Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 5153–5158, April 1999
Immunology, Applied Mathematics

Explaining high alloreactivity as a quantitative consequence of affinity-driven thymocyte selection

VINCENT DETOURS*†‡ AND ALAN S. PERELSON*‡§

*Theoretical Biology and Biophysics, MS K710, and †Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos NM 87545; and ‡Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501

Edited by Robert May, University of Oxford, Oxford, United Kingdom, and approved February 18, 1999 (received for review July 22, 1998)

ABSTRACT **Interactions between $\alpha\beta$ T cell receptors and peptides bound to molecules encoded by the MHC genes underly T cell activation. More than 1% of T cells are activated by foreign (allogenic) MHC molecules, a phenomenon called alloreactivity. Reconciling the high frequency of alloreactivity with the fact that only 1 T cell in 10^4–10^6 responds to a given foreign antigen presented on self MHC has been a long-standing puzzle. We show, by using a quantitative model, that this difference follows from the affinity model of T cell selection. Further, we demonstrate that highly alloreactive pre- and post-selection repertoires can be obtained without assuming germline bias of T cell receptors toward recognition of allele-specific MHC residues. It has been proposed that alloreactivity occurs because self and foreign MHCs bind different subsets of self peptides or alter their conformation differently. We find that such effects decrease rather than increase alloreactivity. Overall, our results show that the affinity model of T cell selection can quantitatively explain both self MHC restriction and high alloreactivity.**

Maturation of T cells in the thymus involves a two-step selection process driven by the affinity of their T cell receptors (TCR) for self peptides presented on proteins encoded by MHC genes. The first step, positive selection (1, 2), disgards thymocytes bearing TCRs with low affinity for MHC–peptide complexes expressed in the thymus. This eliminates T cells that cannot recognize MHC molecules. The second step, negative selection (3, 4), deletes cells with high affinity receptors for thymic MHC–peptide complexes. Thus, removing many self reactive cells. Overall, only 3% of the T cells produced in the thymus have TCRs with the intermediate affinity required to reach the periphery (5).

Because MHC genes are extremely polymorphic, two individuals are very unlikely to express the same set of MHC molecules. As a result of positive selection, T cells are self MHC restricted: they recognize pathogens presented by self MHC molecules but ignore them if presented by foreign MHC molecules (6–14). MHC polymorphism is also the main obstacle to tissue transplantation (15). Typically, 1–24% of T cells are alloreactive (16, 17), i.e., they respond to foreign (allogenic) MHC molecules. Reconciling this high alloresponse frequency with the fact that among naive T cells only 1 in 10^4–10^6 recognizes a given pathogen (18, 19) is a long-standing immunological puzzle. In this paper, we examine quantitatively three hypotheses proposed to explain the high frequency of alloreactivity.

The first hypothesis, due to Matzinger and Bevan (20), suggests that the 2–4 orders of magnitude difference between antigen and MHC response frequencies results from the difference in the diversity of these two types of molecule on the surface of antigen presenting cells (APCs). Because each individual expresses only a few distinct MHC molecules and each MHC molecule associates with a diverse array of peptides, the number of distinct complexes made from a given MHC will greatly exceed the number of complexes made from a given peptide. Consequently, a given MHC will be recognized with high frequency and a given MHC–peptide complex with lower frequency.

The second hypothesis suggests that alloreactivity reflects differences in the presentation of self peptides by self and foreign MHCs rather than differences in the parts of MHC molecules directly accessible by TCRs (21). The foreign APCs used to measure alloreactivity belong to the same species as their self counterparts. Thus, the self and foreign cells should synthesize and process essentially the same proteins. However, each MHC allele encodes a peptide binding motif determining which peptides associate with the MHC molecule (22, 23) and the conformation of the bound peptides (23–25). Thus, self peptides may be perceived as foreign by T cells when presented in the groove of foreign MHC molecules.

The third hypothesis, originally considered by Jerne (26), suggests that alloreactivity resides in our genes (27). It is supported by the finding that the alloreactivity of the preselection repertoire is as high as that of the mature repertoire (27–29).

We present a model of affinity-driven selection of the T cell repertoire and use it to derive expected levels of alloreactivity and self MHC restriction, and to assess the quantitative implications of the three different hypotheses.

MODEL

Minimal Model of Interaction Between TCRs and MHC–Peptide Complexes. The concept of shape space (30) provides a convenient framework with which to represent TCRs and their ligands. As in previous models (reviewed in ref. 31), we represent the "generalized shape" of a protein as a string of digits. The strength of binding of two proteins is then defined as the degree of complementarity between their generalized shapes (Fig. 1).

The affinity between an MHC–peptide complex and a TCR is computed by aligning the strings representing the MHC–peptide complex and the TCR, and then summing all the pairwise digit interactions. As shown in Fig. 1, the central digits of a TCR always contact a peptide, and the extremities MHC. This modeling choice follows from studies according to which TCRs bind MHC–peptide complexes with a common orientation (32–42).

Generation of MHCs, Peptides, and TCRs. Only the interacting portions of TCRs and MHC–peptide complexes are taken into account in the model, not the full structure of these molecules. The set of self MHC molecules consists of n_m random strings, each of l_m digits, representing the polymorphic residues of MHC molecules exposed to TCRs. Essentially all progress in the identification and characterization of self-peptides in alloreactivity has involved MHC class I systems

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office.

