



## A Quantitative Theory of Affinity-driven T Cell Repertoire Selection

VINCENT DETOURS\*<sup>†‡</sup>, RAMIT MEHR<sup>§</sup> AND ALAN S. PERELSON\*<sup>¶</sup>

\**Theoretical Biology and Biophysics, MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A.*, <sup>†</sup>*Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A.*, <sup>‡</sup>*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, U.S.A.* and <sup>§</sup>*Department of Molecular Biology, Princeton University, Princeton, NJ 08544, U.S.A.*

(Received on 9 March 1999, Accepted on 9 July 1999)

Binding of the T cell antigen receptor (TCR) to peptides presented on molecules encoded by major histocompatibility complex (MHC) genes is the key event driving T cell development and activation. Selection of the T cell repertoire in the thymus involves two steps. First, positive selection promotes the survival of cells binding thymic self-MHC-peptide complexes with sufficient affinity. The resulting repertoire is self-MHC restricted: it recognizes foreign peptides presented on self, but not foreign MHC. Second, negative selection deletes cells which may be potentially harmful because their receptors interact with self-MHC-peptide complexes with too high an affinity. The mature repertoire is also highly alloreactive: a large fraction of T cells respond to tissues harboring foreign MHC. We derive mathematical expressions giving the frequency of alloreactivity, the level of self-MHC restriction, and the fraction of the repertoire activated by a foreign peptide, as a function of the parameters driving the generation and selection of the repertoire: self-MHC and self-peptide diversity, the stringencies of positive and negative selection, and the number of peptide and MHC polymorphic residues that contribute to T cell receptor binding. Although the model is based on a simplified digit string representation of receptors, all the parameters but one relate directly to experimentally determined quantities. The only parameter without a biological counterpart has no effect on the model's behavior besides a trivial and easily preventable discretization effect. We further analyse the role of the MHC and peptide contribution to TCR binding, and find that their relative, rather than absolute value, is important in shaping the mature repertoire. This result makes it possible to adopt different physical interpretations for the digit string formalism. We also find that the alloreactivity level can be inferred directly from data on the stringency of selection, and that, in agreement with recent experiments, it is not affected by thymic selection.

© 1999 Academic Press

### 1. Introduction

T and B lymphocytes recognize antigens (e.g. bacteria, viruses) via receptor molecules expressed on their surface (Abbas *et al.*, 1994; Elgert, 1996). Each receptor is constituted of a constant region participating in the signaling

function of the molecule, and a variable region responsible for antigen recognition. Variable regions are generated by rearranging genes from gene segment libraries, thus creating a diverse repertoire of possible receptors, which is believed to cover most of the universe of biological molecules. Each individual cell, however, usually expresses only one allele of its respective receptor genes, and transmits it to its progeny.

<sup>¶</sup> Author to whom correspondence should be addressed.  
E-mail: [asp@lanl.gov](mailto:asp@lanl.gov)

A collection of lymphocytes expressing the same receptors is called a clone. The antigen-recognition potential of each individual animal depends on the number of different clones in its repertoire.

While B lymphocyte receptors bind antigen in its native form, T lymphocytes depend on the processing of the antigen by antigen-presenting cells (APCs). APCs break antigenic proteins into small peptides (of the order of 10 amino acids), and present these peptides on their surface within the grooves of major histocompatibility (MHC, class-I or class-II) molecules. The binding of T cell receptors (TCRs) to these MHC-peptide complexes (MPCs) is a key prerequisite for the activation of a T cell's immune function. MHC genes are extremely polymorphic. Typically, an individual possesses only a small number out of the hundreds MHC alleles identified so far in the human population. Molecules from different MHC alleles differ essentially in the specificity of their peptide-binding groove (Jorgensen *et al.*, 1992). Thus, different individuals present different sets of peptides, making unlikely the evolution of pathogens capable of evading MHC presentation in all individuals of a population.

Since the gene rearrangement process leading to a diverse T cell repertoire is random, it generates T cells potentially harmful to the host because they recognize combination of MHC and self-peptides. Rearrangement also produces cells that are useless because their TCRs interact too poorly with the host's MHC molecules. Efficient and safe function of the immune system is ensured during the development of T cells in the thymus, by the two-step process of thymic selection. First, positive selection (Fowlkes & Schweighoffer, 1995; Jameson *et al.*, 1995) favors the development of T cells whose receptors recognize peptides presented by self-MHC molecules. The resulting selected repertoire is said to be self-MHC-restricted, i.e. positively selected T cells recognize pathogens presented by self-MHC molecules, but ignore them if presented by foreign MHC (Bevan, 1977; Waldmann, 1978; Zinkernagel, 1978; Schwartz, 1984; von Boehmer, 1990). Second, to prevent autoimmunity, negative selection (Nossal, 1994; Sprent & Webb, 1995) deletes from the repertoire T cells recognizing self-peptides presented on self-MHC. The

question arises of how the two opposing forces of positive and negative selection can act by signaling through the same receptor. One major theory proposed to resolve this apparent paradox relies on the concept of different affinity thresholds for positive and negative selection (Alam *et al.*, 1996). Only cells whose receptors bind a self-MPC with affinity above the positive selection threshold receive rescue signals necessary to continue maturation. Cells not receiving these signals because their TCRs have low affinity to thymic MPC die. By contrast, negative selection actively instructs T cells whose receptors bind self-MPC with an affinity exceeding the negative selection threshold to die. Overall, only 3% of the cells produced in the thymus have TCRs with the intermediate affinity required to survive selection (Shortman *et al.*, 1991).

Because MHC genes are extremely polymorphic, two individuals are very unlikely to express the same MHC molecules. Beside controlling self-MHC restriction, MHC polymorphism is also the main obstacle to tissue transplantations (hence the name major *histocompatibility* complex). A graft between two mice identical at all loci except for one codon at the MHC leads to an acute T-cells-mediated rejection. It can be measured *in vitro* that as many as 1–24% of T cells are alloreactive, i.e. they are activated by products of a given foreign MHC allele (Bevan *et al.*, 1976; Ashwell *et al.*, 1986). Alloreactivity has received much attention not only because of its far-reaching therapeutic implications, but also because it is a long-standing immunological puzzle. Its 1–24% frequency is difficult to reconcile with the fact that only one T cell out of  $10^4$ – $10^6$  recognizes a given pathogen (Stockinger *et al.*, 1980; Zinkernagel, 1996).

Recent mutagenesis studies and crystallographic data (Wilson & Garcia, 1997; Davis *et al.*, 1998) have shed considerable light on how TCRs interact with MHC-peptide complexes. In particular, the non-covalent bounds mediating TCR/MHC-peptide interaction have been mapped for several triplets of molecules (Garboczi *et al.*, 1996; Ding *et al.*, 1998; Garcia *et al.*, 1998; Manning *et al.*, 1998), providing important insight into the molecular mechanisms mediating allorecognition (Speir *et al.*, 1998). These studies focused on particular TCR/MHC-peptide

ternary complexes. Thus, they cannot explain, in quantitative terms, observed allorecognition frequencies and self-restriction levels, because these quantities are global macroscopic properties of the T cell repertoire, not properties of any particular TCR. The mathematical model presented in this paper aims at bridging the gap between microscopic molecular properties of TCRs and their ligands, and macroscopic properties of the T cell repertoire.

We propose mathematical expressions\* for deriving expected levels of alloreactivity (Sections 2.5 and 2.7), self-MHC restriction (Section 2.9), and foreign antigen response frequency (Section 2.8) from the parameters controlling the generation and selection of TCRs. This model relies on a minimal representation of MHC molecules, peptides, and TCRs that supports the notions of affinity (Section 2.1), ligand diversity, and ligand size (Section 2.2). Positive and negative selection affinity thresholds are inferred from experimental estimates of the stringencies of the overall selection process, and of negative selection (Section 2.3). The derivation of these expressions reveals two biologically significant results. First, alloreactivity frequency can be inferred directly from data on the stringency of selection, a previously unnoticed relation. Second, selection has no effect on alloreactivity level, a result in accordance with recent experimental data. More biological applications of our theory can be found elsewhere (Detours & Perelson, 1999; Detours *et al.*, 1999).

