*Proc. Natl. Acad. Sci. USA*
Vol. 89, pp. 10896–10899, November 1992
Immunology

# The optimal number of major histocompatibility complex molecules in an individual

MARTIN A. NOWAK*†, KRISTINA TARCZY-HORNOCH‡, AND JONATHAN M. AUSTYN†§

*Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; §Nuffield Department of Surgery, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, United Kingdom; †Wolfson College, University of Oxford, Oxford OX2 6UD, United Kingdom; and ‡St. Hilda's College, University of Oxford, Oxford, United Kingdom

*Communicated by Baruch S. Blumberg, August 17, 1992*

**ABSTRACT** **A straightforward argument is presented to calculate the number of different major histocompatibility complex (MHC) molecules in an individual that maximizes the probability of mounting immune responses against a large number of foreign peptides. It is assumed that increasing the number of MHC molecules per individual, $n$, has three different effects: (*i*) it increases the number of foreign peptides that can be presented; (*ii*) it increases the number of different T-cell receptors (TCRs) positively selected in the thymus; but (*iii*) it reduces the number of TCRs by negative selection. The mathematical analysis shows that $n = 1/f$ maximizes the number of different TCRs that pass through positive and negative selection and that $n = 2/f$ maximizes the probability to mount immune responses against a large fraction of foreign peptides. Here $f$ is the fraction of TCRs deleted by one MHC molecule. Both results depend on approximations that are discussed in the paper. The model presented has implications for our understanding of the evolutionary forces acting on the MHC.**

A major histocompatibility complex (MHC) has been identified in all vertebrate species so far examined. Encoded within this gene complex are molecules that compose or contribute toward MHC heterodimers that can be expressed on cell membranes (1). These molecules bind peptides derived from foreign antigens for presentation to T cells bearing $\alpha\beta$ T-cell receptors (TCRs) during immune responses and are of two structurally related types, class I and class II. Many $\alpha\beta$ TCR$^+$ cytotoxic T cells are MHC class I-restricted, whereas helper T cells are predominantly class II-restricted. The MHC-encoded $\alpha$ chain of class I molecules associates with $\beta_2$-microglobulin, encoded elsewhere in the genome, whereas both the $\alpha$ and $\beta$ chains of class II molecules are encoded within the MHC. For most MHC loci, multiple alleles exist within the species (2). As a consequence of this polymorphism and of alloreactivity, the capacity of a large proportion of the T-cell repertoire of any individual to respond to foreign MHC molecules and/or bound peptides from MHC-disparate individuals of the species (3), MHC disparities give rise to transplantation reactions *in vivo* (allograft rejection and graft-versus-host disease) and the allogeneic mixed leukocyte reaction *in vitro*.

Regarding class I molecules, it is notable that in all vertebrates examined, only two or three appear to function as strong transplantation antigens, even though the MHC may contain as many as 60 class I loci per haploid genome (rats) or perhaps as few as 7 (miniature swine), despite the potential for evolutionary duplication and diversification of individual loci. To date, the best-characterized MHC complexes are in mouse and human, H-2 and HLA, respectively (Fig. 1). Of the 30 to 40 class I MHC loci identified within the H-2 complex of different mouse strains, only those within the K and D regions encode the "classical" class I molecules that are highly polymorphic, are expressed or can be induced on essentially all cell types, are implicated in physiological antigen presentation, and function as strong transplantation antigens; the comparable regions of the HLA complex are designated A, B, and C. In the mouse, only a single locus is present in the D region of some haplotypes, such as H-2b. In other haplotypes, such as H-2d, there is instead a cluster of five genes, but only two loci, D and L, are expressed; the remaining three loci do not encode antigen-presenting molecules and have no known function. In some haplotypes an additional K locus, K2, has been defined, but this locus represents a pseudogene in H-2k mice, for example, and its status is uncertain in other strains. Therefore, completely heterozygotic $F_1$ mice express four to six different classical class I molecules—i.e., maternal and paternal alleles of K, D, and L, where present. The same is true for humans, where, in fact, there are few examples of T cells reactive to C-locus products.

Most mouse class I loci encode the relatively nonpolymorphic "nonclassical" Qa and Tla products (4), only some of which are detectable serologically, that are expressed in an apparently cell-restricted or tissue-specific manner, sometimes as phosphatidylinositol-linked membrane molecules or as secreted products. Homozygous mice of different strains express from ≈20 to 30 Qa and Tla class I loci, but because these loci are considerably less polymorphic than the classical genes, heterozygous mice express fewer than the theoretical maximum of 40–60 molecules.

