1. Antigen recognition in a primary infection. In a minimal model, immune recognition of a new antigen involves three stages. **(A)** Exponential replication of free antigens with rate λ_A. **(B)** Activation of naive B-cells: binding of antigens with association rate k_{on}, dissociation rate k_{off}, and p irreversible activation steps with rate k_{step}. This mechanism yields a total activation rate u_{act} that decays as $u_{\text{act}} \sim (k_{\text{step}}/k_{\text{off}})^p$ in the low-affinity regime (see text). **(C)** Exponential replication of activated B-cells and proliferation of free antibodies with rate λ_B.

rum sensing, and membrane spreading-contraction [21, 22]. Such mechanisms may contribute to proofreading, but the relevance of particular mechanisms for the specificity problem addressed in this article remains unclear. Here we use a minimal p-step model of activation to show that proofreading is essential for specific and timely recognition of pathogens in acute primary infections.

To understand how activation and proofreading act in the face of an exponentially increasing input signal, we map the recognition dynamics onto a generalised Luria-Delbrück process. This resembles the proliferation-mutation dynamics of the classical Luria-Delbrück experiment [23]: the antigen corresponds to the wild-type, activation to mutation, and B-cell lineages to mutant cell lineages. The new feature of the infection dynamics, which has no analogue in the original Luria-Delbrück process, is that each B-cell lineage has a specific antigen binding constant K. This sets the density of B-cell lineages available for activation, $\Omega_0 \sim K^{\beta_{\text{act}}}$, and modulates their activation rate.

In the first part of the paper, we develop the theory of generalised Luria-Delbrück processes for antigen recognition, and we compute the distribution of lineages in activated B-cell repertoires. In the second part, we turn to biological implications of the recognition dynamics. We show that proofreading is required for effective antigen recognition. Our model predicts efficient immune responses, tuned to a balance between speed and potency, at an intermediate number of proofreading steps. Recent data of mouse immune repertoires [11] are shown to be consistent with this prediction.

Activated immune repertoires of different hosts responding to the same antigen show giant fluctuations, similar to mutant populations in a classical fluctuation experiment. Such fluctuations are a hallmark of Luria-Delbrück processes [23, 24, 25, 26]. In a primary immune response, giant fluctuations are generated by "jackpot" clones of large size and high antigen affinity. We derive the underlying statistics of activated repertoires and infer clinically important characteristics of acute primary infections.

Results

Antigen recognition dynamics

Antigen-BCR binding. In the initial phase of an infection, the antigen population grows exponentially with a rate λ_A (Fig. 1A, 2A). For viral pathogens, this process starts with few antigen copies and reaches densities $\rho_A(t)$ of order 10^8/ml within about 5 days, which implies replication factors > 100/d [5, 6, 7, 8]. Free antigens bind to circulating naive B-cells with rate $u_{\text{on}}(t) = b_0 k_{\text{on}} \rho_A(t)$ per cell, where b_0 is the number of BCR per cell and k_{on} is the association rate (Fig. 1B). Association is known to be diffusion-limited with typical rates $k_{\text{on}} \sim 10^6$ $\text{M}^{-1}\text{s}^{-1}$ [27] and B-cells are known to have up to $b_0 = 10^5$ BCRs in their membrane [1]. Therefore, differences in antigen binding affinity between different B-cell lineages result primarily from differences in the dissociation rate k_{off}. Human B-cells have dissociation rates in the range $k_{\text{off}} = 10^{-5} - 10^1 \text{s}^{-1}$ [1]. The corresponding equilibrium binding constant, $K = k_{\text{off}}/k_{\text{on}}$, varies in the range $K = 10^{-11} - 10^{-5}$M. This constant is related to the lineage-specific energy gap between the bound and the unbound state, $K = K_0 \exp(\Delta E)$, where K_0 is a normalisation constant and all energies are measured in units of $k_B T$ (Materials and methods). Importantly, with these parameters, the fraction of antigen-bound B-cells remains small throughout the infection process.

B-cell activation. Upon binding, we assume that B-cells undergo a series of p thermodynamically irreversible steps to activation (Fig. 1B). That is, cells in each intermediate state transform to the next state (with rate k_{step}) or unbind from the antigen (with rate k_{off}), but do not revert to the previous state. The stepwise, stochastic activation dynamics is an inhomogeneous Poisson process with an output activation rate $u_{\text{act}}(t)$. In the relevant regime of low antigen concentration, this rate takes the form

$$u_{\text{act}}(K, t) = \left(1 + \frac{K}{K_{\text{step}}}\right)^{-p} u_{\text{on}}(t) \qquad (1)$$

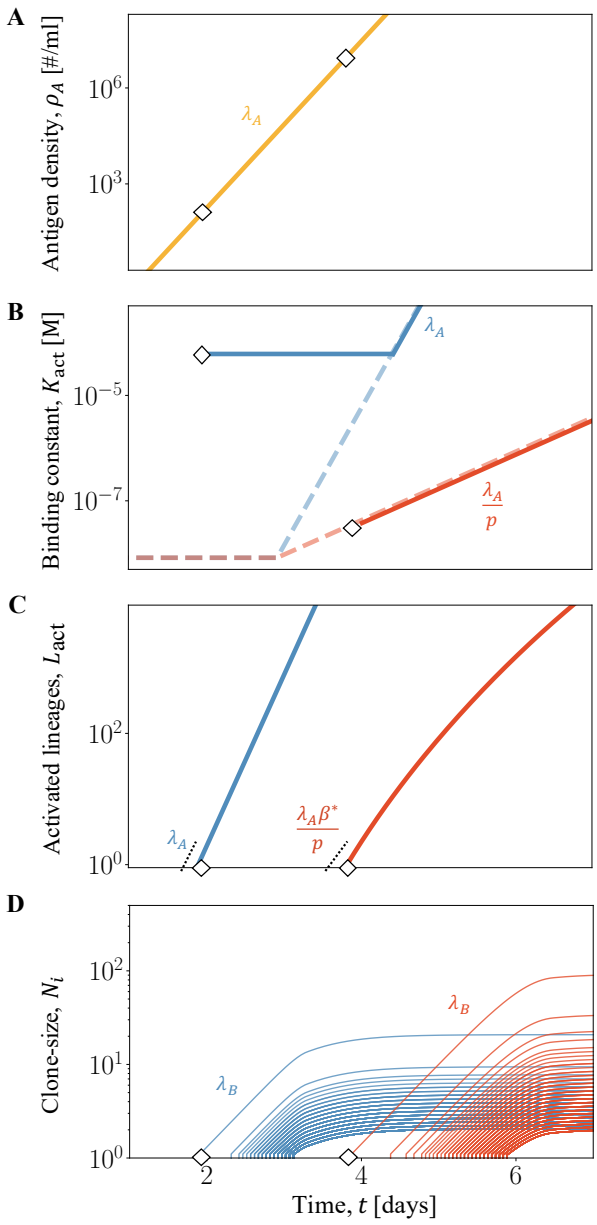

Fig. 2. Generalised Luria-Delbrück replication-activation dynamics. (A) The antigen population, $N_A(t)$, grows exponentially with rate λ_A. Diamonds mark the start of B-cell activation in the low-specificity (LS) and the high-specificity (HS) regime. **(B)** The average binding constant of activation, $K_{\rm act}(t)$, is initially constant (LS, blue) or increases with rate λ_A/p (HS, red). **(C)** The number of activated lineages, $L_{\rm act}(t)$, grows exponentially with initial rates λ_A (LS) and $\beta_{\rm act}\lambda_A/p$ (HS). **(D)** The population of activated lineages, $N_i(t)$ ($i = 1, 2, \ldots,$), shows initially exponential growth with rate λ_B. In the mapping to a classical Luria-Delbrück process, the antigen population corresponds to the wild-type cell population, activation to mutation, and B-cell linages to mutant cell clones under selection. The generalised Luria-Delbrück process couples growth to neutralisation function and repertoire complexity, setting independent growth rates of $K(t)$ and $L(t)$. Analytical results are shown for the following parameters: activation steps, $p = 1$ (LS), $p = 4$ (HS); kinetic parameters, $k_{\rm on} = 10^6 M^{-1} s^{-1}$, $k_{\rm step} = 0.5{\rm min}^{-1}$; number of BCRs per cell, $b_0 = 10^5$; growth rates $\lambda_A = 6{\rm d}^{-1}$, $\lambda_B = 2{\rm d}^{-1}$; repertoire size $L_0 = 10^8$; carrying capacity $\bar{N} = 10^4$.

with $K_{\rm step} = k_{\rm step}/k_{\rm on}$. In the low-affinity regime, the activation rate has the asymptotic form $u_{\rm act}(K,t) \sim (K/K_{\rm step})^{-p} \sim (k_{\rm step}/k_{\rm off})^p$, which can be read off from Fig. 1B: each activation step generates a factor $(k_{\rm step}/k_{\rm off})$ relating the thermodynamic weights of consecutive intermediate states.

Next, we compute the activation probability of a B-cell lineage up to time t, $R(K,t)$. For exponential antigen growth, we find

$$R(t,K) = 1 - \exp\left[-\frac{u_{\rm act}(t,K)}{\lambda_A}\right]; \qquad (2)$$

(Materials and methods). This equation describes a moving front of deterministic activation, $K_1(t) = K_{\rm step}\exp[(\lambda_A/p)(t-t_1)]$, where R reaches values of order 1 (Fig S1). The front starts at affinity $K_{\rm step}^{-1}$ and time $t_1 = \log[(\lambda_A \mathcal{N})/(k_{\rm on} b_0)]/\lambda_A$, where $\mathcal{N}$ is Avogadro's number. With increasing antigen concentration, it moves towards lineages of decreasing antigen affinity at a p-dependent speed (Fig. 2B). Ahead of the front, for $K \gg K_1(t)$, activation of individual lineages is a rare stochastic event. For $p = 1$, activation is asymptotically proportional to the inverse equilibrium constant, or Boltzmann factor, $R \sim K^{-1} \sim \exp(-\Delta E)$. For $p > 1$, the non-equilibrium dynamics of kinetic proofreading generate stronger suppression of activation for weak binders, $R \sim K^{-p}$ [14, 15]. Kinetic proofreading appears to be the simplest mechanism to generate deterministic activation of high-affinity lineages together with strong suppression of low-affinity lineages; mechanisms with reversible antigen-receptor binding have $R \sim K^{-1}$ or remain in the stochastic regime ($R \ll 1$) under the physiological conditions of an early primary infection (Fig. S1, Materials and methods).

Lineage activation marks the onset of the immune response to a new antigen. Activated cells proliferate exponentially with an initial rate λ_B that is comparable to λ_A [28] and shed free antibodies that can neutralise antigens (Fig. 1C). As the activated repertoire grows, cells start to compete for space and resources, including T-cell help [29]. Here we model the clone dynamics as logistic growth, $\dot{N}_j(t) = \lambda_B N_j(t)(1 - \sum_j N_j(t)/\bar{N})$, where $N_j(t)$ is the number of activated cells in clone j and $\bar{N} \sim 10^4$ [28] is a carrying capacity for the total size of the activated repertoire (Fig. 2D; here and below, overbars refer to the repertoire statistics at carrying capacity).

Repertoire response to a given antigen. To characterise the immune repertoire available for primary response against a given antigen, we grade naive B-cell lineages by their affinity to the antigen's binding locus (epitope). We use a simple sequence-specific binding energy model, where epitopes and their cognate BCR are sequence segments, $\mathbf{a} = (a_1, \ldots, a_\ell)$ and $\mathbf{b} = (b_1, \ldots, b_\ell)$. Binding aligns these segments and couples pairs of aligned amino acids, and the

binding energy gap $\Delta E(K) = \log(K/K_0)$ is additive,

$$\Delta E(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^{\ell} \varepsilon(a_k, b_k). \tag{3}$$

For a given antigen, this model defines the density of naive lineages available for activation, $\Omega_0(K) = (K\mathrm{d}/\mathrm{d}K)L_0(K)$, where $L_0(K)$ is the expected number of lineages in an individual with binding constant $< K$ to the epitope $\mathbf{a}$ (Fig. 3). Here we assume that naive repertoires are randomly sampled from an underlying amino acid distribution [3]. Hence, most lineages bind only weakly to a new antigen ($K \sim 0$). The expected minimum binding constant in an individual, K^*, is given by the condition $L_0(K^*) = 1$. This point is to be distinguished from the global minimum for a given antigen, K_m, which often corresponds to a unique BCR genotype $\mathbf{b}_m$ (called the Master sequence). Because individual repertoires cover only a small fraction of the BCR genotype space, the expected maximum antigen affinity remains substantially below the Master sequence ($K^* > K_m$). To characterise the strong-binding tail of the lineage spectrum, we define the micro-canonical entropy, $S(K) = \log \Omega_0(K) + \text{const.}$, and the associated inverse reduced temperature,

$$\beta(K) = \frac{\mathrm{d}S(K)}{\mathrm{d}(\Delta E(K))}. \tag{4}$$

This function measures the exponential increase in lineage density with binding energy, $\Omega_0(K') \sim \exp[\beta(K)\,\Delta E(K')]$, in the vicinity of a given point K. As we will show below, the inverse temperature at maximum binding, $\beta^* = \beta(K^*)$, is a key determinant of the activation dynamics.

