The Impact of Thymic Antigen Diversity on the Size of the Selected T Cell Repertoire

Jose Faro, Santiago Velasco, África González-Fernández and Antonio Bandeira

*J Immunol* 2004; 172:2247-2255; doi: 10.4049/jimmunol.172.4.2247

http://www.jimmunol.org/content/172/4/2247

---

**References**  This article cites 49 articles, 22 of which you can access for free at: http://www.jimmunol.org/content/172/4/2247.full#ref-list-1

**Subscription**  Information about subscribing to *The Journal of Immunology* is online at: http://jimmunol.org/subscription

**Permissions**  Submit copyright permission requests at: http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**  Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/alerts
The Impact of Thymic Antigen Diversity on the Size of the Selected T Cell Repertoire

Jose Faro, Santiago Velasco, África González-Fernández, and Antonio Bandeira

The TCR repertoire of a normal animal is shaped in the thymus by ligand-specific positive- and negative-selection events. These processes are believed to be determined at the single-cell level primarily by the affinity of the TCR-ligand interactions. The relationships among all the involved variables are still unknown due to the complexity of the interactions and the lack of quantitative analysis of those parameters. In this study, we developed a quantitative model of thymic selection that provides estimates of the fractions of positively and negatively selected thymocytes in the cortex and in the medulla, as well as upper-bound ranges for the number of selecting ligands required for the generation of a normal diverse TCR repertoire. Fitting the model to current estimates of positive- and negative-selected thymocytes leads to specific predictions. The results indicate the following: 1) the bulk of thymocyte death takes place in the cortex, and it is due to neglect; 2) the probability of a thymocyte to be negatively selected in the cortex is at least 10-fold lower than in the medulla; 3) <60 ligands are involved in cortical positive selection; and 4) negative selection in the medulla is constrained by a large diversity of selecting ligands on medullary APCs. The Journal of Immunology, 2004, 172: 2247–2255.

In the thymus, two key processes of TCR repertoire selection take place. First, a broad and random repertoire of TCRs is generated (1). Second, in subsequent steps of development, survival or death of the immature thymocyte will occur according to the strength of the TCR-mediated interaction with the stroma cells: only those thymocytes expressing TCRs with weak/medium reactivity to the self-peptide/MHC molecules will be positively selected and complete maturation. The emerging mature population will be largely purged of cells either not reactive (death by neglect) or having strong reactivity to self-ligands (negative selection) (2, 3).

It has been proposed that the intrinsic affinity of the TCR to the respective ligand is the key parameter that determines cell fate (4). An alternative hypothesis is that selection is driven by the avidity of the interaction (5, 6), which is assumed to be directly proportional to the ligand and TCR densities and to the affinity of binding. It now appears that, although ligand and receptor densities affect the probabilities of selection, the affinity of ligand-TCR binding is the major parameter differentiating selecting from non-selecting cases involved in thymic selection is further increased by evidence showing divergent changes in how TCR ligation is handled by the intracellular signaling machinery at distinct stages of maturation. As they mature, thymocytes may alter their sensitivity to ligands by using the kinetic proofreading concept (7) coupled to other mechanisms like serial engagement (8).

To be built into any quantitative model of selection is the fact that any given TCR has a degree of cross-reactivity with distinct MHC-peptides (9, 10). Estimations of the number of different ligands recognized by a single TCR can vary several orders of magnitude depending on the experimental methodology used (11). Consequently, it is difficult to estimate the minimal number of different MHC-peptides capable to select a diverse T cell repertoire (3, 10, 12).

The fate of immature thymocytes is also dependent on their stage of maturation and on the type of the APC. Cortical thymic epithelial cells (cTECs) (3) efficiently mediate positive but not negative selection (13, 14), whereas the reverse is the case for hemopoietic-derived APCs, mainly dendritic cells (15). This may be due in part to differential functional properties intrinsic to each cell type. Thus, interactions with cTECs, which express high levels of CD54 molecules and lack CD80/CD86 molecules, favor positive selection over deletion (16). But the differential behavior observed could also be related to the different levels of ligand diversity. Although medullary stroma cells (epithelial and hemopoietic) can express all or most bloodborne Ags (17), the peptide repertoire expressed by cTECs may be more restricted (18) and even to a certain extent nonoverlapping as recently suggested (19). To what extent the number of different MHC-peptides also determines whether one selection process predominates over the other is thus an essential issue.

Thymocytes may need more than one round of positive signaling by APCs to survive and/or complete their maturation (10, 20, 21), but this remains unknown (20). The complexity of the events involved in thymic selection is further increased by evidence showing divergent changes in how TCR ligation is handled by the intracellular signaling machinery at distinct stages of maturation. As they mature, thymocytes may alter their sensitivity to ligands by...
(up-tuning of thresholds (22); developmental changes in the signaling machinery (23)) and may change their reactivity to a given ligand (23).