§To whom reprint requests should be addressed. e-mail: asp@lanl.gov.

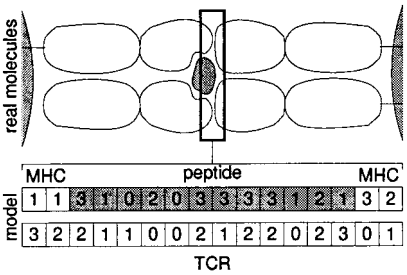

FIG. 1. Digit-string representation of MHC–peptide and TCR interaction. Only the interface between TCRs and MHC–peptide complexes (framed region in upper diagram) are taken into account in the model. MHC–peptide complexes are constructed by inserting a peptide string in an MHC string. TCRs are sequences of digits chosen randomly from $\{0, 1, 2, \ldots, d_{max}\}$ ($d_{max} = 3$ in the figure and 255 in the calculations). The interaction strength, $I(i, j)$, between digits i and j is a measure of their complementarity (66). Affinity, K, is the sum of interaction strengths of contacting digits in aligned strings. Formal definitions of I and K are given in the *Appendix*.

(43). Thus, we focus on MHC class I, although class II can be analyzed in an analogous way. There are three class I loci in mice (44). Alloreactivity and self restriction experiments use inbred mouse strains (45), and thus only one allele is present at each locus. Therefore, we set $n_m = 3$. (See ref. 46 for an estimate of optimal MHC polygenicity.)

Each MHC string "presents" (Fig. 1) n_p peptides strings of l_p random digits. About 10^3–10^4 different peptides can be eluted from molecules of a given MHC allele (47–50). Thus, unless specified otherwise, we set $n_p = 10^4$.

To explore binding motifs, we either assume that the conformations of a peptide induced by the grooves of two MHC molecules from different alleles are so different that the peptide appears to TCRs as two totally unrelated peptides, or that motifs do not influence peptide presentation at all. These two extreme hypotheses are implemented by either (*i*) associating different random peptide strings with each MHC or (*ii*) allowing any peptide to associate with all MHCs. Case *i* also applies if the sets of peptides presented by two MHCs are nonoverlapping. Both alternatives imply a self environment composed of $n_m \times n_p$ MHC–peptide complexes. However, we generate $n_m \times n_p$ distinct self peptides strings in case *i* and only n_p in case *ii*. The later alternative may be unrealistic because motifs appear to have an impact (23). Investigating it is nevertheless necessary to quantify the effect of motifs on alloreactivity.

The number of MHC polymorphic residues, l_p, and peptide residues in contact with TCRs, l_m, are set from crystallographic data. It is assumed that these parameters are the same for all class I loci. The structure of TCR/MHC–peptide complex A6/HLA-A2–Tax (36) reveals 7 peptide and 5 MHC polymorphic residues in contact with TCR A6, which gives $l_p = 7$ and $l_m = 5$. Performing a similar measurement for B7/HLA-A2–Tax (41), 2C/H2-K^b–dEV8 (42), and 2C/H2-L^b–QL9 (51) gives an average of 5.75 peptide and 3.5 MHC residues in contact with the TCR. Since l_m and l_p must be integers, we set $l_m = 4$ and $l_p = 6$. A6, B7, and 2C are all known to be positively selected when expressed in the relevant MHC background. Consequently, the above estimate might not reflect a property of the preselection repertoire. Counting solvent-accessible peptide and MHC polymorphic residues in a class I MHC–peptide crystal structure leads to $l_m = 12$ and $l_p = 5$ (52). This approach is independent of any selection-induced bias, but it has its own caveat because only part of the solvent-accessible surface of the MHC–peptide complex is covered by the TCR (35, 36, 41, 42). In the absence of conclusive data, both $(l_m, l_p) = (4, 6)$ and $(l_m, l_p) = (12, 5)$ are investigated. TCRs are modeled as strings of $l = l_m + l_p$ random digits.

Selection Thresholds and Stringency of Selection. Selection is implemented by introducing two affinity thresholds, K_P and K_N ($K_P < K_N$). Clones binding at least one self MHC–peptide complex with affinity $K \geq K_P$ survive positive selection. Negative selection deletes clones binding one or more self MHC–peptide complexes with $K > K_N$. The values of K_P and K_N are inferred from experimental data by considering the fractions of clones surviving the different stages of selection (see Fig. 2). The fraction of clones reaching the periphery is $f = f_P f_N$, where f_P is the fraction of clones surviving positive selection and f_N is the fraction of positively selected clones that survive negative selection.

About two-thirds of positively selected thymocytes are deleted by negative selection (28, 29, 53–56). Interestingly, probabilistic models of clonal deletion based on the hypothesis that evolution optimizes the size of the repertoire predicted $f_N = 0.37$ (57–59). This estimate will be used here.

Three percent of T cells produced in the thymus reach the periphery (5). However, the fraction of clones, which our model deals with, and the fraction of cells differ because a significant portion of mature T cells divide before emigrating to the periphery (60–62). Scollay *et al.* (61) suggest that one division occurs before emigration to the periphery. Division also occurs earlier in clonal development, with the fraction of CD4$^+$CD8$^+$TCR$^+$ cells that proliferate estimated as being 1.5- to 2-fold larger than the fraction of dividing mature thymocytes (60, 63–65). Overall these data suggest that TCR$^+$ cells go through 2–3 divisions in the thymus. In the absence of more precise information, we assume that two divisions occur on average and hence each clone consists, on average, of four cells. If 3% of thymocytes survive selection, the fraction of clones reaching the periphery is $f = \frac{1}{4} \times 3 = 0.75\%$.