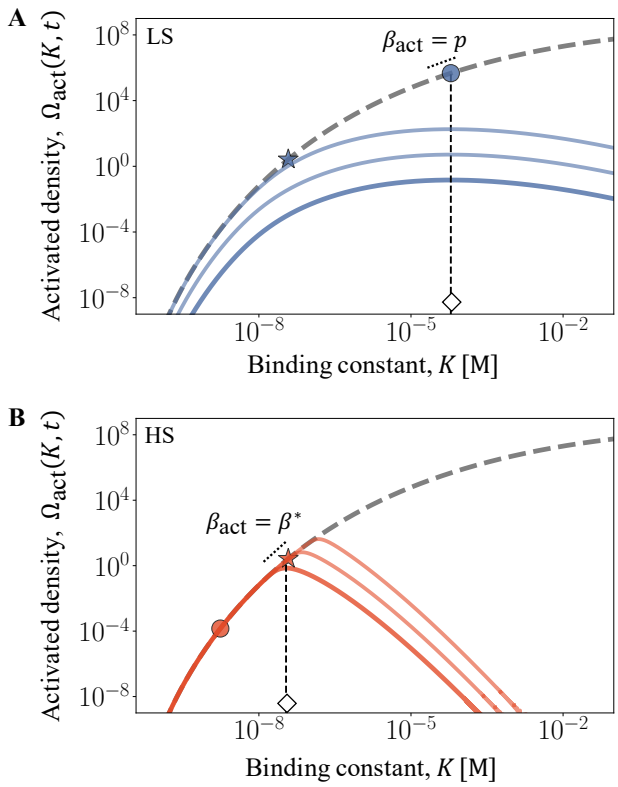

Fig. 3. B-cell response repertoires. Density of naive B-cell lineages, $\Omega_0(K)$ (gray); density of activated lineages, $\Omega(K,t)$, at the start of activation($t = t_{\mathrm{act}}$, thick) and at two subsequent time points (thin) in the **(A)** LS regime ($p = 1$) and **(B)** HS regime ($p = 4$). Specific values of K are marked on the spectral function: K_p (dot); strongest antigen binding in typical individuals, K^* (star); start of activation, K_{act} (open diamond) with initial inverse activation temperature, β_{act} (dotted line). Energy model: TCRen; other parameters as in Fig. 2.

Repertoire response to diverse antigens. How comparable are the response repertoires of different antigens? To address this question, we evaluate the lineage spectrum $\Omega_0(K)$ for a random sample of antigenic epitopes $\mathbf{a}$ (Fig. S2). The interaction energy matrix $\varepsilon(a,b)$ of our main analysis is proportional to the TCRec matrix $\varepsilon(a,b)$ originally inferred for T cell receptors [30]; similar spectra are obtained from the Miyazawa-Jernigan matrix [31] and from normally distributed random energies. For a given antigen, the spectral density depends on broad statistical features of the energy matrix, including the binding length ℓ and the variance of interaction energies, σ_ε^2 (Materials and methods). Here we determine these parameters from observed binding constants $K^* \sim 10^{-7}$M and $K_m \sim 10^{-11}$M of high-affinity antibodies generated in primary infections and of ultra-potent antibodies, respectively [32, 1].

Remarkably, these physiological constraints generate a consistent ensemble of response repertoires. First, the distributions of inferred binding lengths and of the rms. energy variation per site are strongly peaked around values $\ell \sim 20$ and $\sigma_\varepsilon/(\ell^{1/2}) \sim 1$, respectively, which are in tune with known examples (Fig. S2). Second, the lineage densities $\Omega_0(K)$ depend only weakly on the antigen sequence $\mathbf{a}$ and have a nearly universal shape (Fig. S2). In other words, the antigen-averaged spectral density $\Omega_0(K)$ captures the response repertoire available in a typical primary infection. In particular, response repertoires of different antigens with similar K^* and K_m have similar inverse activation temperatures, $\beta^* = (2.5 \pm 0.3)$.

Repertoire activation. The spectral density of naive lineages and the recognition function $R(K,t)$ determine the time-dependent density of activated lineages,

$$\Omega_{\mathrm{act}}(K,t) = \Omega_0(K)\, R(K,t). \tag{5}$$

In Fig. 3, we plot $\Omega_{\mathrm{act}}(K,t)$ at subsequent times for two different numbers of activation steps, with and without proofreading ($p = 1, 4$). 4AB shows simulations of the stochastic B-cell clone population dynamics in a typical individual host. The spectral density of activated lineages is strongly peaked; its high-affinity flank is given by the density of naive lineages, its low-affinity flank by the proofreading-dependent activation dynamics. This function determines two repertoire summary statistics: the expected number of activated lineages, $L_{\mathrm{act}}(t) = \int \Omega_{\mathrm{act}}(K,t)\,\mathrm{d}K/K$, and their

average binding constant $K_{\rm act}(t) = \int K\Omega_{\rm act}(K,t)\,{\rm d}K/K$. Activation starts at an expected time $t_{\rm act}$ given by the condition $L_{\rm act}(t_{\rm act}) = 1$. This sets the initial binding constant $K_{\rm act} \equiv K_{\rm act}(t_{\rm act})$ (marked by diamonds in Fig. 3) and the inverse activation temperature $\beta_{\rm act} \equiv \beta_{\rm act}(K_{\rm act})$ (marked by a tangent line). Importantly, activation has two dynamical regimes.

In the *low-specificity* (LS) regime, for small values of p, activation is peaked on the low-affinity flank of the spectral function (Fig. 3A). The starting point $K_{\rm act} = K_p$ is determined by the condition $\beta_{\rm act} = p$, which follows from the asymptotic form $R \sim K^{-p}$ given by equation (2). In this regime, the number of activation steps, p, determines the specificity of recognition; the activation probability and clone size of individual lineages remains small (Fig. 4A). The LS activation dynamics is characterised by

$$\begin{aligned} \beta_{\rm act} &= p, \\ K_{\rm act}(t) &= K_p, \\ L_{\rm act}(t) &= \exp\left[\lambda_A(t - t_{\rm act})\right], \end{aligned} \qquad \text{(LS)} \qquad (6)$$

as shown in Fig. 2BC; details are given in Materials and methods. This regime ends at a crossover point $p = \beta^*$, where K_p reaches the expected minimum binding constant, K^*.

In the *high-specificity* (HS) regime, for $p > \beta^*$, activation starts at $K_{\rm act} = K^*$ and $\beta_{\rm act} = \beta^*$, then follows the deterministic front $K_{\rm act}(t) = K_1(t)$ (Fig. 3B). Hence, lineages are activated deterministically and in order of decreasing antigen affinity. In this regime, the spectral density of the naive B-cell repertoire determines the specificity of recognition; high-affinity lineages reach substantial clone size (Fig. 4B). We find in the HS activation dynamics

$$\begin{aligned} \beta_{\rm act} &= \beta^*, \\ K_{\rm act}(t) &= K^* \exp\left[\tfrac{\lambda_A}{p}(t - t_{\rm act})\right], \\ L_{\rm act}(t) &= \exp\left[\tfrac{\lambda_A \beta^*}{p}(t - t_{\rm act})\right] \end{aligned} \qquad \text{(HS)} \qquad (7)$$

as shown Fig. 2BC; details are given in Materials and methods. In the next section, we will show that these regimes generate drastically different immune responses. We note that the growth rates λ_A/p and $\beta^*\lambda_A/p$ in equation (7) are specific to generalised Luria-Delbrück processes as defined in this paper (Fig. 2). The analogue of $L_{\rm act}(t)$ in a classical Luria-Delbrück process is the number of mutant clones. Given a constant molecular clock of mutations, this number always grows with rate λ_A, proportionally to the wild-type population size.

Biological consequences

Efficient immune response by kinetic proofreading. As shown in Fig. 3, kinetic proofreading in the HS regime generates a specific profile of B-cell lineages generated in a primary infection: deterministic activation of high-affinity lineages is coupled to strong suppression of low-affinity lineages. To quantify the impact of this profile on immune function, we evaluate the antiserum potency

$$\mathcal{Z}(t) = \sum_j \frac{N_j(t)}{K_j(t)}, \qquad (8)$$

as well as the contributions of individual lineages, $z_j(t) = N_j(t)/K_j$ (the index $j = 1, 2, \ldots$ orders clones by decreasing size). The function $\mathcal{Z}(t)$ sums the antigen affinities of all activated cells and determines the fraction of neutralised virions; it reaches a saturation value $\bar{\mathcal{Z}}$ at carrying capacity of the activated repertoire. We measure potency relative to a reference value $\mathcal{Z}_0$ describing a hypothetical repertoire with homogeneous binding constant K^*. In Fig. 4C, we plot the population average $\langle\bar{\mathcal{Z}}\rangle$ as a function of the number of proofreading steps (here and below, population averages are denoted by brackets). Potency comes close to the reference value in the HS regime, but quickly drops with decreasing p in the LS regime. Without proofreading ($p = 1$), $\langle\bar{\mathcal{Z}}\rangle$ is about 20fold lower than at $p = 4$, argued below to be the approximate number of proofreading steps in human B-cell activation. The difference between activation regimes is even more pronounced for the potency contribution of the largest clone, $\langle\bar{z}_1\rangle$ (Fig. 4C). In the HS regime, where the largest clone is likely also the clone of highest affinity, $\langle\bar{z}_1\rangle$ contributes a substantial fraction of the total potency. There is again a rapid drop in the LS regime; without proofreading, $\langle\bar{z}_1\rangle$ is about 1000fold lower than at $p = 4$.

Another striking difference between the activation regime is in the activation probability of individual lineages, as given by the recognition function $R(K,t)$. In the HS regime, almost all high-affinity lineages get activated (R reaches values close to 1); in the LS regime, most available high-affinity lineages do not get activated and are waisted for pathogen suppression ($R \ll 1$). Together, we conclude that the lineage activation profile generated by kinetic proofreading in the HS regime is a prerequisite for a potent, specific, and efficient primary immune response.

Repertoire-tuned proofreading. Fig. 4AB shows two further functional differences between the activation regimes. In the LS regime, a large number of lineages gets activated, but these clones reach only small population frequencies at carrying capacity, $\bar{x}_j \equiv \bar{N}_j/\bar{N}$. In the HS regime, activation gets increasingly focused on few high-frequency and high-affinity lineages. We can describe the diversity of activated repertoires by the Shannon entropy $\Sigma = -\sum_j \bar{x}_j \log \bar{x}_j$. In the LS regime, the population average $\langle\Sigma\rangle$ is large and varies only weakly; in the HS regime, $\langle\Sigma\rangle$ drops substantially with increasing p (Fig. 4C). Subsequent to activation, a part of the B-cells undergoes affinity maturation in germinal centers. This mutation-selection process produces high-affinity plasma B-cells, as well as a diverse set of memory B-cells. In the HS regime, the larger repertoire diversity found close to the crossover point ($p = \beta^*$) serves both channels of affinity maturation: it