All parameters of this model relate directly to measurable biological quantities (Detours & Perelson, 1999), except for one controlling the discretization of the model. We show that this parameter has no effect on model output as long as it is chosen large enough (Section 2.10). We also show that the output of the model depends on relative rather than absolute contributions of MHCs and peptides to TCR binding (Section 2.11). This finding allows a greater flexibility concerning the interpretation of the formalism used to represent receptors. Finally, the validity of the proposed mathematical expressions is

assessed by comparing them with explicit simulations of the model (see the appendix).

## 2. Model and Results

### 2.1. GENERALIZED SHAPE AND MATCH SCORE

The concept of *generalized shape space* introduced by Perelson & Oster (1979) provides a convenient framework with which to represent receptors and their ligands. The underlying idea is that the non-covalent binding of two proteins can be described with a relatively small number of parameters, such as their geometric shape, their charge and their hydrophobicity. As in previous simulation studies (reviewed by Perelson & Weisbuch, 1997), we model 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). Digits belong to  $\{0, 1, 2, \dots, d_{max}\}$ . Each pair of digits  $(x, y)$  is assigned a digit interaction strength given by

$$I(x, y) = x \oplus y, \quad (1)$$

where  $\oplus$  consists in applying the “exclusive or” operator on the binary representations of  $x$  and  $y$ , and interpreting the result as a decimal number (Detours *et al.*, 1996). This definition makes fast simulations possible, because  $\oplus$  is implemented efficiently in computer hardware. For example, if  $d_{max} = 255$ , then four binary digits fit in a 32 bit integer, which make possible the calculation of interaction strengths between four pairs of digits in only one CPU cycle. Digits are equiprobable in our model, thus the probability distribution of  $I$  between pairs of random digits is

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

$p_I$  is the only feature of  $I$  that affects the outputs of our model. Thus,  $\oplus$  could be replaced by any operator resulting in a similar distribution. Non-binary digit strings have also been used by Lancet *et al.* (1993), Weisbuch & Oprea (1994), and Smith *et al.* (1997); these authors however used different definitions for interaction strength.

\* A software package in C language implementing these expressions and related simulations can be down-loaded from <ftp://cell.lanl.gov/pub/detours/abs-lab-1.1.tar.gz>.

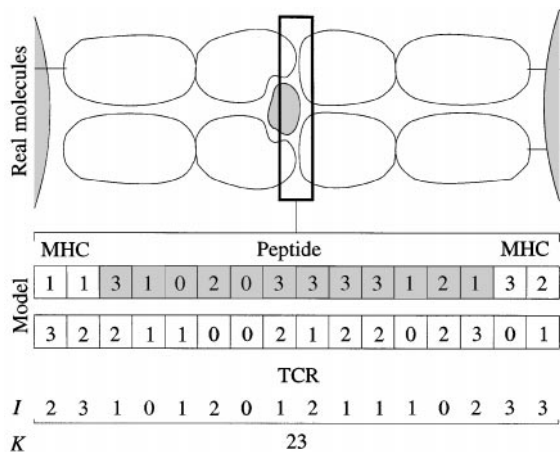


FIG. 1. Digit-string representation of MHC-peptide and TCR interaction. Only the interface between TCRs and MHC-peptide complexes (framed region in upper diagram) are taken into account in the model. MHC-peptide complexes are constructed by inserting an  $l_p$  digits peptide string in an  $l_m$  digits MHC string. TCRs are sequences of  $l_m + l_p$  digits chosen randomly from  $\{0, 1, 2, \dots, d_{max}\}$  ( $d_{max} = 3$  in the figure and 255 in the calculations). The interaction strength,  $I(x, y)$ , between digits  $x$  and  $y$  is a measure of their complementarity (see text). Affinity,  $K$ , is the sum of interaction strengths of contacting digits,  $I(x_i, y_i)$ , in aligned strings. Digits are arranged in a linear fashion in the graphic for visual convenience; however no geometrical hypothesis is included in the equations implementing this binding scheme [eqns (1) and (3)].

Mutation at one position in an antigenic peptide ligand can result in an enhanced, decrease of unchanged T cell response (Vergelli *et al.*, 1997). Experiments in which multiple amino-acid substitutions are made reveal that the integration of the effects of individual residues determines the outcome of TCR ligation (Vergelli *et al.*, 1997). For example, response is unchanged if the antigenic peptide is affected either by two neutral substitutions, or by the combination of enhancing and disrupting substitutions. The additive contribution of individual amino acids is introduced into our model by defining the match score, or affinity  $K$ , between two-digit string proteins  $[x_1, \dots, x_l]$  and  $[y_1, \dots, y_l]$  as the sum of digit interactions:

$$K([x_1, \dots, x_l], [y_1, \dots, y_l]) = \sum_{i=1}^l I(x_i, y_i). \quad (3)$$

Our implementation of  $K$  provides fine grained affinity distributions (Detours *et al.*, 1996), a

feature essential to set selection thresholds to realistic values (Section 2.10). The number of possible match scores between strings of  $l$  digits is  $ld_{max} + 1$ .

The match scores are always computed on aligned strings, so the central digits of TCRs always contact a peptide, and digits at the extremities, a MHC. This modeling choice follows from studies according to which TCRs bind MHC-peptide complexes with a common orientation (Sun *et al.*, 1995; Brock *et al.*, 1996; Sant'Angelo *et al.*, 1996; Garcia *et al.*, 1996; Garboczi *et al.*, 1996; Sim *et al.*, 1996; Chang *et al.*, 1997; Smith & Lutz, 1997; Turner *et al.*, 1997; Ding *et al.*, 1998; Garcia *et al.*, 1998).

The string representation provides a convenient visual representation and a simple indexing scheme for digits. However, the model is independent of this particular geometrical arrangement, because the order in which interaction strengths are added in eqn (3) is not important. In particular, the model supports digits arranged in a two-dimensional array mimicking the solvent-accessible surface of proteins [as proposed by Weinand (1990) and Lancet *et al.* (1993)].

## 2.2. TCRs, MHCs, AND PEPTIDES

Our model describes residues at the interface between TCRs and MHC-peptide complexes, not the full structure of these molecules. MHC and peptide are random strings of  $l_m$  and  $l_p$  digits, respectively. We denote by  $\gamma$  the match score between the extremities of a random TCR in contact with MHC (i.e. the digits facing MHC segments in Fig. 1), and a random MHC string. The distribution of the sum of two discrete independent random variables  $X$  and  $Y$ , with probability distributions  $p_X(\cdot)$  and  $p_Y(\cdot)$ , respectively, is  $p_X * p_Y$  (Chung, 1979, p. 179), where  $*$  denotes the convolution operator. The notation  $p_X^l$  stands for the convolution of  $p_X$   $l$ -times by itself. Since digits are independent in the model, the probability distribution of  $\gamma$  is given by

$$p_\gamma = p_I^{l_m}. \quad (4)$$

Similarly, the distribution of  $\theta$ , the match scores between random peptides and random TCRs is

$$p_\theta = p_I^{l_p}. \quad (5)$$

It is assumed that  $l_m$  and  $l_p$  are the same for all MHCs and peptides. This assumption is reasonable because we restrict our analysis to class-I MHC, which present peptides of 8–9 amino acids (Madden, 1995).

The number of MHC alleles expressed in an individual is denoted by  $n_m$ . The groove of MHC molecules are characterized by allele-specific motifs that peptides sequences need to fit in order to be presented (Madden, 1995). For example, only peptides with amino acids tyrosine and methionine at the third and eighth positions, respectively, associate with murine class-I allele  $K^d$  (Falk *et al.*, 1991). We call  $n_p$  the number of self-peptides of distinct sequences that can bind to molecules encoded by a given MHC allele. How different from the point of view of TCRs are the subsets of self-peptides presented by MHC molecules of different alleles? We explore two extreme hypothesis regarding this issue.

- (i) The sets of peptides associated with each MHC are non-overlapping in the model. This represents the hypothetical *in vivo* situation in which binding motifs are so specific that the probability that a given peptide fits the motifs of two MHC alleles is very low. Another possibility is that the conformations of a peptide induced by the grooves of two MHC molecules from different alleles are so different that this peptide appears to TCRs as two totally unrelated peptides. Under this hypothesis, we say the effect of binding motif is *maximal*.
- (ii) The same set of peptides is presented by all MHCs in the model. This would occur *in vivo* if binding motifs do not influence peptide presentation at all, i.e. MHC motifs a non-specific and have no effect on peptide conformation.

