Supporting Information

Mayer et al. 10.1073/pnas.1421827112

SI Text

Appendix A: Probability Distribution of the Time of First Recognition

To calculate the cost of not recognizing an antigen a, we need to find the distribution of times when a successful encounter takes place. The probability of having the first recognition of antigen a by receptor r in the time between t and $t+\mathrm{d}t$ reads

$$H_a(t)\mathrm{d}t = \lambda_a(t)\mathrm{d}t \cdot \sum_r P_r f_{r,a} \times \lim_{N\to\infty} \prod_{i=1}^{N} \left(1 - \lambda_a(t_i)\frac{t}{N}\sum_r P_r f_{r,a}\right),$$

where the first term is the probability of having an encounter between t and $t+\mathrm{d}t$, the second term is the probability of this encounter being successful, and the third term is the probability of there not being any prior recognition events. For the calculation of the last term we have decomposed the time leading up to t into N intervals of length t/N. Taking the $N\to\infty$ limit yields

$$H_a(t) = \lambda_a(t)\tilde{P}_a e^{-\int_0^t dt'\lambda_a(t')\tilde{P}_a}, \quad [S1]$$

where we have used the shorthand notation $\tilde{P}_a = \sum_r P_r f_{r,a}$ for the probability that a randomly chosen receptor recognizes antigen a.

Appendix B: Convexity of the Expected Cost

We show that the cost function $\langle F\rangle$ is a convex function of its argument $\{P_r\}$ (the receptor distribution). We start by introducing an alternative expression of $\overline{F}_a$, obtained by integration by parts:

$$\overline{F}_a = \int_0^\infty \mathrm{d}m F_a'(m) e^{-m\tilde{P}_a} + F(0). \quad [S2]$$

We calculate the derivatives of this average cost with respect to $\tilde{P}_a$:

$$\frac{\mathrm{d}\overline{F_a}}{\mathrm{d}\tilde{P}_a} = -\int_0^\infty \mathrm{d}m\, m F_a'(m) e^{-m\tilde{P}_a} \quad [S3]$$

$$\frac{\mathrm{d}^2\overline{F_a}}{\mathrm{d}\tilde{P}_a^2} = \int_0^\infty \mathrm{d}m\, m^2 F_a'(m) e^{-m\tilde{P}_a}. \quad [S4]$$

Because by assumption $F_a'(m)$ is positive, the second derivative of $\overline{F}_a$ with respect to $\tilde{P}_a$ is positive. This establishes the convexity of $\overline{F}_a$ as a function of $\tilde{P}_a$. Because $\langle F\rangle = \sum_a Q_a \overline{F}_a$ (with $Q_a \geq 0$), it is a convex function of $\{\tilde{P}_a\}$. Therefore, it is also a convex function of $\{P_r\}$, as $\{P_r\}$ and $\{\tilde{P}_a\}$ are linearly related.

Appendix C: Biological Motivation of Power-Law Cost Functions

In the main text we have developed a general framework for discussing the antigen–receptor recognition process. To fully specify the model we need to choose an effective cost function $F_a(m) = F_a(t_a(m))$. In the main text we derive optimal receptor distributions for a number of effective cost functions, including power-law functions $F(m) = m^\alpha$. Here we sketch plausible scenarios motivating that choice.

Consider an organism being infected with an antigen a. As long as there is no immune reaction, the antigen divides inside its host and thus increases its population size. If the initial population size is small, it is reasonable to assume exponential growth.

The more antigens there are at the time of the immune reaction, the more damage they can potentially do. Likewise, the more antigens there are, the higher the rate of encounters. These two quantities are also expected to grow exponentially in time:

$$F_a(t) = F_a(0)e^{\nu_a t}, \quad [S5]$$

$$\lambda_a(t) = \lambda_a(0)e^{\nu_a' t}. \quad [S6]$$

The two exponents may be different in general, because the number of pathogenic agents that cause the harm may grow differently than the number of antigens that can be recognized by the immune system. This difference could for example stem from the fact that both the pathogen's antigenic exposure and its virulence are cooperative effects and thus scale as a power of the number of invading individuals. Using $m_a(t) = \lambda_a(0)(e^{\nu_a' t} - 1)/\nu_a'$, and eliminating time $t \approx \ln[m_a/\lambda_a(0)]/\nu_a'$ (for t large compared with $1/\nu_a'$), we rewrite the effective cost function in terms of the number of encounters,

$$F_a(m) = F_a(0)\left(\frac{m}{\lambda_a(0)}\right)^{\nu_a/\nu_a'} \propto m^\alpha, \quad [S7]$$

with $\alpha = \nu_a/\nu_a'$.

Appendix D: Analytical Optimization

1. Optimality Conditions. In the following we give optimality conditions for the optimization problem defined in the main text, which are used for the following analytical determination of optimal receptor distributions. These conditions, called Karush–Kuhn–Tucker conditions (1), are derived from a generalization of the method of Lagrange multipliers to inequality as well as equality constraints.

The Lagrangian for the optimization problem is

$$\mathcal{L}(P,\lambda,\nu) = \langle F\rangle(P) + \lambda\left(\sum_r P_r - 1\right) - \sum_r \nu_r P_r, \quad [S8]$$

with

$$\langle F\rangle = \sum_a Q_a \overline{F}_a. \quad [S9]$$

λ is a Lagrange multiplier enforcing the normalization constraint and ν_r are Lagrange multipliers enforcing the nonnegativity constraint. The optimal P^* is an extremum of this Lagrangian. Therefore, the stationarity conditions

$$\left.\frac{\partial\langle F\rangle}{\partial P_r}\right|_{P*} + \lambda^* - \nu_r^* = 0, \quad [S10]$$

with

$$\frac{\partial\langle F\rangle}{\partial P_r} = \sum_a Q_a \overline{F}'_a(\tilde{P}_a) f_{r,a}, \quad \textbf{[S11]}$$

must hold for some value of λ^* and ν_r^* that enforce the constraints. The inequality constraint $P_r \geq 0$ further requires that

$$\nu_r^* \geq 0 \quad \textbf{[S12]}$$

$$\nu_r^* P_r^* = 0, \quad \textbf{[S13]}$$

where Eq. **S13** is known as the complementary slackness condition. It requires the Lagrange multipliers associated with the nonnegativity to be zero unless the constraint is active, i.e., unless the corresponding receptor probability is zero.