Because of the complexity and nonlinearity of the interactions involved, it becomes very difficult to assess intuitively the relationship among all these parameters and how they influence the thymic output in terms of repertoire and cell numbers. As with many other biological processes, an appropriate way to deal with such complexity is to develop testable mathematical models (24). Detours et al. (25) recently developed an affinity-driven T cell selection model, the analysis of which led them to conclude that thymic selection is driven by $10^{-10}$ self-peptides. In that model, the cortex and the medulla are lumped into a single compartment, obscuring the interpretation of this result. More recently, a different model was proposed based on the assumption that thymocyte selection and mature T cell activation are equally driven by the triggering rate of TCRs (26). The authors concluded that, to build a repertoire that is sensitive to foreign peptides while retaining nonreactivity to self, it is sufficient that negative selection removes only 1% of positively selected cells.

In this study, we present a quantitative model of thymic selection that discriminates selective events occurring in the cortex and in the medulla. The model was developed independently of specific information on affinity distributions or affinity thresholds, and is solely based on experimentally determined fractions of selected thymocyte populations.

The Model

The general principle

The key strategy of the present study was to explore the possibility of estimating the maximal number of cortical and medullary selecting ligands from experimentally determined fractions of selected thymocytes in normal mice. To address this issue, we modeled the relationship between the number of selecting ligands and the probability of selection of a given thymocyte. The model is designed to be robust, given the present degree of experimental uncertainty, because to describe those relationships, it does not require specifying several aspects of the biology of the system (impact of coreceptors and cosignaling molecules, intracellular signaling cascades). Nevertheless, as discussed below, this will not affect our calculations. In any case, as far as our model is concerned, affinity and off-rate are interchangeable, without affecting the considerations or the results.

Assumption B5 provides an essential simplification. At first sight, this assumption appears inaccurate, because it excludes additive and antagonistic effects that may result from the interplay between two or more ligands expressed on the same APC. However, the a priori inaccuracy of this assumption will not significantly alter our estimations. In respect to additive effects, we find that assumption B5 has the value of providing an upper bound for the expected number of selecting ligands, because additivity will only decrease the number of ligands required for selection. In Appendix E.2, we provide a formal justification for this statement. In addition, it has been shown recently that the natural pool of MHC molecules contains at most two to three peptides, structurally related to the agonist of a given TCR, and capable of promoting positive selection of the corresponding thymocytes (28). It is thus reasonable to consider, in terms of random probabilities, that interactions with single peptides may still be the case for the majority of the selected thymocytes. In terms of our model, the ability to neglect additive effects of selecting peptides depends on the location of the positive- and negative-selection thresholds in the tail of the hypothetical affinity distribution (Fig. 1).

As for the possibility of antagonist interference, the quantitative requirements of antagonistic activity suggest that this phenomenon may have an impact in thymic selection even lower than additivity (see Appendix E.2 for detailed arguments).

From the assumptions B2 and B3, it follows that thresholds $T_1$ and $T_2$ define affinity constants $K_1$ and $K_2$, respectively. This is illustrated in Fig. 1. These affinity thresholds divide the frequency distribution of thymocyte populations in three meaningful areas: a high ($K_2 < K_1$), a medium ($K_1 < K < K_2$), and a low ($K < K_1$) affinity area. These areas are the probabilities of a thymocyte being negative, positive, and insufficiently signaled, respectively, upon interaction with a single Ag. One can thus assign a probability to each signaling event both in the cortex ($P_{c}^{+} = P_{c}^{-}$) and in the medulla ($P_{m}^{+} = P_{m}^{-}$; see Fig. 1, left panel) and in the medulla ($P_{m}^{+}$, $P_{m}^{-}$, and $P_{c}^{+}$; see Fig. 1, right panel). As mentioned above, if serial engagement interferes with affinity, it will

Table 1. Parameters of the model

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{c}, n_{m}$</td>
<td>Average number of distinct peptides per APC</td>
</tr>
<tr>
<td>$r_{c}, r_{m}$</td>
<td>Average number of different interacting APCs</td>
</tr>
<tr>
<td>$z_{c}, z_{m}$</td>
<td>Minimum number of positive interactions with APCs leading to thymocyte positive selection</td>
</tr>
<tr>
<td>$P_{c}^{+}, P_{c}^{-}, P_{c}^{+}, P_{m}^{+}, P_{m}^{+}, P_{m}^{-}, F_{c}, F_{m}, F_{c}, F_{c}$</td>
<td>Null (N), positive (+) and negative (−) signaling probabilities of TCR/ligand interactions</td>
</tr>
<tr>
<td>$z_{c}, z_{m}$</td>
<td>Fractions of selected thymocytes, respectively, in the cortex ($F_{c}$), the medulla ($F_{m}$) and total ($F$)</td>
</tr>
</tbody>
</table>

* c, Cortex; m, medulla.
decrease the signaling at high-affinity interactions as proposed by Rachmilewitz and Lanzavecchia (8). That will restrict, rather than increase, the range of values of the probability parameters for selection and, hence, will not affect our calculations, which were performed for larger ranges of those probabilities. Importantly, assumption B4 as well as the difference in TCR densities between cortical and medullary thymocytes (the former expressing ~10-fold less molecules) are implicitly incorporated in the model by considering the cortical probabilities \( p_{c,N} \) and \( p_{c,m} \) independently of the respective medullary ones, \( p_{m,N} \) and \( p_{m,m} \). Moreover, the possibility that the sensitivity of thymocytes to ligands is modulated by a tunable threshold mechanism after TCR-ligand experience (22) can be implicitly incorporated as well, at least in its simplest form.