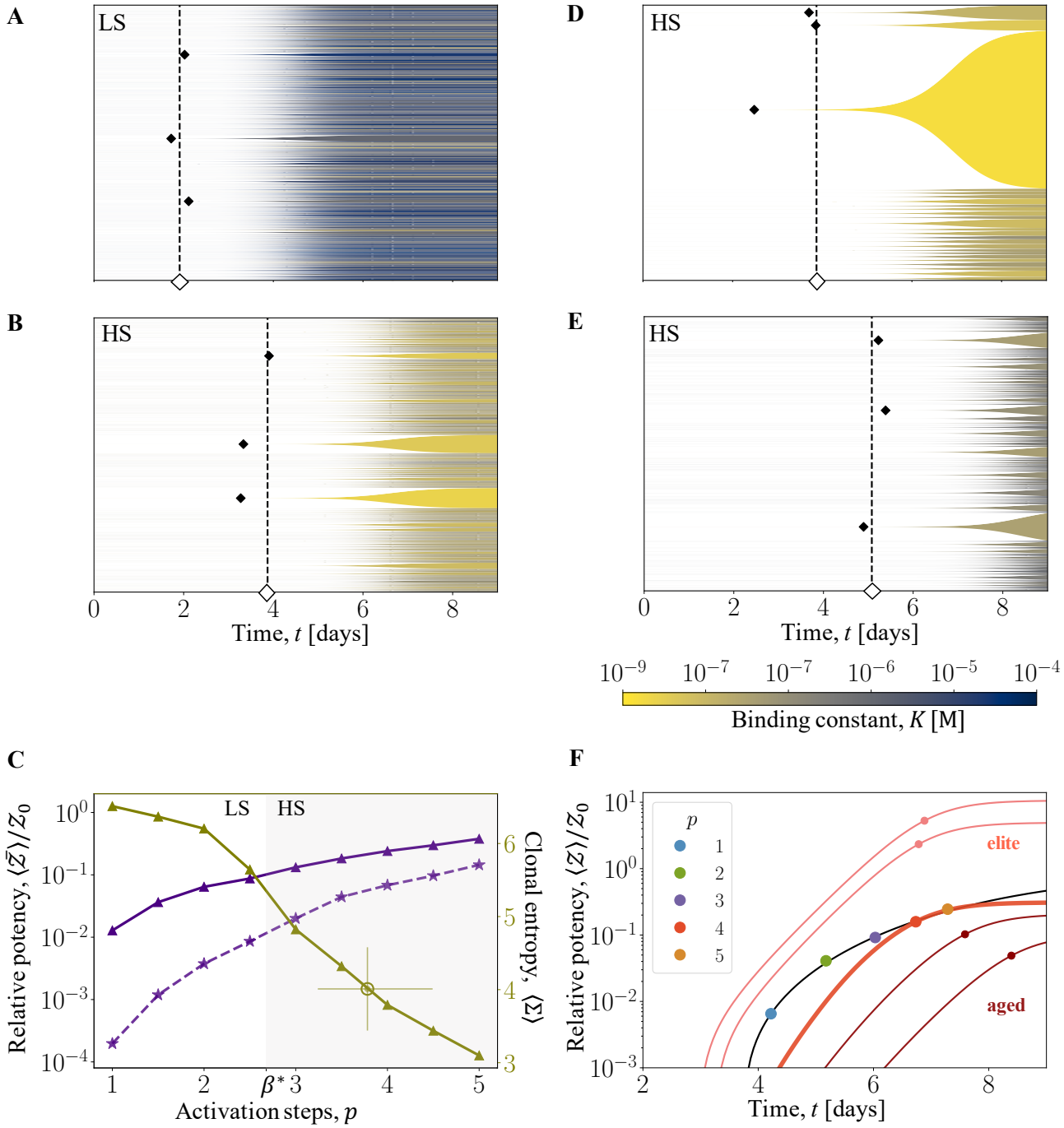

Fig. 4. Proofreading determines potency, speed, and diversity of recognition. (A, B) Activated B-cell repertoires in randomly sampled individuals in the LS regime ($p = 1$) and in the HS regime ($p = 4$). Muller plots with filled areas representing individual clones (height: time-dependent population size, shading: antigen binding constant, K, as given by color bar); activation times of the first few clones (filled diamonds) are shown together with the expectation value $t_{\rm act}$ (open diamond). **(C)** Antiserum potency, $\langle \bar{\mathcal{Z}} \rangle / K_0$, potency component of the largest clone, $\langle \bar{z}_1 \rangle$, and Shannon entropy of the activated repertoire, Σ. Population averages are shown as functions of the number of activation steps, p, at carrying capacity and relative to a reference value $\mathcal{Z}_0 = 1$M. Shading marks the HS regime ($p \gtrsim \beta^*$), an open circle with error bars gives the empirical entropy inferred from the repertoire data of ref. [11]. **(D)** Activated B-cell repertoire in an elite neutraliser occurring at population frequency 10^{-3}. This repertoire is marked by the early activation of a high-affinity "jackpot" clone. **(E)** Activated B-cell repertoire of an aged individual with a 10x reduced number of naive lineages. The onset of activation, $t_{\rm act}$ (black dashed line), is delayed with respect to a full repertoire (panel B). **(F)** Time-dependent antiserum potency, $\langle \mathcal{Z} \rangle(t)$. The population average is shown in the HS regime ($p = 4$, thick red line); the half-saturation point (t_{50}, Z_{50}) is marked by a dot. The family of half-saturation points for different values of p characterises the tradeoff between potency and speed of typical immune responses, which defines a Pareto line (black line). The potency $\mathcal{Z}(t)$ of typical elite neutralisers (at population frequencies 10^{-3} and 10^{-4}, light red lines) is above, that of aged immune systems (at 10x and 100x lineage reduction, dark red lines) is below the population average (HS regime, $p = 3$). Model parameters as in Fig. 2.

facilitates the search for mutational paths towards high-affinity BCR genotypes in plasma cells, and it provides diverse input for memory cell formation [33].

In Fig. 4F we plot the time-dependent, population-averaged potency for different values of p. The response time t_{50}, where $\langle \mathcal{Z} \rangle(t)$ reaches the half-saturation point $\mathcal{Z}_{50} = \langle \bar{\mathcal{Z}}/2 \rangle$, is marked by dots. The p-dependent increase in potency is coupled to an increased time delay of activation, caused by the sequence of intermediate steps and by the constraint to high-affinity lineages. In the HS regime, increasing p yields a diminishing return of potency, while t_{50} continues to increase proportionally to p. Similarly, efficient proofreading requires a sufficiently small activation rate. In the HS regime, for $k_{\text{step}} < K^* k_{\text{on}}$, decreasing k_{step} yields a diminishing return of potency, while t_{50} continues to increase proportionally to $1/k_{\text{step}}$ (Fig. S3). The tradeoff between potency and speed of immune response defines a Pareto surface. This tradeoff, together with the entropy pattern, suggests that optimal immune response is tuned to the spectral density of the naive repertoire: the number and rate of proofreading steps are in the HS regime close to the crossover point $p = \beta^*$ and $k_{\text{step}} = K^* k_{\text{on}}$, respectively.

While there is no direct evidence of proofreading in B-cell activation to date, available data suggests that immune response is in the tuned regime. Specifically, using recent data of clonal B-cell populations in early germinal centres of mice [11], we infer a repertoire entropy $\Sigma = 4.0 \pm 0.6$ (circle in Fig. 4C see Materials and methods), which is consistent with proofreading in the range $p = 3.7[3.2, 4.5]$. This argument is further supported by the inferred clone size statistics, to which we now turn.

Clone size and affinity statistics. B-cell immune repertoires are known to have broad variation of clone sizes, which can be described by power-law distributions [34, 35, 36]. Luria-Delbrück processes provide a simple explanation for such power laws: they relate observables that depend exponentially on time (Fig. 2). First, consider the relation between clone size and probability of occurrence in an individual's repertoire. More lineages are activated later ($L \sim \exp[(\lambda_A \beta_{\text{act}}/p)\, t]$), but these clones reach smaller size ($\bar{N} \sim \exp[-\lambda_B t]$). This relates clone size to rank,

$$\langle \bar{N}_j \rangle \sim j^{-\zeta} \tag{9}$$

with

$$\zeta = \frac{\lambda_B p}{\lambda_A \beta_{\text{act}}} \tag{10}$$

and $\beta_{\text{act}} = \min(p, \beta^*)$; see Materials and methods for details. The index $j = 1, 2, \ldots$ again orders the clones in an individual's repertoire by size. The cumulative distribution aggregated over individuals has the form $\Phi(N) \sim N^{-1/\zeta}$, which is equivalent to equation (9), and spans 3 orders of magnitude in size (Fig. S4, Materials and methods). Simulations confirm these power laws; the clone-rank statistics in a set of randomly chosen individuals follows the same pattern (Fig. 5A, S4). In the HS regime, the exponent ζ increases monotonically as a function of p, reflecting the increasing bias to large clone size generated by proofreading (Fig. 5D).

To test the prediction of equation (9), we use again the mouse repertoire data reported in ref. [11]. The size-rank statistics of clones can be fit to a power law with exponent $\zeta = 0.60 \pm 0.06$ (insert in Fig. 5A). The resulting estimate $p = 4.3 \pm 0.4$ (Fig. 5D) is consistent with the value inferred from the clonal entropy, confirming our inference of proofreading at an intermediate number of steps. Remarkably, clone size distributions extracted from data of human memory B-cell repertoires [37] show power laws with a similar exponent, $\zeta = 0.57 \pm 0.12$ [36]. Given that memory cells are in the same affinity range than activated naive cells [38], this may point to a common dynamical origin; however, our present model contains only the primary activation dynamics and is not directly applicable to memory cells. In refs. [35, 36], the power laws of memory repertoires have been attributed to long-term selection by recurrent infections.

Generalised Luria-Delbrück processes include phenotype (here, antigen affinity) statistics and predict new power laws observable in repertoire data. In the HS regime, activation occurs on a moving front, as given by equation (7). This relates affinity to rank,

$$\langle K_l \rangle \sim l^{1/\beta^*} \qquad \text{(HS)}, \tag{11}$$

where the index l orders clones by decreasing affinity (Fig. 5B, Materials and methods). Equation (11) is again equivalent to a power law in the spectral density, $\Omega_{\text{act}}(K) \sim K^{-\beta^*}$, and is consistent with the affinity-rank statistics in randomly sampled individuals. The exponent $1/\beta^*$ equals the activation temperature of the naive repertoire, as given by equation (4). By combining equations (9) and (11), we obtain a power-law relation between size and affinity,

$$\langle \bar{N}_j \rangle \sim \langle K_j \rangle^{-\zeta\beta^*} \qquad \text{(HS)}, \tag{12}$$

as shown in Fig. 5C. In the HS regime, the size and affinity rankings coincide up to fluctuations, because both are related to time: high-affinity clones get activated before low-affinity clones. The resulting potency-rank relation, $\langle \bar{z}_j \rangle \sim j^{-\zeta - 1/\beta^*}$, is shown in Fig. S4. In the LS regime, the size-affinity correlation is lost. Large clones have affinities of order K_p, high-affinity clones have small size and show a faster decline of affinity with rank than in the HS regime (Fig. 3A, 5BC). Hence, empirical observations of this correlation can provide specific evidence for activation by proofreading in the HS regime.

Elite neutralisers. Generalised Luria-Delbrück immune activation shows particularly pronounced variation between hosts. In the HS regime, a subset of *elite neutralisers* is distinguished by early activation of a single high-affinity clone.