Defining activation of selected T cells is a prerequisite for studying the peripheral repertoire. A clone is considered activated if the affinity of its TCR for a MHC–peptide complex is greater than K_N. The repertoire is self tolerant by construction since no clones that have an affinity larger than K_N for a self MHC–peptide survive negative selection.

The number of digits in the alphabet, d_{max} (see Fig. 1), has no effect on the model's behavior as long as d_{max} is chosen large enough. If d_{max} is too small, only a reduced number of affinity values are generated by the model (66), and it is not possible to find selection thresholds compatible with physiological values of f, f_P, and f_N. Increasing d_{max} from 255 to 1,023 changes the model's outputs (defined below) by at most 3%.

Analyzing the Model. The model can be analyzed by using computer simulations or a mathematical approach. Simulations proceed in three steps. First, a set of self MHC–peptide

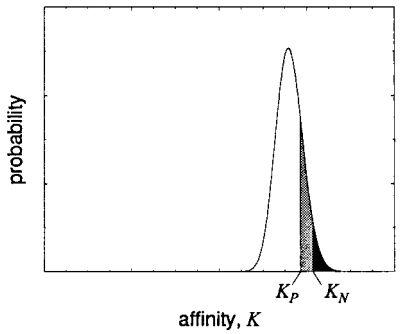

FIG. 2. Setting selection thresholds. Diagram is not drawn to scale to keep it readable. Distribution of the maximum affinity between a TCR of the preselection repertoire and $n_m \times n_p$ random MHC–peptide complexes is plotted (see *Appendix* for mathematical derivation). The selection thresholds K_P and K_N are set such that the fraction of TCRs with maximal affinity greater that K_P is f_P (gray and black areas) and the fraction of TCRs with maximal affinity between K_P and K_N is f (gray area). The fraction of the *preselection* repertoire deleted by negative selection is shaded in black and is equal to $f_P(1 - f_N)$.

complexes is constructed. Then random TCRs are generated, and those satisfying the affinity selection criteria are kept in the repertoire. Finally, sets of foreign MHC and foreign peptides are generated and alloreactivity, self MHC restriction, and the foreign peptide response frequency of the selected repertoire measured. A simulation is the computational equivalent of a set of measurements made on a particular animal. There are 10^7–10^8 T cell clones in a mouse, and thus at least $10^7/f \sim 10^9$ TCRs must be generated, and submitted to selection to simulate the repertoire of one animal. Selection of one TCR requires the calculation of its affinity with each MHC–peptide complex. Since there are 3 MHC loci and 10^4 self peptides, 3×10^{13} affinities need to be evaluated for the generation of one animal's repertoire, making repeated simulations untractable.

Alternatively, mathematical expressions for the average alloreactivity, self MHC restriction, and peptide response frequency can be derived (see *Appendix*). Since such calculations do not rely on the actual selection of a repertoire, they are easily carried out. The results they provide correspond to averages over all simulation outcomes possible for a given parameter set, but they give no information about the variability between different TCR repertoires. For example, in the case of foreign peptide response frequency, simulation outcome depends on the self MHC–peptide complexes generated, on the TCRs submitted to selection, and on the foreign peptide used to challenge the resulting repertoire. The expressions for this quantity give averages over all possible combinations of self MHC–peptide complexes, preselection TCRs, and foreign peptides. Since we are interested in average properties of the repertoire, a mathematical approach will be used here.

RESULTS

Affinity-Driven Selection Can Produce a Self MHC Restricted Repertoire. Self MHC restriction has been estimated by comparing the effector activity against foreign peptides presented on self MHC and foreign MHC (6, 8–12). Effector functions are not represented in our model, but we assume that their intensity is proportional to the number of responding clones.

Assuming that MHC binding changes peptide conformation (binding motif case *i*), the *response frequency*, R, to a given foreign peptide is defined as the fraction of clones activated by this peptide when presented in combination with one of the n_m self MHC molecules. The mathematical model in the *Appendix* gives $R = 1.3 \times 10^{-5}$ if the contribution to the interaction with TCRs of peptides and MHC polymorphic residues are, respectively, $l_m = 4$ and $l_p = 6$, whereas $R = 1.1 \times 10^{-5}$ if $(l_m, l_p) = (12, 5)$. Both estimates of R are consistent with the experimental range 10^{-6}–10^{-4} (18, 19).

The model predicts that the response frequency to a foreign peptide presented on allogenic MHC molecules, $R_a = 9.4 \times 10^{-7}$ if $(l_m, l_p) = (4, 6)$. Thus, the *restriction ratio* $r = R/R_a = 14$, i.e., 14 times as many clones are activated by foreign peptides if presented on self MHCs as compared to foreign MHCs. Assuming a uniform clone size distribution, the measurements of Stockinger *et al.* (18) give a restriction ratio, r, of 6–10. By contrast, if $(l_m, l_p) = (12, 5)$, $R_a = 1.7 \times 10^{-5}$, and the repertoire is better at recognizing peptides presented by *foreign* MHC molecules than on self MHC. Since MHC restriction is well established, we conclude that $(l_m, l_p) = (12, 5)$ is an unrealistic parameter choice. Thus, according to our model, peptides must contribute to a substantially greater fraction of the interaction with TCRs than MHC polymorphic residues to account for self restriction.

