Tregs self-organize into a computing ecosystem and implement a sophisticated optimization algorithm for mediating immune response

Robert Marsland III[a,1], Owen Howell[a], Andreas Mayer[b], and Pankaj Mehta[a,c,1]

[a]Department of Physics, Boston University, Boston, MA 02215; [b]Lewis-Sigler Institute, Princeton University, Princeton, NJ 08540; and [c]Faculty of Computing and Data Science, Boston University, Boston, MA 02215

Regulatory T cells (Tregs) play a crucial role in mediating immune response. Yet an algorithmic understanding of the role of Tregs in adaptive immunity remains lacking. Here, we present a biophysically realistic model of Treg-mediated self-tolerance in which Tregs bind to self-antigens and locally inhibit the proliferation of nearby activated T cells. By exploiting a duality between ecological dynamics and constrained optimization, we show that Tregs tile the potential antigen space while simultaneously minimizing the overlap between Treg activation profiles. We find that for sufficiently high Treg diversity, Treg-mediated self-tolerance is robust to fluctuations in self-antigen concentrations but lowering the Treg diversity results in a sharp transition—related to the Gardner transition in perceptrons—to a regime where changes in self-antigen concentrations can result in an autoimmune response. We propose an experimental test of this transition in immune-deficient mice and discuss potential implications for autoimmune diseases.

adaptive immunity | Tregs | optimization | ecology | biophysics

The adaptive immune system of humans and other mammals solves a challenging computational problem with amazing reliability. Using only the information contained in the binding affinities between certain macromolecules, the system must distinguish potentially pathogenic cells from its own cells, to eliminate the former without harming the latter. While understanding how the immune system accomplishes this feat is fascinating even from a purely theoretical point of view, this problem also has many urgent practical implications since an increasing number of autoimmune diseases and allergies are thought to stem from an inability to accurately distinguish self-antigens and foreign antigens (1).

An important step forward in the effort to understand adaptive immunity came with the discovery of regulatory T cells (Tregs) (2, 3). Like all T cell phenotypes, Tregs express T cell receptors (TCRs) on their surface, which bind to antigen peptides displayed on the surface of other cells via the major histocompatibility complex (MHC). But unlike in conventional T cells, TCR binding and activation in Tregs have the effect of suppressing T cell proliferation and cytokine production. This Treg-mediated suppression of self-activation complements negative selection in the thymus against self-reactive T cell lineages. In fact, it has been experimentally shown that even after undergoing negative selection in the thymus, T cells can raise a full immune response against native tissues if Tregs are artificially removed from the immune system (2). For all these reasons, Tregs are thought to play a critical role preventing autoimmune responses.

The problem of distinguishing self from nonself is made even harder by the requirement that the immune system must be able to reliably respond to even small amounts of foreign antigen but be robust to potentially large fluctuations in the concentrations of self-antigens. Specifically, even in the absence of foreign antigens, the immune system needs to tolerate fluctuations in the relative abundances of self-antigens, as development, circadian cycles, and other natural variations in biological activities lead to very different protein expression patterns even in healthy cells.

In this paper, we propose a simple dynamical model for the interactions between conventional T cells, Tregs, and antigens, which captures the essential aspects of known Treg biology. We then use this minimal model to understand how the immune system can achieve the stringent requirements described above, while simultaneously providing an algorithmic interpretation for Treg-mediated adaptive immunity. This work builds on a longstanding tradition of ecological modeling in theoretical immunology (4–6) and is an application of this approach to Treg dynamics. Finally, we note that for notational brevity, in this paper we always use T cells to mean conventional T cells (nonregulatory T cells).

Model Development

We now present a minimal model for Treg-mediated self-tolerance. As shown in Fig. 1, our model has three basic components: conventional T cells, Tregs, and antigens. Upon binding an antigen, T cells enter an active state in which they begin to proliferate and to release various interleukin signaling molecules. These interleukins stimulate other immune cells, including Tregs which in turn suppress T cell proliferation (Fig. 1A). It is known that the activation of a sufficiently large number of T cells triggers an immune response, a process likely mediated through a quorum-sensing mechanism (7). We do not explicitly include this last process in the minimal model presented in the main

---

Significance

Understanding the origins and mechanisms underlying autoimmune diseases represents a major conceptual problem in immunology. It is now clear that regulatory T cells (Tregs) play an important role in suppressing autoimmune responses, but we lack an algorithmic understanding of how they function. Here we present a simple mathematical model that organizes known biology and suggests that Tregs implement a sophisticated optimization algorithm to prevent autoimmune responses. We propose experimental tests of these ideas and discuss implications of our results for autoimmune diseases.

---

[1]To whom correspondence may be addressed. Email: pankajm@bu.edu or marsland@bu.edu.
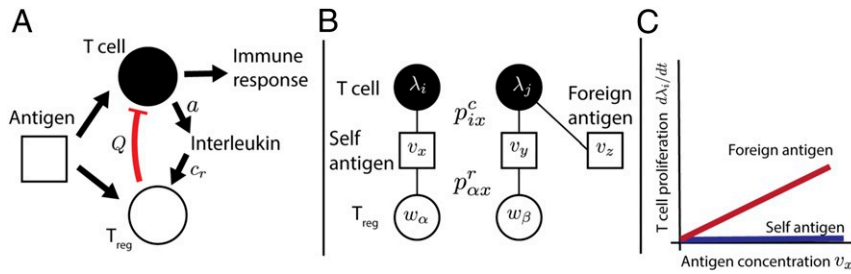
Fig. 1. A minimal model of Treg-mediated self-tolerance. (*A*) Binding to antigen stimulates proliferation of conventional T cells ("T cell"), which in turn initiates an immune response. Antigen binding also stimulates proliferation of Tregs, which locally inhibit T cell proliferation, with inhibition strength Q. Treg proliferation also requires a sufficient local concentration of interleukin, which is produced at rate a by activated T cells. (*B*) Idealized "whitelist" network of T cell and Treg cross-reactivity functions. Each edge represents the interaction of T cell i (p^c_{ix}) or Treg α ($p^r_{\alpha x}$) with antigen x. In this example, each kind of Treg binds only to one kind of self-antigen, and no Tregs bind to the foreign antigen. We model the growth rate of the population λ_i of T cells of type i as a function of the cross-reactivities, the abundance v_x of each antigen, and the abundances w_α of the Tregs. (*C*) The ideal network ensures that T cell proliferation rates are insensitive to self-antigen concentrations v_x and remain zero as levels of various proteins naturally fluctuate. At the same time, foreign antigens that do not bind to the Tregs can cause net proliferation of T cells and produce an immune response.

text. Instead, we focus on the initial phase of the response and ask whether T cells will begin to proliferate in response to a foreign ligand or a change in the concentration of self-ligands. We assume that sufficient proliferation will result in immune response through the processes downstream of the signaling pathways we study here. In *SI Appendix*, section I, we show that the minimal model presented in the main text can be derived from a more biologically realistic mechanistic model that includes additional components.