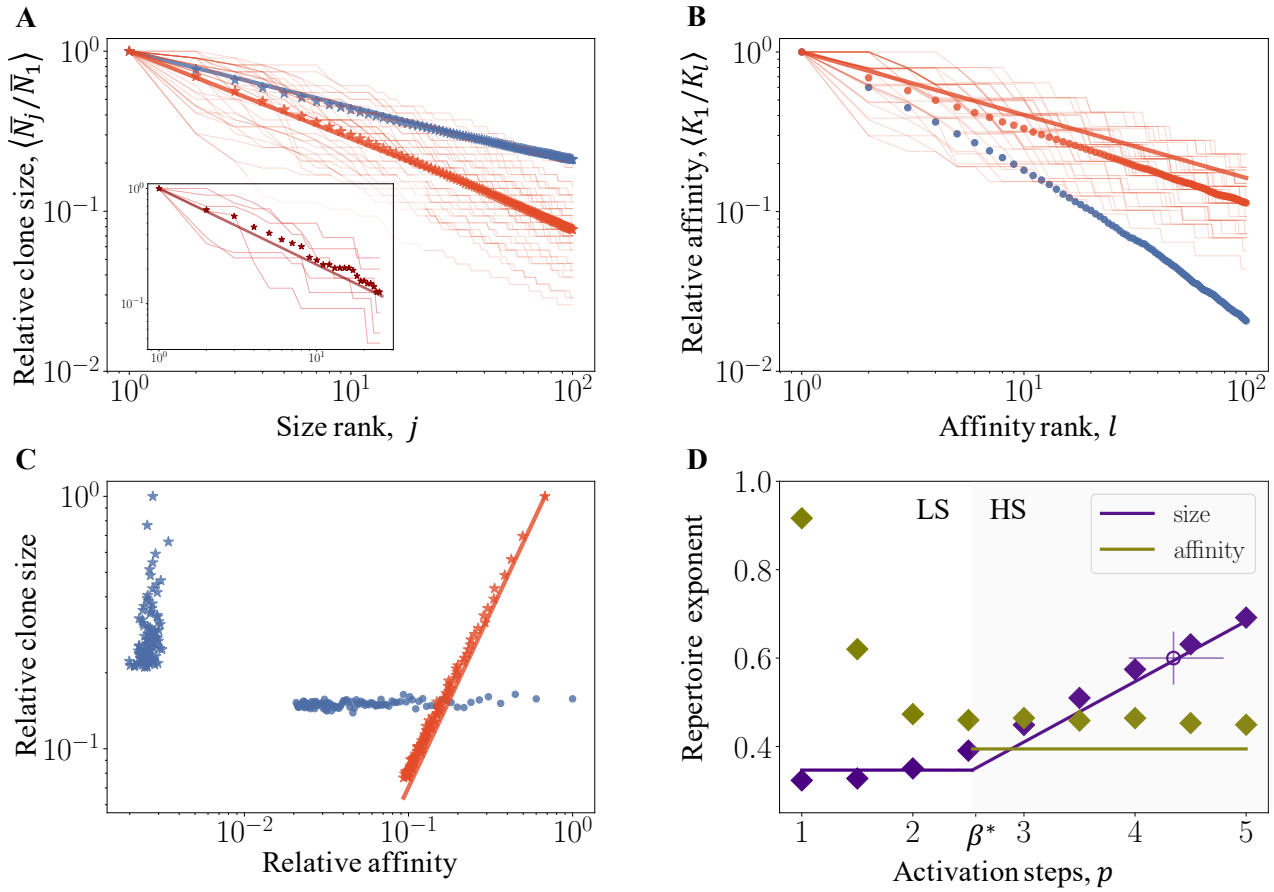

Fig. 5. Clone size and affinity statistics of activated repertoires. **(A)** Clone size statistics. Simulation results for the average relative clone size, $\langle \bar{N}_j/\bar{N}_1 \rangle$, are plotted against the size rank, $j = 1, \ldots 100$, for $p = 1, 4$ (blue and red, respectively). Lines mark the expected power law, $\langle \bar{N}_j \rangle \sim j^{-\zeta}$, as given by equation (9). Thin lines give rankings in a set of randomly chosen individuals. Insert: empirical rank-size relation inferred from the repertoire data of ref. [11]. **(B)** Affinity statistics. Simulation results for the average relative antigen binding constant, $\langle K_l/K_1 \rangle$, are plotted against the affinity rank, $l = 1, \ldots, 100$. A line marks the expected power law in the HS regime, $\langle K_l \rangle \sim l^{1/\beta_{\rm act}}$, as given by equation (11). **(C)** Size-affinity correlation. HS regime: average relative affinity, $\langle K_j/K_1 \rangle$ vs. average relative clone size, $\langle \bar{N}_j/\bar{N}_1 \rangle$ for the largest clones ($j = 1, \ldots, 100$; red stars), together with the expected power law correlation, $\langle \bar{N}_j \rangle \sim \langle K_j \rangle^{-\zeta/\beta^*}$ (red line). LS regime: $\langle K_j/K_1 \rangle$ vs. $\langle \bar{N}_j/\bar{N}_1 \rangle$ for the largest clones ($j = 1, \ldots, 100$; blue stars), and $\langle K_l/K_1 \rangle$ vs. $\langle \bar{N}_l/\bar{N}_1 \rangle$ for high-affinity clones ($l = 1, \ldots, 100$; blue circles), indicating loss of the size-affinity correlation. **(D)** Empirical repertoire exponents from size and affinity rankings are plotted as functions of the number of activation steps, p (diamonds); the high-specificity regime is marked by shading. Lines mark power laws emerging from the Luria-Delbrück proliferation-activation dynamics; their exponents ζ and β^* depend on the proliferation rates of antigen and activated B-cells, the activation temperature, and the number of proofreading steps. An open circle with error bars gives ζ inferred from the repertoire data of ref. [11]. Model parameters as in Fig. 2.

This jackpot clone dominates the activated immune repertoire and generates exceptionally high potency. Fig. 4D shows an example of the activation dynamics that occurs in one of 10^3 individuals, which is to be compared with the pattern in typical individuals (Fig. 4B). Such elite neutralisers are ahead of the Pareto surface of typical immune responses (Fig. 4F). The cumulative distribution $\Phi(\mathcal{Z})$, which gives the fraction of responders with saturation potency $> \mathcal{Z}$, displays two regimes of elite neutralisers (Fig. S4). In the pre-asymptotic regime, the jackpot clone takes a large part but not all of the repertoire ($\bar{N}_1 < \bar{N}$). The pre-asymptotic potency distribution turns out to be dominated by clone size fluctuations, which implies $\Phi(\mathcal{Z}) \sim \mathcal{Z}^{-1/\zeta}$ (Materials and methods). In the asymptotic regime, the jackpot clone dominates the repertoire ($\bar{N}_1 \simeq \bar{N}$); hence, $\Phi(\mathcal{Z})$ is proportional to the naive density of high-affinity clones, $\Phi(\mathcal{Z}) \sim L_0(\bar{N}/\mathcal{Z})$. For example, one in 10^5 individuals has a primary response with potency 100x above average, comparable to a memory immune response carried by affinity maturated B-cell lineages.

Immune ageing. Recent results indicate that the most prominent effect of immune ageing is a decrease in the overall size and diversity of the repertoire [39, 40, 41, 42]. Our model predicts two effects of this decrease: primary immune responses come later and with reduced potency. Simulations of the activation dynamics in an aged repertoire in the HS regime show that the activation is delayed and antigen affinities are reduced compared to a full-size repertoire (Fig. 4E, to be compared with Fig. 4B). The time-dependent potency remains behind the Pareto line

and reaches a reduced value at carrying capacity (Fig. 4F). For a 10fold decrease in repertoire size, t_{50} increases by about 1d and $\langle \bar{\mathcal{Z}} \rangle$ drops to half of its full-size value. Fig. S3 shows the full dependence of potency on repertoire size. Given a reference point in the HS regime, a reduction of size always has a sizeable effect, while an increase eventually induces a cross-over to the LS regime and a diminishing return of potency.

Discussion

A potent adaptive immune response to a new antigen requires the specific activation of immune cells with high affinity to the antigen. Here we have developed a minimal, biophysical model for immune recognition by naive B-cells in a primary infection (Fig. 1). We have shown that active processes of antigen-receptor binding – kinetic proofreading – are an essential part of the initial recognition dynamics: a highly specific response to a new antigen requires proofreading with depth above a threshold value, $p > \beta^*$. Available data of mouse B-cell repertoires are consistent with proofreading close to the specificity threshold, $p \approx \beta^*$, which amounts to at least 3 consecutive proofreading steps (Fig. 4 and 5). As shown by our model, immune responses tuned to this point are close to a functional optimum: they balance potency and speed, and they generate a diverse set of activated clones for subsequent affinity maturation.

Our model maps the proliferation-activation dynamics of a primary infection onto a generalised Luria-Delbrück process (Fig. 2). In this process, fluctuations are coupled to immune function: jackpot clones have large size and high antigen affinity, resulting in high neutralisation potency (Fig. 4). The underlying statistics of activated immune repertoires is characterised by statistical power laws with two basic exponents: the clone size exponent ζ and the inverse activation temperature β^* (Fig. 5). These exponents characterise the proofreading-dependent specificity of activated immune repertoires, as given by equations (9) and (11). Importantly, our model predicts ζ and β^* in terms of independently measurable quantities and without fit parameters. The statistics of activated B-cell repertoires shape clinically important characteristics of primary infections, including the potency drop of aged responders and the increased response of elite neutralisers (Fig. 4).

Power-law statistics of lineages and the occurrence of elite neutralisers are commonly ascribed to antigen-mediated selection on immune repertoires, often through multiple exposures [43, 35, 44, 45, 36]. Here we have shown that similar features can already emerge in the primary immune response to an acute infection, prior to any antigen-mediated selection effects. Following an acute infection, a part of the activated B-cell repertoire is further processed by affinity maturation. How this step reshapes the repertoire statistics will be studied in a subsequent publication.

Repertoire sequencing combined with neutralisation assays can test our model and probe adaptive immune systems in new ways. By recording the power-law rank-size relation of linages in early post-infection B-cell repertoires and measuring the antigen binding constant of these clones, we can extract the corresponding power laws and infer the central parameters of our model: the specificity threshold, β^* and the number of proofreading steps, p (Fig. 5). The parameter β^* measures the density of B-cell lineages close to the maximum-affinity lineage. This density depends on the size of the naive B-cell repertoire and on the complexity of the antigen-receptor binding motif. Both quantities emerge as key determinants of primary immune responses. In contrast, the parameter p characterises the molecular dynamics of antigen recognition. As we have shown, p-step proofreading generates an effective inverse temperature $\beta \approx p$ that measures the specificity gain in a dense repertoire. At the point of optimal recognition, $p \sim \beta^*$, antigen recognition dynamics matches repertoire complexity. From this point, increasing p at a constant repertoire size L produces a diminishing return of proofreading; conversely, increasing L at constant p produces a diminishing return of repertoire size.

The link between repertoire size and antigen recognition machinery has implications for the macro-evolution of adaptive immune systems. The size of the total B-cell repertoire varies drastically across vertebrates, ranging from $\sim 3 \times 10^5$ cells in zebrafish [34] to $\sim 10^{11}$ in humans [1, 2, 3]. Because potency and timeliness of immune responses are likely to be under strong selection, our model says that the functional balance of tuned proofreading, $p \approx \beta^*$, is also a maximum of fitness. Evolutionary changes of the repertoire size are then predicted to occur together with changes of the recognition machinery, which includes number and rate of proofreading steps. Tracing these co-evolutionary dynamics by comparative cross-species studies may provide an avenue to better understand the evolution of complex immune systems.

Acknowledgements

We thank D. Valenzano and A. Ryabova for discussions, M. Karmakar and M. Meijers for a careful reading of the manuscript, and all members of the Lässig lab for input. This work was partially supported by Deutsche Forschungsgemeinschaft (grant CRC 1310, to ML).

Materials and methods

Antigen-receptor interaction models. Here we model the binding (free) energy between an antigen with epitope sequence $\mathbf{a} = (a_1, \ldots, a_\ell)$ and a B-cell receptor (BCR) sequence $\mathbf{b} = (b_1, \ldots, b_\ell)$ as an additive function,

$$\Delta E(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^{\ell} \varepsilon(a_k, b_k). \tag{13}$$

This function includes entropy contributions of non-translational degrees of freedom; i.e., rotations and elastic deformations of the molecules involved. We use three established models of amino acid interactions $\varepsilon(a, b)$: the TCRec model originally inferred for T cell receptors [30], the Miyazawa-Jernigan matrix [31], and normally distributed random energies. In each case, we introduce a scale factor that is inferred from measured BCR-antigen binding energies (see below). The zero point of ΔE, by definition, corresponds to a reference antigen-BCR pair with equilibrium binding constant $K_0 = 1\text{M}$. In this gauge, the binding energy and the dissociation constant K of an arbitrary pair are related by

$$\frac{K(\mathbf{a}, \mathbf{b})}{K_0} = \exp\left[\frac{\Delta E(\mathbf{a}, \mathbf{b})}{k_B T}\right], \tag{14}$$

where k_B is Boltzmann's constant. In the main text and below, we express all energies in units of $k_B T$ at a fixed, physiological temperature T.

Density of BCR states. To characterise the naive B-cell repertoire available for response to a given antigen $\mathbf{a}$, we use the density of lineage states defined by a unique BCR sequence $\mathbf{b}$,

$$\Omega_0(\Delta E) = \frac{\mathrm{d}}{\mathrm{d}\Delta E} L_0(\Delta E), \tag{15}$$

where $L_0(\Delta E)$ is the expected number of lineages in an individual with binding constant $K < \exp(\Delta E)$ to the epitope $\mathbf{a}$ (we suppress the dependence on $\mathbf{a}$ in this paragraph). By equation (14), this form is equivalent to the definition given in the main text, $\Omega_0(K) = (K\mathrm{d}/\mathrm{d}K)L_0(K)$. We further define the micro-canonical entropy

$$S(\Delta E) = \log \Omega_0(K) \tag{16}$$

and the associated micro-canonical inverse temperature,

$$\beta(\Delta E) = \frac{\mathrm{d}S(\Delta E)}{\mathrm{d}\Delta E}, \tag{17}$$

a parameter that is independent of the physiological temperature T appearing in equation (14). Because we measure energies in units of $k_B T$, the parameter β gives the inverse temperature in units of $(k_B T)^{-1}$. To compute these micro-canonical quantities, we evaluate the canonical partition function,

$$Z(\beta) = \sum_{\mathbf{b}} \exp[-\beta \Delta E(\mathbf{a}, \mathbf{b})] = \prod_{i=1}^{\ell} \sum_{b} \exp[-\beta \varepsilon(a_i, b)], \tag{18}$$

which depends on β as an independent parameter. This function defines the canonical binding energy,

$$\Delta E_c(\beta) = -\frac{\mathrm{d}}{\mathrm{d}\beta} \log Z(\beta), \tag{19}$$

which is an expectation value in the ensemble (18), and the associated entropy,

$$S_c(\beta) = \log Z(\beta) + \beta \Delta E_c(\beta). \tag{20}$$

We invert the relation (19) to write the inverse temperature as a function of the binding energy, $\beta(\Delta E_c)$, and we substitute this function into equation (20) to obtain $S_c(\Delta E_c)$. Upon equating $\Delta E = \Delta E_c$, these functions provide an excellent approximation to their micro-canonical counterparts $\beta(\Delta E)$ and $S(\Delta E)$, as given by equations (16) and (17). Thus, the canonical formalism provides an efficient way to compute the density of states, $\Omega_0(\Delta E)$, for the system at hand.