**Results**

**Experimental and theoretical constraints**

With the above assumptions, we built a mathematical model, detailed in Appendix A, to analyze the effects of the described parameters on thymic selection. Before this analysis, we applied the compartmental structure of the model to existing experimental data on total selected fractions of thymocytes, and derived thereof theoretical estimations of fractions of selected thymocytes in the cortex and the medulla. Using these estimations, the model established general relationships between the total number of selecting ligands in each compartment (\( n_N \times r_N \) and \( n_m \times r_m \)) and the corresponding probabilities of negative selection by a single ligand (\( p_{c,N} \) and \( p_{c,m} \)). Furthermore, by applying the model’s equations to experimental data obtained from mice expressing a single class II MHC-peptide, we obtained minimal estimates for the probabilities of positive signaling (\( p_{c+} \) and \( p_{m+} \)).

**Fractions of differentially selected thymocytes in the cortex and in the medulla.** Approximately 5% of daily produced thymocytes are positively selected (3). Estimates of this fraction as well as the fractions of neglected and negatively selected thymocytes (see Table II) do not distinguish between cortical and medullary thymocytes (29–31). Only one study determined the fraction of negatively selected thymocytes in the medulla (29) (see also Table II and Appendix B). Using those experimental values in Equations A5, we estimated the remaining fractions of thymocytes in the two major thymic compartments (for details, see Appendix C). The theoretical values found are shown in Table II together with the experimental data. Our calculations indicate that 75–90% of the thymocytes die of neglect in the cortex. Indeed, ~80% of CD4+8+ thymocytes from MHC class I+II+ mice do not up-regulate the activation markers CD69 and CD5 after interacting with MHC+ thymic stroma in reaggregate cultures (21, 32). This has been interpreted as indicating that those thymocytes receive very weak activation signals, if from MHC+ APCs, and thus die by neglect. Our results also indicate that in the cortex 10–20% of thymocytes are positively selected, whereas <5% are negatively selected.

In the medulla, 50–70% of the fraction of positively selected thymocytes that escaped cortical deletion are negatively selected, whereas 25–50% will complete maturation. At this point, the fraction of neglected cells cannot be estimated more precisely than as being <25%.

In conclusion, in agreement with previous observations (13, 33), our calculations indicate that the bulk of cell death in the thymus is due to neglect in the cortex (75–90%), whereas negative selection in the medulla is responsible only for 5–14% of the total cell loss.

**General boundaries for the number of selecting Ags in the cortex and the medulla.** According to the present model, the cortical and medullary fractions of negatively selected thymocytes are related to the respective probabilities of negative signaling (\( p_{c-} \) and \( p_{m-} \)) and the respective numbers of selecting ligands (\( n_N \times r_N \) and \( n_m \times r_m \)) in the following way (see Appendix A):\( 1 - F_{c-} = (1 - p_{c-})^{n_N \times r_N} \) and \( 1 - F_{m-} = (1 - p_{m-})^{n_m \times r_m} \). Substituting the theoretical estimations of \( F_{c-} \) and \( F_{m-} \) (see Table II) in the respective Equations A5, taking logarithms at both sides of the equations and using the well-known approximation \( \log(1 - p) \approx -p \) for \( |p| \ll 1 \), one is led to the following boundaries for the total number of Ags screened in normal mice by a positively selected thymocyte in the cortex (\( n_N \times r_N \)) and later in the medulla (\( n_m \times r_m \)):

\[
\begin{align*}
n_N \times r_N & \leq \frac{0.5}{p_{c-}} \quad \text{and} \quad 0.7 \leq n_m \times r_m \leq \frac{1.2}{p_{m-}} \quad \text{(i.e., } n_N \times r_N \approx \frac{1}{p_{c-}}) \quad (1)
\end{align*}
\]

Clearly, the maximum value predicted for the number of Ags in the cortex is relevant only if the probability \( p_{c-} \) is not too low; for instance, Equation 1, left, predicts that for \( p_{c-} < 10^{-3} \), the total number of selecting ligands \( n_N \times r_N \) must be <5000, whereas if \( p_{c-} < 10^{-5} \), then \( n_N \times r_N \) is <500. In contrast, the boundaries for \( n_m \times r_m \) (Equation 1, right) define a quite narrow range, and hence, it must be relevant for virtually any value of the probability of negative selection \( p_{m-} \); for example, if \( p_{m-} = 10^{-4} \), it follows that medullary thymocytes should interact with \( 10^4 \) different ligands. A relationship very similar to that between \( n_N \times r_N \) and \( p_{c-} \) has been found by Mason (11) for optimized TCR cross-reactivity, although he made no distinction between cortical and medullary negative selection.