Physiological conditions lie between these two extreme hypotheses. In case (i), each one of the  $n_m$  MHC is associated with a different set of  $n_p$  self-peptides. Thus, selection is driven by  $n_m$  sets of  $n_p$  distinct self-peptides. In case (ii), all  $n_m$  self-MHC associate with the same set of  $n_p$  self-peptides. Thus, although they are  $n_m \times n_p$  self-MHC peptide complexes, selection is driven by only  $n_p$  self-peptides. Alternative (ii) is clearly

unrealistic, because motifs do have an impact (Madden, 1995). Investigating it is nevertheless necessary to quantify the effect of motifs on the properties of the repertoire.

Our goal is to measure self-restriction and alloreactivity, which depend, by definition, on MHC polymorphism, and on the specificity of TCRs. Therefore, MHC polymorphism-independent effects do *not* need to be part of the model, which legitimates the following simplifications. The effect of T cell coreceptors is omitted. Conserved MHC residues are not represented, i.e. the  $n_m$  MHC segments are interpreted as the polymorphic parts of MHC molecules accessible to TCRs. To our knowledge, there is no evidence for a germline-encoded bias of TCRs towards recognition of some particular peptides, and bias towards recognition of MHC most likely results from interaction with MHC-conserved residues (Merkenschlager *et al.*, 1997; Zerrahn *et al.*, 1997; Jameson & Bevan, 1998), which are not taken into account here. Thus, assuming that pre-selection TCRs are random is justified in the context of the model [see Detours & Perelson (1999) for a more extensive discussion of this issue].

### 2.3. POSITIVE AND NEGATIVE 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  $K > K_N$ . Stated formally, if  $\Omega$  is the set of all self-MHC-peptide complexes, the necessary and sufficient condition for a clone to be part of the peripheral repertoire is

$$K_P \leq \max_{K \in \Omega} (K) \leq K_N. \quad (6)$$

$K_P$  and  $K_N$  are related to experimental data by considering the fractions of clones surviving the different stages of selection (Fig. 2).

The fraction of clones allowed to reach the periphery is

$$f = f_P \cdot f_N \quad (7)$$

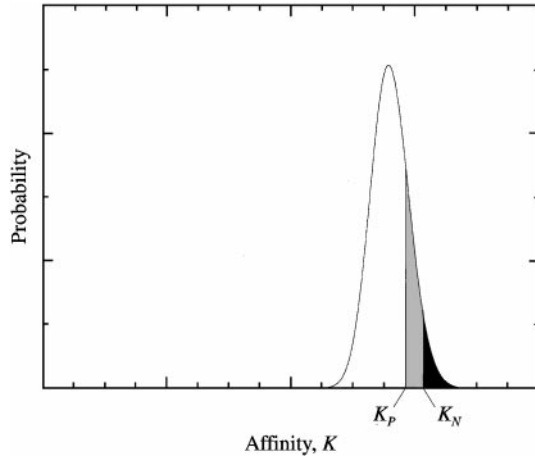


FIG. 2. Setting selection thresholds. To keep it readable, the diagram is not drawn to scale. Distribution of the maximum affinity between a TCR in the pre-selection repertoire and  $n_m \times n_p$  random MHC-peptide complexes is plotted (see text for mathematical derivation). The selection thresholds  $K_p$  and  $K_N$  are set such that the fraction of TCRs with best affinity greater than  $K_p$  is  $f_p$  (grey and black areas) and the fraction of TCRs with best affinity between  $K_p$  and  $K_N$  is  $f$  (grey area). The black area represents the fraction of TCRs deleted by negative selection.

where  $f_p$  is the fraction of clones surviving positive selection (a similar parameter is used by Nowak *et al.*, 1992), and  $f_N$  is the fraction of positively selected clones that survive negative selection (Nowak *et al.*, 1992; De Boer & Perelson, 1993; Whitaker & Renton, 1993; Nemazee, 1996). Parameters  $f$  and  $f_N$  can be inferred from recent experimental data (Detours & Perelson, 1999).

In order to relate eqns (7) and (6) to data, we now proceed to derive expressions for the distribution of maximal match score between a random TCR and the set of all  $n_m \times n_p$  self-MHC-peptide complexes under the hypotheses (i) and (ii).

Let  $\delta$  be the maximal match score between a random TCR and a set of MHC-peptide complexes made out a random MHC combined with  $n_p$  random peptides. According to eqn (3), the match score between two strings is the sum of match scores between their subparts, thus  $\delta$  is the sum of  $\gamma$  and the maximum of  $\theta$  over  $n_p$  peptides. The distribution of the first quantity is given in eqn (4). We define  $M_{X,n}(\cdot)$  to be the maximum of  $n$  independent random variables with identical distribution  $p_X$  [ $M_{X,n}(\cdot)$  is derived in Chung, 1979, p. 128]. Accordingly, the distribution of the

maximum of  $\theta$  over  $n_p$  peptides is  $M_{\theta,n_p}$ , because self-peptide strings are random and independent of each other in the model.

Since there are no binding motifs in the model [their effect is implemented implicitly under (i)], peptide match scores are not allele dependent, and are therefore independent of MHC match score. Thus,

$$p_\delta = p_\gamma * M_{\theta,n_p}. \quad (8)$$

Under hypothesis (i), different MHCs present different self-peptides. Thus  $\omega$ , the best match score between a random TCR and a set of MHC-peptide complexes made out of  $n_m$  MHCs and  $n_m$  sets of  $n_p$  random peptides, is simply equal to the maximum of  $\delta$  over  $n_m$  independent trials:

$$p_\omega = M_{\delta,n_m}. \quad (9)$$

We call  $\omega'$  the equivalent of  $\omega$  under hypothesis (ii). All  $n_m$  MHCs present the same  $n_p$  peptides under this hypothesis; consequently,

$$p_{\omega'} = M_{\gamma,n_m} * M_{\theta,n_p}. \quad (10)$$

Taken together eqns (6–9), specify a unique pair of selection thresholds ( $K_p$ ,  $K_N$ ) such that, for hypothesis (i),

$$f_p = \sum_{z=K_p}^l p_\omega(z) \quad \text{and} \quad f = \sum_{z=K_p}^{K_N} p_\omega(z). \quad (11)$$

Substituting  $\omega'$  for  $\omega$  gives selection thresholds under hypothesis (ii).

#### 2.4. T CELL ACTIVATION

Defining activation of selected T cells is a prerequisite for studying alloreactivity and antigen response frequency. A clone is considered activated by a set of MHC-peptide complexes,  $\Gamma$ , if the match score between its TCR and at least one MHC-peptide complex in  $\Gamma$  is greater than  $K_N$ . In other words, activation occurs if

$$K_N < \max_{K \in \Gamma} (K). \quad (12)$$

TABLE 1  
 Notation

Notation	Definition
$\oplus$	Exclusive or
*	Convolution
$p_X$	Probability distribution of $X$
$M_{X,n}$	Probability distribution of the maximum of $X$ over $n$ independent trials
$I(a, b)$	Interaction strength between digits $a$ and $b$
$K(x, y)$	Affinity between strings $x$ and $y$
$n_m$	Number of MHC class-I loci
$n_p$	Number of self-peptides
$l_m$	Number of MHC digits
$l_p$	Number of peptide digits
$f$	Fraction of selected clones
$f_N$	Fraction of positively selected clones that survive negative selection
$d_{max}$	Largest digit
$\gamma$	Affinity between a MHC and a pre-selection TCR
$\theta$	Affinity between a peptide and a pre-selection TCR
$\omega$	Maximum affinity of a pre-selection TCR over a set of $n_m \times n_p$ MHC-peptide complexes
$\phi$	Maximum affinity of a selected TCR over all self-MHCs
$\psi$	Maximum affinity of a selected TCR over all self-peptides
$\eta$	Affinity between a selected TCR and a self-MHC
$f_P$	Fraction of positively selected clones
$K_P$	Affinity threshold for positive selection
$K_N$	Affinity threshold for negative selection
$a_0$	Pre-selection alloreactivity
$a$	Post-selection alloreactivity
$R$	Fraction of clones responding to a foreign peptide presented on self-MHC
$R_a$	Fraction of clones responding to a foreign peptide presented on foreign MHC
$r$	Self-MHC restriction ratio

Note: Symbols may be primed when referring to hypothesis that binding motifs have no effect (case (ii)).