The three conditions may be reformulated as

$$\left.\frac{\partial\langle F\rangle}{\partial P_r}\right|_{P*} + \lambda^* \geq 0 \quad \textbf{[S14]}$$

$$\left(\left.\frac{\partial\langle F\rangle}{\partial P_r}\right|_{P*} + \lambda^*\right) P_r = 0. \quad \textbf{[S15]}$$

For all receptors that are present in the optimal repertoire ($P_r^* > 0$) these conditions imply

$$\left.\frac{\partial\langle F\rangle}{\partial P_r}\right|_{P*} = -\lambda^*. \quad \textbf{[S16]}$$

If a receptor is not present in the optimal repertoire ($P_r^* = 0$), then the less stringent condition holds:

$$\left.\frac{\partial\langle F\rangle}{\partial P_r}\right|_{P*} \geq -\lambda^*. \quad \textbf{[S17]}$$

We note here that $\partial\langle F\rangle/\partial P_r \leq 0$ (because more receptors always yield a lower cost), so that $\lambda^* \geq 0$.

These two conditions can be explained as follows: If a repertoire is optimal, all changes allowed by the constraints will lead to a higher cost; i.e., moving receptors from one type to another will not yield an improvement. All partial derivatives of the cost with respect to the receptor probabilities should thus be equal to the same value (Eq. **S16**). If there are already no receptors of a certain type, i.e., $P_r = 0$, we get a less stringent condition. We can no longer remove receptors away from this type r, but only add some to it, at the expense of other receptor types. The increase in cost due to the depletion of these other types should be higher than the gain of moving them to type r. The partial derivatives of the cost with respect to the receptors that are not present in the repertoire must thus be larger than the partial derivatives of the present receptors, which are given by $-\lambda^*$ (Eq. **S17**).

2. Solution for Uniquely Specific Receptors. We now solve Eqs. **S16** and **S17** for a repertoire of uniquely specific receptors (no cross-reactivity). Eq. **S11** becomes

$$\frac{\partial\langle F\rangle}{\partial P_r} = Q_r \overline{F}'_r(P_r), \quad \textbf{[S18]}$$

where we have used the fact that in the absence of cross-reactivity $\tilde{P}_a = P_a$. If all optimal receptor probabilities are positive, then we can insert this relationship into Eq. **S16** to obtain

$$Q_r \overline{F}'_r(P_r^*) = -\lambda^* \quad \textbf{[S19]}$$

and thus

$$P_r^* = h_r \frac{-\lambda^*}{Q_r}, \quad \textbf{[S20]}$$

where $h_r = \overline{F}'^{(-1)}_r$ denotes the inverse function of $\overline{F}'_r$. Because that function $\overline{F}'_r$ is always negative, h_r must take a negative argument.

For some cost functions, solving this equation may yield some negative receptor probabilities. In these cases some of the nonnegativity constraints need to be active. Setting $P_r = 0$ when Eq. **S20** is negative yields the correct optimal distribution under the nonnegativity constraint. We verify that for these r, Eq. **S17** is satisfied by $P_r = 0$, because

$$Q_r \overline{F}'_r(P_r = 0) \geq Q_r \overline{F}'_r\left[h_r\left(\frac{-\lambda^*}{Q_r}\right)\right] = -\lambda^*, \quad \textbf{[S21]}$$

where we have used the fact that $\overline{F}'_r$ is an increasing function of its argument (due to the positivity of its derivative; compare Eq. **S4**), and $h_r(-\lambda^*/Q_r) \leq 0$.

In summary, the solution to the optimization problem is

$$P_r^* = \max\left\{h_r\left(\frac{-\lambda^*}{Q_r}\right), 0\right\}, \quad \textbf{[S22]}$$

where the value of λ^* is fixed by the normalization condition $\sum_r P_r = 1$.

In Table S1 we give the explicit expressions of $\overline{F}_a$ and h_a, for the particular choices of the cost function $F(m)$ considered in the main text.

3. Solution for Cross-Reactive Receptors. The previous results can be generalized to cross-reactive receptors in a continuous space, using Fourier transforms. This generalization will lead up to the results presented in the *Cross-Reactivity Dramatically Reduces Diversity in the Optimal Repertoire* section of the main text and notably the Gaussian case discussed therein.

a. Deconvoluting the optimality conditions in Fourier space. We consider a continuous receptor–antigen space and we assume a translation invariant cross-reactivity function $f_{r,a} = f(r-a)$. We write the optimality condition Eq. **S16**,

$$\int \mathrm{d}a\, Q(a)\overline{F}'\left[\tilde{P}^*(a)\right] f(r-a) = -\lambda^*, \quad \textbf{[S23]}$$

where in continuous space the coverage is defined as

$$\tilde{P}(a) = \int \mathrm{d}r\, P(r) f(r-a). \quad \textbf{[S24]}$$

We note that both expressions involve integrals, which are convolutions with the cross-reactivity kernel. Because the convolution of a constant is also a constant, a solution of

$$Q(a)\overline{F}'\left(\tilde{P}^*(a)\right) = -\lambda', \quad \text{with } \lambda' > 0, \quad \textbf{[S25]}$$

is also a solution of Eq. **S23**. As in the case of uniquely specific receptors, we can solve this equation for $\tilde{P}^*(a)$,

$$\tilde{P}^*(a) = h\left[\frac{-\lambda'}{Q(a)}\right], \quad \textbf{[S26]}$$

where $h = \overline{F}'^{(-1)}$ as in Eq. **S20**. If there was no cross-reactivity, there would be no difference between P and $\tilde{P}$, and we would be done. Here we need to perform a deconvolution to obtain the optimal receptor distribution P from the optimal coverage $\tilde{P}$. We

do so in Fourier space, where the convolution turns into a product. Deconvolution is therefore much simpler in Fourier space as it corresponds to a division

$$\mathcal{F}[\tilde{P}] = \mathcal{F}[P]\mathcal{F}[f] \quad \Leftrightarrow \quad \mathcal{F}[P] = \frac{\mathcal{F}[\tilde{P}]}{\mathcal{F}[f]}, \qquad \textbf{[S27]}$$

where we have defined the Fourier transform of a function $g(x)$ as $\mathcal{F}[g](k) = \int_{-\infty}^{\infty} dx g(x) e^{ikx}$. To calculate the optimal receptor distribution we insert Eq. **S26** into Eq. **S27** and perform an inverse Fourier transform $\mathcal{F}^{-1}[\tilde{g}](x) = (1/2\pi)\int_{-\infty}^{\infty} dk \tilde{g}(k) e^{-ikx}$ to obtain

$$P^* = \mathcal{F}^{-1}\left[\frac{\mathcal{F}[h(-\lambda'/Q)]}{\mathcal{F}[f]}\right]. \qquad \textbf{[S28]}$$