**Impact of Ag diversity on thymic selection**

What is the number of distinct selecting Ags in the cortex and in the medulla of a normal thymus? Although some researchers suggested that a diverse T cell repertoire could be selected by very few Ags (10, 12)—perhaps the MHC molecules themselves—others contended that many different Ags are necessary for the selection of a diverse set of TCRs (10, 12, 34, 35). The latter view was...
questioned by the intuitive argument that any gain in positive selection provided by a large variety of selecting Ags would necessarily be counteracted by a similar increase in negative selection. However, it is our view that it is not possible to predict even qualitatively how those selection processes may vary with increasing numbers of selecting Ags without additional ad hoc assumptions. On equally intuitive grounds, it is possible to conceive that the relationship between selection and peptide diversity is not linear, and that there exists a window of ligand diversity in which positive selection prevails over deletion.

**Impact of antigenic diversity on thymocyte selection.** We thus studied the dependence of the different fractions of surviving or dying cortical and medullary thymocytes on the number of selecting Ags per APC \( (n_c \text{ and } n_m) \), while varying the number of rounds required for positive selection \( (z_c, z_m) \). The results show that, first, the fraction of negatively selected thymocytes increases steadily, with decreasing rate, as the number of peptides per APC increases (Fig. 2A, solid lines). As expected from Equation A4, this increase is affected only by \( r_c, p_m \). Conversely, the neglected fraction decreases steadily when the number of ligands increase. In most of the cases analyzed, the impact of the numbers of encountered APCs, \( r_c, (r_m) \), on thymocyte selection is very similar to that of \( n_c, (n_m) \). This is so not only for the fraction of negatively selected thymocytes (as expected from the model’s equations) but also for the neglected and positively selected fractions.

Second, in contrast, the effect of the number of peptides per APC on the fractions of positively selected thymocytes is biphasic and has a maximum (Fig. 2A, dotted lines). Note that this is to be expected from the above fact that, as the number of different ligands increases (see Fig. 2A, right panel), the fraction of negatively selected thymocytes increases steadily concomitantly with a steady decrease of the fraction of neglected cells. This latter becoming virtually insignificant, from this point on the positively and negatively selected fractions of thymocytes become interdependent. This behavior is dramatically modulated by parameters such as \( z_c, z_m \) or \( p_c, p_m \). Notice that, for certain values of \( z \), there is no number of ligands that obeys simultaneously the fraction of positively and negatively selected fractions of thymocytes (e.g., in Fig. 2A, bottom left panel).

An important consequence of the distinct impact of antigenic diversity on the fractions of positively and negatively selected thymocytes is that there is an optimum value for the number of peptides per APC in respect to the ratio of positively to negatively selected thymocytes in the cortex \( (F_{c+}/F_{c-}) \) and in the medulla \( (F_{m+}/F_{m-}) \). This is shown in Fig. 2B. As expected, the particular value of the optimum also depends on the model parameters (see, for instance, the impact of parameters \( z_c, z_m \) on the attained maxima). However, as discussed below, rather than optimizing the number of peptides per APC, in the cortex the immune system seems to limit the selected repertoire by reducing signaling (sub-optimal number of peptides, gray squares located at values of number of selecting ligands lower than those giving a maximum ratio). In contrast, in the medulla it seems to further limit the selected repertoire by excess of signaling (supraoptimal number of peptides).

In conclusion, this analysis provides a rationale for the view that there is a range of increasing peptide diversity in which positive selection increases faster than negative selection. Based on different theoretical grounds, an optimum for ligand diversity has also been predicted by other authors (11, 36, 37).

**Estimating normal antigenic diversity in the cortex and the medulla.** Next, we analyzed the impact of each parameter on the model’s behavior under the special case of imposing the require-

![FIGURE 2. Impact of model parameters on the frequencies of selected thymocytes. A, Dependence of the distinct fractions of dying and surviving thymocytes in the cortex and the medulla on the number of selecting Ags per APC \( (n_c \text{ and } n_m) \) and on the required minimal number of interactions for positive selection \( (z_c, z_m) \). Expected frequencies of neglected (dashed lines), positively selected (dotted lines), and negatively selected (solid lines) thymocytes are shown. Horizontal bands labeled Null (light gray), + (white), and − (darker gray) indicate, respectively, the theoretically estimated frequency ranges of default apoptosed, and positively and negatively selected thymocytes (see Table II). Ranges of admissible values for the numbers of selecting ligands \( (n_c \text{ and } n_m) \) are indicated by a bar below each graph. Note that, for the parameter values used, the case \( z = 10 \) has no range of \( n_c \) compatible with experimental data. B, The ratio of positively to negatively selected thymocyte fractions as a function of the number of selecting peptides per APC. Shown here is the fact that, both in the cortex and the medulla, this ratio has a biphasic behavior. Horizontal lines correspond to the cases \( F_{c+} = F_{c-} \text{ and } F_{m+} = F_{m-} \). Gray bands indicate the curve regions that fit data in Table II. Note that the values for \( z \) were chosen in order that the ordinate scale facilitates the visual location of the intersection of the curves with the constant line \( y = 1 \), and of the gray bands. Parameter values in A and B were as follows: \( r_c = r_m = 50 \) APC, \( p_c = 10^{-2}, p_c = 10^{-4}, p_m = 2 \times 10^{-2} \text{ and } p_m = 10^{-5} \).