Absolute restriction would be observed for pathogens whose peptides cannot be presented by foreign MHC. However, the possibility that the repertoire appears absolutely restricted to foreign MHC because of failure of self MHCs to present peptides of the pathogen is equiprobable. These effects would cancel each other out when considering average restriction over many experimental systems. Consequently, Ir-gene defects need *not* to be taken into account in our calculation of self restriction.

Assuming that binding motifs have no effect gives response frequencies, R and R_a, n_m times larger because a foreign peptide can be presented by all n_m MHCs in any given haplotype. However, the resulting restriction ratio is not affected.

Affinity-Driven Selection Accounts for High Postselection Alloreactivity and Implies That Peptide Binding Motifs Decrease It. Alloreactivity is the fraction of the repertoire responding to foreign MHC molecules presenting peptides that we assume are in the set of self peptides. Peptide binding motifs determine which self peptides associate with particular MHC molecules, and in what conformation. To assess the quantitative impact of this effect, we compare alloreactivity computed assuming a maximal effect of binding motifs (case *i*), with its value computed assuming no effect (case *ii*). If motifs cause alloreactivity, then there should be a higher alloreactivity level under the first hypothesis.

According to our model, binding motifs *decrease* alloreactivity. The alloreactivity, a, is equal to 2% when motifs have no effect [1.4% if $(l_m, l_p) = (12, 5)$], and 1.3% when their effect is maximal [irrespectively of (l_m, l_p)]. This somewhat counterintuitive result can be explained as follows. The affinity between a selected TCR and self MHC–peptide is larger than the average affinity between TCRs and random MHC–peptide complexes because of positive selection (not shown). So, any random change in self MHC–peptide complexes will, in general, decrease the affinity toward its average value. It does not matter whether the change in MHC–peptide complex occurs at the level of peptide or MHC residues, because this distinction is absent when considering the overall TCR/MHC–peptide binding affinity. This analysis is independent of whether binding motifs control the peptide sequences associating with MHC, peptides conformations, or both.

Overall, the model shows that affinity-driven selection account for alloreactivity levels of 1.3–2%, but is not compatible with the notion that binding motifs are the cause of these high levels.

The Affinity Model and Data on the Stringency of Selection Imply High Preselection Alloreactivity. Alloreactivities of the mature and preselection repertoires are very similar (27–29). Is this compatible with the affinity-driven selection hypothesis? Since TCRs are produced at random in our model, self and foreign MHC complexes are equivalent, and both appear as sets of random strings from the point of view of the preselection repertoire. This is also true of self and foreign peptides. Thus, we define preselection alloreactivity as the fraction of TCRs in the preselection repertoire with affinity greater than K_N for at least one of the $n_m \times n_p$ random MHC–peptide complexes. As shown Fig. 2, this quantity equals¶ $f_P(1 - f_N)$. Using $f_N = 37\%$ and $f = 0.75\%$, the values deduced earlier, we conclude that the alloreactivity of the preselection repertoire should equal 1.3%. Experimental estimates of preselection alloreactivity are $5.7 \pm 2\%$ (29) and $2.7 \pm 2.8\%$ (28). The latter estimate is compatible with our calculations. Corresponding postselection alloreactivity estimates are $5.4 \pm 2.8\%$ (29) and $3 \pm 2.3\%$ (28). Thus, in both cases pre- and postselection alloreactivities are similar. The same is true in our model with the preselection alloreactivity, 1.3%,

¶Surprisingly, the above formula is independent of model parameters such as n_m, n_p, l_m, and l_p, which control MHC and peptide length and diversity. This by no means implies that those parameters have no influence on preselection alloreactivity *in vivo*. Rather, it suggests that they make their influence felt by changing the stringencies of positive and negative selection.

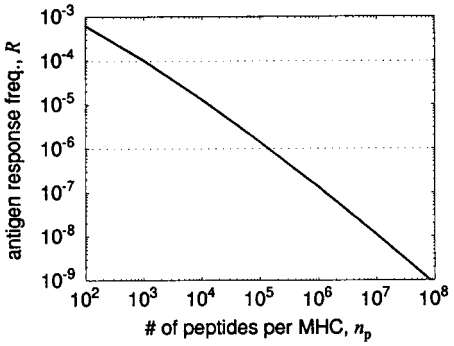

FIG. 3. Effect of self peptide diversity, n_p, on foreign peptide response frequency, R. Selection thresholds have been adjusted for each value of n_p to keep f, f_P, and f_N at their physiological levels.

and postselection alloreactivity of 1.3% or 2%, depending on the effect of binding motifs.

High pre- and postselection alloreactivities were obtained assuming random TCRs. Thus, the hypothesis put forward by Jerne (26) that alloreactivity is the consequence of a genetic bias of TCRs toward allele-specific MHC residues is not necessary in the context of affinity-driven selection. No conclusion can be drawn about conserved MHC residues bias, because those are not represented in the model.