This result is valid only as long as the above quantity is positive and normalizable, as we shall see below.

b. The Gaussian case. In this section we apply the general results of the previous section to a concrete example. To find the optimal receptor distribution analytically we use Eq. **S28**, we assume the antigen distribution and cross-reactivity function are Gaussian

$$Q(a) = \frac{1}{\sqrt{2\pi\sigma_Q^2}} \exp\left(\frac{-a^2}{2\sigma_Q^2}\right), \qquad \textbf{[S29]}$$

$$f(r-a) = \exp\left[\frac{-(r-a)^2}{2\sigma^2}\right], \qquad \textbf{[S30]}$$

and we take

$$F(m) = m^{\alpha}. \qquad \textbf{[S31]}$$

Inserting h from Table S1 into Eq. **S28** allows us to write

$$P^* \propto \mathcal{F}^{-1}\left[\frac{\mathcal{F}[Q^{1/(1+\alpha)}]}{\mathcal{F}[f]}\right] \qquad \textbf{[S32]}$$

as an equivalent equation determining the optimal repertoire. We can calculate the modified antigen distribution as

$$Q(a)^{1/(1+\alpha)} \propto \exp\left(-\frac{a^2}{2(1+\alpha)\sigma_Q^2}\right). \qquad \textbf{[S33]}$$

The Fourier transform of a Gaussian function of variance σ^2 is a Gaussian function of variance $1/\sigma^2$ (2). Therefore, we have

$$\mathcal{F}\left[Q^{1/(1+\alpha)}\right](q) \propto \exp\left[\frac{-(1+\alpha)\sigma_Q^2 q^2}{2}\right], \qquad \textbf{[S34]}$$

$$\mathcal{F}[f](q) \propto \exp\left[\frac{-\sigma^2 a^2}{2}\right], \qquad \textbf{[S35]}$$

from which

$$\frac{\mathcal{F}[Q^{1/(1+\alpha)}]}{\mathcal{F}[f]} \propto \exp\left\{\frac{-\left[(1+\alpha)\sigma_Q^2 - \sigma^2\right]q^2}{2}\right\} \qquad \textbf{[S36]}$$

follows. Taking the inverse Fourier transform and normalizing, we obtain

$$P^*(r) = \frac{1}{\sqrt{2\pi\left[(1+\alpha)\sigma_Q^2 - \sigma^2\right]}} \exp\left(-\frac{r^2}{2\left[(1+\alpha)\sigma_Q^2 - \sigma^2\right]}\right). \qquad \textbf{[S37]}$$

Normalization is possible only for $\sigma < \sigma_Q\sqrt{1+\alpha} \equiv \sigma_c$. In the limit $\sigma \to \sigma_c$ the Gaussian converges to a Dirac delta function. Intuition suggests that a Dirac delta function centered on the peak position should remain optimal for further increases in σ. To prove this assertion we note that a Dirac delta function is zero everywhere, except in one point. Because all but one receptor probabilities are at the boundary defined by the nonnegativity constraints, we need only to check Eq. **S17**. We compute the left-hand side of Eq. **S23** as a function of r,

$$\int dp\, Q(a)\overline{F}'[\tilde{P}^*(a)]f(r-a) \propto -\exp\left\{\frac{-r^2\left[\sigma^2 - (1+\alpha)\sigma_Q^2\right]}{2\sigma^2\left(\sigma^2 - \alpha\sigma_Q^2\right)}\right\}, \qquad \textbf{[S38]}$$

and note that it has a minimum for $r = 0$. This shows that the partial derivatives of the expected cost at $r \neq 0$ are greater than at $r = 0$, implying that Eq. **S17** holds.

The cost of the optimal repertoires as a function of the cross-reactivity width σ is given by

$$\langle F\rangle(P^*) = \left(\frac{\sigma_Q}{\sigma}\right)^{\alpha} \begin{cases} (1+\alpha)^{(1+\alpha)/2} & \text{if } \sigma < \sigma_c, \\ \dfrac{(\sigma/\sigma_Q)^{\alpha}}{\sqrt{1-\alpha(\sigma_Q/\sigma)^2}} & \text{otherwise.} \end{cases} \qquad \textbf{[S39]}$$

Both expressions give the same cost at the transition $\sigma = \sigma_c$. After multiplying by $(\sigma/\sigma_Q)^{\alpha}$ to compare at constant recognition capability $\int f = \sqrt{2\pi}\sigma$, this expression is constant for $\sigma < \sigma_c$ and grows for $\sigma > \sigma_c$.

c. General argument for peakedness. A simple argument can help us understand why cross-reactivity generically leads to peaked optimal solutions. The convolution with a kernel is a smoothening operation, represented by a low-pass filter in the Fourier domain. The optimal solution in the absence of the nonnegativity constraints requires that $\tilde{P}_a = h(Q_a)$. As $\tilde{P}_a$ is the low-pass filtered version of P_r, the high-frequency components of $h(Q_a)$ will be magnified by the deconvolution. These high-frequency wiggles can lead to negative values of $\mathcal{F}^{-1}[h(Q_a)]$, which are not allowed, leading us to set many values of $P(r)$ to zero. This effect results in a peaked solution. Because the size of the cross-reactivity kernel is inversely proportional to the cutoff frequency in the Fourier domain, we expect the spacing of the peaks to be related to the size of the cross-reactivity kernel.