Antigen-receptor ensembles. To compare the response repertoires for different antigens, we evaluate the BCR lineage density $\Omega_0(K, \mathbf{a})$ for a random sample of epitope sequences $\mathbf{a}$. For a given amino acid interaction matrix and a given antigen, our energy model has two free parameters, the binding length ℓ and the scale factor of the energy, which sets the energy variance $\sigma_\varepsilon^2 = \sum_k [\sum_b \varepsilon^2(a_k, b)/20 - (\sum_b \varepsilon(a_k, b)/20)^2]$. Here we calibrate these parameters by tuning the minimum binding constant expected in an individual repertoire and the global minimum binding constant to observed values of typical high-affinity antibodies generated in primary infections and of ultra-potent antibodies, $K^* \approx 10^{-7}$M and $K_m \approx 10^{-11}$M [32, 1]

The resulting ensemble of response repertoires has the following properties (Fig. S2): (i) The distributions of inferred binding lengths and of the rms. energy variation per site are strongly peaked around values $\ell \sim 20$ and $\sigma_\varepsilon/\ell^{1/2} \sim 1$. (ii) A higher energy variance per site can be traded for a shorter binding length, consistent with a constraint on the total energy variance σ_ε^2. (iii) The lineage densities $\rho_0(K)$ depend only weakly on the antigen sequence $\mathbf{a}$ and have similar inverse activation temperatures, $\beta^* \equiv \beta(K^*) = 2.5 \pm 0.3$.

Activation of B-cells by kinetic proofreading. An early infection is characterised by low densities of antigens and B-cells. Accordingly, we model the activation of individual B-cells upon functional binding with a single antigen. We assume that activation of antigen-bound cells requires a chain of p irreversible steps with characteristic rate k_{step} (Fig. 1). Hence, the B-cell activation rate takes the form

$$u_{\text{act}}(K, t) = u_{\text{on}}(t)\, p_a(K). \tag{21}$$

Here, the association rate per B-cell, $u_{\text{on}}(t) = b_0 k_{\text{on}} \rho_A(t)$ is proportional to the number of receptors per cell, b_0, the diffusion-limited association rate to a single B-cell receptor, k_{on}, and the antigen density, $\rho_A(t)$. The probability of activation after association is given by

$$p_a = \left(\frac{k_{\text{step}}}{k_{\text{step}} + k_{\text{off}}}\right)^p = \left(\frac{1}{1 + K/K_{\text{step}}}\right)^p, \tag{22}$$

where $k_{\text{off}} = K k_{\text{on}}$ and $k_{\text{step}} = K_{\text{step}} k_{\text{on}}$. At each intermediate state, the antigen can dissociate or undergo the next activation step; these alternatives are independent Poisson processes with rates k_{off} and k_{step}, respectively. Hence, the next activation step occurs before dissociation with probability $k_{\text{step}}/(k_{\text{step}} + k_{\text{off}})$.

The probability that a lineage gets activated up to time t is

$$R(K, t) = 1 - \exp\left[\int_0^t u_{\text{act}}(K, t')\, \mathrm{d}t'\right] = 1 - \exp\left[-\frac{u_{\text{act}}(K, t)}{\lambda_A}\right], \tag{23}$$

as given by equation (2) of the main text. Here we have used that antigens proliferate exponentially, $\rho_A(t) \sim \exp(\lambda_A t)$. At early times, activation is association-limited and rare for all K,

$$R(K, t) \simeq \frac{b_0 k_{\text{on}} \rho_A(t)}{\lambda_A (1 + K/K_{\text{step}})^p} \tag{24}$$

$$< \exp\left[-\frac{(t_1 - t)}{\lambda}\right], \qquad (t < t_1) \tag{25}$$

where

$$t_1 = \frac{1}{\lambda_A} \log\left(\frac{\lambda_A \mathcal{N}}{k_{\text{on}} b_0}\right), \tag{26}$$

and $\mathcal{N}$ is Avogadro's number. For $t > t_1$, we have

$$R(K, t) \simeq \begin{cases} 1 & (t > t_1, K < K_1(t)) \\ \left(\dfrac{K}{K_1(t)}\right)^{-p} & (t > t_1, K > K_1(t)), \end{cases} \tag{27}$$

with

$$K_1(t) = K_{\text{step}} \exp\left[\frac{\lambda_A}{p}(t - t_1)\right]. \tag{28}$$

That is, deterministic activation of B-cell lineages occurs along a moving front, $K_1(t)$. Ahead of the front, activation is strongly suppressed by proofreading.

Alternative models of B-cell activation. To highlight the specific role of kinetic proofreading in the activation of naive B-cells, we compare the proposed activation mechanism to alternative mechanisms with reversible binding kinetics (Fig. S1). The corresponding rates govern transitions between unbound and intermediate antigen-bound B-cell states (marked by grey shading in Fig. 1 and Fig. S1). We note that all activation mechanisms have at least one irreversible step: the last transition to exponential proliferation and antibody production (marked by green shading).

In an early primary infection, the activation of B-cells takes place under specific physiological conditions: (i) The antigen density and, hence, the equilibrium occupancy of B-cells remains low ($\rho_A/K \lesssim 10^{-1}$, given $\rho_A \approx 10^{11}/\mathrm{l} \lesssim 10^{-13}\mathrm{M}$ and $K \gtrsim K^* \approx 10^{-7}\mathrm{M}$). These conditions differ drastically from the densities in confined spaces, e.g., lymph nodes and germinal centers, which are relevant for the binding kinetics of presented antigens. (ii) The association kinetics of virions and plasma B-cells is believed to be diffusion-limited; i.e., it takes place at a homogeneous rate k_{on} [46]. This excludes mechanisms for specific recognition by formation of immunological synapses, which have been proposed for presented antigen and operate by modulation of an activation-limited rate k_{on} [47, 48, 49, 17, 50]. (iii) For efficient activation, the rate k_{step} cannot be smaller than all other transition rates, as it is usually assumed in models of kinetic proofreading [14, 15]. Tuned rates discussed in the main text are of order $k_{\mathrm{step}} = 4 \cdot 10^{-3}\mathrm{s}^{-1}$, which implies $k_{\mathrm{step}} \gtrsim k_{\mathrm{off}}$ for high-affinity B-cell lineages. In this regime, the activation dynamics at a given antigen density ρ_A is close to a non-equilibrium steady state, even if the binding kinetics satisfies detailed balance. We consider two specific classes of models with diffusion-limited association and reversible binding kinetics:

- Activation via an excited intermediate state. This process is a reversible analogue of the proofreading dynamics discussed in the main text. Allowed transitions are between the unbound state and the primary bound state (with rates $k_{\mathrm{on}}, k_{\mathrm{off}}$) and between the primary bound state and the excited intermediate state (with rates k_e^+, k_e^-), as shown in Fig. S1. Using detailed balance, the antigen-bound states have reduced binding energies $\Delta E = \log(K/K_0)$ and $\Delta E_e = \Delta E + \Delta\Delta E_e$ respectively, where $K = k_{\mathrm{off}}/k_{\mathrm{on}}$ and $\Delta\Delta E_e = \log(k_e^-/k_e^+)$. Like the proofreading model, this reversible model has a deterministic activation front given by equations (26) and (28). However, the activation probability is asymptotically proportional to the equilibrium occupancy of the intermediate state,

$$R(K,t) \simeq b_0 k_{\mathrm{step}}\, \rho_A(t) \exp\left(-\Delta E_e\right) \sim \frac{\rho_A(t)}{K} \qquad (K \gg K_1(t)), \tag{29}$$

  leading to weak suppression of low-affinity lineages [14, 15].

- Activation by cooperative binding. In this process, activation requires binding of two or more virions to receptors of the same B-cell, which has been observed for antigens actively transported to lymph nodes and presented to B-cells [48, 49, 51]. A bound state of p virions has the reduced binding energy $\Delta E_p = (p\Delta E + \Delta\Delta E_p) = \log(K_p/K_0)$, where $\Delta E_p = \log(K/K_0)$ is the single-particle binding energy and $\Delta\Delta E_p$ is the contribution of cooperative binding. Fig. S1 shows the case $p = 2$, where $\Delta\Delta E_p = \log(k'_{\mathrm{off}}/k_{\mathrm{off}})$. In the cooperative binding model with detailed balance, the asymptotic activation probability is proportional to the equilibrium occupancy of the p-virion bound state,

$$R(K,t) \simeq b_0 k_{\mathrm{step}}\, \rho_A^p(t) \exp\left(-\Delta E_p\right) \sim \frac{\rho_A^p(t)}{K^p} \qquad (K_p \gg K_{\mathrm{step}}), \tag{30}$$

  with $K_{\mathrm{step}} = k_{\mathrm{step}}/k_{\mathrm{on}}$. For $p > 1$, this model leads to stronger suppression of low-affinity lineages; however, at the low antigen concentrations of an early infection, even high-affinity naive lineages do not reach deterministic activation ($R \ll 1$ for $\rho_A(t)/K \lesssim 10^{-1}$).

We conclude that the kinetic proofreading mechanism of B-cell activation introduced in the main text is the simplest model to generate deterministic activation of high-affinity lineages together with strong suppression of low-affinity lineages under the physiological conditions of an early primary infection.

Activation statistics of B-cell repertoires. Given the density of naive B-cell lineages, $\Omega_0(K)$, and the activation probability $R(t,K)$, we can evaluate the density of activated lineages,

$$\Omega_{\mathrm{act}}(K,t) = \Omega_0(K)\, R(K,t), \tag{31}$$

and the total number of activated lineages, $L_{\mathrm{act}}(t) = \int \Omega_{\mathrm{act}}(K,t)\, \mathrm{d}(\log K)$. Two activation regimes emerge:

- Low-specificity (LS) regime. In this regime, the specificity of activation is limited by the number of proofreading steps. According to equation (24), the function Ω_{act} is strongly peaked at a value K_p defined by the condition

$\beta(K_p) = p$ (Fig. 3A). Integrating this function yields the expected number of activated lineages at time t,

$$\begin{aligned} L_{\rm act}(t) &= \int \Omega_{\rm act}(K,t)\,{\rm d}(\log K) \\ &\approx \frac{b_0 k_{\rm on}\exp\left[\lambda_A t\right]}{\mathcal{N}} \int \frac{\Omega_0(K)}{(1+K/K_{\rm step})^p}\,{\rm d}(\log K) \\ &= \exp\left[\lambda_A(t-t_{\rm act})\right] \qquad \text{(LS)}, \end{aligned} \tag{32}$$

where $t_{\rm act}$ is given by the condition $L_{\rm act}(t_{\rm act}) = 1$.

- High-specificity (HS) regime. In this regime, the specificity of activation is limited by the complexity of the naive repertoire. According to equation (27), the function $\Omega_{\rm act}$ is peaked around the moving front, $K_1(t)$ (Fig. 3B). In this case, we obtain

$$\begin{aligned} L_{\rm act}(t) &= \int \Omega_{\rm act}(K,t)\,{\rm d}(\log K) \\ &= \Omega_0(K_1(t)) \times \int \frac{\Omega_0(K)}{(\Omega_0(K_1(t))(1+K/K_{\rm step})^p}\,{\rm d}(\log K) \qquad (33) \\ &= \exp\left[\frac{\lambda_A \beta^*}{p}(t-t_{\rm act})\right] \qquad \text{(HS)}. \qquad (34) \end{aligned}$$

  Here we use equations (16) and (28), and we note that the integrand in (33) has a peak value of order 1 and depends only weakly on t.

These results are given in equations (6) and (7) of the main text. In the activation dynamics discussed here, we assume that genetically distinct B cell lineages are also distinguishable in terms of their antigen binding affinity. Specifically, in our energy models, random mutations generate a binding energy change of order $k_B T$. If the sequence-energy map is highly degenerate, multiple activations occurring in sequence clusters of very similar antigen affinity can generate new scaling regimes.

Statistics of clone size. Here we compute the cumulative distribution function (CDF) of clone size, $\Phi(N,t)$, which is defined as the fraction of activated clones with size $> N$ at time t. Given exponential growth with rate λ_B, this function is given by the fraction of lineages activated before a time $t'(N)$,

$$\Phi(N,t) = \frac{L_{\rm act}(t'(N))}{L_{\rm act}(t)} \qquad \text{with } t'(N) = t - \frac{\log N}{\lambda_B}. \tag{35}$$

Using equations (32) and (34), we obtain

$$\Phi(N,t) = \exp\left[\zeta(t'(N) - t_{\rm act})\right] \sim N^{-1/\zeta}, \tag{36}$$

where the exponent ζ is defined in equation (10) of the main text. In a similar way, we compute the expected size of the j-th largest clone, $\langle N_j\rangle(t)$ ($j = 1, 2, 3, \dots$). We write $\langle N_j\rangle(t) \sim \exp[\lambda_B(t - t_j)]$, where the activation time t_j is given by the condition $L_{\rm act}(t_j) = j$. Using again equations (32) and (34), we have $t_j \sim \zeta \log j$ and

$$\langle N_j\rangle(t) \sim j^{-\zeta}. \tag{37}$$

Equations (36) and (37) are related by Zipf's law. Both are independent of t and, hence, valid also for the saturation clone sizes, $\bar{N}_j$, as used in equation (9) of the main text.