Self Peptide Diversity Has a Very Small Impact on Alloreactivity, but Displays a Strong Inverse Correlation with Antigen Response Frequency. One possible explanation of alloreactivity is that many more MHC–peptide complexes are made from a given MHC allele product than from a given peptide sequence (20). If this is the case, then one would expect that increasing the number of different self peptides per MHC, n_p, would increase alloreactivity, a (see Fig. 3).

We find that if, as suggested above, the effect of binding motifs is maximal, then a would be equal to the preselection alloreactivity, i.e., 1.3%, and be independent of n_p. If binding motifs have no effect, then increasing n_p *decreases* alloreactivity. It is 5% when the number of self peptides is 100 and falls to 1.3% when n_p is 10^8. Since our goal is to derive the consequences of affinity-driven selection under physiological conditions, selection thresholds were adjusted for each n_p value to keep f, f_P, and f_N at their physiological levels (see *Model*). Both 5% and 1.3% are in agreement with experimentally determined ranges of a. Thus, low as well as high self peptide diversity is consistent with high alloreactivity frequency.

The puzzle of alloreactivity does not only lie in its high frequency but in the fact that it is 2–4 orders of magnitude larger than the antigen response frequency. Examining the foreign peptide response frequency, R, for different peptide diversities (Fig. 3), we find that R decreases almost linearly as n_p is increased. When $n_p = 100$, the response frequency of the postselection repertoire, R, is 6.1×10^{-4}, whereas it is 8.1×10^{-10} when $n_p = 10^8$. By contrast, the alloreactivity, a, decreases at most by a factor 5 over the same interval in n_p. Thus, peptide diversity has a major influence on the difference between MHC and peptide response frequency. Interestingly, values of R in the experimental range, 10^{-6}–10^{-4}, can be obtained only if n_p lies between 10^3 and 10^5 (Fig. 3). Hence, our model agrees with the notion presented by Bevan (67) that selection is driven by 10^3–10^4 different self peptides.

DISCUSSION

Previous attempts to explain alloreactivity have relied on nonmathematical arguments, and thus could not rigorously address its fundamental quantitative nature. The model presented here gives *quantitative* estimates of self restriction and alloreactivity. The mathematical procedure gives average results over a very large number of antigenic systems and self/foreign haplotypes pairs, whereas experimental studies have been confined to a small number of systems.

Experiments based on the comparison between allogenic and syngenic immune responses demonstrated strong restriction in some instances (6, 8, 10–12) but weak or absent restriction in others (8, 68–72). Thus, it is difficult to draw any conclusion on the *average* level of self restriction from experimental data. Our model shows that the repertoire could recognize peptide presented on self MHCs 14 times more frequently than peptide presented on foreign MHC. Since this prediction concerns average behavior, it is compatible with absolute restriction or no restriction, for particular self/foreign haplotype combinations. Measures of restriction based on precursor frequencies relate directly to our model in which only the fraction of responding clones is measurable. Using limiting dilution analysis, Stockinger *et al.* (18) estimated the self restriction ratio to be 6–10 (see also ref. 13), a value comparable to our estimate.

It has been proposed that alloreactivity occurs because self and foreign MHC molecules present different subsets of self peptides, or present the same self peptides in different conformations (21). According to the affinity model, self restriction is possible only if positive selection improves the interaction between TCRs and self MHC–peptide. At higher affinities interaction with both self peptides and self MHC is enhanced. Thus, TCRs in the selected repertoire have, on average, stronger interaction with self than with nonself peptides. Any alteration of self peptides induced by foreign MHC will therefore lower the average affinity rather than increase it. Accordingly, our model predicts that the average alloreactivity of the selected repertoire over many experimental systems should be 2% in the absence of binding motifs, and 1.3% if their effect is maximal. Hence, peptide binding motifs decrease alloreactivity. Both 1.3% and 2% are within the experimental range of 1–24% (16, 17).

Our calculations indicate that alloreactivity and self peptide diversity are inversely related. However, the negative impact of high diversity is small: alloreactivity in the range 1–24% could result from a repertoire of 100 as well as from a repertoire of 10^8 self peptides. By contrast, we found a much stronger inverse correlation between self peptide diversity and antigen response frequency. Response frequencies in the range 10^{-6}–10^{-4} (18, 19) only occur in our model if thymic selection is driven by 10^3–10^5 self peptides. These results show that the argument of Matzinger and Bevan (20) is quantitatively sound. A small number of distinct MHC molecules associate with a diverse array of peptides. Thus the number of distinct complexes made from a given MHC greatly exceeds the number of complexes made from a given peptide, hence the larger response frequency in the first case.

We found that the affinity model implies a preselection alloreactivity of 1.3%, compatible with some experimental measurements (28). These later data have been interpreted as evidence for a germline bias of TCRs toward MHC recognition (27–29). To explain alloreactivity, Jerne (26) postulated that each clone is specific for one of the many MHC alleles present in the species. Our analysis shows that this postulate is unnecessary in the context of affinity-driven selection. The estimate of 1.3% has been obtained by assuming that TCR residues in contact with allele-specific portions of MHC molecules are totally random, thus precluding a germline bias. The issue of bias toward conserved MHC residues cannot be addressed with the current version of the model.

Overall, our results show that affinity-driven selection of thymocytes is in quantitative agreement with experimental estimates of foreign antigen response frequency, self restriction, and alloreactivity.