Appendix E: Numerical Optimization

We numerically minimize the cost function subject to the normalization and nonnegativity constraints by using a fast projected-gradient algorithm. In the following we provide details on this numerical algorithm. To facilitate notations let us define the function to minimize as $g(x)$, where x is a vector in a Euclidean space, and the convex set C is defined by the constraints. In these notations the problem we want to solve can be stated as

$$\min_{x \in C} g(x). \qquad \textbf{[S40]}$$

Given an arbitrary starting point $x_0 \in C$ the algorithm performs the iterative procedure

$$y^{k+1} = x^k + \omega^k\left(x^k - x^{k-1}\right), \quad \text{[S41]}$$

$$x^{k+1} = \mathcal{P}\left(y^{k+1} - s^k \nabla g\left(y^{k+1}\right)\right), \quad \text{[S42]}$$

where ∇ denotes the gradient. Here $\mathcal{P}$ denotes a projection onto C, ω^k is an extrapolation step size, and s^k is the step size taken in the direction of the gradient. The extrapolation step size has to be chosen carefully to ensure the faster convergence of this method with respect to an ordinary gradient method. Following ref. 3 we use

$$\omega_k = \frac{k}{k+3}. \quad \text{[S43]}$$

The step size s is determined by backtracking (4): We iteratively decrease s by multiplication by $\beta < 1$ until $g(z) \le g(y^k) + (z - y^k)\cdot\nabla g(y^k) + (1/2s)(z-y^k)^2$, where $x\cdot y$ denotes the inner dot product between x and y, and $z = \mathcal{P}(y^k - s\nabla g(y^k))$. In practice we determine s in this way at the first step of the optimization and then keep it fixed based on this initial estimate.

The projection of a point y onto a convex set C is defined by the following quadratic programming problem:

$$\mathcal{P}(y) = \underset{x\in C}{\operatorname{argmin}} \frac{1}{2}(x-y)^2. \quad \text{[S44]}$$

If the convex set is a simplex as is the case for our problem, efficient algorithms fortunately exist for solving this problem. We use the algorithm described in ref. 5.

To stop the iteration one needs to define suitable stopping criteria. As the problem is convex we can establish a lower bound for the cost by solving a linear programming problem as follows:

$$g_{lb} = g\left(x^k\right) + \min_{x\in C}\left[\left(x - x^k\right)\cdot\nabla g\left(x^k\right)\right] \le g(x^*). \quad \text{[S45]}$$

The linear programming problem $\bar{x}^k = \operatorname{argmin}_{x\in C}\nabla g(x^k)^T(x - x^k)$ is solved explicitly (6) by

$$\bar{x}^k = e_{i^*},\ i^* = \underset{i}{\operatorname{argmin}}\left(\nabla g\left(x^k\right)\right)_i, \quad \text{[S46]}$$

where e_i denotes the ith unit vector. We can use this lower bound to define a stopping criterion for the numerical optimization

$$\frac{g\left(x^k\right) - g_{lb}}{g_{lb}} < \epsilon. \quad \text{[S47]}$$

For all reported numerical results we have chosen $\epsilon = 10^{-8}$.

To minimize finite size effects in the simulations we have used periodic boundary conditions in the receptor/antigen space. The discretization steps used in the figures are listed below:

Step	Figure
0.5σ	Fig. 5
0.1σ	Fig. 3 and Figs. S2–S4
0.05σ	Fig. 4

All source code associated with this paper is available online at dx.doi.org/10.5281/zenodo.16796.

Appendix F: Tiling Properties: Radial Distribution Function and Power Spectral Density of the Receptor Distribution

We analyze the tiling structure of the peaks in the optimal distribution P_r^* found in Fig. 3 of the main text. A useful technique, borrowed from condensed matter physics, is to measure the radial distribution function (7), $g(R) = \langle P(r)P(r')\rangle_{|r-r'|=R}$, where $|r-r'|$ is the distance between points r and r'. Fig. S2*A* presents $g(R)$ for P^* in two dimensions. The initial drop at small r indicates that peaks in P^* are rarely close—i.e., peaks in the optimal repertoire tend to repel each other. This exclusion, which operates over the range of strong cross-reactivity, is a sensible way to distribute resources, as it limits redundant protection against the same pathogens. The damped oscillation of the peaks of $g(R)$ confirms that the receptors in P^* are organized into a disordered tiling pattern. A similar radial distribution function is seen in high-density random packings of hard spheres where the spheres must cover as much space as possible but exclude each other. In both cases, the tiling ensures uniform coverage of space at large scales.

To quantify the regularity of the tiling, we calculate the normalized power spectral density of the 2D pattern, $S(q) = \sum_{r,r'} P_r P_{r'} e^{iq(r-r')}/\sum_r P_r^2$, where q is a wave vector. Large (small) $|q|$ correspond to short (long) distances in antigen shape space. When P_r is made of Dirac delta peaks of uniform heights, $S(q)$ coincides with the structure factor familiar in physics and satisfies $S(q\to\infty) = 1$. Fig. S2*B* shows $S(q)$ averaged over many realizations of the antigen landscape and over all directions of q so that it depends only on its modulus $|q|$. $S(q)$ approaches 1 for large q, showing that the precise local positions of the peaks are random. (The small departure from 1 is attributable to numerical discretization.) $S(q)$ is very low for small q, indicating that the number of receptors contained in any given large area of the shape space is very reproducible, providing uniform coverage—a property called hyperuniformity in the context of jammed materials (8–10). For our optimal repertoires small-scale fluctuations (large q) get smoothed out by cross-reactivity and can be tolerated, whereas at large scales the fluctuations track variations in the antigenic landscape to provide smooth coverage (Fig. S3).

Appendix G: Non-Gaussian, Long-Tailed, and Nonuniform Cross-Reactivity Functions

To assess the impact of different assumptions about the nature of cross-reactivity on the results we performed a number of simulations with different kernel functions.