A typical human immune system has been estimated to contain about $N_c \sim 10^6$ distinct lineages of conventional T cells and a similar number N_r of Treg lineages, each carrying a different TCR, specific to a different set of antigens (8). Labeling each T cell lineage by i ($i = 1, 2 \ldots N_c$) and each Treg lineage by α ($\alpha = 1, 2, \ldots N_r$), we can encode this diversity in the cross-reactivity functions p^c_{ix} and $p^r_{\alpha x}$, which quantify the strength of interaction of conventional T cells and Tregs with possible antigens x, respectively. At the most basic level, the index x simply represents a unique amino acid sequence that could be displayed on the cell surface. But since T cells are known to respond to antigens in a tissue-specific manner (9), x can more abstractly be thought of as indexing possible tissue–antigen pairs (see *SI Appendix* for more details). For brevity, we often refer to these antigen–tissue pairs by the shorthand antigen. One can visualize the cross-reactivity functions as an interaction network, with nodes corresponding to T cells, Tregs, and antigens and edges representing the interaction strengths (see Fig. 1*B* for a particularly simple example).

Our aim is to use these cross-reactivity functions to model the dynamics of the number of cells λ_i of conventional T cell lineage i and the number of cells w_α of Treg lineage α. In general, these abundances will depend on antigen concentrations. We use v_x to denote the abundance of antigen x. We assume that the timescales on which self-antigen concentrations change are much slower than the Treg/T cell dynamics, so these v_x will be treated as fixed quantities when we analyze T cell and Treg dynamics or find possible steady states of T cell and Treg abundances. It turns out that this assumption is not essential to our main results, as shown in *SI Appendix*, Fig. S3, because the independence of growth rates from v_x in the emergent tiling phase implies that the same solution exists regardless of the variations in antigen levels.

In our minimal dynamical model, T cells of lineage i can be activated at a rate proportional to the cross-reactivity functions p^c_{ix} times the antigen concentration v_x. When activated, T cells proliferate at a rate ρ. As shown in Fig. 1*B*, T cell proliferation is suppressed by Tregs. Experiments indicate that Treg-mediated suppression of T cell proliferation is highly localized (10, 11).

This is incorporated in our model by an antigen-specific suppression level Q_x that is proportional to the abundance of Tregs activated by antigen x, with a constant of proportionality b. With these assumptions, conventional T cell abundances can be described using the differential equation

$$\frac{d\lambda_i}{dt} = \lambda_i \sum_x p^c_{ix} v_x (\rho - Q_x)$$
$$Q_x = b \sum_\alpha p^r_{\alpha x} w_\alpha, \qquad [1]$$

where the first term $p^c_{ix} \lambda_i v_x$ gives the abundance of T cells activated by each tissue-specific antigen x, and the second term $(\rho - Q_x)$ is the growth rate of activated cells.

Experiments also indicate that T cell activation stimulates proliferation of nearby Tregs (3). One potential mechanism for this interaction is the local production of interleukin signals by activated T cells. Tregs are known to be particularly sensitive to interleukin levels and to rapidly take up interleukin from their environment (12). We denote the local interleukin concentration in the vicinity of cells displaying a particular tissue-specific antigen concentration x by IL_x. The change in number of Tregs from lineage α, w_α, can be written as the product of the abundance of Tregs bound to a tissue-specific antigen x (given by $p^r_{\alpha x} v_x w_\alpha$) and an interleukin-dependent local proliferation rate $c_r \text{IL}_x$ (with proportionality constant c_r). We also assume that the in absence of interleukin Tregs die at a rate m. These Treg dynamics can be summarized in the differential equation

$$\frac{dw_\alpha}{dt} = w_\alpha \sum_x p^r_{\alpha x} v_x [c_r \text{IL}_x - m]$$
$$\text{IL}_x = \frac{a \sum_j p^c_{jx} \lambda_j}{c_r \sum_\beta p^r_{\beta x} w_\beta}. \qquad [2]$$

In *SI Appendix*, section I we show that Eqs. **1** and **2** can be derived from a realistic mechanistic model in the limit where interleukin dynamics are assumed to be fast compared to T cell and Treg proliferation.

Surprisingly, we can rewrite these dynamics in a slightly different way that makes no explicit reference to antigens. The central objects in this formulation are the "overlap kernels" $\phi_{i\alpha}$—which measure the similarity between the activation profiles of a T cell from lineage i and a Treg from lineage α—and

$\phi_{\alpha\beta}$—which measures the overlap between Tregs from lineages α and β—with

$$\phi_{i\alpha} = \sum_x v_x p_{\alpha x}^r p_{ix}^c$$

$$\phi_{\alpha\beta} = \sum_x v_x p_{\alpha x}^r p_{\beta x}^r. \qquad [3]$$

Note that these overlaps depend on the cross-reactivity functions as well as the antigen concentrations. In *SI Appendix*, we describe an approximation that is exact in the "emergent tiling" regime discussed below, which allows the dynamics of Eqs. **1** and **2** to be written entirely in terms of the overlaps:

$$\frac{d\lambda_i}{dt} = \lambda_i r_i \left[\rho - b r_i^{-1} \sum_\alpha \phi_{i\alpha} w_\alpha \right]$$

$$\frac{dw_\alpha}{dt} = \frac{m}{\sum_\beta w_\beta \bar{p}_\beta} w_\alpha \left[\kappa_\alpha - \sum_\beta \phi_{\alpha\beta} w_\beta \right], \qquad [4]$$

with $\kappa_\alpha = a \sum_j \lambda_j \phi_{j\alpha}/m$, $r_i = \sum_x v_x p_{ix}$, and $\bar{p}_\beta = \sum_x p_{\beta x}^r$. Surprisingly, the explicit dependence on antigens has completely disappeared from Eq. **4**. Instead, all information about antigen concentrations appears only through the overlap kernels. This is similar to the "kernel trick" in machine learning where all information about overlaps in a feature space can be encoded in a kernel function (13).

These equations also naturally lend themselves to an ecological interpretation in terms of consumer resource models and generalized Lotka–Volterra models. We can view T cells as exponentially growing "resources," with growth rate $r_i^{-1}\rho$, that are consumed by Tregs, with the consumption rate depending on the "resource utilization function" $\phi_{i\alpha}$. Note that the Treg dynamics in Eq. **4** take the form of a Lotka–Volterra equation. Tregs grow at a rate κ_α that depends on how many resources they consume but also compete with other Tregs. The strength of competition depends on the overlap kernel $\phi_{\alpha\beta}$. Ecologically, $\phi_{\alpha\beta}$ can be thought of as the "niche overlap" between Tregs in antigen space. The mechanistic origins of this competition can be traced to the fact that higher niche overlaps mean Tregs are more likely to be colocalize and hence more likely to compete for interleukins produced by T cells at a given tissue/antigen (Fig. 1).