ment to fit all the experimental and the theoretical constraints determined in this study (Table II and Appendix D). In this way, we obtained ranges of admissible values for each parameter as a function of the other parameters. This procedure is illustrated in Fig. 2A, where the admissible values for \( n_c \) are those for which the curves representing the fractions \( F_{c+}, F_{c-}, F_{m+} \), and \( F_{m-} \) lay within the respective horizontal bands. Note that those bands correspond to the estimated ranges in Table II. Remarkably, the number of selecting ligands obtained for the cortex (Fig. 2B, left, □) is always lower than the corresponding optimum number. In contrast, for the medulla the reverse is true, that is, the number of selecting ligands is always larger than the optimum (Fig. 2B, right, thick gray lines). If the optima for the cortex and the medulla were similar, the above finding would indicate that the number of selecting ligands per APC might be lower in the cortex than in the medulla.

The results of a systematic computer analysis using this procedure are summarized in Table III. They indicate that, in the cortex,
there is a direct correlation between the minimum number of positive-signaling cell interactions required for positive selection ($z_c$) and the number of different selecting peptides per cortical APC ($n_c$). However, there is an inverse correlation between this number, $n_c$, and both the average number of interacted cortical APCs ($r_c$) and the probability of positive signaling by a single ligand $p_{c+}$. In contrast, the value of $p_{m-}$ has very little impact on these parameters for $p_{m-}$ $\leq 10^{-4}$. Conversely, in the medulla, the values of $n_m$ that are consistent with experimental observations are quite independent of the parameters $p_{m+}$ and $z_m$, but are highly dependent on $p_{m-}$ and $r_m$, as expected from Equation 1, right.

Thus, the present analysis delineates those scenarios that entail large, medium, and small numbers of selecting Ags ($n \times r$) in terms of a few key parameters: $p_{c+}$ and $z_c$ in the cortex, and $p_{m-}$ and $z_m$ in the medulla.

In mice expressing class II molecules exclusively on cTECs (K14 mice), negative selection in the medulla is absent (13). In these mice, the fraction of mature CD4$^+$ thymocytes is reduced 2- to 3-fold as compared with normal mice and the ratio single-positive (SP) CD4/CD8 is $\sim 1$ (38). Interestingly, the density of class II molecules in cTECs is $<10\%$ the one in normal mice (T. Laufer, unpublished observations). In the present model, a 10- to 15-fold reduction in ligand density in the cortex implies a 10- to 15-fold increase in the corresponding threshold affinities, and a decrease in the values of $p_{c+}$ and $p_{c-}$ of at least 22- and 84-fold, respectively. In turn, such decrease in $p_{c+}$ and $p_{c-}$ implies $\geq 20$-fold decrease of the fraction $F_{c+}$. Thus, according to the model, it is expected that, in K14 mice, $F_{c+}^* \approx 0.005-0.01$. In contrast, a 2- to 3-fold reduction of the total fraction of positively selected cells implies that, in those mice, $F_{c+}^* = F_{c+}/3$ to $F_{c+}/2 = 0.007-0.025$. Thus, the reduction of cortical positive selection in these mice accounts by itself for the reduction observed in the number of CD4 SP thymocytes. This implies that, in the medulla of K14 mice, either ($F_{mN} + F_{m-}$) = 0, or ($F_{mN} + F_{m-}$) $< F_{m+}$. That is, not only negative selection is negligible in the medulla of K14 mice, but also in these mice, and a fortiori in normal mice, death by neglect in the medulla is negligible. Survival of mature CD4 thymocytes in the absence of class II-dependent signals could be due to weak interactions with MHC class I molecules or nonclassical MHC molecules (39).

The above estimation that $F_{c+}^* \approx F_{c+} = 0.007-0.025$ in K14 mice, amounts to an 8- to 15-fold reduction relative to the estimated values for $F_{c+}$ in normal mice (see Table II). This prompted us to ask under what parameter regimen of the model it verifies the condition that a 10-fold reduction of $p_{c+}$ and $p_{c-}$ leads to an 8- to 15-fold reduction of $F_{c+}$. To answer that question, we performed a systematic analysis for the cortex, similar to that summarized in Table III, using the same parameter ranges, but this time setting $p_{c+}$ to 10-fold lower values, $p_{c+} = 0$ and $F_{c+} = 0.007-0.025$. Strikingly, that condition is fulfilled only for $p_{c+} < 0.1$, $r_c < 30$, $z_c = 1$, and $n_c = 1-2$.