Empirical clone size statistics in early germinal centers. We analyze sequencing data of early germinal centers (GCs) from Tas et al. [11] as a proxy for the initial population of activated B-cell clones. We use CDR3 sequence counts to estimate the clone size of the different B-cell lineages taking part in the response. The dataset contains samples from 6 dissected lymph nodes, each of which contains 2 GCs. We calculate the expected scaled clone size of the j-th largest clone found in each lymph node, N_j/N_1 (thin lines in Fig. 4A, insert) and the average value over all 6 lymph nodes (thick line in Fig. 4A, insert). We also calculate the clonal entropy $\Sigma = -\sum_j x_j \log x_j$ with $x_j = N_j/N$ and $N = \sum_j N_j$ (the sum runs over all clones in the dataset).

Statistics of antigen affinity. We now evaluate the CDF of antigen binding constants, $\Phi(K,t) = L_{\rm act}(K,t)/L_{\rm act}(t)$, in the HS regime. For the high-affinity tail of this function, $K \gtrsim K^*$, activation occurs deterministically, which implies $L_{\rm act}(K,t) \simeq L_0(K)$. Recalling that $\Omega_0(K) = (K{\rm d}/{\rm d}K)L_0(K)$, we obtain $\Phi(K) \sim L_0(K) \sim \Omega_0(K)$. Hence,

$$\Phi(K,t) \sim K^{\beta^*} \qquad \text{(HS)}. \tag{38}$$

Using Zipf's law, as for the clone size, we obtain the expectation value of the l-th lowest binding constant,

$$\langle K_l \rangle(t) \sim l^{1/\beta^*} \qquad \text{(HS)}, \tag{39}$$

as given in equation (11) of the main text. Again, this relation is independent of t and valid also at the saturation point. Because activation occurs on the moving front $K_1(t)$, the clone rankings by size and affinity are equivalent up to fluctuations. Hence, by combining equations (37) and (39), we obtain a power law relating size and affinity

$$\langle \bar{N}_j \rangle \sim \langle K_j \rangle^{-\zeta\beta^*} \qquad \text{(HS)}. \tag{40}$$

In the LS regime, there is no clear power law relation between affinity and rank (Fig. 5B). High-affinity activated clones span the range between K^* and K_p and show a faster decline of affinity with rank than in the HS regime (Fig. 5BC). Hence, empirical exponents fitted to affinity-rank data take values $> 1/\beta^*$ (Fig. 5D).

Potency statistics and elite neutralisers. In the HS regime, equations (37) and (39) also determine the statistics of single-clone potencies $\bar{z}_j = \bar{N}_j/K_j$,

$$\langle \bar{z}_j \rangle \sim j^{-\zeta-1/\beta^*} \qquad \text{(HS)}. \tag{41}$$

For typical individuals, many clones contribute to the total potency $\bar{\mathcal{Z}} = \sum \bar{z}_i$. However, a characteristic of Luria-Delbrück-Delbrück models is the existence of giant fluctuations. In the HS regime of the model proposed here, there is a set of elite neutralisers singled out by early activation of their first clone. These "jackpot" clones have simultaneously high affinity and large size, which takes a sizeable fraction of the total activated repertoire, $\bar{N}_1 \lesssim \bar{N}$ (Fig. 4D). The CDF of potency, $\Phi(\mathcal{Z})$, is defined as the fraction of individuals with $\bar{\mathcal{Z}} < \mathcal{Z}$. For $\bar{\mathcal{Z}} \gg \langle \bar{\mathcal{Z}} \rangle$, potency is dominated by jackpot clones, $\bar{Z} \simeq \bar{z}_1$. We find two scaling regimes (Fig. S4C). In the pre-asymptotic regime ($\bar{N}_1 < \bar{N}$), size fluctuations of the jackpot clone are dominant, and we can write $\bar{z}_1 \approx \bar{N}_1/K^*$. Hence, by equation (36), the CDF of potency takes the form

$$\Phi(\mathcal{Z}) \sim \mathcal{Z}^{-1/\zeta} \qquad \text{(HS, pre-asymptotic regime)}. \tag{42}$$

In the asymptotic regime ($\bar{N}_1 \approx \bar{N}$), affinity fluctuations are dominant, and we have $\bar{z}_1 \approx \bar{N}/K$. Hence,

$$\Phi(\mathcal{Z}) \sim L_0(\bar{N}/\mathcal{Z}) \qquad \text{(HS, asymptotic regime)}. \tag{43}$$

Model parameters and numerical simulations. In the analytical and numerical analysis, we use the following empirical parameters for the activation process and the B-cell repertoire: (i) Size of the naive repertoire: $L_0 = 10^9$ lineages [2, 4, 3, 1]. (ii) Growth rate of the antigen population: $\lambda_A = 6{\rm d}^{-1}$ [5, 6, 7, 8]. (iii) Proliferation rate of activated B cells: $\lambda_B = 2{\rm d}^{-1}$ [28]. (iv) Activation step rate: $k_{\rm step} = 0.5{\rm min}^{-1}$. As discussed in the main text, this value is tuned to a repertoire size $L_0 = 10^9$. (v) Antigen-BCR diffusion-limited association rate, $k_{\rm on} = 10^6{\rm M}^{-1}s^{-1}$ [27]. (vi) Carrying capacity of activated B cells: $\bar{N} = 10^4$ cells [28]. In the optimality analysis of tuned repertoires Fig. S3, we vary p, $k_{\rm step}$, and L_0 around the LS-HS crossover point given by $p \sim \beta^*(L_0)$, $k_{\rm step} \approx k_{\rm on}K^*(L_0)$.

To simulate a primary B-cell response, we start by creating an initial population of L_0 B-cell lineages. Each B-cell lineage is represented by a BCR sequence of length ℓ, randomly drawn with the set of 20 amino acids. Assuming a deterministic expanding antigen concentration as defined in the main text, we calculate the non-homogeneous activation rate of each B-cell lineage, $u_{\rm act}(t,K)$, as given in equation 21. Here we use the fact that the time for the first event, t_1, in a non-homogeneous Poisson process with rate $u(t)$ is distributed according to

$$P(t_1) = \frac{1}{\mathcal{N}} u(t_1) \exp\left[-\int_0^{t_1} u(t)dt\right]. \tag{44}$$

with $\mathcal{N} = \int_0^\infty P(t)dt$. We sample then the activation time of each B-cell lineage by sampling uniformly distributed random numbers $r \sim \mathcal{U}[0,1]$ and then using the inverse of the cumulative version of equation 44. Once we have all activation times, we proceed to determine the clone size of each of the B-cell clones. Here we integrate $L_{\rm act}$ coupled differential equations

$$\dot{N}_{\mathbf{b}}(t) = \lambda_B N_{\mathbf{b}}(t)\left(1 - \frac{\sum_{\mathbf{b}} N_{\mathbf{b}}(t)}{\bar{N}}\right), \tag{45}$$

assuming all clones start with clone size $N_{\mathbf{b}}(0) = 1$. We neglect all B-cell clones whose final clone size is smaller than $\bar{N}_{\mathbf{b}} = 2$, corresponding to less than one cell division. All notebooks, codes and input files used in this work are available at `https://github.com/rmorantovar/primary_immune_response` .

References

1. Altan-Bonnet, G., Mora, T. & Walczak, A. M. Quantitative immunology for physicists. *Physics Reports* **849,** 1–83. ISSN: 03701573 (2020).
2. Glanville, J. *et al.* Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 20216–20221. ISSN: 00278424 (48 2009).
3. Elhanati, Y. *et al.* Inferring processes underlying B-cell repertoire diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370.** ISSN: 14712970 (1676 2015).
4. Morbach, H., Eichhorn, E. M., Liese, J. G. & Girschick, H. J. Reference values for B cell subpopulations from infancy to adulthood. *Clinical and Experimental Immunology* **162,** 271–279. ISSN: 13652249 (2 2010).
5. Smith, A. M., Adler, F. R. & Perelson, A. S. An accurate two-phase approximate solution to an acute viral infection model. *Journal of Mathematical Biology* **60,** 711–726. ISSN: 03036812 (5 2010).
6. Pawelek, K. A. *et al.* Modeling within-host dynamics of influenza virus infection including immune responses. *PLoS Computational Biology* **8,** 1–13. ISSN: 15537358 (6 2012).
7. Goyal, A., Reeves, D. B., Cardozo-Ojeda, E. F., Schiffer, J. T. & Mayer, B. T. Viral load and contact heterogeneity predict sars-cov-2 transmission and super-spreading events. *eLife* **10,** 1–27. ISSN: 2050084X (2021).
8. Sender, R. *et al.* The total number and mass of SARS-CoV-2 virions. *Proceedings of the National Academy of Sciences of the United States of America* **118,** 1–9. ISSN: 10916490 (25 2021).
9. Goodnow, C. C., Vinuesa, C. G., Randall, K. L., MacKay, F. & Brink, R. Control systems and decision making for antibody production. *Nature Immunology* **11,** 681–688. ISSN: 15292908 (8 2010).
10. Lam, J. H. & Baumgarth, N. The Multifaceted B Cell Response to Influenza Virus. *The Journal of Immunology* **202,** 352–359. ISSN: 0022-1767 (2 2019).
11. Tas, J. M. *et al.* Visualizing antibody affinity maturation in germinal centers. *Science* **351,** 1048–1054. ISSN: 10959203 (6277 2016).
12. Cancro, M. P., Gerhard, W. & Klinman, N. R. The diversity of the influenza-specific primary B-cell repertoire in BALB/c mice*. *Journal of Experimental Medicine* **147,** 6577–6581. ISSN: 15409538 (3 1979).
13. Perelson, A. S. & Oster, G. F. Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-non-self discrimination. *Journal of Theoretical Biology* **81,** 645–670. ISSN: 10958541 (4 1979).
14. Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences of the United States of America* **71,** 4135–4139. ISSN: 00278424 (10 1974).
15. Ninio, J. Kinetic amplification of enzyme discrimination. *Biochimie* **57,** 587–595. ISSN: 61831638 (5 1975).
16. Mckeithan, T. W. Kinetic proofreading in T-cell receptor signal transduction. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 5042–5046. ISSN: 00278424 (11 1995).
17. Goldstein, B., Faeder, J. R. & Hlavacek, W. S. Mathematical and computational models of immune-receptor signalling. *Nature Reviews Immunology* **4,** 445–456. ISSN: 14741733 (6 2004).
18. François, P., Voisinne, G., Siggia, E. D., Altan-Bonnet, G. & Vergassola, M. Phenotypic model for early T-cell activation displaying sensitivity, specificity, and antagonism. *Proceedings of the National Academy of Sciences of the United States of America* **110,** E888–E897. ISSN: 00278424 (10 2013).
19. Pettmann, J. *et al.* The discriminatory power of the t cell receptor. *eLife* **10,** 1–42. ISSN: 2050084X (2021).
20. Tsourkas, P. K., Liu, W., Das, S. C., Pierce, S. K. & Raychaudhuri, S. Discrimination of membrane antigen affinity by B cells requires dominance of kinetic proofreading over serial engagement. *Cellular and Molecular Immunology* **9,** 62–74. ISSN: 16727681 (1 2012).
21. Harwood, N. E. & Batista, F. D. New Insights into the Early Molecular Events Underlying B Cell Activation. *Immunity* **28,** 609–619. ISSN: 10747613 (5 2008).

22. Butler, T. C., Kardar, M. & Chakraborty, A. K. Quorum sensing allows T cells to discriminate between self and nonself. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 11833–11838. ISSN: 00278424 (29 2013).

23. Luria, S. E. & Delbrück, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28,** 491–511. ISSN: 0016-6731 (6 1943).

24. Lea, D. E. & Coulson, C. A. The distribution of the numbers of mutants in bacterial populations. *Journal of Genetics* **49,** 264–285. ISSN: 00221333 (1949).

25. Mandelbrot, B. A population birth-and-mutation process, I: explicit distributions for the number of mutants in an old culture of bacteria. *Journal of Applied Probability* **11,** 437–444. ISSN: 0021-9002 (3 1974).

26. Kessler, D. A. & Levine, H. Large population solution of the stochastic luria-delbruck evolution model. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 11682–11687. ISSN: 00278424 (29 2013).