Results

We now analyze the implications of these dynamics in greater detail. Our analysis exploits the ecological interpretation described above by making use of our recently discovered mapping between ecological dynamics and constrained optimization (14–16). This allows us to naturally give an algorithmic interpretation of the computations performed by Tregs and identify a phase transition in the behavior of Treg-mediated self-tolerance as a function of Treg diversity.

Tregs Minimize Niche Overlap in Antigen Space. If Treg and conventional T cell dynamics are fast compared to the rate at which antigen concentrations change, we can focus on analyzing the steady-state abundances of Tregs and T cells. In this case, we can set the left-hand side of equations in Eq. **4** equal to zero and the resulting steady-state equations have a natural interpretation in terms of constrained optimization. In *SI Appendix*, section II, we show that the steady states of the dynamical equations Eq. **4**

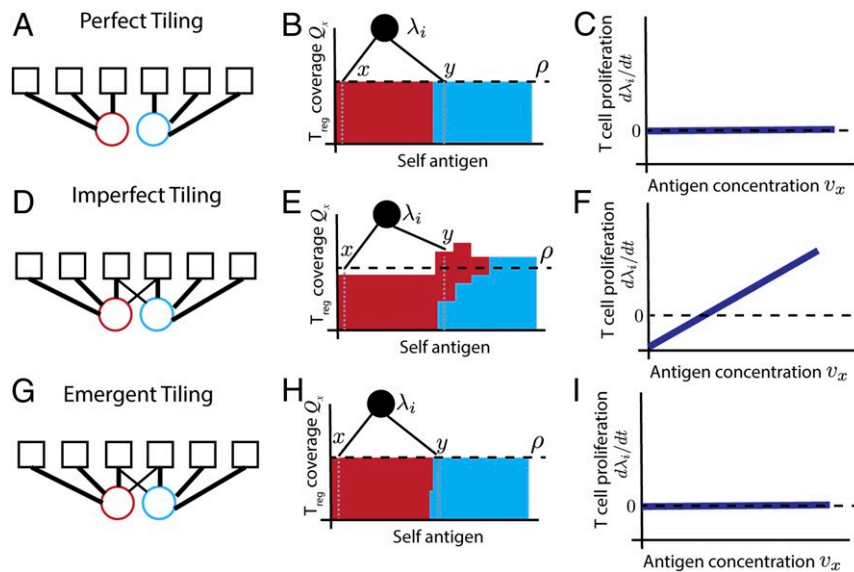are equivalent to the solutions of the following the constrained optimization problem:

$$\operatorname*{argmin}_{\mathbf{w}} \frac{1}{2} \sum_{\alpha,\beta} w_\alpha \phi_{\alpha\beta} w_\beta$$

$$\text{subject to}: \; r_i^{-1} \sum_\alpha \phi_{i\alpha} w_\alpha \geq \frac{\rho}{b} \; \text{and} \; w_\alpha \geq 0. \qquad [5]$$

The T cell concentrations λ_i play the role of generalized Lagrange multipliers in the Karush–Kuhn–Tucker (KKT) conditions that enforce the inequality constraints above.

This optimization problem has a very beautiful biological interpretation. To see this, note that the $\phi_{\alpha\beta}$ is a measure of the similarity between Tregs. Thus, this optimization tells us that Treg populations self-organize to minimize overlaps in the activation profile of Tregs. However, when performing this optimization, one must ensure that no T cell lineage i can be activated in the absence of foreign ligands to prevent undesirable autoimmune responses. This last condition is represented in the inequality constraint $r_i^{-1} \sum_\alpha \phi_{i\alpha} w_\alpha \geq \frac{\rho}{b}$ which simply states that Tregs must be able to suppress the proliferation of any T cell lineage i. Equivalently, from an ecological perspective, this means that Tregs must be able to cover the potential space of activated T cells while simultaneously minimizing niche overlap between Tregs. This is analogous to niche partitioning and species packing in ecological systems (17, 18), with Treg lineages playing the roles of species and T cells playing the role of resources.

Emergent Tiling Is Required for Robust Self-Tolerance. The previous results relied on the overlap kernel formulation of our model. We now reanalyze these dynamics from the perspective of antigens. As discussed in the Introduction, an important property of an effective immune system is that it must reliably respond to foreign antigens but also be robust to fluctuations in the concentrations of self-antigens. One simple, but biophysically unrealistic, way of achieving this robustness is depicted in Fig. 1B, which depicts a simple "whitelist" scenario for Treg-mediated self-tolerance, with a specialized Treg for each self-antigen and all Tregs having the same binding affinity. In such a scenario, the sums over x in Eqs. **1** and **2** are unnecessary, since each equation contains only one term. Setting $d\lambda_i/dt = dw_\alpha/dt = 0$ then yields uniform equilibrium Treg coverage $Q_x = \rho$ and a uniform interleukin profile $IL_x = m/c_r$ over all antigens x that interact with a surviving T cell and Treg. As can be immediately seen from the equations, these conditions guarantee that proliferation rates $d\lambda_i/dt$ and dw_α/dt remain zero for all choices of the antigen concentrations v_x. A direct consequence of this is that Treg-mediated self-tolerance is robust against fluctuations in the antigen abundances. Nonetheless, if a foreign antigen is introduced that does not interact with any of the Tregs, T cells that bind to the new antigen will still proliferate at the original rate ρ since there is no specialized Treg to inhibit their growth.

Real immune systems, however, cannot achieve this one-to-one correspondence between self-antigens and Tregs because of biophysical constraints stemming from the nature of TCR–peptide interactions (19). Fig. 2 illustrates three other possible scenarios of response to fluctuations in self-antigen concentrations (Fig. 2 A–C). In the first scenario, each Treg binds with equal strength to a different set of nonoverlapping antigens. These sets perfectly tile the set of self-antigens, covering all of them without gaps or overlaps. Since each antigen interacts with only one Treg (although Tregs will generically interact with multiple antigens), a set of Treg abundances can still easily be found for which the proliferation rate vanishes everywhere. However, just as in the original whitelist example, this scenario can only be

Marsland et al.
Tregs self-organize into a computing ecosystem and implement a sophisticated optimization algorithm for mediating immune response