The repertoire is self-tolerant by construction since no clone having a match score larger than  $K_N$  to a self-MHC-peptide complex can survive negative selection in the model.

## 2.5. PRE-SELECTION ALLOREACTIVITY

Pre-selection alloreactivity,  $a_0$ , is defined as the fraction of clones in the pre-selection repertoire that are activated by any MHC-self-peptide complexes. However, the pre-selection repertoire cannot distinguish self-peptides from random peptides, since the ability to do so is acquired during selection. Thus, under (i), it follows from eqns (9) and (12) that

$$a_0 = \sum_{z > K_N} p_\omega(z), \quad (13)$$

which because of eqns (11) and (7) simplifies to

$$a_0 = f_P \cdot (1 - f_N). \quad (14)$$

The same argument applies under (ii), thus pre-selection alloreactivity is the same under either hypothesis. It is worth noting that the above expression is independent of the model's parameters controlling MHCs and peptides. 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 does not imply that those parameters have no influence on pre-selection alloreactivity *in vivo*. Rather, it suggests that they make their influence felt by changing the stringencies of positive and negative selection.

## 2.6. MATCH SCORES BETWEEN SELECTED TCRS AND SELF-MHC-PEPTIDE COMPLEXES

Assessing the effect of selection on the distribution of match scores between TCRs and self-MHCs, and between TCRs and self-peptides, is a necessary step towards the study of the post-selection repertoire. Let  $\phi$  be the *best* match score of a given *selected* TCR over all self-MHCs under hypothesis (ii).  $p_\phi(k)$  is the probability that a TCR recognizing self-MHCs with the 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  $\phi + z$  lies within the selection window. Therefore,

$$p_\phi(k) = \frac{1}{f} M_{\gamma, n_m}(k) \sum_{z=K_p-k}^{z \leq K_N-k} M_{\theta, n_p}(z). \quad (15)$$

The distribution  $p_\psi$  of the best match score of a given selected TCR over all self-peptides,  $\psi$ , is obtained by swapping  $\gamma$  and  $\theta$ , and  $n_m$  and  $n_p$  in the above equation.

Under hypothesis (i), match scores with self-MHC,  $\eta$ , rather than *best* match scores will be relevant. They can be partitioned into two groups: either the MHC drove the selection of the TCR at hand, or it did not. Since positive selection is stringent *in vivo*, it is very unlikely that more than one MHC has a match score large enough to mediate positive selection. Therefore, we neglect these situations. As a result, the probability that a selected TCR is matched against an MHC that drove selection is  $1/n_m$ , and the probability that it is not is  $(1 - 1/n_m)$ . In the first case, the match score distribution is given by eqn (15). In the second case, the only constraint on  $\eta$  is that no self-peptides can be found with a match score,  $z$ , for the TCR such that  $K_p \leq \eta + z$ . Combining these two scenarios together leads to

$$p_\eta(k) = \frac{1}{n_m} \left[ p_\phi(k) + \frac{n_m - 1}{1 - f_p} p_\gamma(k) \sum_{z=0}^{z < K_p - k} M_{\theta, n_p}(z) \right]. \quad (16)$$

## 2.7. POST-SELECTION ALLOREACTIVITY

Under (ii), self- and foreign-MHCs present the same subset of self-peptides; thus the post-

selection alloreactivity,  $a'$ , under this hypothesis is

$$a' = \sum_{z > K_N} [M_{\gamma, n_m} * p_\psi](z). \quad (17)$$

Under (i), different subsets are presented by self- and foreign-MHCs. The peptides presented by foreign MHC were, by definition, not present during selection, and for this reason cannot be distinguished from random peptides by the post-selection repertoire. Thus, the argument of Section 2.5 applies again

$$a = \sum_{z > K_N} p_\omega(z) = a_0. \quad (18)$$

In conclusion, the model predicts that if binding motifs have a maximal effect, then pre- and post-selection alloreactivities are identical. Interestingly, recent experimental comparisons between these quantities lead to a similar result. Merkschlager *et al.* (1997) measured that the alloreactivity of the murine repertoire selected under normal condition, and selected by anti-CD3 antibodies in MHC deficient fetal thymic organ culture was  $3 \pm 2.3$  and  $2.7 \pm 2.8\%$ , respectively. Zerrahn *et al.* (1997) used non-specific anti-TCR $\alpha\beta$  instead of anti-CD3 antibodies and found that the alloreactivity was  $5.7 \pm 2\%$  prior to selection and  $5.4 \pm 2.8\%$  after.

## 2.8. FOREIGN PEPTIDE RESPONSE FREQUENCY

Since self-MHCs associate with non-overlapping sets of peptides under hypothesis (i), the foreign peptide for which the response frequency is measured is presented by one MHC only. It follows from eqns (12) and (16) that the fraction of clones responding to a foreign peptide presented by any one of the  $n_m$  self-MHCs is

$$R = \sum_{z > K_N} [p_\eta * p_\theta](z). \quad (19)$$

Under (ii), any peptide presented by one MHC is presented by all of them. Thus, antigen response frequency is

$$R' = \sum_{z > K_N} [p_\phi * p_\theta](z). \quad (20)$$

2.9. SELF-MHC RESTRICTION RATIO

The extent of self-MHC restriction has typically been estimated by comparing the effector activity against foreign peptides presented by self-MHC with the activity against foreign peptides presented by foreign MHC (Zinkernagel & Doherty, 1974; Bevan & Fink, 1978; Miller, 1978; Sprent, 1978; Waldmann, 1978; Zinkernagel, 1978). There are no effector functions in our model but it is reasonable to assume that response intensity is proportional to the number of responding clones, which is measurable in the model.

The response frequency to a foreign peptide presented on foreign MHC,  $R_a$ , is derived as in the previous section, except that the presenting MHCs appear as random to the selected TCRs. Thus in the case of hypothesis (i),

$$R_a = \sum_{z > K_N} [p_\gamma * p_\theta](z), \quad (21)$$

and for (ii),

$$R'_a = \sum_{z > K_N} [M_{\gamma, n_m} * p_\theta](z). \quad (22)$$

Inbred mice are more susceptible to diseases. This phenomena, called Ir-genes defect, is due to the fact that these animals are homozygous and therefore express a reduced array of MHC molecules, which limits the diversity of the peptides their APCs can present (Janeway & Travers, 1995). Absolute restriction would be observed *in vivo* for peptides that cannot be presented by foreign MHC. On the other hand, the possibility that the repertoire appears absolutely restricted to foreign MHC due to failure of self-MHCs to present the peptide is equiprobable. These effects would cancel each other out when considering average restriction over many experimental systems. Thus, we do not take Ir-genes defects into account in our calculation of self-restriction. The *self-MHC restriction ratio* of the selected repertoire is defined for hypotheses (i) and (ii), as

$$r = \frac{R}{R_a}, \quad \text{and} \quad r' = \frac{R'}{R'_a}. \quad (23)$$

This definition is identical to the one proposed by Stockinger *et al.* (1980) if we assume that the

clone size distribution is uniform in the naive T cell repertoire. As far as we know there is no reason to assume otherwise.

2.10. DISCRETENESS OF THE MODEL

Seven parameters (Table 2) determine the outputs of the model (Section 2.5–2.9). All can be estimated from biological data (Detours & Perelson, 1999), except one, the number of possible digits,  $d_{max}$ . This latter parameter controls the discretization of the model.

Figure 3 shows that  $d_{max}$  has no significant effect on the calculated values  $R$ ,  $a'$ , and  $r$ , long it is chosen large enough ( $\geq 31$ ). Low values of  $d_{max}$  result in a coarse discretization of affinity distribution  $p_\omega$  (Section 2.1) and subsequently in the impossibility of adjusting selection thresholds [eqn (11) and Fig. 3] as to implement the biological values of  $f$  and  $f_N$  (Table 1). For example, if  $d_{max} = 7$ , then the best approximation that can be found for  $f = 7.5 \times 10^3$  using eqn (11) is  $2.5 \times 10^2$ . Obviously  $d_{max}$  has no effect at all on  $a$  [see eqn (14)].