APPENDIX

The mathematical expressions used to analyze the model are briefly presented here.‖ They give results in agreement with simulations of the model (not shown).

Preliminaries. Let X and Y be two discrete independent random variables with probability distributions $p_X(.)$ and $p_Y(.)$, respectively. The distribution of $X + Y$ is $p_X \star p_Y$ (see ref. 73, p. 179), where $\star$ denotes the convolution operator. The convolution of p_X by itself l times is written p_X^l. We define $M_{X,n}(.)$ to be the maximum of n independent random variables with identical distribution p_X ($M_{X,n}(.)$ is derived in ref. 73, p. 128).

Match Scores. Let $\{0, 1, 2, \ldots, d_{\max}\}$ be a set of digits. The interaction strength between digits x and y is by definition

$$I(x, y) = x \oplus y,$$

where $\oplus$ consists in applying the "exclusive or" operator on the binary representations of x and y and interpreting the result as a decimal integer. For example, $I(1, 3) = 2$. Since all digits are equiprobable in our model, the distribution of interaction strengths is

$$p_I(i) = \begin{cases} \frac{1}{d_{\max}+1} & \text{if} \quad i \in \{0, 1, 2, \ldots, d_{\max}\} \\ 0 & \text{otherwise.} \end{cases}$$

The match score, or affinity, between digit-strings $[x_1, \ldots, x_l]$ and $[y_1, \ldots, y_l]$ is defined by

$$K([x_1, \ldots, x_l], [y_1, \ldots, y_l]) = \sum_{i=1}^{l} I(x_i, y_i).$$

We denote by γ the match score between random TCRs and random MHCs. Its probability distribution is given by

$$p_\gamma = p_I^{l_m}.$$

Similarly, the distribution of θ, the match scores between random peptides and random TCRs, is $p_\theta = p_I^{l_p}$.

Selection Thresholds. The maximal match score between random TCRs and random MHC–peptide complexes, ω, governs selection (Fig. 2). Two hypothesis are explored: (*i*) different MHCs present nonoverlapping sets of peptides; (*ii*) the same peptides are presented by all MHCs. We assume that peptide and MHC digits are independent. Thus, under hypothesis *i*,

$$p_\omega = M_{\delta,n_m}$$

with $p_\delta = p_\gamma \star M_{\theta,n_p}$. Under hypothesis *ii*,

$$p_{\omega'} = M_{\gamma,n_m} \star M_{\theta,n_p}.$$

Assuming hypothesis *i*, there is a unique pair (K_P, K_N) satisfying

$$f_P = \sum_{z=K_P}^{l_m+l_p} p_\omega(z), \quad \text{and} \quad f = \sum_{z=K_P}^{K_N} p_\omega(z).$$

Thresholds under hypothesis *ii* are obtained by substituting ω' in place of ω in the above equation.

‖A software package in C language implementing these expressions and related simulations can be downloaded from `ftp://ftp-t10.lanl.gov/pub/detours/abs-lab-1.1.tar.gz`.

Distribution of Match Scores for Self MHCs and Self Peptides. Let ϕ be the *best* match score of a given *selected* TCR over all self MHCs. $p_\phi(k)$ is the probability that a TCR recognizing self MHCs with best match score k is generated *and* selected. The probability of the first event is equal to $M_{\gamma,n_m}(k)$. The second event occurs if the maximum match score over all self peptides, z, is such that $k + z$ lies within the selection window. Therefore, assuming hypothesis *ii*

$$p_\phi(k) = \frac{1}{f} M_{\gamma,n_m}(k) \sum_{z=K_P-k}^{z \le K_N-k} M_{\theta,n_p}(z).$$

The distribution p_ψ of the best match score of a given selected TCR over all self peptides, ψ, is obtained by swapping γ and θ, and n_m and n_p in the above equation. The distribution of match scores between a *selected* TCR and a self MHC under hypothesis *i* is

$$p_\eta(k) = \frac{1}{n_m}\left[p_\phi(k) + \frac{n_m - 1}{1 - f_P} p_\gamma(k) \sum_{z=0}^{z<K_P-k} M_{\theta,n_p}(z) \right].$$

This expression neglects (very unlikely) situations when more that one self MHC drives positive selection.

Alloreactivity. Alloreactivity is defined as the fraction of clones responding to a foreign MHC haplotype in combination with self peptides. Under hypothesis *i*, different self peptides are presented by self and foreign MHCs. Together with the definition of the selection thresholds, this implies that the alloreactivity, a, is given by

$$a = \sum_{z>K_N} p_\omega(z) = f_P(1 - f_N).$$

Under hypothesis *ii*, self peptides driving selection also drive alloreactivity. Hence the alloreactivity, now called a', is given by

$$a' = \sum_{z>K_N} [M_{\gamma,n_m} \star p_\psi](z).$$

Preselection alloreactivities are identical under hypotheses *i* and *ii* and equal to a.