PNAS | 3 of 8
https://doi.org/10.1073/pnas.2011709118

Fig. 2. Emergent tiling of self-antigen space required for robust self-tolerance. (*A*) Cross-reactivity network with perfect tiling of self-antigens by Tregs. Each self-antigen binds to exactly one Treg receptor, and all nonzero affinities are equal. (*B*) Local Treg-mediated suppression in this scenario, with the height of each colored region representing the suppression level contributed by a given Treg in the vicinity of a given self-antigen. The two Treg lineages are shown in the same colors as in *A*, and antigens are arranged along the horizontal axis. Perfect tiling ensures that the suppression strength is uniform across all self-antigens, exactly canceling the basal proliferation rate ρ. Also shown is a conventional T cell (black circle) that binds to two of the antigens (x and y). (*C*) Net proliferation rate of T cell from *B*. Rate is shown as a function of the concentration v_x of antigen x, assuming that the concentration v_y of antigen y is held fixed. (*D*) Generic cross-reactivity network, with nonuniform affinities and overlap between the cross-reactivity functions of different Tregs. (*E*) Possible result for Treg-mediated suppression, which is no longer uniform across antigens. (*F*) Proliferation rate of T cell from *E* as a function of v_x at fixed v_y. (*G*) Generic cross-reactivity network, as in *D*. (*H*) Treg coverage under emergent tiling, where a set of Treg abundances w_α is found that restores uniform suppression levels, despite the heterogeneity and overlaps in the cross-reactivity network. Note that this solution is generically possible only at much higher levels of Treg diversity for the given number of antigens. (*I*) Proliferation rate of T cell from *H* as a function of v_x at fixed v_y.

achieved by requiring a biophysically implausible fine tuning of TCR–peptide binding. In the absence of such fine tuning, the system becomes sensitive to fluctuations in self-antigen abundances v_x since generically the Treg coverage of antigen space will be uneven (Fig. 2 *D–F*).

These examples suggest that the only biophysically plausible way for the immune system to achieve robustness to fluctuations in self-antigens while maintaining sensitivity to foreign antigens is to tune the relative abundances of the Tregs and T cells to produce uniform coverage and a uniform interleukin profile, with $Q_x = \rho$ and $IL_x = m/c_r$ at every antigen x, despite the biophysically unavoidable overlaps between the Treg cross-reactivity functions (Fig. 2 *D–F*). We call such a tiling of the antigen space an emergent tiling. Note that emergent tiling is generically impossible in the simple cross-reactivity network sketched here, with many antigens and only two Tregs, and the schematic is meant only to convey the idea of uniform coverage in the presence of overlaps. We now proceed to investigate the conditions for emergent tiling without fine tuning in more complex networks.

Emergent Tiling Is Possible above a Threshold Level of Treg Diversity. Inserting the full expressions for IL_x and Q_x from Eqs. **1** and **2** into the conditions for uniform coverage discussed above ($Q_x = \rho$, $IL_x = m/c_r$), we find that sufficient conditions for a stable emergent tiling solution can be written as

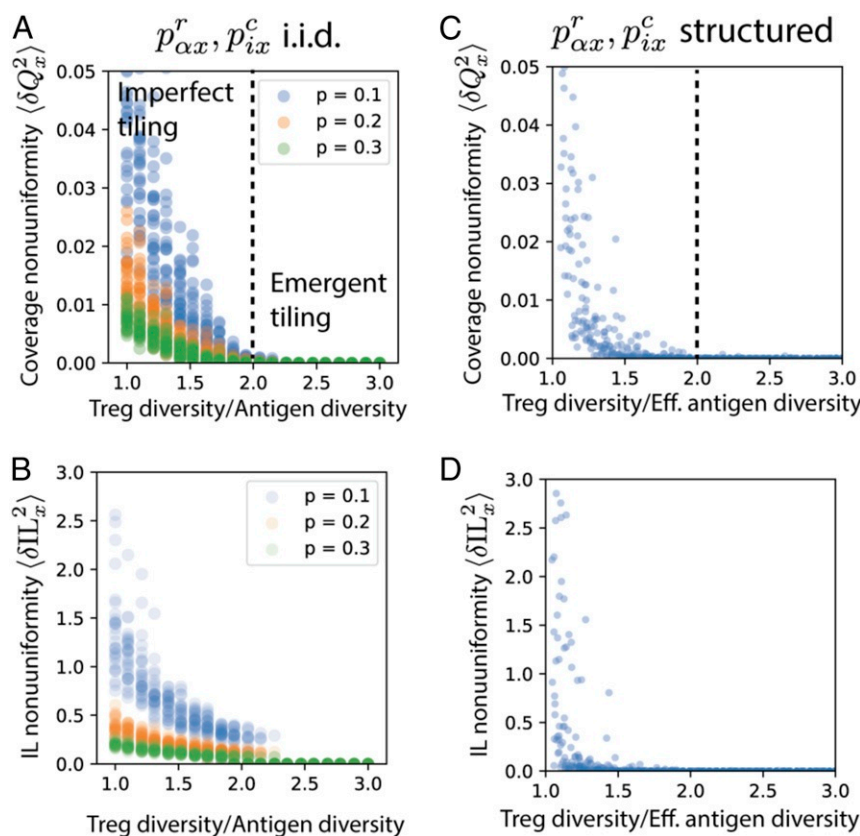$$\sum_{\alpha} p^r_{\alpha x} w_\alpha = \frac{\rho}{b} \quad \text{and} \quad \sum_{j} p^c_{jx} \lambda_j = \frac{m\rho}{ab}, \qquad [6]$$

where as before α runs over Treg lineages, j over T cell lineages, and x over possible self-antigens. Note that we have used the first condition to simplify the expression for IL_x in the second one. When a solution to these equations exists, it

also solves the optimization problem stated in Eq. **5**, but this is not immediately clear from the formulation in terms of overlaps. In *SI Appendix*, section II, we perform a series of duality transformations to rewrite the optimization problem in such a way that the equations in Eq. **6** naturally arise. In general, whether these conditions can be satisfied will depend on the dimensionality of the space we are working in (i.e., the number of Tregs N_r, the number of T cells N_c, the number of tissue-specific antigens N_a, and the structure of the cross-reactivity function).

The question of whether the Treg dynamics achieve emergent tiling is thus reduced to the question of whether the equations in Eq. **6** have solutions. We begin by investigating the simplest kind of cross-reactivity function, where the entries of the matrix $p^r_{\alpha x}$ and p^c_{ix} are independently drawn from a Bernoulli distribution. (As we discuss below, real cross-reactivity functions are much more complicated than this but it is helpful theoretically to start with this simple model to gain intuition.) This problem has been considered in various contexts including the theory of perceptrons (a simple model of statistical learning) and more recently in the context of ecology (20–22). In fact, it is possible to show that in high dimensions when $N_r, N_c, N_a \gg 1$ a solution generically exists if there are at least twice as many T cell and Treg lineages as types of antigens: $N_r/N_a > 2$ and $N_c/N_a > 2$ and no solution exists if the opposite is true. These two regimes are separated by a phase transition known as the Gardner transition in the statistical physics literature (20, 22, 23). We give a brief overview of this relation in *SI Appendix*, section IIID.