In summary, the model predicts that a single encounter with APCs may be enough for positive selection of cortical thymocytes and that, in the cortex, $<60$ different MHC/peptide ligands are involved in the selection process, whereas it is compatible with several thousands of selecting ligands in the medulla. As mentioned above, it was estimated that thymic selection is driven by several thousands of peptides (11, 25). This result is in line with ours if interpreted as reflecting medullary, but not cortical, selection.

**Estimating minimum values for the probabilities of positive signaling ($p_{c+}$ and $p_{m+}$) driven by a single peptide.** Transgenic mice expressing a single (A$^*$/Epil$^+$ mice; Ref. 34) or predominant (Ea1$^+$ mice; Refs. 40 and 41) class II MHC-peptide complex have been produced. Ea1$^+$ mice contain a full compartment of CD4 T cells, but 2-9% of the class II molecules harbored endogenous peptides instead of the transgenic peptide (42). In A$^*$/Epil$^+$ mice, in which endogenous peptides bound to MHC class II peptides are undetectable, it was estimated using different semiquantitative approaches that the diversity of the mature T cell repertoire ($F_{c+}$) varies between 10 and $50\%$ of the one in normal mice (34, 43), that is, $F_{c+}^* = 0.2-2.5\%$. Applying these values to the model Equation A5, middle, and assuming that the ligand density in the thymus of those transgenic mice is comparable with the most abundant natural ligands in wild-type mice, it can be estimated that the probabilities $p_{c+}$ and $p_{m+}$ must be, in both cases, higher than 0.002 (see Appendix D). However, note that a value for $p_{m+} \leq 0.01$ implies that $p_{c+}$ is $>0.2$, that is, $>20\%$ of cortical thymocytes must be signaled positively by any random ligand. This is rather unlikely, because the full set of cortical ligands do not signal $>20\%$ cortical thymocytes. Hence, one can safely assume a minimum value for each probability of 0.01, that is, the model predicts that at least 1% of cortical and medullary thymocytes are positively signaled after interacting with any given MHC-peptide expressed at high density.

However, these probabilities of positive signaling may be underestimations of the corresponding values in wild-type mice if one considers the possibility that low abundant peptides, themselves unable to select positively, may contribute to the positive signal delivered by abundant peptides, in a manner similar to that of costimulatory molecules (44, 45).

**Discussion**

**Heterogeneity of APC and diversity of the peptide repertoire**

Recently, Rudensky and colleagues (46) have shown that the intracellular proteolytic environment of cTECs and of hemopoietic-derived APCs, is quite different. This indicates that different sets of
peptides and/or different concentrations of peptides are likely to be expressed by each of these cell types. The possibility that the two repertoires have in fact a relative low degree of overlap is reinforced by the recent observation that a high degree of medullary deletion occurs in catenin L^+ mice (19). This is further supported by the drastic differences observed in I-A^b mice between cTECs and hematopoietic-derived APCs, which show inverse levels of expression of major peptides like class II-associated invariant chain peptide (high level in thymic epithelial cells) and Eαβ2-8s (high level in bone marrow (BM)-derived APCs) (18).

If all APCs within a given compartment express the same set of peptides and with similar densities, this implies that thymocytes interact (although perhaps repeatedly) with only one type of cortical APC and only one type of medullary APC, i.e., r_c, r_m = 1. In this case, Equations A4 can be approximated by the following:

\[
\log(1 - F_c) \approx -p_c \times n_c \quad \text{and} \quad \log(1 - F_m) \approx -p_m \times n_m.
\]

That is, in this scenario, the model predicts a linear correlation between \(\log(1 - F_c)\) and \(n_c\) and between \(\log(1 - F_m)\) and \(n_m\), with \(p_c\) and \(p_m\) being the respective proportionality constants. Following this approach, \(p_c\) and \(p_m\) could be estimated from the slopes of the above linear relationships. The generation of mouse lines expressing 2–10 different peptides on the membrane of all APCs, will allow the testing of this prediction by measurement of the frequencies of negatively selected thymocytes in the cortex (\(F_c\)) and the medulla (\(F_m\)). This must be done in conjunction with quantitative analyses of the membrane densities of total MHC and distinct MHC-peptides molecules in the various cell types involved in selection. Also, the possible interference of the other classes of MHC molecules in the selection process should be excluded by using the correspondent knockout animals. To our knowledge, these two aspects have not been addressed in depth in mice expressing single class II MHC-peptide molecules.

Survival of medullary thymocytes

Analysis of data derived from K14 mice within the framework of the present model indicates that death by neglect in the medulla is negligible, and that most medullary thymocytes either survive or are negatively selected. Class I^+II^+ mice do not have thymocytes and do not develop a thymic medulla (9). However, K14 thymic APCs are normal in the expression of class I molecules. Thus, if medullary thymocytes require for survival only weak TCR-MHC interactions, they can be signaled by either class I or class II molecules, irrespectively of whether they are CD4 or CD8 SP cells. This ancillary hypothesis, supported by the present model, can be tested in experiments in which chimeric mice are generated by reconstituting K14 mice with BM-derived cells from class I^+II^+ mice or class I^+II^+ mice. In the first case, there should be no thymocyte survival in the medulla, whereas in the second case the K14 phenotype should be recovered.