The number of the amino acids is only 20, but the local interaction strength associated with any particular pair of residues depends on the distance separating them, which in turn depends on the local conformation of the interacting molecules. The inter-residue distances are distributed on a continuous scale, thus local free energies must also be. Since the distribution of local interaction strengths becomes finer-grained as  $d_{max}$  is increased (Section 2.1), setting  $d_{max}$  to large values is also justified biologically.

TABLE 2

*All parameters are set from biological data (Detours & Perelson, 1999), except  $d_{max}$  (see text)*

Name	Definition	Value
$n_m$	Number of MHC class-I loci	3
$n_p$	Number of self peptides	$10^4$
$l_m$	Number of MHC digits	4
$l_p$	Number of peptide digits	6
$f$	Fraction of selected clones	0.75%
$f_N$	Fraction of positively selected clones that survive negative selection	37%
$d_{max}$	Largest digit	255

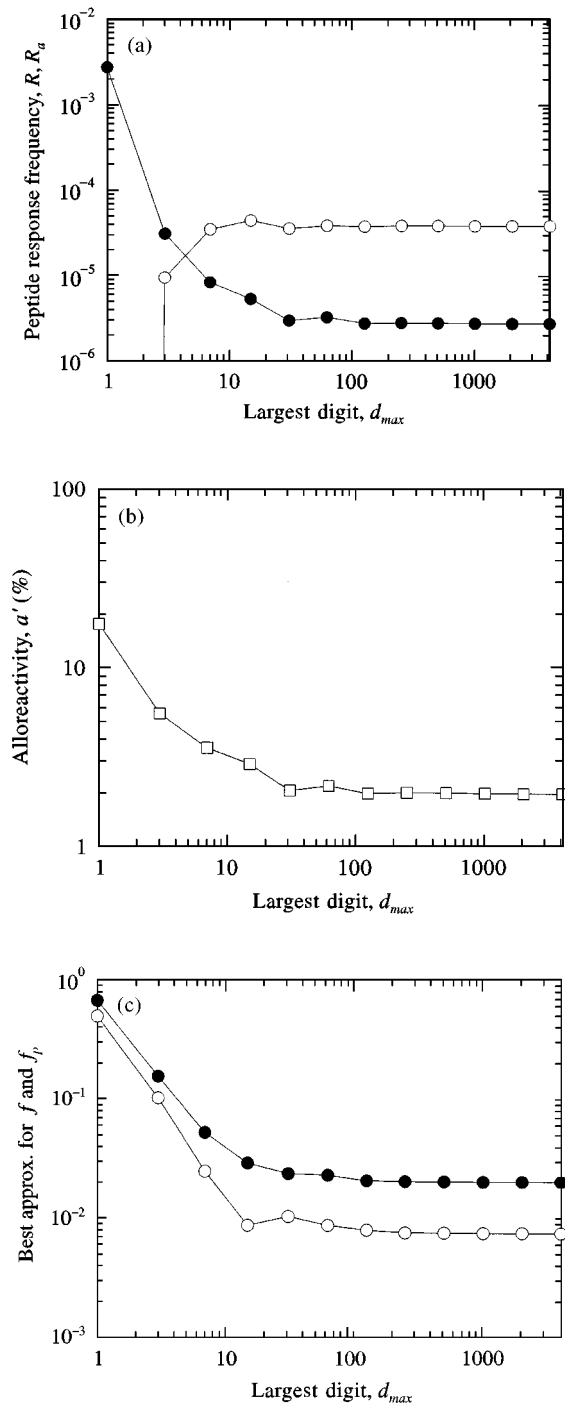


FIG. 3.  $d_{max}$  has no effect if chosen large enough. (a) Response frequencies against foreign peptide presented on self-MHC,  $R$  (○), and on foreign MHC,  $R_a$  (●). (b) Alloreactivity in the absence of binding motifs effects,  $a'$ . (c) Best approximations for the fraction of clones surviving positive selection using,  $f_p$  (●), and the fraction of clones surviving both positive and negative selection,  $f$  (○). All curves are stationary beyond  $d_{max} = 100$ .

#### 2.11. TOTAL RECEPTOR SIZE HAS A WEAK EFFECT WHEN RELATIVE CONTRIBUTION OF MHC AND PEPTIDE STRINGS TO BINDING IS KEPT CONSTANT

Although it may seem natural to view digits as amino-acid residues, one may also interpret digit strings as approximating continuous mean force potentials. Doing so would lead to assign large values to  $l_p$  and  $l_m$ .

Figure 4 shows that changing  $l_m$  and  $l_p$  does not have a drastic effect on the model as long as the relative contribution of peptide to TCR binding,  $c_p = l_p/(l_m + l_p)$ , is kept constant. Foreign peptide response frequencies,  $R$ , and  $R_a$  hardly depend on  $l$  [Fig. 4(a)]. The restriction ratio,  $r$ , converges toward 10 when  $l$  is increased, a decrease of 30% from the ratio obtained for very small strings. Such variation is acceptable in the context of the present study because experimental estimates of self-restriction vary over a much wider range (Matzinger, 1993; Detours & Perelson, 1999). The pre-selection alloreactivity,  $a_0$ , and the post-selection alloreactivity under the hypothesis of maximal motif effect,  $a$ , are independent of  $l_m$  and  $l_p$  [eqns (14) and (18)]. Post-selection alloreactivity assuming the absence of motif effect,  $a'$ , increase sublogarithmically with  $l$ , and stays in the experimental range 1–24% as  $l$  is varied in the range [4, 3000] [Fig. 4(b)]. Not enough data could be collected to determine whether  $a'$  converges toward a finite number. Taken together, these data suggest that the model accommodates both the residue and the mean force potential interpretations of digit strings.

### 3. Discussion

Interactions between T cells and APCs are key events in immune responses and T cell repertoire development. The core assumption of our model is that the affinity between TCR and MHC-peptide complex on APC determines the outcome of these interactions. Translating this hypothesis into a mathematically and computationally tractable quantitative framework requires a drastically simplified representation of protein-protein binding.

Tractability is achieved in the present study by modeling proteins as digit strings (Lancet *et al.*, 1993; Weisbuch & Oprea, 1994; Smith *et al.*,

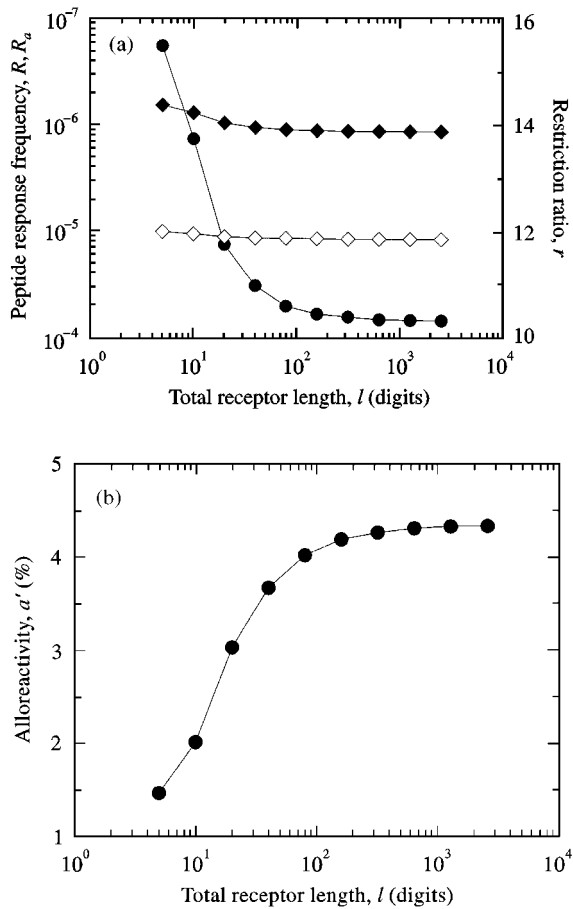


FIG. 4. Weak influence of total receptor length when  $c_p$  is constant. Total receptor length  $l = l_m + l_p$  is varied, while relative contribution of peptide to binding,  $c_p$ , is kept constant ( $c_p$  is set using values of  $l_m$  and  $l_p$  given in Table 1). (a) Foreign peptide response frequencies,  $R$  (black  $\diamond$ ), and  $R_a$  ( $\diamond$ ); restriction ratio  $r$  ( $\bullet$ ). (b) Alloreactivity in the absence of motif effect,  $a'$ .