In Fig. 3 *A* and *B*, we plot the equilibrium deviations from uniform Treg coverage $\langle \delta Q_x^2 \rangle = \frac{1}{N_a} \sum_x (Q_x - \rho)^2$ and from uniform interleukin concentration $\langle \delta IL_x^2 \rangle = \frac{1}{N_a} \sum_x (IL_x - m/c_r)^2$ as a function of the ratio N_r/N_a, for randomly generated cross-reactivity functions with $N_a = 100$ antigens, with $p^r_{\alpha x}$ sampled

Marsland et al.
Tregs self-organize into a computing ecosystem and implement a sophisticated optimization
algorithm for mediating immune response

Fig. 3. A phase transition from imperfect tiling to emergent tiling as a function of Treg diversity. (*A*) Mean-squared deviation $\langle \delta Q_x^2 \rangle$ of Treg coverage from the ideal uniform solution $Q_x = \rho$ versus the ratio N_r/N_a of the number of Treg lineages to the number of antigens. Equilibrium Treg and T cell abundances were obtained by convex optimization as described in the main text and *SI Appendix*. N_a and the number of conventional T cell lineages N_c were held fixed at 100 and 1,000, respectively, while N_r was swept from 100 to 300. The elements of the cross-reactivity functions $p_{\alpha x}^r$ and p_{ix}^c were sampled independently from Bernoulli distributions with success probability p from 0.1 to 0.3. (*B*) Mean-squared deviation $\langle \delta IL_x^2 \rangle$ of local interleukin levels from the ideal uniform solution $IL_x = m/c_r$, for the same set of simulations. (*C*) Same as in *A*, but with structured cross-reactivity functions $p_{\alpha x}^r$ and p_{ix}^c encoding a one-dimensional shape space, as described in the main text and *SI Appendix*. The horizontal axis indicates the ratio of the effective number of Tregs N_r^{eff} to N_a, where N_r^{eff} is defined as the number of singular values of the cross-reactivity matrix $p_{\alpha x}^r$ that exceed a cutoff threshold of $\epsilon = 10^{-6}$. (*D*) Mean-squared deviation of local interleukin levels for the same set of simulations.

from Bernoulli distributions with three different probabilities of interaction. To make these plots, we solved the corresponding constrained optimization problem (Eq. **5**) using standard numerical techniques for convex optimization (*SI Appendix*, section IV). We see that in all cases, the nonuniformity vanishes near the predicted transition point $N_r/N_a = 2$.

To determine the number of Treg lineages N_r needed to achieve emergent tiling, we must estimate the number of distinct self-antigens N_a. The maximum number of cleavage points for creating a peptide for display on the MHC is the total number of codons ($\sim 10^7$) in the coding regions of the human genome. Since some of these peptides are redundant, and only a fraction can be successfully cleaved and loaded onto MHCs, a conservative upper bound would have 10% of the possible cleavage sites result in displayed peptides, yielding $N_a \sim 10^6$. This number is the same order of magnitude as the observed Treg diversity, which has been estimated at 3.5×10^6 (8). It is therefore not implausible that $N_r/N_a > 2$ in real immune systems.

However, biologically realistic cross-reactivity functions differ significantly from the independent and identically distributed (i.i.d.) cases described above since antigens with similar shapes bind to similar sets of receptors. Receptor affinities in Tregs and conventional T cells are shaped through a complex process of positive and negative selection in the thymus. For example, experiments suggest that Tregs may in fact have higher affinity for self-antigens than conventional T cells (24). In general,

quantitatively understanding cross-reactivity functions in different cell types is an important open problem. In light of this incredible and unknown complexity, we sought to ask whether the intuition above also holds for slightly more biologically realistic scenarios. To do so, we used a toy model for cross-reactivity based on antigen low-dimensional shape spaces that have been extensively used in the statistical physics literature to model TCR repertoires (25). Fig. 3 *C* and *D* shows numerical simulations of a one-dimensional shape space with $N_a = 5,000$ antigens. Each of $N_r = 500$ Tregs and $N_c = 500$ conventional T cells binds to a group of similar antigens, with $p_{\alpha x}^r = e^{-(x-x_\alpha)^2/2\sigma^2}$, $p_{ix}^c = e^{-(x-x_i)^2/2\sigma^2}$, where the center of the group x_α or x_i is randomly chosen for each Treg α and T cell i, and the width σ is the same for all lineages in a given simulation run (see *SI Appendix* for details). Since the antigens are now "correlated," instead of the absolute number of antigens it is useful to define an effective antigen dimension N_a^{eff}. We define N_a^{eff} as the effective rank of the cross-reactivity matrix $p_{\alpha x}^r$, that is, the number of singular values that exceed a cutoff threshold $\epsilon = 10^{-6}$. The effective number of antigens decreases from $N_a^{\mathrm{eff}} = 484$ to 97 as σ varies from 10 to 100. When we plot the nonuniformity $\langle \delta Q_x^2 \rangle$ and $\langle \delta IL_x^2 \rangle$ as a function of N_r/N_a^{eff}, we see a very similar pattern to the i.i.d. case, with reliable emergent tiling for $N_r/N_a^{\mathrm{eff}} > 2$. As a further check on our model, we also ran simulations when the cross-reactivities are drawn from a five-dimensional shape space

PHYSICS

IMMUNOLOGY AND
INFLAMMATION

(*SI Appendix*, section IV and Figs. S4 and S5). Once again, we found an emergent tiling phase when $N_r/N_a^{\text{eff}} > 2$. We also ran simulations to check what happens when we allow the "antigen concentrations" v_x to vary in time. *SI Appendix*, Fig. S3 shows simulations from our full model when v_x is allowed to rapidly oscillate in time. Somewhat surprisingly, we find that the dynamics still self-organize into an emergent tiling phase despite the fact that Treg and T cell abundances no longer reach a steady state. Collectively, these results suggest that the existence of the emergent tiling phase is a robust feature of this class of models. Finally, we note that for more biologically realistic choices of the cross-reactivity matrices while the exact ratio of Treg to antigen diversity at which the transition to the emergent tiling phase occurs may change, general arguments from statistical mechanics suggest that such a phase will exist even in these more complex settings.

Proposed Experimental Test of Emergent Tiling Transition. A key prediction of our previous analysis is that lowering the Treg diversity results in a sharp transition to a regime where changes in self-antigen concentrations can result in an autoimmune response (Fig. 3). We now propose an experimental test of this prediction by repurposing a classical immunological experimental design previously used to discover Treg function (Fig. 4). In the original experiments T cells (including both Tregs and conventional T cells) were transferred from the spleen and/or lymph nodes of a mouse with a functioning thymus to another mouse from a strain ("athymic nude") homozygous for a mutation that renders it congenitally incapable of producing any kind of T cell (2, 10). If all of the T cells are transferred together, the recipient mouse remains healthy. But if the Tregs are eliminated from the population, with only the conventional T cells injected into the recipient, severe autoimmune syndromes result in multiple organs.