Homologues of Qa and Tla products have yet to be identified in the human, although three nonclassical class I molecules, HLA-E, -F, and -G, have been identified in one haplotype (8). In the mouse, further MHC-linked class I-related loci have been defined: the Hmt subfamily and the less homologous Mb1 subfamily and the CD1 family, the location of which has not been defined. Although their function is largely obscure, some of the mouse Qa, Tla, and CD1 molecules may be involved in antigen presentation to T cells expressing $\gamma\delta$ TCR, but their limited polymorphism suggests they are likely to bind limited numbers or types of antigen. To date, there is only one clear example of $\alpha\beta$ TCR recognition of a peptide associated with a nonclassical class I molecule, Tla (9). Thus, an exceedingly limited number of MHC class I loci have been selected during evolution for a role in conventional antigen presentation to $\alpha\beta$ TCR$^+$ T cells.

A similar situation pertains to class II loci. The MHC class II region of the mouse, H-2I, encodes the A$\alpha$, A$\beta$, E$\alpha$, and E$\beta$ chains, the respective chains of which associate to form the IA and IE isotypes. Even so, some inbred mouse strains and ≈20% of wild mice do not express IE molecules due to defective or deleted E$\alpha$ or E$\beta$ loci (IE$^-$ strains). The $\alpha$ and $\beta$ loci (designated A and B) known to be expressed in the class

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.

Abbreviations: TCR, T-cell receptor; MHC, major histocompatibility complex.