First, we investigated the family of kernel functions defined by $f(r-a) = \exp[-(|r-a|/\eta)^\gamma]$ (Fig. S4 *A–D*, *Left*). By changing the parameter γ we can go from an exponential ($\gamma = 1$) via a Gaussian $\gamma = 2$ to a top-hat kernel ($\gamma\to\infty$). Up to $\gamma = 2$ all such kernels have positive Fourier transforms, whereas for $\gamma > 2$ the Fourier transforms also take negative values. The positive definiteness has been shown to be an important property in a related problem in ecology (11). Second, we also investigated how long-tailed kernel functions change the optimal repertoire by considering the functional form $f(r-a) = 1/(1 + (|r-a|\eta)^2)$ (Fig. S4 *A–D*, *Right*).

Dropping one further assumption, we investigated the influence of varying the width of the cross-reactivity function between receptors (Fig. S5). The width of the cross-reactivity was drawn randomly from a log-normal distribution with different coefficients of variation (corresponding to different amounts of scatter in the width). Biologically, the overall stimulatory capacity of receptors is constrained, and we rescaled the cross-reactivity so that all receptors had the same overall stimulatory potency.

Appendix H: Excluding Strongly Self-Binding Receptors

The presence of self-antigens that should not be recognized puts constraints on which receptors the repertoire might contain. As a first step to understand how such a requirement interacts with the trade-off considered in this paper we analyzed a simple model: A number of self-antigens are picked at random positions. The repertoire is not allowed to have receptors that are too highly reactive to any of the self-antigens. In practice this is ensured by adding a constraint to the optimization that none of the receptors in the repertoire can have a distance smaller than σ to any self-antigen. Introducing this constraint changes the optimal repertoire,

but key features such as the fragmentation of the repertoire and the tiling are conserved (Fig. S7).

Appendix I: Model for Receptor Dynamics

Here we describe our model for competitive receptor dynamics. We then show how, in a mean-field limit where antigen encounters are very frequent, this model reduces to a system of differential equations for the population dynamics.

At every step we update the number of receptors according to

$$\Delta N_r = \Delta t \cdot N_r \left[A\left(\sum_{r'} N_{r'} f_{r',a} \right) f_{r,a} - d \right], \qquad \text{[S48]}$$

where the antigen a is drawn randomly with probability Q_a and Δt is a parameter determining how much the repertoire changes per step.

In the limit where Δt is small the dynamics cycle through different encountered pathogens so fast that they effectively become the following dynamics:

$$\frac{\mathrm{d}N_r}{\mathrm{d}t} = N_r \left[\sum_a Q_a A\left(\sum_{r'} N_{r'} f_{r',a} \right) f_{r,a} - d \right]. \qquad \text{[S49]}$$

These dynamics are of mean-field type; i.e., they neglect the effect of the stochasticity in the encounter of pathogens.

Appendix J: The Stable Fixed Point of the Mean-Field Population Dynamics Minimizes the Cost Function

In this section we show that the stable fixed point $\{N_r^*\}$ of the population dynamics, Eq. **S49**, gives a probability distribution $P_r = N_r/N_{\mathrm{tot}}$ (with $N_{\mathrm{tot}} = \sum_r N_r$) that minimizes the cost $\langle F \rangle$. For this correspondence to be exact, the availability function of the dynamics and the effective cost function of the optimization must be related by

$$A(\tilde{N}_a) = -c' \overline{F}' \frac{\tilde{N}_a}{N_{\mathrm{st}}}, \qquad \text{[S50]}$$

where $\tilde{N}_a = \sum N_r f_{r,a}$, and N_{st} is the total number of receptors N_{tot} at the fixed point.

A fixed point is characterized by $\mathrm{d}N_r/\mathrm{d}t = 0$. If $N_r > 0$, this translates into

$$\sum_a Q_a A\left(\sum_{r'} N_{r'} f_{r',a} \right) f_{r,a} - d = 0. \qquad \text{[S51]}$$

Using the correspondence between availability and cost function given by Eq. **S50**, we rewrite this condition as

$$\sum_a Q_a \overline{F}'(\tilde{P}_a) f_{r,a} = -c' d, \qquad \text{[S52]}$$

which is equivalent to the optimality condition Eq. **S16**, with the identification $\lambda^* = c' d$.

For $N_r = 0$ we need to work a bit harder to show that the optimality condition at the boundary Eq. **S17** is satisfied. Here the key assumption establishing the minimization of the cost function is the stability of the fixed point. A fixed point is stable if the real parts of the Jacobian's eigenvalues are all negative. The Jacobian reads

$$\begin{aligned} J_{r,r'} = {} & \delta_{r,r'} \left(\sum_a Q_a A\left(\sum_{r''} N_{r''} f_{r'',a} \right) f_{r,a} - d \right) \\ & + N_r \sum_a Q_a A'\left(\sum_{r''} N_{r''} f_{r'',a} \right) f_{r,a} f_{r',a}. \end{aligned} \qquad \text{[S53]}$$

We remark that for $N_r = 0$ the rth row of the Jacobian is nonzero only on the diagonal. That value on the diagonal is an eigenvalue of the Jacobian and must be negative:

$$\sum_a Q_a A\left(\sum_{r'} N_{r'} f_{r',a} \right) f_{r,a} - d < 0. \qquad \text{[S54]}$$

Again we replace $A(\sum_r N_r f_{r,a})$ by $-\overline{F}'_a(\tilde{P}_a)$ according to Eq. **S50** to obtain

$$\sum_a Q_a \overline{F}'(\tilde{P}_a) f_{r,a} > -c' d, \qquad \text{[S55]}$$

which is equivalent to the optimality condition at the boundary Eq. **S17**, provided that $\lambda^* = c' d$.

Appendix K: Cost Function as a Lyapunov Function of the Mean-Field Dynamics

Here we show rigorously that, when the availability function is scale invariant, as in the case for the simple cost function $F(m) = m^\alpha$, the dynamics must converge toward a fixed point. This fixed point is unique and corresponds to the optimal of the cost $\langle F \rangle$, as we have shown in the previous section.