In the decades since these original studies were carried out, additional techniques have been developed that make it possible to modulate and assess the impact of Treg repertoire diversity. First, it was shown that Tregs can undergo clonal expansion in vitro after coculture with dendritic cells (26). Specific Treg clones can be expanded using known receptor–antigen pairs, or all of the Tregs can be expanded together, nonspecifically, by loading the dendritic cells with antibodies against CD3, one of the components of the TCR, which effectively acts like a universal antigen. Second, repertoire diversity can be directly measured with high-throughput sequencing of the TCR genes (8, 27, 28).

These two techniques together allow one to envision experiments with Treg ecology analogous to existing methods in microbial ecology (29). In particular, one could reduce the Treg repertoire diversity by imposing a population bottleneck. Clonal expansion could then be applied to bring the total cell count back to the original level, before injecting the cells into the recipient organism. Receptor sequencing of the postexpansion cells makes it possible to quantify the final Treg diversity and calibrate the bottleneck to achieve a range of diversity levels.

Our theory predicts that emergent tiling is required for robust self-tolerance and that emergent tiling is possible only if the Treg diversity exceeds a certain threshold, which is larger than the bare minimum level required to simply cover all of the self-antigens. If the diversity is reduced to sufficiently low levels, the Treg populations can only achieve imperfect tiling, and conventional T cell proliferation can be induced by natural fluctuations in self-antigen concentrations. Measures of autoimmune activity, such as concentrations of antibodies against self-peptides from various organs (10), should therefore show a strong negative correlation with the number of distinct Treg clones injected into the athymic mouse, with a sharp Treg diversity threshold below which there is an autoimmune response.

Comparison with Existing Data. Our model also makes several qualitative predictions that can be compared with existing experimental observations. First, we predict that autoimmune disorders can be caused by genetic mutations that significantly restrict the Treg TCR repertoire but otherwise leave the immune system intact. This prediction has been confirmed in nonobese diabetic (NOD) mice, a standard animal model for type I diabetes (30). These mice spontaneously develop an autoimmune disorder whereby the immune system destroys the insulin-producing β cells in the pancreas. An assay of thymic TCR repertoires from these mice revealed that the diversity of insertions/deletions
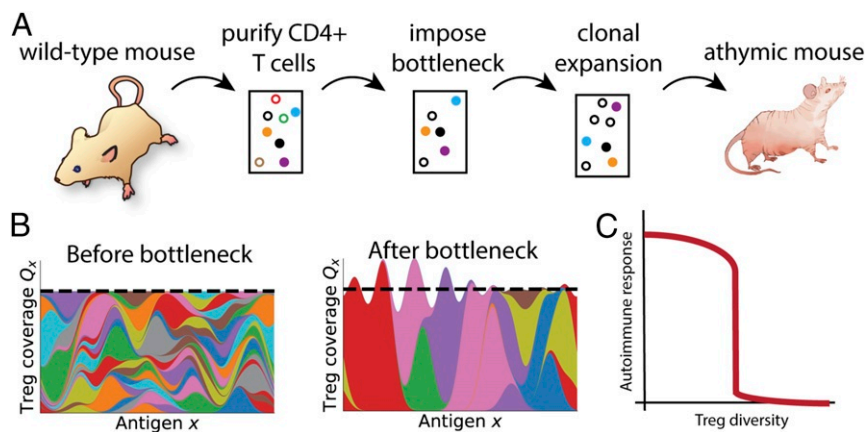
Fig. 4. Proposed experimental test for emergent tiling transition. (*A*) T cells (circles) are purified from a wild-type mouse, and Tregs are selectively depleted from the sample following standard protocols (e.g., antibody plus complement). The number of Treg clonotypes (open circles) remaining after depletion will depend on the size of the resulting population bottleneck, allowing a variety of levels of Treg diversity to be generated. The total Treg cell count is then restored to its original level using standard protocols for clonal expansion of Tregs. These Tregs, along with the rest of the T cells from the original sample, are then introduced into athymic nude mice, which lack native T cells. After an appropriate waiting period, the mouse is evaluated for an autoimmune response, for instance by measuring levels of antibodies against various self-peptides. (*B*) Schematic of Treg coverage before and after the population bottleneck. As in Fig. 2, each colored region represents a different Treg, and the height of the region represents the contribution of that Treg to the total suppression Q_x. The wild-type mouse has sufficient Tregs to achieve emergent tiling (*Left*), but this is no longer possible if the Treg diversity is driven below the threshold (*Right*). (*C*) Schematic of predicted autoimmune activity (e.g., self-antibody levels) as a function of Treg diversity. Our model predicts a sharp transition at a critical diversity level, which is significantly higher than what is required simply to cover all of the self-antigens.

in one subfamily of α-chain V domains was between five and eight times smaller for Tregs than for conventional T cells, while the diversity levels of the two lineages were similar (within a factor of 2) for wild-type (C57BL/6) mice. Furthermore, the NOD mice expressed only between 5 and 8 of the 49 functional α-chain J domains in the mouse genome, compared to more than 20 in each of the wild-type mice, resulting in a further repertoire restriction. These data suggest that the genetic defects of the NOD mice lead to excessively stringent thymic selection criteria for commitment to the Treg lineage, resulting in a Treg repertoire size below the threshold for emergent tiling.

The relationship between Treg diversity and self-tolerance has also been tested using transgenic mice engineered to eliminate all variation in the TCR β chain (31). In a similar setup to the one proposed above, a strain of mice that is congenitally deficient in Tregs (due to deletion of the Treg interleukin receptor) is injected with Tregs purified from either a wild-type mouse or one of the low-TCR-diversity transgenic mice. While the Tregs from the wild-type mouse reliably prevented autoimmune pathologies, most of the grafts from the low-diversity transgenic mice resulted in some autoimmune symptoms. The low-diversity Treg injection was sufficient, however, to keep 5 of 16 mice healthy for the duration of the experiment. This suggests that the constricted Treg repertoire of the transgenic mice may have been sufficient to provide full coverage of self-antigens but too small to achieve emergent tiling. Our model predicts that diversity levels in this range can provide some short-term protection against autoimmunity, especially in a highly controlled laboratory setting, but that this protection is easily lost due to changes in relative antigen abundances v_x.