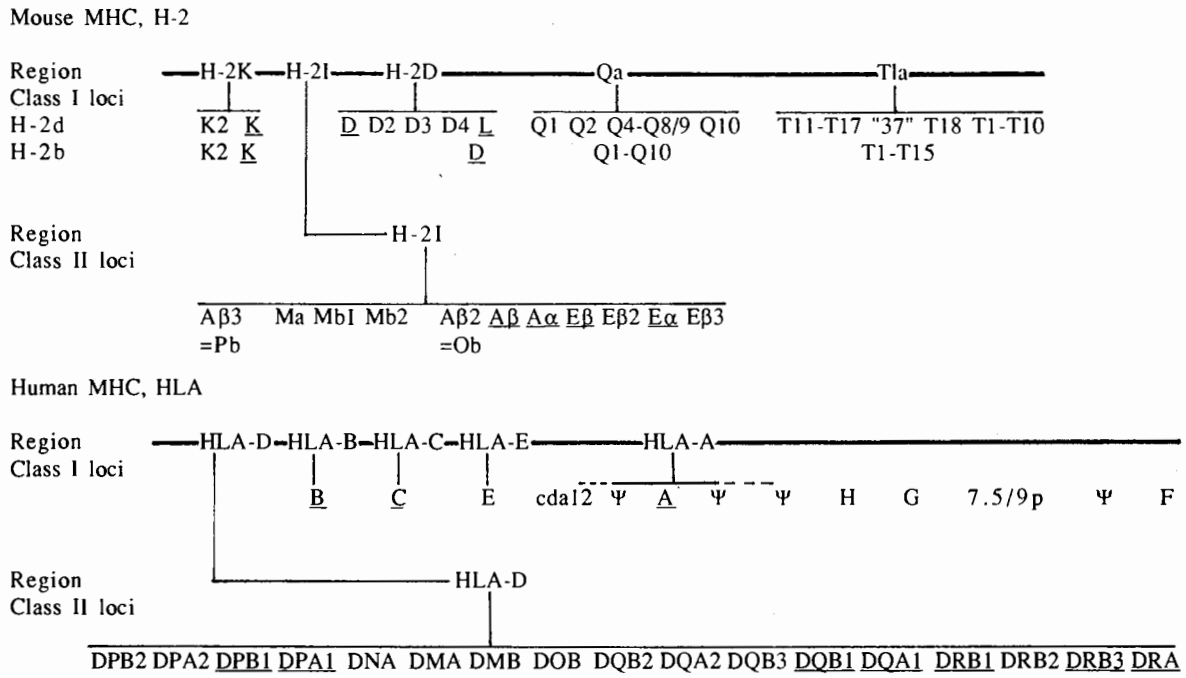


FIG. 1. Class I and class II loci of mouse and human MHC, H-2 and HLA. For each species, the MHC regions are indicated, and the order of class I loci is shown above that of the class II loci (1, 4–7). The classical class I loci and class II loci encoding molecules known to be involved in antigen presentation to T cells are underlined. Class I loci of two strains of mice are illustrated: the classical loci of the H-2b haplotype are similar to k, and those of d are similar to q, r, and v. In addition to the three nonclassical HLA class I loci that can be expressed (*HLA-E*, *-F*, and *-G*), several pseudogenes have been mapped (e.g., *HLA-H* and Ψ). Known class II pseudogenes or gene fragments are mouse Aβ3 (*Pb*) and human *DPA2*, *DPB2*, *DQB3*, and *DRB2*; as yet no translation products have been identified from DNA and DNB transcripts, and the *DQA2* and *DQB2* genes are not known to be expressed. *DMA* and *DMB* correspond to the *RING6* and *RING7* loci (5). The following is not shown: the mouse class I-related loci of the Hmt and Mb1 subfamilies and the CD1 family (4); two human class I-like loci, 1.7p mapping to the left of HLA-B, and HLA-X to the left of HLA-E (7); the human class II loci DRB4 and DRB5, either of which can be expressed instead of DRB3 in some haplotypes, although in others all three are absent; and a pseudogene mapping between the DRB loci and DRA. The maximum number of known classical class I and class II MHC molecules that could be expressed in homozygotic and heterozygotic ($IE^+$) mice is a total of 7 and 18, respectively; the corresponding numbers for humans are 15 and 43. It is important to note that these values do not take into account the recently defined class II-related genes (see text). These numbers should be compared with the computed values in Table 1.

II region of the human MHC, HLA-D, are *DPA1* and *DPB1* (encoding the DP molecule); *DQA1* and *DQB1* (for DQ); and *DRA*, *DRB1*, and *DRB3* or *-4* or *-5* if present (for one or two DR molecules) (10). The expression of class II molecules is more complicated than that of class I because of the potential for heterodimer formation between $\alpha$ and $\beta$ chains of different haplotypes and/or isotypes and because the human *DRA* and mouse E$\alpha$ loci are essentially invariant. The predominantly expressed class II molecules are probably isotype-matched as well as allele-matched. Allele-mismatched combinations have been detected functionally, using T-cell clones that recognize unique $F_1$ determinants, as well as biochemically on the surface of transfected cells, but these are rare and their formation depends on the haplotypes involved. Mixed isotypes have also been detected in gene-transfected and tumor cell lines, but the general consensus is that these molecules are, at best, of only minor physiological significance for antigen presentation by normal cells.

The combinations of different class II molecules that could potentially be expressed in heterozygotic ($IE^+$) mice are as follows: haplotype (maternal, m, or paternal, p)-matched and isotype (A or E)-matched ($A^m_\alpha A^m_\beta$, $A^p_\alpha A^p_\beta$) = 2; haplotype mismatched and isotype-matched ($A^m_\alpha A^p_\beta$, $A^p_\alpha A^m_\beta$) = 2; haplotype-matched and isotype-mismatched ($A^m_\alpha E^m_\beta$, $A^p_\alpha E^p_\beta$) = 2; haplotype-mismatched and isotype-mismatched ($A^m_\alpha E^p_\beta$, $A^p_\alpha E^m_\beta$) = 2; invariant $E_\alpha$ with isotype-matched or isotype mismatched $\beta$ chains ($E_\alpha E^m_\beta$, $E_\alpha E^p_\beta$, $E_\alpha A^p_\beta$, $E_\alpha A^m_\beta$) = 4. Thus heterozygotic mice could potentially express a maximum of 12 of these MHC class II molecules (4 in $IE^-$ strains), and homozygous mice could express 4 ($IE^+$ strains) or 2 ($IE^-$ strains) molecules. In humans a similar analysis leads to the following results for individuals heterozygotic at all loci: haplotype-matched and isotype (P or Q)-matched = 4; haplotype-mismatched and isotype-matched = 4; haplotype-matched and isotype-mismatched = 12; invariant DR$\alpha$ with isotype-matched DR$\beta$1 chains or with isotype-mismatched $\beta$ chains (i.e., DP$\beta$ or DQ$\beta$) = 6, this total increasing to 8 in individuals that also express one of the DR$\beta$3, -4, or -5 products. Thus, heterozygotic humans could potentially express a maximum of 40 of these MHC class II molecules (12 in homozygotes).