$A(x)$ is scale invariant if a function v exists such that $A(\gamma x) = v(\gamma) A(x)$. In this case we will see that the changes of relative frequencies P_r in the repertoire over time depend only on the total number of receptors through a prefactor. Below we derive the equations governing these dynamics and then prove that these dynamics are ensured to converge to a stable fixed point. We do so by showing that the dynamics admit the expected cost $\langle F \rangle$ as a Lyapunov function, i.e., a function that continually decreases under the dynamics.

For ease of notation we rewrite Eq. **S49** as

$$\frac{\mathrm{d}N_r}{\mathrm{d}t} = N_r [\pi_r(N) - d], \qquad \text{[S56]}$$

where N is shorthand for $\{N_r\}$, and $\pi_r = \sum_a Q_a A(\sum_{r'} N_{r'} f_{r',a}) f_{r,a}$ is the growth rate of receptor type r. The relative frequencies $P_r = N_r/N_{\mathrm{tot}}$ evolve according to

$$\frac{\mathrm{d}P_r}{\mathrm{d}t} = \frac{1}{N_{\mathrm{tot}}} \frac{\mathrm{d}N_r}{\mathrm{d}t} - \frac{N_r}{N_{\mathrm{tot}}^2} \frac{\mathrm{d}N_{\mathrm{tot}}}{\mathrm{d}t} \qquad \text{[S57]}$$

$$= P_r \left[\pi_r(N) - \sum_{r'} P_{r'} \pi_{r'}(N) \right]. \qquad \text{[S58]}$$

If A is scale invariant, so is π_r and $\pi_r(N) = \pi_r(N_{\mathrm{tot}} P) = v(N_{\mathrm{tot}}) \pi_r(P)$. Then the equations further simplify to

$$\frac{\mathrm{d}P_r}{\mathrm{d}t} = v(N_{\mathrm{tot}}) P_r \left[\pi_r(P) - \sum_{r'} P_{r'} \pi_{r'}(P) \right], \qquad \text{[S59]}$$

$$= v(N_{\mathrm{tot}}) P_r (\pi_r - \overline{\pi}), \qquad \text{[S60]}$$

where $\overline{\pi} = \sum_r P_r \pi_r$.

We can now write how the expected cost $\langle F \rangle$ evolves in time:

$$\frac{\mathrm{d}\langle F \rangle}{\mathrm{d}t} = \sum_r \frac{\partial \langle F \rangle}{\partial P_r} \frac{\mathrm{d}P_r}{\mathrm{d}t} \qquad \text{[S61]}$$

$$= v(N_{\mathrm{tot}}) \sum_r P_r \left[\sum_a Q_a \overline{F}'_a(\tilde{P}_a) f_{r,a} \right] (\pi_r - \overline{\pi}) \qquad \text{[S62]}$$

$$= -\frac{v(N_{\mathrm{tot}})}{c'} \sum_r P_r \left[\sum_a Q_a A\left(N_{\mathrm{st}} \tilde{P}_a\right) f_{r,a} \right] (\pi_r - \overline{\pi}) \qquad \textbf{[S63]}$$

$$= -\frac{v(N_{\mathrm{tot}}) v(N_{\mathrm{st}})}{c'} \sum_r P_r \pi_r (\pi_r - \overline{\pi}) \qquad \textbf{[S64]}$$

$$= -\frac{v(N_{\mathrm{tot}}) v(N_{\mathrm{st}})}{c'} \sum_r P_r (\pi_r - \overline{\pi})^2 \leq 0. \qquad \textbf{[S65]}$$

This proves that the cost always decreases with time, i.e., is a Lyapunov function of the dynamics. Therefore, the dynamics will reach a stable fixed point at steady state, which is guaranteed to be the global minimum of the expected cost $\langle F \rangle$.

1. Boyd SP, Vandenberghe L (2004) *Convex Optimization* (Cambridge Univ Press, Cambridge, UK).
2. Bronshtein IN, Semendyayev KA, Musiol G, Muehlig H (2007) *Handbook of Mathematics* (Springer, Berlin Heidelberg), Vol 3.
3. Parikh N, Boyd S (2013) Proximal algorithms. *Found Trends Optim* 1(3):123–231.
4. Beck A, Teboulle M (2009) A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J Imaging Sci* 2(1):183–202.
5. Duchi J, Shalev-Shwartz S, Singer Y, Chandra T (2008) Efficient projections onto the l_1 ball for learning in high dimensions. *Proceedings of the International Conference on Machine Learning* (ICML, Helsinki), pp 272–279.
6. Bertsekas DP (1999) *Nonlinear Programming* (Athena Scientific, Belmont, MA).
7. Chaikin PM, Lubensky TC (2000) *Principles of Condensed Matter Physics* (Cambridge Univ Press, Cambridge, UK), Vol 1.
8. Torquato S, Stillinger FH (2003) Local density fluctuations, hyperuniformity, and order metrics. *Phys Rev E Stat Nonlin Soft Matter Phys* 68(4 Pt 1):041113.
9. Donev A, Stillinger FH, Torquato S (2005) Unexpected density fluctuations in jammed disordered sphere packings. *Phys Rev Lett* 95(9):090604.
10. Berthier L, Chaudhuri P, Coulais C, Dauchot O, Sollich P (2011) Suppressed compressibility at large scale in jammed packings of size-disperse spheres. *Phys Rev Lett* 106(12):120601.
11. Pigolotti S, López C, Hernández-García E (2007) Species clustering in competitive Lotka-Volterra models. *Phys Rev Lett* 98(25):258101.

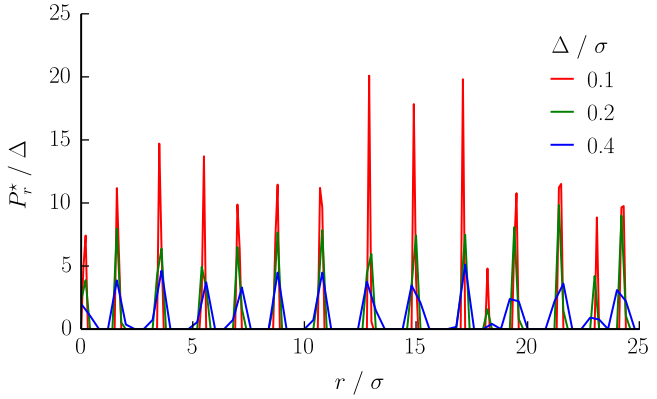

Fig. S1. Solving the optimization problem with a finer and finer discretization step suggests that the peaks found in the optimal receptor distributions converge to true Dirac delta functions. Starting from a problem with a discretization step of $\Delta = 0.1\sigma$, we construct coarse-grained versions of it by downsampling the antigen distribution two- and fourfold, yielding $\Delta = 0.2\sigma$ and 0.4σ, respectively. The resulting coarse-grained optimization problems are then solved, and the optimal distributions P_r^*/Δ are represented (after appropriate normalization by the step size). The random antigen distribution is log-normal with coefficient of variation $\kappa = 0.25$.