Response Frequency to a Random Peptide. Under hypothesis *i*, the probabilities, R and R_a, that a foreign peptide triggers activation when combined with a self MHC or a foreign MHC, respectively, are

$$R = \sum_{z>K_N} [p_\eta \star p_\theta](z) \quad \text{and} \quad R_a = \sum_{z>K_N} [p_\gamma \star p_\theta](z).$$

Under hypothesis *ii*, the foreign peptide is presented by all the MHCs a a given haplotypes; therefore,

$$R' = \sum_{z>K_N} [p_\phi \star p_\theta](z) \quad \text{and} \quad R'_a = \sum_{z>K_N} [M_{\gamma,n_m} \star p_\theta](z).$$

We thank Ramit Mehr and Bernhard Sulzer for useful discussions, Catherine Macken for help with the notation, and Ronald Schwartz for helpful comments. Portions of this work were performed under the auspices of the U.S. Department of Energy. This work was supported by National Institutes of Health Grants RR06555 and AI28433 (to A.S.P.) and Defense Advanced Research Planning Agency Grant ONR N00014-95-1-0975.

1. Fowlkes, B. J. & Schweighoffer, E. (1995) *Curr. Opin. Immunol.* **7**, 188–195.
2. Jameson, S. C., Hogquist, K. A. & Bevan, M. J. (1995) *Annu. Rev. Immunol.* **13**, 93–126.
3. Nossal, G. J. (1994) *Cell* **76**, 229–239.
4. Sprent, J. & Webb, S. R. (1995) *Curr. Opin. Immunol.* **7**, 196–205.
5. Shortman, K., Egerton, M., Spangrude, G. J. & Scollay, R. (1990) *Semin. Immunol.* **2**, 3–12.