From the above, between 4 and 12 class II molecules could potentially be expressed in completely heterozygotic $IE^+$ mice (only 2–4 in $IE^-$ strains) and 8–40 in humans, but the physiologically relevant number is thought to be much closer to the lower estimate. An additional set of MHC class II-related genes has also been defined (5, 6), the mouse *Ma*, *Mb1*, and *Mb2*, and the human *RING6* and *RING7* (*HLA-DMA* and *-DMB*) loci (Fig. 1). These loci could encode additional cell-surface heterodimers, but their function awaits definition. Furthermore, an unusual H-2 class II product, H-2O, containing the O$\beta$ (formerly A$\beta$2) chain paired to a specific O$\alpha$ chain, is expressed in thymic medulla and peripheral B cells (11), which also express the E$\beta$3 chain-encoding gene, but the function of these molecules is likewise unclear at present. In addition, the class II regions of both mouse and human contain a number of other loci and homologous gene segments that cannot be, or are not known to be, expressed (Fig. 1 legend), and this situation is likely to pertain to other species.

Because of their role in antigen presentation, it might seem advantageous for an individual to express more, rather than

fewer, MHC molecules to increase the probability of presenting any foreign peptide to T cells and, hence, of inducing an immune response to a particular pathogen. In addition, the intrathymic development and expansion of $\alpha\beta$ TCR$^+$ T cells apparently requires interactions with self-MHC molecules alone, in a process termed positive selection, the mechanism of which is poorly understood; hence, an increase in the number of different MHC molecules expressed would presumably increase the diversity of the T-cell repertoire and thus, also, the probability of recognizing foreign peptides. However, as discussed above, the number of MHC molecules expressed in any individual is tightly constrained. It seems probable that this constraint is from the capacity of MHC molecules to render portions of the T-cell repertoire unavailable or unusable during tolerance induction (related arguments have appeared elsewhere—e.g., see ref. 2).

The capacity of MHC molecules to bind peptides from normal self-components necessitates the existence of mechanisms, collectively termed negative selection, to prevent autoreactivity. In the mouse, at least two such mechanisms have been demonstrated, clonal deletion and functional inactivation or clonal anergy (12), which result in negative selection of potentially autoreactive $\alpha\beta$ TCR$^+$ T cells (and there is some evidence that $\gamma\delta$ TCR$^+$ T cells can be negatively selected by Tla molecules; ref. 13).

Therefore, were the number of MHC molecules expressed by any individual to increase, more self-peptides would be presented, and this presentation would result in deletion or inactivation of more of the total T-cell repertoire. This situation, in turn, would reduce the potential for inducing an immune response to any given organism.

The following mathematical analysis allows an estimation of the optimal number of different MHC molecules expressed per individual in order to achieve the maximum capacity to present and recognize foreign peptides with the minimum loss of functional T cells through negative selection.

Let $T_0$ denote the total potential number of different TCRs that can be generated in the thymus before positive or negative selection. Let $h$ be the fraction of (T cells bearing) TCRs that is positively selected by one MHC molecule. The number of different TCRs that are positively selected by $n$ different MHC molecules is then given by $T_0[1 - (1 - h)^n]$. Let $f$ denote the fraction of (positively selected T cells bearing) TCRs that is deleted, anergized, or otherwise rendered tolerant by one MHC molecule. The fraction of TCRs remaining after negative selection by $n$ MHC molecules is $(1 - f)^n$. The total number of TCRs that pass through positive and negative selection exerted by $n$ different MHC molecules is then given by

$$T_n = T_0[1 - (1 - h)^n](1 - f)^n. \quad [1]$$

$T_n$ has a maximum for

$$n = -\frac{1}{\log(1-h)}\log\left(1 + \frac{\log(1-h)}{\log(1-f)}\right). \quad [2]$$

This maximum represents the largest possible TCR repertoire in the periphery (after positive and negative selection). Assuming $\log(1 - h) \approx -h$ and $\log(1 - f) \approx -f$, we obtain

$$n = \frac{1}{h}\log\left(1 + \frac{h}{f}\right). \quad [3]$$

For $h >> f$ this is approximately $n = (\log h - \log f)/h$; for $h = f$ we have $n = \log 2/f$; and for $h << f$ we have $n = 1/f$. It is tempting to assume that $h$ is smaller than $f$ because only a small fraction of the randomly generated TCRs may be positively selected by a given MHC molecule, but a larger fraction of those TCRs that are positively selected may be deleted afterward because of autoreactivity against MHC plus self-peptide. If this assumption is correct, then $n = 1/f$ maximizes the variety of TCRs that pass through positive and negative selection to compose the functional TCR pool in the periphery. This is an interesting result.