1997). This formalism differs from previous bit strings approaches (Farmer *et al.*, 1986; Perelson, 1988; De Boer & Perelson, 1991; Seiden & Celada, 1992; Perelson & Weisbuch, 1997) in that it makes it possible to control the resolution of affinity distributions (Detours *et al.*, 1996). Fine-grained distributions are essential to set selection thresholds at realistic values in our model. The cost of this new control is an additional parameter for the size of the digit alphabet. We have shown that the lack of a biological counterpart for this parameter is not a problem for the modeling of affinity-based selection, since it has no effect provided that resolution is high enough to set selection thresholds in accordance with experimental data.

Although it seems natural to interpret digits as amino acids, alternative views are possible. We show that the relative contributions of MHCs and peptides to TCRs binding, rather than their absolute size, determines the model's behavior. Thus, these molecules can be represented by very long strings. Since digit discretization has no effect if the number of digits is large enough, this result implies that quantitative conclusions derived from the model are unchanged if digit-strings are interpreted as approximation for protein mean force potentials, rather than amino-acid strings. No topological assumptions are included in the equations used to compute affinity, beside the well-established fact that some parts of the TCR always contact the peptide whereas others always contact the MHC (Section 2.1). Consequently, our conclusions would remain the same if digits were arranged in two-dimensional arrays [as in Weinand, (1990) and Lancet *et al.* (1993)], to approximate three-dimensional mean force at the contact surface. Whatever the degree of realism of receptor representation, only affinity distributions ultimately matter in our model. Our results hold for any scheme in which the overall binding free energy is the sum of independent free energies gained through local contacts. Biologically reasonable binding schemes not respecting this independence condition (e.g. De Boer & Perelson, 1991; Detours *et al.*, 1996) give nearly Gaussian affinity distributions (Percus & Percus, 1994). Using an explicit receptor representation is useful mostly to facilitate reasoning.

Whether affinity drives T cell selection and activation is an unresolved question. This quantity correlates with the outcome of thymic selection (Alam *et al.*, 1996), and with the intensity of mature T cell lytic activity (Sykulev *et al.*, 1994). Yet, lysis level did not correlate with ligand affinity in another system (al-Ramadi *et al.*, 1995). The concept of serial triggering (Valitutti *et al.*, 1995) according to which a given MHC-peptide complex may be reused sequentially by different TCRs may change the current view of T cell signaling. High affinity is mediated by fast on-rate and/or slow off-rates. In the serial triggering model, slow off-rates prevent reuse of MHC-peptide complexes, and therefore do not provide maximal signaling. Thus, T cell selection

and activation may be driven by a function of on- and off-rates more complex than the ratio assumed in the affinity model. The validity of our theory depends on whether such function is normally distributed.

A transgenic T cell clone can be positively or negatively selected by decreasing or increasing the concentration of the peptides it recognizes (Ashton-Rickardt *et al.*, 1994; Ashton-Rickardt & Tonegawa, 1994; Sebzda *et al.*, 1994), and very low concentrations of such peptides do not mediate positive selection (Ashton-Rickardt, 1993; Ashton-Rickardt *et al.*, 1994). The number of triggered TCRs, which depends on MHC-peptides density on APCs determine the outcome of activation in mature cells (Viola & Lanzavecchia, 1996). Taken together, these data suggest that *avidity*, the combined effect of the affinity (or any other quantity characteristic of TCR/MHC-peptides interaction) and the concentration of the ligands, controls T cell signaling. Omitting this feature in our theory may undermine estimates of self-restriction. Elliott (1993) remarked that positive selection could occur because of low-affinity ligands present at high density. This may decrease the level of restriction. A more sophisticated mathematical model including the concept of avidity is needed to assess the quantity of such ligands, and whether their impact on average restriction level is significant or not.

The primary goal of our model was to frame the issue of alloreactivity in rigorous quantitative terms. The formal derivation of expressions for alloreactivity reveals two interesting implications of the affinity-driven selection hypothesis. The selection of thymocytes in a MHC-deficient background through antibodies binding constant portions of TCR-CD3 complexes shed some light on the properties of the repertoire prior to selection. Of particular interest is the finding that its alloreactivity is similar to that of the mature repertoire (Merkenschlager *et al.*, 1997; Zerrahn *et al.*, 1997). We find that if MHC-binding motifs have a strong impact on the sequences or conformation of the self-peptides driving selection, a reasonable assumption (Madden, 1995), then affinity-based selection implies that pre- and post-selection alloreactivities are identical. This result applies for any parameter setting of the model. Interestingly, pre-/post-alloreactivity

frequency can be inferred directly from the stringencies of positive and negative selection (Sections 2.5 and 2.7). This remarkably simple relation,  $a = f_P \cdot (1 - f_N)$ , was obtained by pointing out a previously unnoticed connection between the selection thresholds and data on the stringency of selection (Section 2.3).

Our model complements experimental approaches in three ways. Most experimental evidence for self-restriction comes from measuring the reactivity of the repertoire against a *few* particular antigens (Zinkernagel & Doherty, 1974; Bevan & Fink, 1978; Miller, 1978; Sprent, 1978; Waldmann, 1978; Zinkernagel, 1978). In transgenic mice experiments, the fate of a *single* T cell clone is analysed as a function of the selecting environment (von Boehmer, 1990). The small number of antigens and TCR specificities studied, and the discrepancies among the data produced (Matzinger, 1993), suggest that these experiments should be repeated on a large number of systems to obtain reliable *averages* of the level of restriction. Unfortunately, doing so is technically difficult. The mean field analysis presented here provides average results over all possible combinations of self/foreign-MHC and foreign peptide that can be constructed in the model.

Alloreactivity, antigen response frequency, and self-MHC restrictions are global properties of the repertoire, not properties of any single clone. Their emergence, however, ultimately lies in the details of the interactions between individual pairs of TCR and MHC-peptide complexes. For example, although alloreactivity is determined by the structure of TCRs, there is no way its high frequency can be fully explained by investigating any single one of them. The mathematical model presented in this paper establishes a connection between microscopic properties of TCR/MHC-peptide interactions, and macroscopic properties of the T cell repertoire such as alloreactivity.

Finally, the phenomena of T cell repertoire selection has been investigated from a variety of perspectives. Our model attempts to unify several of them. More specifically, it explains how the diversity and structural properties of ligands, and the stringencies of selection processes, interact to shape the mature repertoire.

Portions of this work were performed under the auspices of the U.S. Department of Energy. It was supported by NIH grants RR06555 and AI28433 (A.S.P.), and by NIH grant GM20964-25 for the study of genetics and regulation of autoimmunity, and by NIH grant AI10227-01 (R.M.).