27. Pecht, I., Givol, D. & Sela, M. Dynamics of hapten-antibody interaction. Studies on a myeloma protein with anti-2, 4-dinitrophenyl specificity. *Journal of Molecular Biology* **68,** 241–247. ISSN: 00222836 (2 1972).

28. Bocharov, G. A. & Romanyukha, A. A. Mathematical model of antiviral immune response III. Influenza a virus infection. *Journal of Theoretical Biology* **167,** 323–360. ISSN: 00225193 (4 1994).

29. Schwickert, T. A. *et al.* A dynamic T cell-limited checkpoint regulates affinity-dependent B cell entry into the germinal center. *Journal of Experimental Medicine* **208,** 1243–1252. ISSN: 00221007 (6 2011).

30. Karnaukhov, V. K. *et al.* Predicting TCR-peptide recognition based on residue-level pairwise statistical potential. *bioRxiv.* `https://www.biorxiv.org/content/early/2022/02/19/2022.02.15.480516` (2022).

31. Miyazawa, S. & Jernigan, R. L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *Journal of Molecular Biology* **256,** 623–644. ISSN: 00222836 (3 1996).

32. Eisen, H. N. Affinity enhancement of antibodies: how low-affinity antibodies produced early in immune responses are followed by high-affinity antibodies later and in memory B-cell responses. *Cancer immunology research* **2,** 381–392. ISSN: 23266074 (5 2014).

33. Shinnakasu, R. *et al.* Regulated selection of germinal-center cells into the memory B cell compartment. *Nature Immunology* **17,** 861–869. ISSN: 15292916 (7 2016).

34. Weinstein, J. A., Jiang, N., White, R. A., Fisher, D. S. & Quake, S. R. High-throughput sequencing of the zebrafish antibody repertoire. *Science* **324,** 807–810. ISSN: 00368075 (5928 2009).

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

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

37. Briney, B., Inderbitzin, A., Joyce, C. & Burton, D. R. Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature* **566,** 393–397. ISSN: 14764687 (7744 2019).

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

39. Gibson, K. L. *et al.* B-cell diversity decreases in old age and is correlated with poor health status. *Aging Cell* **8,** 18–25. ISSN: 14749718 (1 2009).

40. Weiskopf, D., Weinberger, B. & Grubeck-Loebenstein, B. The aging of the immune system. *Transplant International* **22,** 1041–1050. ISSN: 09340874 (11 2009).

41. Wang, C. *et al.* Effects of Aging, Cytomegalovirus Infection, and EBV Infection on Human B Cell Repertoires. *The Journal of Immunology* **192,** 603–611. ISSN: 0022-1767. `https://doi.org/10.4049/jimmunol.1301384` (2 Jan. 2014).

42. Bradshaw, W. J., Poeschla, M., Placzek, A., Kean, S. & Valenzano, D. R. Extensive age-dependent loss of antibody diversity in naturally short-lived turquoise killifish. *eLife* **11.** ISSN: 2050084X (2022).

43. Doria-Rose, N. A. *et al.* Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature* **509,** 55–62. ISSN: 14764687 (7498 2014).

44. Nourmohammad, A., Otwinowski, J. & Plotkin, J. B. Host-Pathogen Coevolution and the Emergence of Broadly Neutralizing Antibodies in Chronic Infections. *PLoS Genetics* **12,** 1–23. ISSN: 15537404 (7 2016).

45. Yan, L. & Wang, S. Shaping Polyclonal Responses via Antigen-Mediated Antibody Interference. *iScience* **23,** 1–16. ISSN: 25890042 (10 2020).

46. Hearty, S., Leonard, P. & O'Kennedy, R. in *Antibody Engineering: Methods and Protocols, Second Edition* (ed Chames, P.) 411–442 (Humana Press, Totowa, NJ, 2012). ISBN: 978-1-61779-974-7. `https://doi.org/10.1007/978-1-61779-974-7_24`.

47. Qi, S. Y., Groves, J. T. & Chakraborty, A. K. Synaptic pattern formation during cellular recognition. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 6548–6553. ISSN: 00278424 (12 2001).

48. Batista, F. D., Iber, D. & Neuberger, M. S. B cells acquire antigen from target cells after synapse formation. *Nature* **411,** 489–494. ISSN: 00280836 (6836 2001).

49. Carrasco, Y. R., Fleire, S. J., Cameron, T., Dustin, M. L. & Batista, F. D. LFA-1/ICAM-1 interaction lowers the threshold of B cell activation by facilitating B cell adhesion and synapse formation. *Immunity* **20,** 589–599. ISSN: 10747613 (5 2004).

50. Fleire, S. J. *et al.* B cell ligand discrimination through a spreading and contraction response. *Science* **312,** 738–741. ISSN: 00368075 (5774 2006).

51. Gonzalez, S. F. *et al.* Trafficking of B cell antigen in lymph nodes. *Annual Review of Immunology* **29,** 215–233. ISSN: 07320582 (2011).

Supplementary Figures

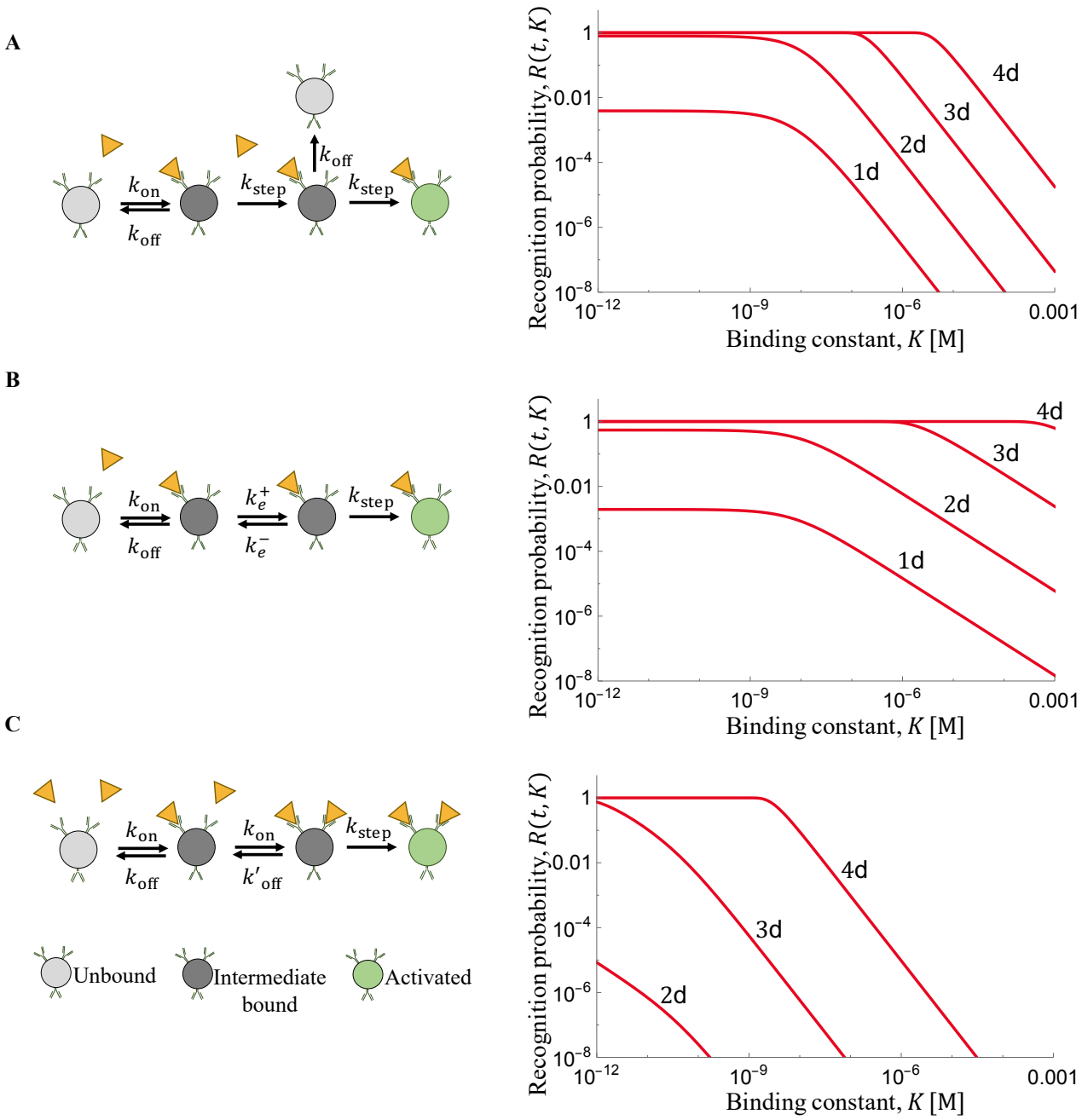

Fig. S 1. Mechanisms of B-cell activation. A schematic of the activation process (left) and the recognition probability $R(t, K)$ ($t = 1\text{d}, \ldots, 4\text{d}$ after the start of the infection) are shown for the kinetic proofreading model discussed in the main text and for two models with reversible binding kinetics. **(A)** Activation by kinetic proofreading (as in Fig. 1). This process has p intermediate bound states and p irreversible activation steps ($p > 1$, shown here for $p = 2$). For antigen affinities of typical naive B-cells ($K \gtrsim 10^{-7}M$), kinetic proofreading generates fast, deterministic recognition of high-affinity lineages ($R \simeq 1$) and strong suppression of low-affinity lineages ($R \sim K^{-p}$ for $K \gg K_1(t)$). **(B)** Activation via an intermediate excited state. This process is similar to the kinetic proofreading model, but the binding kinetics is reversible. It generates fast, deterministic of high-affinity lineages but only weak suppression of low-affinity lineages ($R \sim K^{-1}$ for $K \gg K_1(t)$). **(C)** Activation by cooperative binding. In this process, activation requires simultaneous binding of p virions to the same B-cell ($p > 1$, shown here for $p = 2$). It generates strong suppression of low-affinity lineages ($R \sim K^{-p}$ for $K \gg K_{\text{step}}$), but even high-affinity naive lineages do not reach deterministic activation ($R \ll 1$ for $K \sim 10^{-7}M$). Kinetic parameters: $k_{\text{on}} = 10^6 s^{-1} M^{-1}$; $\delta = 10^{-2}$; $k_{\text{step}} = 0.5\text{min}^{-1}$; $k_{\text{step}}^- = 0.1 k_{\text{step}}$.