Table 1. The optimal number of different MHC molecules per individual, $n$, to maximize the probability $P$ in Eq. **4**

| | $f$* | | | |
|---|---|---|---|---|
| $q$* | 0.2 | 0.1 | 0.05 | 0.02 |
| $10^{-6}$ | 12 | 29 | 72 | 229 |
| $10^{-8}$ | 9 | 19 | 39 | 102 |
| $10^{-10}$ | 9 | 19 | 39 | 99 |

The optimal number maximizes the size of the repertoire of foreign peptides that can be both presented by MHC molecules and recognized by TCRs. This maximum arises from the consideration that more MHC molecules can expand the T-cell repertoire (positive selection) and present a larger variety of foreign peptides but will, on the other hand, delete too many TCRs (negative selection).

*Parameters $q$ and $f$ denote, respectively, the probability that (a T cell bearing) a given TCR binds to an arbitrary peptide–MHC complex, and the fraction of TCRs deleted (or rendered tolerant) by one MHC molecule. For this numerical example, other parameter values were chosen as $T_0 = 10^{10}$, $h = 0.0001$, and $p = 0.001$.

In the following we do not use any of the above approximations but proceed directly with Eq. **1**. Let $p$ denote the fraction of all foreign peptides that can bind to a single MHC molecule. If there are $n$ different MHC molecules, then a given peptide will bind to exactly $i$ of them with the probability $\binom{n}{i}p^i(1 - p)^{n-i}$. The probability that at least one of these complexes will be recognized by at least one out of the $T_n$ different TCRs in the periphery is given by $1 - (1 - q)^{iT_n}$. Here $q$ is the probability that a given TCR binds to a given peptide–MHC complex. We can now express the probability that a given foreign peptide will be presented by an MHC molecule and that the resulting complex will be recognized by a TCR, as

$$P = \sum_{i=1}^{n}\binom{n}{i}p^i(1-p)^{n-i}[1 - (1 - q)^{iT_n}]. \quad [4]$$

To derive Eqs. **1** and **4** we have assumed that the parameters $h, f$, and $p$ are the same for all different MHC molecules under consideration. This assumption may be an approximation because different MHC molecules probably can, for example, select or delete TCRs with different efficiency, and some may bind more peptides than others (14). In this context the parameters $h, f$, and $p$ represent average quantities.

The sum in Eq. **4** can be calculated exactly. We obtain $P = 1 - [1 - p + p(1 - q)^{T_n}]^n$. This equation describes a one-humped curve for the probability $P$ as a function of $n$. We are interested in the value of $n$ that maximizes $P$. Table 1 shows this optimal value of $n$ for a number of different parameter values $q$ and $f$. In this analysis, $n$ represents the total number of class I and class II molecules because these molecules select TCRs from a common pool.

To derive an analytic result for the numeric values we need some approximations. First, we may assume that the product $pn$ is well below 1. Note that $pn$ is the average number of different MHC molecules that bind to a given peptide. If $pn << 1$, then it is unlikely that the same peptide will be presented by more than one MHC molecule. In this case we can write

$$P \approx pn[1 - (1 - q)^{T_n}]. \quad [5]$$

It seems also reasonable to assume that $hn$ is much smaller than 1. Note that $hn$ is the average number of different MHC molecules that positively select one given TCR. If $hn << 1$,

then it is rare that a given TCR is positively selected by more than one MHC molecule. We obtain from Eq. **1**:

$$T_n \approx T_0 hn(1-f)^n. \quad [6]$$

From $\frac{\partial P}{\partial n} = 0$ (using Eqs. **5** and **6**) we obtain that the optimal number of different MHC molecules per individual is indirectly proportional to $f$, the fraction of TCRs deleted by a single MHC molecule in the thymus:

$$n = \frac{k}{f}, \quad [7]$$

where $k = 1 + [(1-q)^{-T_n} - 1]/q^{T_n}$. The factor $k$ still depends on $n$, so Eq. **7** is transcendental in $n$. Using the approximation $\log(1-q) \approx -q$, we can write $k = (e^{qT_n} - 1)/q^{T_n} + 1$. The factor $k$ is close to 2 for small values of the product $qT_n$; for example, we have $k = 2.05$ for $qT = 0.1$ and $k = 2.71$ for $qT = 1$. The product $qT_n$ is the average number of different TCRs that bind to (and hence permit an effective immune response against) a given peptide–MHC complex. If this product is small, we find

$$n = 2/f. \quad [8]$$

This result is rather surprising and simple; the optimal number of MHC molecules is simply twice the reciprocal of the fraction of TCRs deleted by a single MHC molecule. This approximation should be compared with the computed values in Table 1.