Testing the affinity hypothesis of thymocyte selection

As shown above, the affinity thresholds determine all the critical probabilities for each stage of maturation. Interestingly, when all APCs in the cortex and the medulla express a single MHC-peptide, the thymocyte fractions \(F_{c+}, F_{c-}, F_{m+}, F_{m-}\) reduce to the probabilities \(p_{c+}, p_{c-}, p_{m+}, p_{m-}\), respectively (see Equations A3 and A4). Consequently, these basic probabilities can be estimated experimentally, for example by comparing mice expressing a single MHC-peptide in all thymic cells to mice expressing it under the K14 promoter, that is, only in cTECs. Comparable results should be obtained with many different peptides.

Potential implications for thymic selection during development

During ontogeny, there is a dramatic increase of the thymus (9). This corresponds in the model to an increase of the number of different APCs (\(r_c\) and/or \(r_m\)), suggesting some interesting testable possibilities. For instance, if the fractions of selected thymocytes in the newborn and the adult thymus are comparable, then the model predicts that the number of Ags per APC (\(n_c\) and \(n_m\)) must decrease with time concomitantly with the increase of \(r_c\) and \(r_m\). Alternatively, it might be the case that, during the ontogeny of thymic medulla, the asymptotic increase of \(r_m\) is unparalleled by the other parameters, which remain virtually unchanged. In this scenario, the model predicts that, except for large values of \(z_m\), there would be an initial increase of \(F_m\) until it reaches a large maximum, followed by a monotonous decrease as the size of the medulla (\(r_m\)) increases further, whereas \(F_c\) will increase steadily until reaching adult values. These fractions will follow a behavior similar to that of the corresponding thymocyte fractions in Fig. 2A, top right panel, but with \(r_m\) in the horizontal axis, instead of \(n_m\). This last possibility is attractive, because contrary to the cortex, the medulla forms and expands as a consequence of interactions between TCR^+ thymocytes and stroma cells, a process that continues in mice at least during the first week after birth (9). This scenario would also explain why thymocyte deletion is defective in neonatal mice (47).

In summary, the present model makes a number of testable predictions, among them the estimation that negative selection in the medulla is at least 10-fold higher than in the cortex, and that the number of selecting ligands in the cortex, capable of selecting a normal TCR repertoire, is <60.

We can now extend our studies to the impact of the collective contributions of nonabundant peptides of selecting-range affinity in contributing to the signaling level of TCR/ligand (44) in a manner similar to that of costimulatory molecules. In this respect, there are additional biological constrains, some developmentally programmed (48) and others adaptively acquired (22, 23, 44), which differentially affect affinity thresholds of cortical and medullary thymocytes. A more detailed integration of all these aspects within the more general, mechanistic boundaries provided by the present model is currently in progress.

Finally, the poor deletion capacity of cortical epithelial cells may not simply reflect the low diversity of the set of peptides expressed, but also an intrinsic poor ability of these stromal cells to induce negative selection of thymocytes reacting with high avidity to self-ligands due to the expression of CD54, but not CD80/CD86, costimulatory molecules (16). Analysis of this critical issue offers the possibility to incorporate in the model the selection of natural regulatory T cells.

Appendix A

The computation of the probabilities that a thymocyte interacting with a cortical APC be negatively signaled, \(f_{c-}\), and insufficiently signaled, \(f_{cN}\), follows readily from the independence-of-events assumption, whereas the probability of being positively signaled (\(f_{c+}\)) follows from the general relationship \(f_{c+} + f_{c-} + f_{cN} = 1\). Thus, we have

\[
f_{c-} = 1 - (1 - p_{c-})^{r_c}, \quad f_{cN} = (p_{cN})^{r_c}; f_{c+} = (1 - p_{c-})^{r_c} - (p_{cN})^{r_c},
\]

(A1)

where \(p_{cN}\) is the probability that a randomly picked cortical thymocyte interacting with a given Ag be insufficiently signaled, and \(p_{c-}\) is the probability that the interaction results in negative selection.
Consider a cortical thymocyte interacting with $r_c$ APCs. Because these are also independent events, it follows that the probability that such a thymocyte not be negatively selected in the cortex is as follows:

$$f_{cN} = \sum_{i=0}^{r_c-1} \binom{r_c}{i} f_{cN}^i (1-p_{cN})^{r_c-i}$$

The second term on the right-hand side of Equation A2 is the probability that a thymocyte interacting with $r_c$ APCs experiences at least $z_c$ cell encounters, resulting in positive signaling. This can be identified, therefore, with the fraction of thymocytes positively selected in the cortex, $F_{c+}$. In terms of $p_{cN}$ and $p_{c-}$, this fraction reads as follows:

$$F_{c+} = \sum_{i=0}^{r_c} \binom{r_c}{i} p_{cN}^i (1-p_{c-})^{r_c-i}$$

Note also that the first term in Equation A2 can be interpreted as the fraction of thymocytes that die of neglect in the cortex (denoted $F_{cN}$). A similar calculation gives for the medullary compartment the corresponding probabilities $f_{mN} = f_{mN}$ as well as $F_{m+}$ and $F_{mN}$.