In this same series of Treg transfer experiments, the authors of ref. 31 also explored some additional features of the Treg population dynamics. By quantifying the frequencies of individual Vα2 CD3 sequences before injection into the host and at several time points afterward, they showed that the relative population sizes of Treg clones changed dramatically over the course of 3 mo (31). Many of the dominant clones at the end of the experiment were extremely rare at the beginning, while initially dominant clones were nearly driven to extinction. This repertoire reshaping agrees with the predictions of our model, where the high interleukin levels in Treg-deficient mice lead to rapid proliferation of the exogenous Tregs, with the proliferation rates of different clones highly dependent on the distribution of displayed self-antigens within the host.

Finally, our model makes predictions about the results of injecting exogenous Tregs into an animal whose own Treg populations are already established. If the recipient has high enough Treg diversity to achieve emergent tiling, the uniform interleukin profile barely meets the minimum level required for proliferation. This means that no new Tregs will be able to proliferate, regardless of their binding specificity. If the recipient is not in the emergent tiling phase, however, some of the local interleukin levels will be sufficient to support net proliferation, and novel Tregs specific for the corresponding antigens will be able to invade. This prediction is consistent with experiments on TCR transgenic mice with Treg diversity reduced to about half the level of the wild type (32). When Tregs from wild-type mice were injected into these transgenic mice, they proliferated and eventually made up more than 20% of the overall Treg population in the recipient. But when Tregs from the transgenic mice were injected into the wild type, no proliferation was observed.

All these qualitative observations admit of many possible explanations, but together they demonstrate the power of our model for providing a single underlying biologically realistic mechanism that is parsimonious with a wide range of experimental data on Treg function.

Discussion

In this work, we have constructed a minimal model of Treg-mediated self-tolerance that is consistent with known biological and experimental facts. Specifically, our model reflects the empirical observation that Tregs suppress the proliferation of conventional T cells via a highly local mechanism (10, 11), which depends on the binding specificity of the Treg TCR (33, 34). This feature of the biology makes the overlap $\phi_{i\alpha}$ between binding profiles of Tregs and conventional T cells a crucial quantity in the analysis of Treg function. Our model also accounts for the dependence of Treg proliferation on local concentrations of interleukins generated by activated conventional T cells (3, 12). It has long been recognized that this stimulatory effect of conventional T cells on Tregs completes a homeostatic feedback loop (figure 2 of ref. 3). But the fact that interleukin is internalized and degraded after binding to a Treg's interleukin receptor (12) allows us to formulate a more specific hypothesis: that Treg population dynamics can be modeled using classical resource competition theory, with local interleukin pools acting as scarce limiting resources. This analogy highlights the role of the overlap $\phi_{\alpha\beta}$ between Treg binding profiles as a second crucial quantity and enables us to analyze the Treg behavior in terms of the ecological concepts of species packing and niche partitioning (17).

We have shown that the resulting immune dynamics have a natural interpretation in terms of optimization. We also find that for sufficiently high Treg diversities, these simple dynamics allow Tregs to self-organize into a state that allows the immune system to retain sensitivity to foreign ligands while simultaneously being robust to fluctuations in the concentrations of self-antigens.

The high Treg diversity required for emergent tiling helps explain the otherwise surprising fact that Tregs contain a similar number of distinct TCRs as conventional T cells, even though the Tregs need only to interact with peptides from the human genome, while the conventional T cells must cover all possible pathogens (8). Even though a much smaller number of Treg lineages would be sufficient to cover all of the self-antigens, emergent tiling in our simple models requires $N_r/N_a > 2$, with $N_a \sim 10^6$ in the human immune system.

This theory raises a number of further questions, which we do not address here. The first one concerns the specific requirements for maintaining pathogen sensitivity. While the repertoire of conventional T cells in an ideal immune system should cover all possible foreign antigens, the repertoire of Tregs must have some gaps in coverage to leave the system free to respond to at least some subset of foreign antigens. It has been observed that high-affinity antigen binding is associated with commitment to the Treg phenotype, suggesting a difference in positive and/or negative selection thresholds between Tregs and other phenotypes (3, 35, 36). This seems to be a natural way of achieving this difference in coverage, and understanding the details of how this might be achieved is an important area of future research. Doing so will require us to move beyond the simple shape space models for cross-reactivity used in this work and consider more biophysically realistic models for antigen–TCR binding.

Another interesting question is about the acquisition of tolerance to foreign peptides. The immune system is tolerant to many things that are not presented in the thymus when the Tregs are generated, such as peptides from various foods and commensal microbes (37). Tolerance can also be acquired later in life. There is evidence that Tregs can be generated from other T cell lineages in the periphery and not just in the thymus (3, 37). This raises the possibility that the Treg repertoire may be adaptively repopulated on a slow timescale with new lineages that interact with such nonself peptides.

Our theoretical framework allows us to generate hypotheses about causes of natural autoimmune disorders. One robust observation that has yet to be explained is the de novo onset

Marsland et al.
Tregs self-organize into a computing ecosystem and implement a sophisticated optimization algorithm for mediating immune response

PNAS | 7 of 8
https://doi.org/10.1073/pnas.2011709118

of autoimmune syndromes in people with persistent viral infections (38). A possible reason for this is that the characteristics of a persistent infection somehow do not allow for adaptation via generation of new Tregs that specifically bind to the viral antigen, and so existing Treg populations end up clonally expanding to inhibit the responding conventional T cells with which they share a partial overlap. Since the conventional T cells can proliferate on binding to viral antigen without local Treg-mediated suppression, larger populations of partially overlapping Tregs are required to maintain a global net proliferation rate of zero. But this process naturally breaks the emergent tiling needed for robustness and thus could produce increased sensitivity to fluctuations in self-antigen levels. In the future, it will be interesting to further explore this scenario to better understand whether it can increase our understanding of autoimmune diseases.

Materials and Methods

Please see *SI Appendix* for detailed materials and methods. Briefly, we constructed a detailed, biologically realistic model for the immune dynamics of Tregs, conventional T cells, and interleukins. Our model relied on experimentally supported biological assumptions that we outline in detail in *SI Appendix*. We then derived the minimal model in Eqs. **1** and **2** presented in the main text by assuming interleukin dynamics were fast compared to T cell and Treg proliferation rates and that interleukin consumption of Tregs is much larger than those of conventional T cells. By exploiting the ecological interpretation of this minimal model, we derived a dual representation of this model in terms of optimization (Eq. **5**). We then used Eq. **5** to derive sufficient conditions for emergent tiling using a series of duality transformations. To check these analytic results, we performed numerical simulations of both the dynamical models and corresponding optimization problems.