Thus, the crucial parameter of interest is the fraction $f$. Alloreactivity experiments lead to the crude estimate that 1–7% of the T-cell repertoire is reactive to multiple MHC disparities (refs. 15–17 and the references therein). This result may suggest a rather low value of $f$ and, therefore, a large value for the optimal $n$. It is not quite clear whether estimates based on alloreactivity can reflect the intrathymic situation. Another possibility may be to estimate the fraction $f$ from *in vitro* experiments (18, 19).

An important question is whether natural selection has resulted in an immune system with a number of MHC molecules close to the optimum, as calculated in our model. Is there a strong selective force to mount immune responses against all possible foreign peptides, or is it sufficient instead to recognize one or a few epitopes on a given pathogen? It may be very costly to have a very large number of different MHC molecules on every single cell in an animal. Our estimate gives the optimal number of different MHC molecules that maximizes the probability of recognizing a foreign peptide. This result may represent an upper limit of what we would expect to find in nature.

We thank Robert May, Alan Perelson, Geoffrey Butcher, Robert Lechler, and John Todd for helpful discussions. M.A.N. is the Guy Newton Junior Research Fellow at Wolfson College, Oxford, and a Wellcome Trust Senior Research Fellow in Basic Biomedical Science.

1. Guillemot, F., Auffray, C., Orr, H. T. & Strominger, J. (1988) in *Molecular Immunology*, eds. Homes, B. D. & Glover, D. M. (IRL Press, Oxford), pp. 81–143.
2. Lawlor, D. A., Zemmour, J., Ennis, P. D. & Parham, P. (1990) *Annu. Rev. Immunol.* **8,** 23–63.
3. Lechler, R. I., Lombardi, G., Batchelor, J. R., Reinsmoen, N. & Bach, F. H. (1990) *Immunol. Today* **10,** 83–88.
4. Stroynokski, I. (1990) *Annu. Rev. Immunol.* **8,** 501–530.
5. Kelly, A. P., Monaco, J. J., Cho, S. & Trowsdale, J. (1991) *Nature (London)* **353,** 571–573.
6. Cho, S., Attaya, M. & Monaco, J. J. (1991) *Nature (London)* **353,** 573–576.
7. Trowsdale, J., Ragoussis, J. & Campbell, R. D. (1991) *Immunol. Today* **12,** 443–446.
8. Heinrichs, H. & Orr, H. T. (1990) *Immunol. Res.* **9,** 265–274.
9. Milligan, G. N., Flaherty, L., Braciale, V. L. & Braciale, T. J. (1991) *J. Exp. Med.* **174,** 133–138.
10. Bodmer, J. G., Marsh, S. G. E., Albert, E. D., Bodmer, W. F., Dupont, B., Erlich, H. A., Mach, B., Mayr, W. R., Parham, P., Sasazuki, T., Schreuder, G. M. T., Strominger, J. L., Svejgaard, A. & Terasaki, P. I. (1991) *Hum. Immunol.* **31,** 186–194.
11. Karisson, L., Surh, C. D., Sprent, J. & Peterson, P. A. (1991) *Nature (London)* **351,** 485–488.
12. Mueller, D. L., Jenkins, M. K. & Schwartz, R. H. (1989) *Annu. Rev. Immunol.* **7,** 445–480.
13. Dent, A. L., Matis, L. A., Hooshmand, F., Widacki, S. M., Bluestone, J. A. & Hedrick, S. M. (1990) *Nature (London)* **343,** 714–719.
14. Rothbard, J. B. & Gefter, M. L. (1991) *Annu. Rev. Immunol.* **9,** 527–566.
15. Wilson, D. B., Blyth, H. & Nowell, P. C. (1968) *J. Exp. Med.* **128,** 1157–1181.
16. Ford, W. L., Simmonds, S. J. & Atkins, R. C. (1975) *J. Exp. Med.* **141,** 681–696.
17. Bevan, M. J. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 2094–2098.
18. Swat, W., Ignatowicz, L., Boehmer, H. & Kisielov, P. (1991) *Nature (London)* **351,** 150–153.
19. Boehmer, H. & Kisielow, P. (1990) *Science* **248,** 1369–1373.