6. Zinkernagel, R. M. & Doherty, P. C. (1974) *Nature (London)* **248**, 701–702.
7. Bevan, M. J. (1977) *Nature (London)* **269**, 417–418.
8. Bevan, M. J. & Fink, P. J. (1978) *Immunol. Rev.* **42**, 3–19.
9. Miller, J. F. (1978) *Immunol. Rev.* **42**, 76–107.
10. Sprent, J. (1978) *Immunol. Rev.* **42**, 108–137.
11. Waldmann, H. (1978) *Immunol. Rev.* **42**, 202–223.
12. Zinkernagel, R. M. (1978) *Immunol. Rev.* **42**, 224–270.
13. Schwartz, R. H. (1984) in *Fundamental Immunology*, ed. Paul, W. E. (Raven, New York), 1st Ed., pp. 379–438.
14. von Boehmer, H. (1990) *Annu. Rev. Immunol.* **8**, 531–556.
15. Auchincloss, H., Sykes, M. & Sachs, D. H. (1998) *Fundamental Immunology*, ed. Paul, W. E. (Lippincott–Raven, New York), 3rd Ed., pp. 1175–1235.
16. Bevan, M. J., Langman, R. E. & Cohn, M. (1976) *Eur. J. Immunol.* **6**, 150–156.
17. Ashwell, J. D., Chen, C. & Schwartz, R. H. (1986) *J. Immunol.* **136**, 389–395.
18. Stockinger, H., Pfizenmaier, K., Hardt, C., Rodt, H., Rollinghoff, M. & Wagner, H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7390–7394.
19. Zinkernagel, R. M. (1996) *Science* **271**, 173–178.
20. Matzinger, P. & Bevan, M. J. (1977) *Cell. Immunol.* **29**, 1–5.
21. Sherman, L. A. & Chattopadhyay, S. (1993) *Annu. Rev. Immunol.* **11**, 385–402.
22. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. & Rammensee, H. G. (1991) *Nature (London)* **351**, 290–296.
23. Madden, D. R. (1995) *Annu. Rev. Immunol.* **13**, 587–622.
24. Chen, W., McCluskey, J., Rodda, S. & Carbone, F. R. (1993) *J. Exp. Med.* **177**, 869–873.
25. Chattopadhyay, S., Theobald, M., Biggs, J. & Sherman, L. A. (1994) *J. Exp. Med.* **179**, 213–219.
26. Jerne, N. K. (1971) *Eur. J. Immunol.* **1**, 1–9.
27. Jameson, S. C. & Bevan, M. J. (1998) *Curr. Opin. Immunol.* **10**, 214–219.
28. Merkenschlager, M., Graf, D., Lovatt, M., Bommhardt, U., Zamoyska, R. & Fisher, A. G. (1997) *J. Exp. Med.* **186**, 1149–1158.
29. Zerrahn, J., Held, W. & Raulet, D. H. (1997) *Cell* **88**, 627–636.
30. Perelson, A. S. & Oster, G. F. (1979) *J. Theor. Biol.* **81**, 645–670.
31. Perelson, A. S. & Weisbuch, G. (1997) *Rev. Mod. Phys.* **69**, 1219–1267.
32. Sun, R., Shepherd, S. E., Geier, S. S., Thomson, C. T., Sheil, J. M. & Nathenson, S. G. (1995) *Immunity* **3**, 573–582.
33. Brock, R., Wiesmuller, K. H., Jung, G. & Walden, P. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 13108–13113.
34. Sant'Angelo, D., Waterbury, G., Preston-Hulburt, P., Yoon, S. T., Medzhitov, R., Hong, S. & Janeway, C. J. (1996) *Immunity* **4**, 367–376.
35. Garcia, K. C., Degano, M., Stanfield, R. L., Brunmark, A., Jackson, M. R., Peterson, P. A., Teyton, L. & Wilson, I. A. (1996) *Science* **274**, 209–219.
36. Garboczi, D. N, Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E. & Wiley, D. C. (1996) *Nature (London)* **384**, 134–141.
37. Sim, B., Zerva, L., Green, M. & Gascoigne, N. (1996) *Science* **273**, 963–966.
38. Chang, H., Smolyar, A., Spoerl, R., Witte, T., Yao, Y., Goyarts, E. C., Nathenson, S. G. & Reinherz, E. L. (1997) *J. Mol. Biol.* **271**, 278–293.
39. Smith, K. D. & Lutz, C. T. (1997) *J. Immunol.* **158**, 2805–2812.
40. Turner, S., Jameson, S. C. & Carbone, F. R. (1997) *J. Immunol.* **159**, 2312–2317.
41. Ding, Y., Smith, K. J., Garboczi, D. N., Utz, U., Biddison, W. E. & Wiley, D. C. (1998) *Immunity* **8**, 403–411.
42. Garcia, K. C., Degano, M., Pease, L. R., Huang, M., Peterson, P. A., Teyton, L. & Wilson, I. A. (1998) *Nature (London)* **279**, 1166–1172.
43. Frelinger, J. A. & McMillan, M. (1996) *Immunol. Rev.* **154**, 45–58.
44. Janeway, C. A. & Travers, P. (1995) *Immunobiology* (Garland, New York).
45. Klein, J. (1982) *Immunology, The Science of Self-Nonself Discrimination*. (Wiley, New York).
46. Nowak, M. A., Tarczy-Hornoch, K. & Austyn, J. M. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10896–10899.
47. Chicz, R. M., Urban, R. G., Lane, W. S., Gorga, J. C., Stern, L. J., Vignali, D. A. & Strominger, J. L. (1992) *Nature (London)* **358**, 764–768.
48. Cox, A. L., Skipper, J., Chen, Y., Henderson, R. A., Darrow, T. L., Shabanowitz, J., Engelhard, V. H., Hunt, D. F. & Slingluff, C. L., Jr. (1994) *Science* **264**, 716–719.
49. Hunt, D. F., Michel, H., Dickinson, T. A., Shabanowitz, J., Cox, A. L., Sakaguchi, K., Appella, E., Grey, H. M. & Sette, A. (1992) *Science* **256**, 1817–1820.
50. Engelhard, V. H. (1994) *Annu. Rev. Immunol.* **12**, 181–207.
51. Manning, T. C., Schlueter, C. J., Brodnicki, T. C., Parke, E. A., Speir, J. A., Garcia, K. C., Teyton, L., Wilson, I. A. & Kranz, D. M. (1998) *Immunity* **8**, 413–425.
52. Silver, M. L, Guo, H., Strominger, J. L. & Wiley, D. C. (1992) *Nature (London)* **360**, 367–369.
53. Ignatowicz, L., Kappler, J. & Marrack, P. (1996) *Cell* **84**, 521–529.
54. Surh, C. D., Lee, D. S., Fung-Leung, W. P., Karlsson, L. & Sprent, J. (1997) *Immunity* **7**, 209–219.
55. Tourne, S., Miyazaki, T., Oxenius, A., Klein, L., Fehr, T., Kyewski, B., Benoist, C. & Mathis, D. (1997) *Immunity* **7**, 187–195.
56. van Meerrwijk, J. P. M, Marguerat, S., Lees, R. K., Germain, R. N., Fowlkes, B. J. & MacDonald, H. R. (1997) *J. Exp. Med.* **185**, 377–383.
57. De Boer, R. J. & Perelson, A. S. (1993) *Proc. R. Soc. London B* **B252**, 343–369.
58. Whitaker, L. & Renton, A. M. (1993) *J. Theor. Biol.* **164**, 531–536.
59. Nemazee, D. (1996) *Immunol. Today* **17**, 25–29.
60. Ernst, B., Surh, C. D. & Sprent, J. (1995) *J. Exp. Med.* **182**, 961–971.
61. Scollay, R. & Godfrey, D. I. (1995) *Immunol. Today* **16**, 268–273.
62. Penit, C. & Vasseur, F. (1997) *J. Immunol.* **159**, 4848–4856.
63. Egerton, M., Scollay, R. & Shortman, K. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2579–2582.
64. Shortman, K., Vremec, D. & Egerton, M. (1991) *J. Exp. Med.* **173**, 323–332.
65. Lucas, B., Vasseur, F. & Penit, C. (1994) *J. Immunol.* **153**, 53–62.
66. Detours, V., Sulzer, B. & Perelson, A. S. (1996) *J. Theor. Biol.* **183**, 409–416.
67. Bevan, M. J. (1997) *Immunity* **7**, 175–178.
68. Swain, S. L., Trefts, P. E., Tse, H. Y. & Dutton, R. W. (1977) *Cold Spring Harbor Symp. Quant. Biol.* **41**, 597–609.
69. Howard, J. (1980) *Nature (London)* **286**, 15–16.
70. Doherty, P. C. & Bennink, J. R. (1980) *Scand. J. Immunol.* **12**, 271–280.
71. Matzinger, P. (1993) *Immunol. Rev.* **135**, 81–117.
72. Nanda, N. K. & Sercarz, E. E. (1995) *Intl. Immunol.* **7**, 353–358.
73. Chung, K. L. (1979) *Elementary Probability with Stochastic Processes* (Springer, New York).