1. L. Wang, F. S. Wang, M. E. Gershwin, Human autoimmune diseases: A comprehensive update. *J. Intern. Med.* **278**, 369–395 (2015).
2. S. Sakaguchi, K. Fukuma, K. Kuribayashi, T. Masuda, Organ-specific autoimmune diseases induced in mice by elimination of T cell subset. I. Evidence for the active participation of T cells in natural self-tolerance; deficit of a T cell subset as a possible cause of autoimmune disease. *J. Exp. Med.* **161**, 72–87 (1985).
3. S. Sakaguchi, T. Yamaguchi, T. Nomura, M. Ono, Regulatory T cells and immune tolerance. *Cell* **133**, 775–787 (2008).
4. R. J. De Boer, A. S. Perelson, T cell repertoires and competitive exclusion. *J. Theor. Biol.* **169**, 375–390 (1994).
5. S. A. Levin, B. Grenfell, A. Hastings, A. S. Perelson, Mathematical and computational challenges in population biology and ecosystems science. *Science* **275**, 334–343 (1997).
6. A. Mayer, V. Balasubramanian, T. Mora, A. M. Walczak, How a well-adapted immune system is organized. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5950–5955 (2015).
7. T. C. Butler, M. Kardar, A. K. Chakraborty, Quorum sensing allows T cells to discriminate between self and nonself. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11833–11838 (2013).
8. N. Fazilleau, H. Bachelez, M. L. Gougeon, M. Viguier, Cutting edge: Size and diversity of CD4+ CD25high Foxp3+ regulatory T cell repertoire in humans: Evidence for similarities and partial overlapping with CD4+ CD25- T cells. *J. Immunol.* **179**, 3412–3416 (2007).
9. D. J. Campbell, C. H. Kim, E. C. Butcher, Chemokines in the systemic organization of immunity. *Immunol. Rev.* **195**, 58–71 (2003).
10. S. Sakaguchi, N. Sakaguchi, M. Asano, M. Itoh, M. Toda, Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol.* **155**, 1151–1164 (1995).
11. T. Takahashi *et al.*, Immunologic self-tolerance maintained by CD25+ CD4+ naturally anergic and suppressive T cells: Induction of autoimmune disease by breaking their anergic/suppressive state. *Int. Immunol.* **10**, 1969–1980 (1998).
12. T. R. Malek, The biology of interleukin-2. *Annu. Rev. Immunol.* **26**, 453–479 (2008).
13. C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).
14. P. Mehta, W. Cui, C. H. Wang, R. Marsland III, Constrained optimization as ecological dynamics with applications to random quadratic programming in high dimensions. *Phys. Rev.* **99**, 052111 (2019).
15. O. Howell, W. Cui, R. Marsland, P. Mehta, Machine learning as ecology. *J. Phys. A-Math. Theor.* **53**, 334001 (2020).
16. R. Marsland, W. Cui, P. Mehta, The minimum environmental perturbation principle: A new perspective on niche theory. *Am. Nat.* **196**, 291–305 (2020).
17. R. MacArthur, Species packing, and what competition minimizes. *Proc. Natl. Acad. Sci. U.S.A.* **64**, 1369–1371 (1969).
18. R. MacArthur, Species packing and competitive equilibrium for many species. *Theor. Popul. Biol.* **1**, 1–11 (1970).
19. A. K. Sewell, Why must T cells be cross-reactive?. *Nat. Rev. Immunol.* **12**, 669–677 (2012).
20. E. Gardner, The space of interactions in neural network models. *J. Phys. Math. Gen.* **21**, 257 (1988).
21. H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, New York, NY, 2001).
22. S. Landmann, A. Engel, Systems of random linear equations and the phase transition in MacArthur's resource-competition model. *Europhys. Lett.* **124**, 18004 (2018).
23. E. Gardner, Maximum storage capacity in neural networks. *EPL* **4**, 481 (1987).
24. C. S. Hsieh, H. M. Lee, C. W. J. Lio, Selection of regulatory T cells in the thymus. *Nat. Rev. Immunol.* **12**, 157–167 (2012).
25. G. Altan-Bonnet, T. Mora, A. M. Walczak, Quantitative immunology for physicists. *Phys. Rep.* **849**, 1–83 (2020).
26. S. Yamazaki *et al.*, Direct expansion of functional CD25+ CD4+ regulatory T cells by antigen-processing dendritic cells. *J. Exp. Med.* **198**, 235–247 (2003).
27. A. M. Sherwood *et al.*, Deep sequencing of the human TCRγ and TCRβ repertoires suggests that TCRβ rearranges after $\alpha\beta$ and $\gamma\delta$ T cell commitment. *Sci. Transl. Med.* **3**, 90ra61 (2011).
28. A. Six *et al.*, The past, present, and future of immune repertoire biology – the rise of next-generation repertoire analysis. *Front. Immunol.* **4**, 413 (2013).
29. R. Tecon *et al.*, Bridging the holistic-reductionist divide in microbial ecology. *mSystems* **4**, e00265-18 (2019).
30. C. Ferreira *et al.*, Non-obese diabetic mice select a low-diversity repertoire of natural regulatory T cells. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8320–8325 (2009).
31. D. Adeegbe, T. Matsutani, J. Yang, N. H. Altman, T. R. Malek, CD4+ CD25+ Foxp3+ T regulatory cells with limited TCR diversity in control of autoimmunity. *J. Immunol.* **184**, 56–66 (2010).
32. L. Föhse *et al.*, High TCR diversity ensures optimal function and homeostasis of foxp3+ regulatory tcells. *Eur. J. Immunol.* **41**, 3101–3113 (2011).
33. S. Hori, M. Haury, A. Coutinho, J. Demengeot, Specificity requirements for selection and effector functions of CD25+ 4+ regulatory t cells in anti-myelin basic protein T cell receptor transgenic mice. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8213–8218 (2002).
34. D. Olivares-Villagómez, A. K. Wensky, Y. Wang, J. J. Lafaille, Repertoire requirements of CD4+ T cells that prevent spontaneous autoimmune encephalomyelitis. *J. Immunol.* **164**, 5499–5507 (2000).
35. A. Košmrlj, A. K. Jha, E. S. Huseby, M. Kardar, A. K. Chakraborty, How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16671–16676 (2008).
36. A. Košmrlj, A. K. Chakraborty, M. Kardar, E. I. Shakhnovich, Thymic selection of T-cell receptors as an extreme value problem. *Phys. Rev. Lett.* **103**, 068103 (2009).
37. J. L. Coombes *et al.*, A functionally specialized population of mucosal CD103+ DCs induces Foxp3+ regulatory T cells via a TGF-β–and retinoic acid–dependent mechanism. *J. Exp. Med.* **204**, 1757–1764 (2007).
38. S. Kivity, N. Agmon-Levin, M. Blank, Y. Shoenfeld, Infections and autoimmunity–friends or foes?. *Trends Immunol.* **30**, 409–414 (2009).

8 of 8 | PNAS
https://doi.org/10.1073/pnas.2011709118

Marsland et al.
Tregs self-organize into a computing ecosystem and implement a sophisticated optimization algorithm for mediating immune response