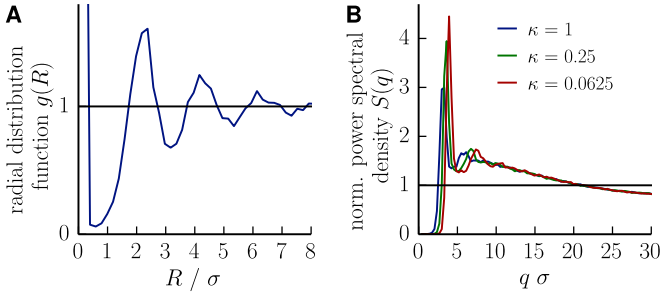

Fig. S2. Radial distribution function and normalized power spectral density of the optimal receptor distribution P_r^* for random environments in two dimensions. (*A*) The radial distribution function of P_r^* shows an exclusion zone around each peak, followed by oscillations characteristic of a local tiling pattern. (*B*) Normalized power spectral density $S(q)$ of P_r^* for different values of the parameter κ quantifying the heterogeneity of the antigenic landscape. The high suppression of fluctuations at large scales (small q) indicates that the pattern has very little fluctuation in the number of receptors used to cover large surface areas.

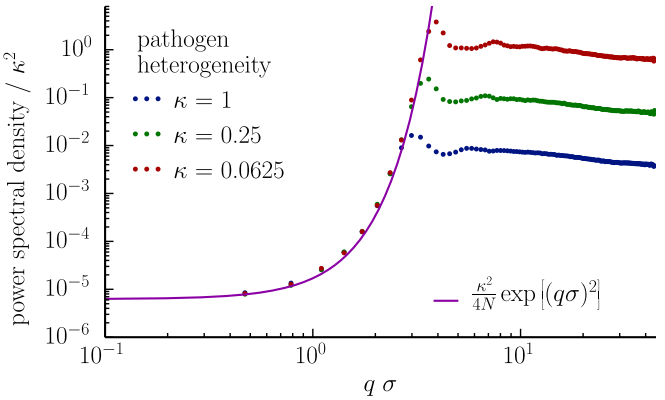

Fig. S3. Power spectral density normalized by the squared antigenic environment heterogeneity index κ: $|\sum_r P_r e^{iqr}|^2/\kappa^2$. The data collapse for different κ shows that the fluctuations at large scale are entirely attributable to those of the antigenic environment and scale with them. At these large scales, the power spectrum of the receptor distribution is approximately given by $\exp[(q\sigma)^2]\kappa^2/4N$. The exponential term stems from the inverse of the Fourier transform of f (Eq. **S28**). In other words, the coverage of the antigenic space exactly tracks the distribution of antigens, with no additional fluctuations due to the random positioning of peaks (which would be present if this positioning was Poisson distributed). This property is called disordered hyperuniformity in the physics of jammed materials (8–10). Parameters are the same as in Fig. 3.

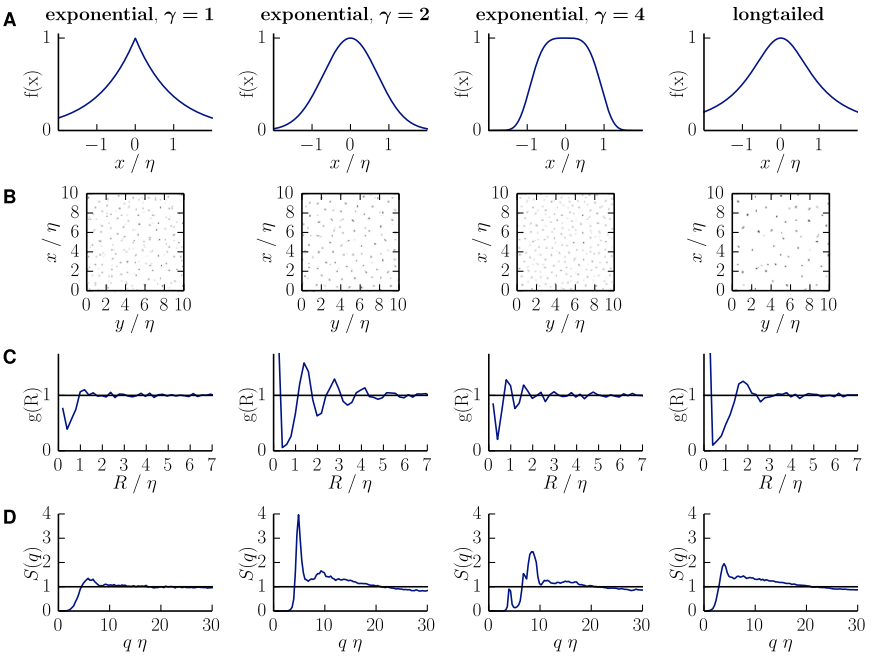

Fig. S4. Influence of the choice of the cross-reactivity kernel $f(a-r)$ on the optimization problem. Regardless of the kernel choice the optimal repertoire is peaked for nonuniform antigen distributions. The details of distribution depend on the cross-reactivity kernel. (*A*) Kernel functions used to describe cross-reactivity. *A*, *Left*, *Center Left*, and *Center Right* show exponential kernels of the form $f(r-a)=\exp[-(|r-a|/\eta)^\gamma]$ with different values of the parameter γ. *A*, *Right* shows a long-tailed kernel of the form $f(r-a)=1/(1+(|r-a|/\eta)^2)$. (*B*) Examples of optimal receptor distributions in two dimensions, for antigenic environments generated as in Fig. 3*B* (with coefficient of variation $\kappa=0.25$). (*C*) Radial distribution function of the optimal distribution. (*D*) Structure factor of the optimal distribution. The results in *C* and *D* are averaged over 10 independent runs. A linear effective cost function $F(m)=m$ is assumed throughout.