## REFERENCES

- ABBAS, A. K., LICHTMAN, A. H. & POBER, J. S. (1994). Cellular and Molecular Immunology. Philadelphia: W. B. Saunders Co.
- AL-RAMADI, B. K., JELONEK, M. T., BOYD, L. F., MARGULIES, D. H. & BOTHWELL, A. L. (1995). Lack of strict correlation of functional sensitization with the apparent affinity of MHC/peptide complexes for the TCR. *J. Immunol.* **155**, 662–673.
- ALAM, S. M., TRAVERS, P. J., WUNG, J. L., NASHOLDS, W., REDPATH, S., JAMESON, S. C. & GASCOIGNE, N. R. (1996). T-cell-receptor affinity and thymocyte positive selection. *Nature* **381**, 616–620.
- ASHTON-RICKARDT, P. G. (1993). The role of peptide in the positive selection of CD8+ T cells in the thymus. *Thymus* **22**, 111–115.
- ASHTON-RICKARDT, P. G., BANDEIRA, A., DELANEY, J. R., VAN KAER, L., PIRCHER, H. P., ZINKERNAGEL, R. M. & TONEGAWA, S. (1994). Evidence for a differential avidity model of T cell selection in the thymus. *Cell* **76**, 651–663.
- ASHTON-RICKARDT, P. G. & TONEGAWA, S. (1994). A differential-avidity model for T-cell selection. *Immunol. Today* **15**, 362–366.
- ASHWELL, J. D., CHEN, C. & SCHWARTZ, R. H. (1986). High frequency and nonrandom distribution of alloreactivity in T cell clones selected for recognition of foreign antigen in association with self class II molecules. *J. Immunol.* **136**, 389–395.
- BEVAN, M. J. (1977). In a radiation chimaera, host H-2 antigens determine immune responsiveness of donor cytotoxic cells. *Nature* **269**, 417–418.
- BEVAN, M. J. & FINK, P. J. (1978). The influence of thymus H-2 antigens on the specificity of maturing killer and helper cells. *Immunol. Rev.* **42**, 3–19.
- BEVAN, M. J., LANGMAN, R. E. & COHN, M. (1976). H-2 antigen-specific cytotoxic T cells induced by concanavalin A: estimation of their relative frequency. *Eur. J. Immunol.* **6**, 150–156.
- BROCK, R., WIESMULLER, K. H., JUNG, G. & WALDEN, P. (1996). Molecular basis for the recognition of two structurally different major histocompatibility complex/peptide complexes by a single T-cell receptor. *Proc. Nat. Acad. Sci. U.S.A.* **93**, 13108–13113.
- CHANG, H., SMOLYAR, A., SPOERL, R., WITTE, T., YAO, Y., GOYARTS, E. C., NATHENSON, S. G. & REINHERZ, E. L. (1997). Topology of T cell receptor-peptide/class I MHC interaction defined by charge reversal complementation and functional analysis. *J. Mol. Biol.* **271**, 278–293.
- CHUNG, K. L. (1979). Elementary Probability with Stochastic Processes. Berlin: Springer-Verlag.
- DAVIS, M. M., BONIFACE, J. J., REICH, Z., LYONS, D., HAMPL, J., ARDEN, B. & CHIEN, Y. (1998). Ligand recognition by  $\alpha\beta$  T cell receptors. *Annu. Rev. Immunol.* **16**, 523–544.
- DE BOER, R. J. & PERELSON, A. S. (1991). Size and connectivity as emergent properties of a developing immune network. *J. theor. Biol.* **149**, 381–424.
- DE BOER, R. J. & PERELSON, A. S. (1993). How diverse should the immune system be? *Proc. Roy. Soc. Lond. B* **252**, 343–369.
- DETOURS, V., MEHR, R. & PERELSON, A. S. (1999). Deriving quantitative constraints under which T cell selection operates from data on the mature T cell repertoire (submitted).
- DETOURS, V. & PERELSON, A. S. (1999). Explaining high alloreactivity as a quantitative consequence of affinity-driven selection of thymocytes. *Proc. Nat. Acad. Sci. U.S.A.* **96**, 5153–5158.
- DETOURS, V., SULZER, B. & PERELSON, A. S. (1996). Size and connectivity of the idiotypic network are independent of the discreteness of the affinity distribution. *J. theor. Biol.* **183**, 409–416.
- DING, Y., SMITH, K. J., GARBOCZI, D. N., UTZ, U., BIDDISON, W. E. & WILEY, D. C. (1998). Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. *Immunity* **8**, 403–411.
- ELGERT, K. D. (1996). Immunology: *Understanding the Immune System*. New York: Wiley-Liss.
- ELLIOTT, J. I. (1993). The identity of the cells that positively select thymocytes. *Immunol. Rev.* **135**, 215–225.
- FALK, K., ROTZSCHKE, O., STEVANOVIC, S., JUNG, G. & RAMMENSEE, H. G. (1991). Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**, 290–296.
- FARMER, J. D., PACKARD, N. H. & PERELSON, A. S. (1986). The immune system, adaption and machine learning. *Physica D* **22**, 187–204.
- FOWLKES, B. J. & SCHWEIGHOFFER, E. (1995). Positive selection of T cells. *Cur. Opin. Immunol.* **7**, 188–195.
- GARBOCZI, D. N., GHOSH, P., UTZ, U., FAN, Q. R., BIDDISON, W. E. & WILEY, D. C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384**, 134–141.
- GARCIA, K. C., DEGANO, M., PEASE, L. R., HUANG, M., PETERSON, P. A., TEYTON, L. & WILSON, I. A. (1998). Structural basis of plasticity in T cell receptor recognition of self peptide-MHC antigen. *Nature* **279**, 1166–1172.
- GARCIA, K. C., DEGANO, M., STANFIELD, R. L., BRUNMARK, A., JACKSON, M. R., PETERSON, P. A., TEYTON, L. & WILSON, I. A. (1996). An alphabeta T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* **274**, 209–219.
- JAMESON, S. C. & BEVAN, M. J. (1998). T cell selection. *Cur. Opin. Immunol.* **10**, 214–219.
- JAMESON, S. C., HOGQUIST, K. A. & BEVAN, M. J. (1995). Positive selection of thymocytes. *Annu. Rev. Immunol.* **13**, 93–126.
- JANEWAY, C. A. & TRAVERS, P. (1995). *Immunobiology*. New York: Garland.
- JORGENSEN, J. L., REAY, P. A., EHRICH, E. W. & DAVIS, M. M. (1992). Molecular components of T-cell recognition. *Annu. Rev. Immunol.* **10**, 835–873.
- LANCET, D., SADOVSKY, E. & SEIDEMANN, E. (1993). Probability model for molecular recognition in biological receptor repertoires: significance to the olfactory system. *Proc. Nat. Acad. Sci. U.S.A.* **90**, 3715–3719.