Let us now denote the fractions of thymocytes that are negatively selected in the cortex and medulla, respectively, $F_{c-}$ and $F_{m-}$. Then, using the general relations $F_{c+} + F_{cN} + F_{c-} = F_{m+} + F_{mN} + F_{m-} = 1$, one has the following:

$$F_{c+} = 1 - (1-p_{c-})^{r_c}$$

Finally, from the assumption that positively selected thymocytes move unidirectionally from the cortex to the medulla to periphery, a straightforward computation shows that the total expected fractions (cortex plus medulla) of thymocytes following default apoptosis ($F_{cN}$), positive selection ($F_{c+}$), and negative selection ($F_{c-}$), are related to the above equations through the following equations:

$$F_N = F_{cN} + F_{c+} \times F_{mN}, F_+ = F_{c+} \times F_{m+}, F_- = F_{c-} + F_{c+} \times F_{m-}$$

Equation A5, right, one has the following: $0.1 > F_{c+} \times F_{m-} > F_{c+} \times 0.5$, which implies the following:

$$F_{c+} < 0.2$$

From the result in Appendix B, one has $F_{mN}, F_{m+} < 0.5$, which together with Equation A5, middle, yields the following: $0.05 \approx F_{c+} = F_{c+} \times F_{m+} < F_{c+} \times 0.5$, and therefore, $0.1 < F_{c+}$

From Equations A5 and C1, one has the following: $0.05 \approx F_{c+} = F_{c+} \times F_{m+} < 0.2 \times F_{m+}$, and therefore, $0.25 < F_{m+}$

This result, together with the result in Appendix B, yields the following: $0.5 > F_{mN} + F_{m+} > F_{mN} + 0.25$, which implies the following: $F_{mN} < 0.25$

Using the result in Appendix B and Equation C2 into Equation A5, right, one has the following: $0.1 \geq F_{c+} = F_{c+} \times F_{m+} > 0.1 \times F_{m+}$, and therefore, $F_{c+} \leq 0.05$

This, together with Equations C1 and C2 and the fact that $F_{cN} + F_{c+} + F_{m-} = 1$, implies that $0.75 < F_{cN} < 0.9$

### Appendix B

In radiation BM chimeras reconstituted with cells from genetically engineered mice not expressing class II MHC molecules on BM-derived cells (29), it has been estimated that the total fraction of positively selected thymocytes is $F_{c+} \approx 0.10–0.15$ (that is, two to three times its value in normal mice ($F_{c+}$)). Considering that medullary thymic epithelial cells induce anergy rather than negative selection (17), one can assume that these chimeric mice $F_{m-} = 0$. If, in addition, it is assumed that $F_{m+} = F_{m+}/(F_{m+} + F_{mN})$ (where $F_{m+}$ and $F_{mN}$ are frequencies in normal mice), and that the corresponding thymocyte frequencies in the cortex are unaffected (i.e., $F_{c+} = F_{c+}$ and $F_{cN} = F_{cN}$), using Equation A5, one has $F_{c+} = F_{c+} F_{m+} = F_{c+} F_{m+}/(F_{m+} + F_{mN}) = F_{c+}/(1 - F_{cN})$. From this relationship and the experimental estimates of $F_{c+}$ and $F_{cN}$, it follows directly that $0.5 < F_{mN} < 0.7$

### Appendix C

The derivation of estimates for the different fractions in the cortex and medulla follows from a straightforward algebraic manipulation of inequalities. Thus, from the result in Appendix B and
antagonize an MHC-bound peptide, the antagonist should occupy a very large fraction of MHC molecules. At best, the few most abundant natural peptides could be natural antagonists for few of those natural peptides represented at <10 copies. Consequently, their repertoire is likely to be very limited. In contrast, the tremendous excess of an antagonist can be reached in viral immune responses because not only does inflammation induce a drastic increase of MHC molecules but also one viral peptide can occupy up to 10% of all MHC molecules.

In addition, there is evidence that, at best, antagonists behave as weak agonists in thymic selection. As suggested elsewhere (23), it looks like in developing thymocytes the intracellular machinery triggered by TCR signaling is still immature and hence not prone to translate antagonistic events. An outstanding example is the E1 peptide case, an antagonist that induces positive and negative selection of the respective TCR-bearing thymocytes in TAP-1−/− and TAP-1+ fetal thymic organ cultures (50). This behavior is typical of a weak agonist (23).

Appendix F
The model of Detours et al. (25) is within the same conceptual framework as ours. In particular, they assume the following: 1) The model of Detours et al. (25) is within the same conceptual framework as ours. In particular, they assume the following: 1)

Acknowledgments
We thank J. Cameiro, A. Coutinho, M. A. R. Marcos, and K. León for discussions and useful suggestions, and Z. Grossman, W. Haas, and J. Borghans for critical reading of the manuscript.

References