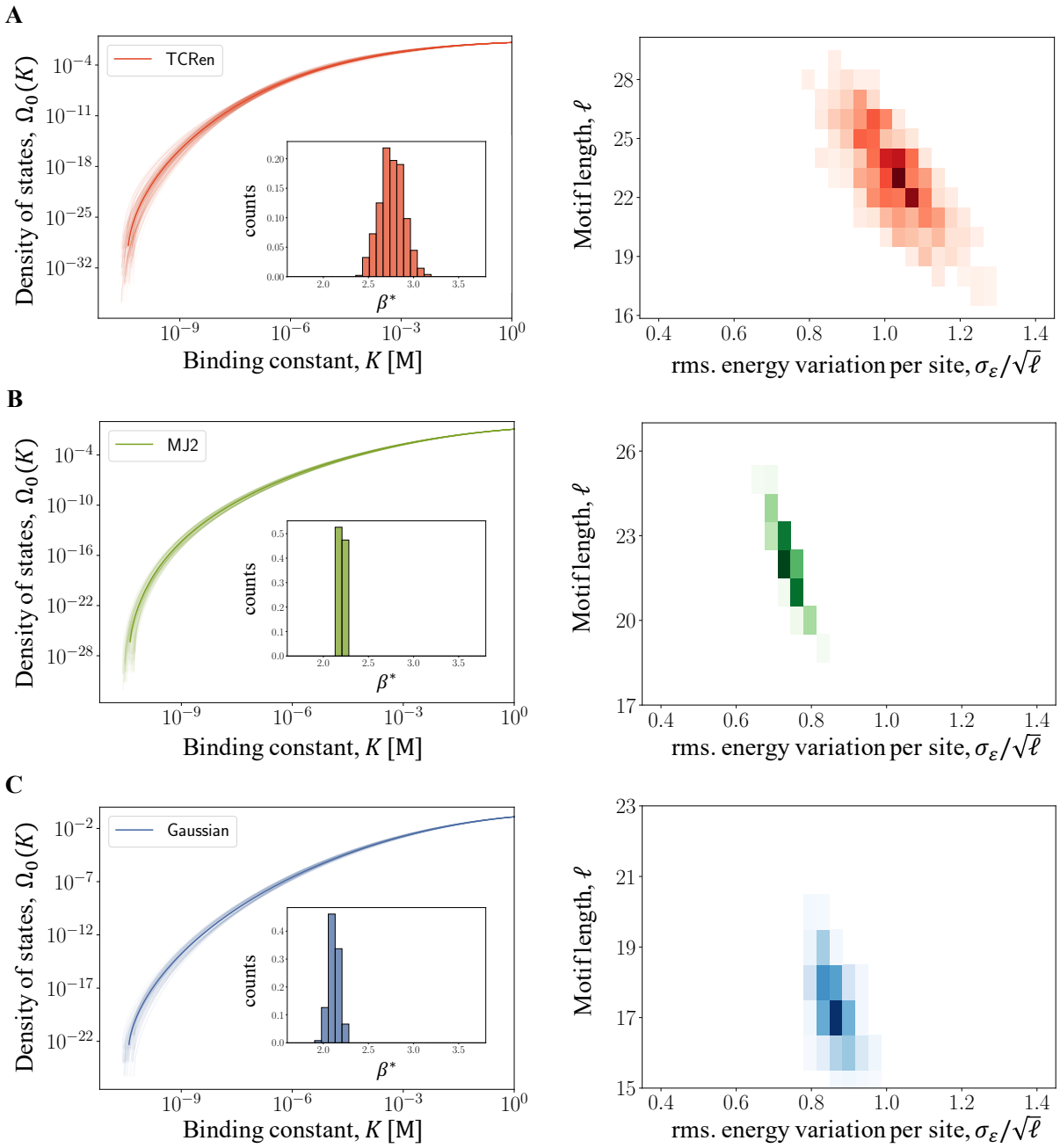

Fig. S 2. Density of states in different energy models. Left: density of naive B-cell lineages, $\Omega_0(K)$, for a sample of antigen epitope sequences with $K_m \approx 10^{-11}M$ and $K^* \approx 10^{-7}M$ (thin lines: individual antigens, thick line: ensemble average); see SI text. Insert: ensemble distribution of the inverse activation temperature β^*, as defined in the main text. Right: joint distribution of the binding length, ℓ, and of the rms. energy variation per site, $\sigma_\varepsilon/\ell^{1/2}$. Binding motifs are randomly sampled from different additive energy models: **(A)** TCRen matrix [30], **(B)** Miyazawa-Jernigan matrix [31], **(C)** Gaussian random energies; details are given in SI Text. Model parameters as in Fig. 2.

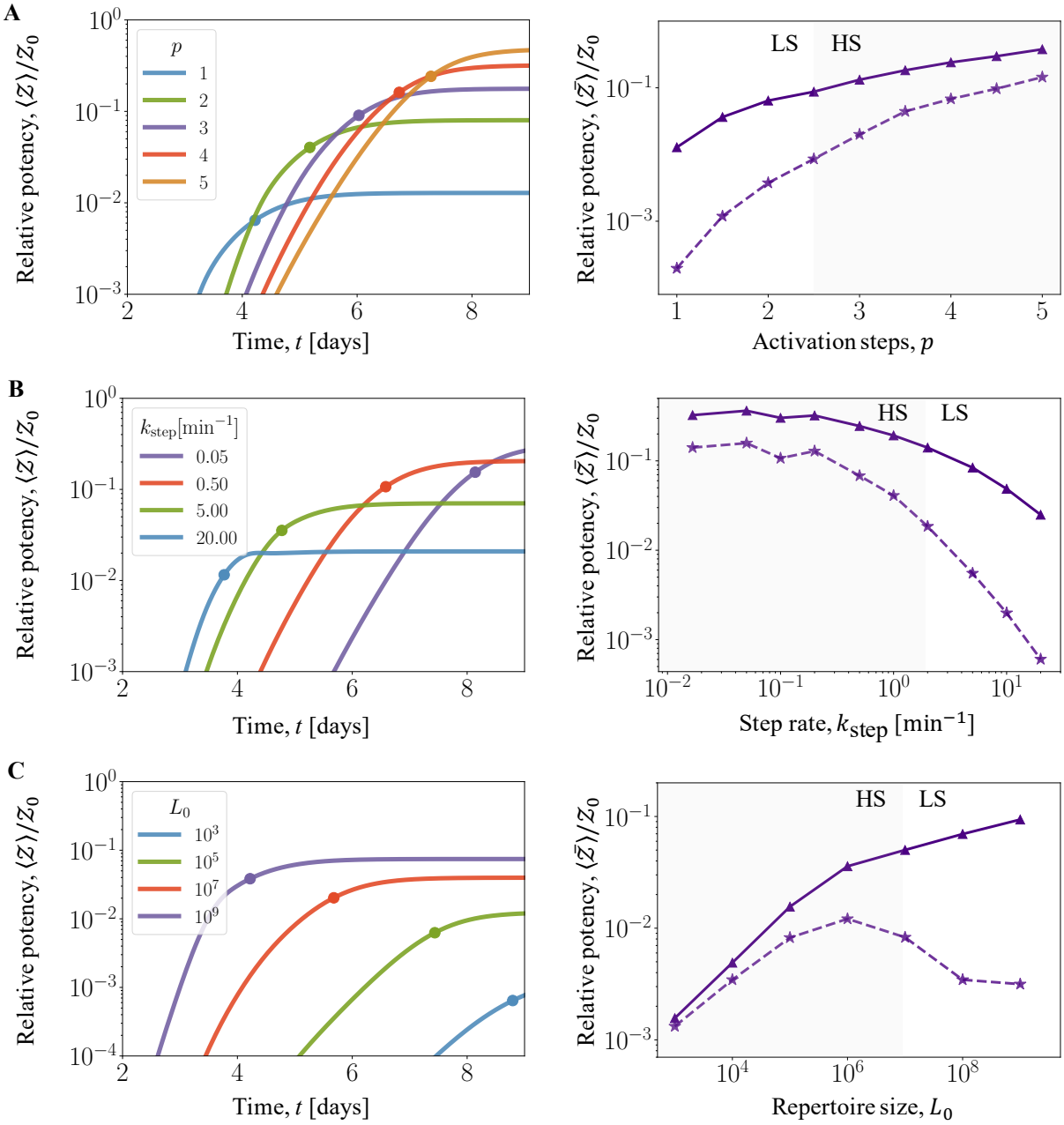

Fig. S 3. Optimality of antigen recognition by tuned proofreading. Antiserum potency is shown for variation of the proofreading parameters p, k_{step}, and of the repertoire size, L_0, around the LS-HS crossover point ($p \sim \beta^*(L_0)$, $k_{\text{step}} \approx k_{\text{on}} K^*(L_0)$, $L_0 = 10^9$), which characterises a tuned repertoire as discussed in the main text. Left: time-dependent, population averaged potency, $\mathcal{Z}(t)$; the half-saturation point (t_{50}, Z_{50}) is marked by a dot. Right: saturation value, $\bar{\mathcal{Z}}$ (thick lines), and the potency component of the largest clone, $\bar{z}_1$ (dashed lines). All potencies are shown in units of the reference value $\mathcal{Z}_0$ defined in the main text; the HS regime is marked by shading. **(A)** Variation of the number of proofreading steps, p, at constant k_{step} and L_0 (cf. Fig. 4EF). In the LS regime ($p \lesssim \beta^*$), potency increases strongly with p. In the HS regime ($p \gtrsim \beta^*$), increasing p yields a diminishing return of potency, while t_{50} continues to increase proportionally to p. **(B)** Variation of the rate of proofreading steps, k_{step}, at constant p and L_0. In the LS regime ($k_{\text{step}} \gtrsim k_{\text{on}} K^*$), potency increases strongly with decreasing k_{step}. In the HS regime ($k_{\text{step}} \lesssim k_{\text{on}} K^*$), decreasing k_{step} yields a diminishing return of potency, while t_{50} continues to increase proportionally to $1/k_{\text{step}}$. **(C)** Variation of the repertoire size, L_0, at constant p and k_{step}. Here we use $p = 2$ and $k_{\text{step}} = 0.05\text{min}^{-1}$, which are tuned for $L_0 = 10^7$. In the HS regime ($L_0 \lesssim 10^7$), potency increases strongly with L_0. In the LS regime ($L_0 \gtrsim 10^7$), increasing L_0 yields a diminishing return of potency and the potency component of the largest clone decreases. Other model parameters as in Fig. 2.

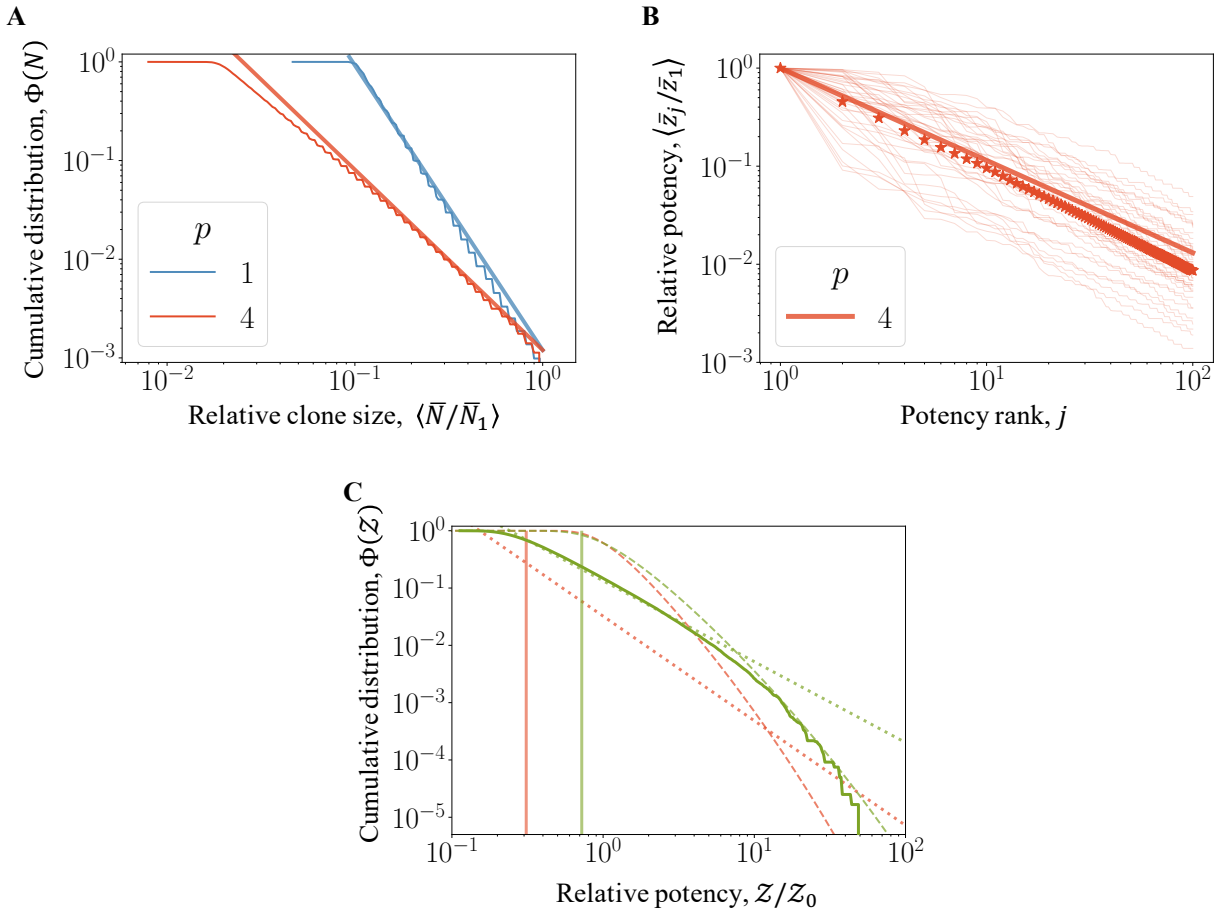

Fig. S 4. Statistics of activated repertoires. (A) Cumulative clone size distribution aggregated over individuals, $\Phi(N) \sim N^{-1/\zeta}$ (blue: LS regime, $p = 1$; red: HS regime, $p = 3$). **(B)** Potency-rank relation in the HS regime, $\langle \bar{z}_j \rangle \sim j^{-\zeta-1/\beta^*}$ for potency-ranked clones ($j = 1, \ldots, 100$). Thin lines give rankings in a set of randomly chosen individuals. **(C)** Cumulative population distribution function of potency in the HS regime, $\Phi(\mathcal{Z})$ ($p = 4$; $L_0 = 10^7$ in green and $L_0 = 10^9$ in red). Elite neutralisers with jackpot clones define a pre-asymptotic regime, $\Phi(\mathcal{Z}) \sim \mathcal{Z}^{-1/\zeta}$ (dotted), and an asymptotic regime, $\Phi(\mathcal{Z}) \sim L_0(\bar{N}/\mathcal{Z})$ (dashed). Vertical lines indicate population average $\langle \bar{\mathcal{Z}} \rangle / \langle \mathcal{Z}_0 \rangle$. Model parameters as in Fig. 2.