- MADDEN, D. R. (1995). The three-dimensional structure of peptide-MHC complexes. *Annu. Rev. Immunol.* **13**, 587-622.
- 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). Alanine scanning mutagenesis of an  $\alpha\beta$  T cell receptor: mapping the energy of antigen recognition. *Immunity* **8**, 413-425.
- MATZINGER, P. (1993). Why positive selection? *Immunol. Rev.* **135**, 81-117.
- MERKENSCHLAGER, M., GRAF, D., LOVATT, M., BOMHARDT, U., ZAMOYSKA, R. & FISHER, A. G. (1997). How many thymocytes auditions for selection? *J. Exp. Med.* **186**, 1149-1158.
- MILLER, J. F. (1978). Restrictions imposed on T-lymphocyte reactivities by major histocompatibility complex: implications for T-cell repertoire selection. *Immunol. Rev.* **42**, 76-107.
- NEMAZEE, D. (1996). Antigen receptor 'capacity' and the sensitivity of self-tolerance. *Immunol. Today* **17**, 25-29.
- NOSSAL, G. J. (1994). Negative selection of lymphocytes. *Cell* **76**, 229-239.
- NOWAK, M. A., TARCZY-HORNOCH, K. & AUSTYN, J. M. (1992). The optimal number of major histocompatibility complex molecules in an individual. *Proc. Nat. Acad. Sci. U.S.A.* **89**, 10896-10899.
- PERCUS, O. E. & PERCUS, J. K. (1994). String matching for the novice. *Am. Math. Monthly* **101**, 944-947.
- PERELSON, A. S. (1988). Toward a realistic model of the immune system. In: *Theoretical Immunology: Part 2* (Perelson, A. S., ed.), pp. 377-401. Reading, MA: Addison-Wesley.
- PERELSON, A. S. & OSTER, G. F. (1979). Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-nonsel self discrimination. *J. theor. Biol.* **81**, 645-670.
- PERELSON, A. S. & WEISBUCH, G. (1997). Immunology for physicists. *Rev. Mod. Phys.* **69**, 1219-1267.
- SANT'ANGELO, D., WATERBURY, G., PRESTON-HULBERT, P., YOON, S. T., MEDZHITOV, R., HONG, S. & JANEWAY, C. J. (1996). The specificity and orientation of a TCR to its peptide-MHC class II ligands. *Immunity* **4**, 367-376.
- SCHWARTZ, R. H. (1984). The role of gene products of the major histocompatibility complex in T cell activation and cellular interactions. In: *Fundamental Immunology*, 1st edn (Paul, W. E., ed.), pp. 379-438. New York: Raven Press.
- SEBZDA, E., WALLACE, V. A., MAYER, J., YEUNG, R. S., MAK, T. W. & OHASHI, P. S. (1994). Positive and negative thymocyte selection induced by different concentrations of a single peptide. *Science* **263**, 1615-1618.
- SEIDEN, P. E. & CELADA, F. (1992). A model for simulating cognate recognition and response in the immune system. *J. theor. Biol.* **158**, 329-357.
- SHORTMAN, K., VREMEC, D. & EGERTON, M. (1991). The kinetics of T cell antigen receptor expression by subgroups of CD4+8+ thymocytes: delineation of CD4+8+3(2+) thymocytes as post-selection intermediates leading to mature T cells. *J. Exp. Med.* **173**, 323-332.
- SIM, B., ZERVA, L., GREEN, M. & GASCOIGNE, N. (1996). Control of MHC restriction by TCR  $v_{\alpha}$  CDR1 and CDR2. *Science* **273**, 963-966.
- SMITH, K. D. & LUTZ, C. T. (1997). Alloreactive T cell recognition of MHC class I molecules: the T cell receptor interacts with limited regions of the MHC class I long  $\alpha$  helices. *J. Immunol.* **158**, 2805-2812.
- SMITH, D. J., FORREST, S., HIGHTOWER, R. R. & PERELSON, A. S. (1997). Deriving shape space parameters from immunological data. *J. theor. Biol.* **189**, 141-150.
- SPEIR, J. A., GARCIA, K. C., BRUNMARK, A., DEGANO, M., PETERSON, P. A., TEYTON, L. & WILSON, I. A. (1998). Structural basis of 2C TCR allrecognition of H-2Ld peptide complexes. *Immunity* **8**, 553-562.
- SPRENT, J. (1978). Role of H-2 gene products in the function of T helper cells from normal and chimeric mice in vivo. *Immunol. Rev.* **42**, 108-137.
- SPRENT, J. & WEBB, S. R. (1995). Intrathymic and extrathymic clonal deletion of T cells. *Cur. Opin. Immunol.* **7**, 196-205.
- STOCKINGER, H., PFIZENMAIER, K., HARDT, C., RODT, H., ROLLINGHOFF, M. & WAGNER, H. (1980). H-2 restriction as a consequence of intentional priming: T cells of fully allogeneic chimeric mice as well as of normal mice respond to foreign antigens in the context of H-2 determinants not encountered on thymic epithelial cells. *Proc. Nat. Acad. Sci. U.S.A.* **77**, 7390-7394.
- SUN, R., SHEPHERD, S. E., GEIER, S. S., THOMSON, C. T., SHEIL, J. M. & NATHENSON, S. G. (1995). Evidence that the antigen receptors of cytotoxic T lymphocytes interact with a common recognition pattern on the H-2Kb molecule. *Immunity* **3**, 573-582.
- SYKULEV, Y., BRUNMARK, A., TSOMIDES, T. J., KAGEYAMA, S., JACKSON, M., PETERSON, P. A. & EISEN, H. N. (1994). High-affinity reactions between antigen-specific T-cell receptors and peptides associated with allogeneic and syngeneic major histocompatibility complex class I proteins. *Proc. Nat. Acad. Sci. U.S.A.* **91**, 11487-11491.
- TURNER, S., JAMESON, S. C. & CARBONE, F. R. (1997). Functional mapping of the orientation for TCR recognition of an H2-Kb-restricted ovalbumin peptide suggests that the beta-chain subunit can dominate the determination of peptide side chain specificity. *J. Immunol.* **159**, 2312-2317.
- VALITUTTI, S., MULLER, S., CELLA, M., PADOVAN, E. & LANZAVECCHIA, A. (1995). Serial triggering of many T-cell receptors by a few peptide-MHC complexes. *Nature* **375**, 148-151.
- VERGELLI, M., HEMMER, B., KALBUS, M., VOGT, A. B., LING, N., CONLON, P., COLIGAN, J. E., MCFARLAND, H. & MARTIN, R. (1997). Modifications of peptide ligands enhancing T cell responsiveness imply large numbers of stimulatory ligands for autoreactive T cells. *J. Immunol.* **158**, 3746-3752.
- VIOLA, A. & LANZAVECCHIA, A. (1996). T cell activation determined by T cell receptor number and tunable thresholds. *Science* **273**, 104-106.
- VON BOEHMER, H. (1990). Developmental biology of T cells in T cell-receptor transgenic mice. *Annu. Rev. Immunol.* **8**, 531-556.
- WALDMANN, H. (1978). The influence of the major histocompatibility complex on the function of T-helper cells in antibody formation. *Immunol. Rev.* **42**, 202-223.
- WEINAND, R. G. (1990). Somatic mutation, affinity maturation and the antibody repertoire: a computer model. *J. theor. Biol.* **143**, 343-382.
- WEISBUCH, G. & OPREA, M. (1994). Capacity of a model immune network. *Bull. Math. Biol.* **56**, 899-921.

WHITAKER, L. & RENTON, A. M. (1993). On the plausibility of the clonal expansion theory of the immune system in terms of the combinatorial possibilities of amino-acids in antigen and self-tolerance. *J. theor. Biol.* **164**, 531–536.

WILSON, I. A. & GARCIA, K. C. (1997). T-cell receptor structure and TCR complexes. *Current Opinion Struct. Biol.* **7**, 839–848.

ZERRAHN, J., HELD, W. & RAULET, D. H. (1997). The MHC reactivity of the T cell repertoire prior to positive and negative selection. *Cell* **88**, 627–636.

ZINKERNAGEL, R. M. (1978). Thymus and lymphohemopoietic cells: their role in T cell maturation in selection of T cells' H-2-restriction-specificity and in H-2 linked Ir gene control. *Immunol. Rev.* **42**, 224–270.

ZINKERNAGEL, R. M. (1996). Immunology taught by viruses. *Science* **271**, 173–178.

ZINKERNAGEL, R. M. & DOHERTY, P. C. (1974). Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248**, 701–702.

APPENDIX

In order to check that the proposed mathematical expressions and the underlying approximation [eqn (16)] are correct, we compared them with computer simulations of the model.

Simulations were run as follows. First a set of self-MHC-peptide complexes was constructed. Then  $n_t$  random TCRs were generated, and those satisfying the affinity selection criteria kept in the repertoire. Finally, a set of  $n_h$  foreign haplotypes, each containing  $n_m$  foreign MHC class-I allele was generated, as well as sets of  $n_f$  foreign peptides. These strings were used to measure alloreactivity, self-MHC restriction, and the foreign peptide response frequency of the selected repertoire. These two latter quantities were then averaged over all of the  $n_h \times n_f$  possible combinations of foreign MHCs and peptides in each simulation. Alloreactivity was averaged over all  $n_h$  foreign haplotypes. Results were further averaged over ten simulations (Table A1) in order to circumvent finite size effects associated with particular self-MHC-peptide environments.

Since simulations are very time-consuming, several parameters have not been set to their

TABLE A1  
*Comparison of theory and simulations. Parameters were set to values given in Table A2*

Property	Theory	Simulations
$a$	0.11	$0.11 \pm 0.03$
$R$	$3.5 \times 10^{-4}$	$3.5 \times 10^{-4} \pm 7 \times 10^{-5}$
$R_a$	$6.9 \times 10^{-5}$	$6.8 \times 10^{-5} \pm 4.9 \times 10^{-5}$
$r$	5.1	$5 \pm 0.16$
$a'$	$3.1 \times 10^{-2}$	$3.2 \times 10^{-2} \pm 1.8 \times 10^{-2}$
$R'$	$1.1 \times 10^{-4}$	$1.2 \times 10^{-4} \pm 3.7 \times 10^{-5}$
$R'_a$	$2.3 \times 10^{-5}$	$2.4 \times 10^{-5} \pm 3 \times 10^{-5}$
$r'$	4.9	$5.1 \pm 0.25$

TABLE A2  
*Toy parameter set*

Name	Definition	Value
$n_m$	Number of MHC class-I loci	3
$n_p$	Number of self peptides	512
$l_m$	Number of MHC digits	4
$l_p$	Number of peptide digits	12
$f$	Fraction of selected clones	2%
$f_N$	Fraction of positively selected clones that survive negative selection	37%
$d_{max}$	Largest digit	255
$n_t$	Number of TCRs undergoing selection*	$5 \times 10^6$
$n_h$	Number of foreign haplotypes*	100
$n_f$	Number of foreign peptides*	100

\*Used in simulations only.

most plausible biological value (Table A2) in order to obtain results within a reasonable time frame. In particular, the pre-selection TCR repertoire size,  $n_t$ , was set to an unrealistically small value, but sufficient to measure model output with acceptable precision. The theory derived here agreed with simulations, suggesting that it is correct.