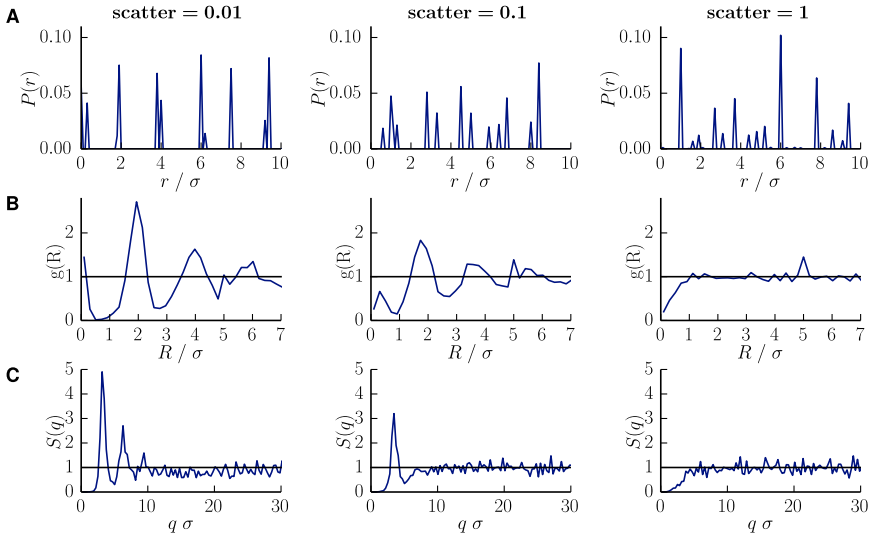

Fig. S5. Influence of varying amounts of nonuniformity in the widths of the cross-reactivity kernel on the optimization problem. (*A*) Examples of optimal receptor distributions in one dimension, for antigenic environments generated as in Fig. 3*B* (with coefficient of variation $\kappa = 0.25$). A linear effective cost function $F(m) = m$ was assumed. (*B*) Radial distribution function of the optimal distribution. (*C*) Structure factor of the optimal distribution. The results in *B* and *C* are averaged over 30 independent runs.

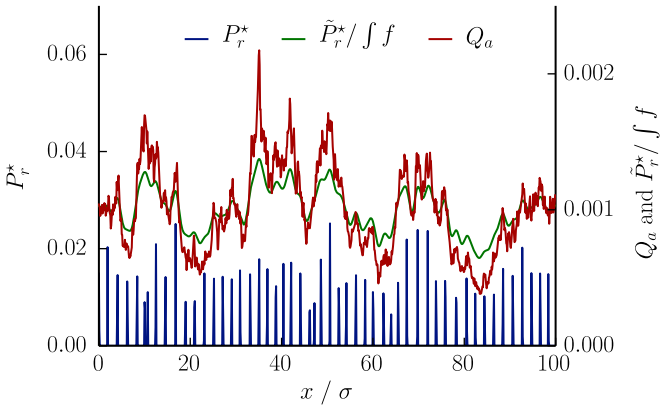

Fig. S6. Adding correlations to the antigen distribution does not change the peakedness of optimal receptor distributions. The result of the optimization is shown for a random antigen landscape with correlations. The antigen distribution is generated by Fourier filtering. First we generate an uncorrelated, normally distributed random series. This series is then filtered to obtain a power spectrum $\propto 1/(1+(10q\sigma)^2)$. Finally, the filtered series is exponentiated to ensure the nonnegativity of the generated values.

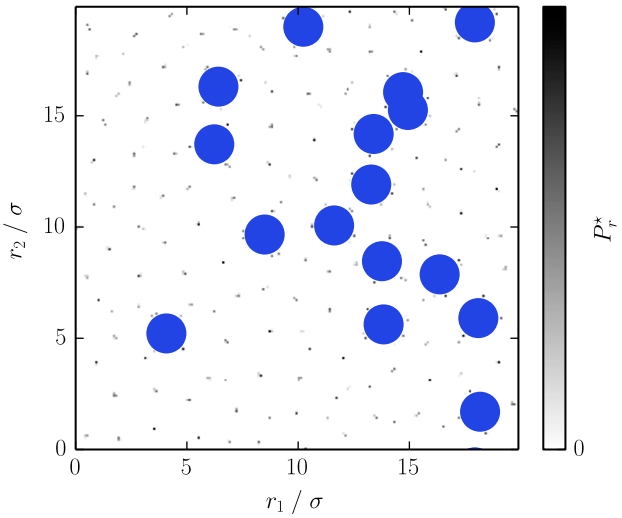

Fig. S7. Effect of exclusion zones around self-antigens on the optimization problem. No receptors are allowed in exclusion regions around self-antigens (shaded in blue).

Table S1. Intermediate results in the derivation of the optimal solution

$F(m)$	$\bar{F}(\tilde{P}_a)$	$h(x)$
m^α	$\Gamma(1+\alpha)/\tilde{P}_a^\alpha$	$(-x/(\alpha\Gamma(1+\alpha)))^{\frac{1}{1+\alpha}}$
$\ln m$	$\gamma - \ln\tilde{P}_a$	$-1/x$
$1-\exp(-\beta m)$	$\beta/(\beta+\tilde{P}_a)$	$\sqrt{-\beta/x}-\beta$
$\Theta(m-m_0)$	$\exp(-m_0\tilde{P}_a)$	$-\ln(-x/m_0)/m_0$

The first column shows several choices of the effective cost function, $F(m)$. For these cost functions the second column shows the average cost of a pathogenic attack, $\bar{F}(\tilde{P}_a)$, and the third column shows the inverse of its derivative, $h=(\bar{F}')^{-1}$. Γ is the Gamma function, γ is Euler's constant, and β and m_0 are positive constants.