Toward a Quantitative Analysis of the Repertoire of Class I MHC-Specific Inhibitory Receptors on Natural Killer Cells

R.E. VANCE and D.H. RAULET

Department of Molecular and Cell Biology, Division of Immunology, and Cancer Research Laboratory, 489 Life Sciences Addition, University of California at Berkeley, Berkeley, CA 94720, USA

K. Kärre et al. (eds.), *Specificity, Function, and Development of NK Cells*

1 Introduction

Recent research establishes that natural killer (NK) cells recognize class I MHC molecules on potential target cells. Unlike T cells, however, recognition of target cell class I molecules by NK cells inhibits their activation and prevents destruction of the target cell. The pattern of recognition exhibited by NK cells suggests that one of their key functions is to destroy self cells that have extinguished or reduced expression of some or all class I molecules.

Deficiency in the expression of all class I molecules, due to mutation of β_2-microglobulin or genes that regulate class I biosynthesis, typically renders cells sensitive to lysis by NK cells (KARRE et al. 1986; LIAO et al. 1991). However, deficiency in the expression of a subset of the cell's MHC class I molecules, rather than all of them, can also render a cell sensitive to NK cells. For example, NK cells from an F_1 ($MHC^{a/b}$) mouse can often destroy parental ($MHC^{a/a}$) cells (BENNETT 1987; CHADWICK and MILLER 1992; CUDKOWICZ and STIMPFLING 1964). A comparable situation has been generated experimentally by creating H-2^b mice transgenic for the D^d class I gene (OHLEN et al. 1989). NK cells from these mice destroy nontransgenic H-2^b target cells. NK cells also often lyse fully allogeneic cells (BENNETT 1987). In all three of these situations the target cell is missing some or all of the class I molecules expressed by the host. These patterns of lysis have led to the "missing-self" hypothesis, which postulates that NK cells destroy target cells that lack some or all class I molecules of the host (LJUNGGREN and KARRE 1990). There are exceptions to this model, in its simplest form, but these exceptions do not negate the model although they do suggest additional complexities to the system.

What is the molecular basis of the missing-self model? Recent work suggests that NK cell recognition is controlled by integrating signals from both activating and inhibitory receptors. With the exception of the NK cell's Fc receptor the activating receptors are poorly characterized and may include members of the NKR-P1 receptor family (RYAN and SEAMAN 1997). In contrast the inhibitory receptors are increasingly well understood and provide an explanation for missing-self recognition. In mice the Ly-49 family of receptors bind class I MHC molecules and thereby inhibit NK cell activity (BROWN et al. 1997; GEORGE et al. 1997; TAKEI et al. 1997). These receptors are dimeric type II membrane proteins that contain a C-terminal carbohydrate recognition domain and comprise a family of approximately ten receptors designated Ly-49A to Ly-49I, encoded by closely linked genes on chromosome 6 in the mouse. The specificities of only a few of the receptors have been investigated. The data suggest that Ly-49A reacts with D^d and D^k, Ly-49C with K^b and D^d, and Ly-49G2 with D^d and L^d. The available evidence suggests that the engagement of Ly-49 receptors inhibits activation mediated by several types of stimulating receptors, including putative NK receptors for tumor cell specific ligands, the NK cell's Fc receptor, and the T cell antigen receptor (CORREA et al. 1994; HELD et al. 1996a; KARLHOFER et al. 1992).

2 The Ly-49 Receptor Repertoire

Monoclonal antibodies reactive with at least four Ly-49 receptors have been generated (BROWN et al. 1997; GEORGE et al. 1997; MASON et al. 1995; ROLAND and CAZENAVE 1992; TAKEI et al. 1997). The JR9-318 and A1 antibodies, among others, bind Ly-49A; the SW-5E6 mAb binds Ly-49C and Ly-49I; and the 4D11 mAb reacts with Ly-49G2. With some of these antibodies the distribution of the corresponding receptors on different NK cells has been investigated and reveals a complex expression pattern. As depicted in Table 1, from 15%–60% of NK cells can react with a given anti-Ly-49 monoclonal antibody (RAULET et al. 1997). These and other analyses have permitted three important conclusions to be made: (a) NK cells commonly express Ly-49 receptors that are apparently irrelevant for the animal in the sense that they fail to react detectably with the host's class I MHC molecules; (b) NK cells commonly coexpress two or more Ly-49 receptors; and (c) unlike the T and B cell receptors, the distribution of Ly-49 receptors to different cells may not involve somatic gene recombination since normal Ly-49 expression occurs in mice harboring mutations in the recombination machinery, and no Ly-49 gene rearrangements have been detected in mature NK cell populations.

The distribution of Ly-49 receptors to different NK cell subsets underlies the capacity of NK cells to discriminate class I *different* target cells, as opposed to the

Table 1. Expression of Ly-49 receptors in MHC-different mice: the "product rule"

	Percentage of NK cells in MHC background					
	B10.D2 (H-2^d)		B6 (H-2^b)		B6-$\beta_2m^{-/-}$	
NK cell subset[a]	Observed[b]	Expected by product rule[c]	Observed[b]	Expected by product rule[c]	Observed[b]	Expected by product rule[c]
Ly-49A$^+$	15.7	–	17.9	–	25.3	–
SW-5E6$^+$	47.9	–	44.5	–	63.5	–
4D11$^+$	43.6	–	48.9	–	56.4	–
Ly-49A$^+$ SW-5E6$^+$	5.0	7.5	5.9	8.0	14.6	16.1
Ly-49A$^+$ 4D11$^+$	5.1	6.8	9.9	8.8	18.7	14.3
SW5E6$^+$ 4D11$^+$	21.1	20.9	21.6	21.8	40.0	35.8
Ly-49A$^+$ SW-5E6$^+$ 4D11$^+$	1.7	3.3	4.7	3.9	13.3	9.1

[a] Refers to cells that express the indicated receptor regardless of whether they express other receptors. 4D11 reacts at least with Ly-49G2; SW-5E6 reacts at least with Ly-49C and Ly-49I.

[b] Values are averages of percentages from at least four determinations. Data derived from HELD et al. 1996b.

[c] Calculated by multiplying the observed component frequencies, e.g., %Ly-49A+ SW5E6+= (0.157×0.479)×100.

simpler task of detecting class I *deficient* target cells. Consider the destruction of $H\text{-}2^{a/a}$ target cells by $H\text{-}2^{a/b}$ NK cells. $H\text{-}2^{a/b}$ mice can harbor various NK cell subsets, including those with inhibitory Ly-49 receptors for $H\text{-}2^{a}$ and not $H\text{-}2^{b}$, a separate set with inhibitory receptors for both $H\text{-}2^{a}$ and $H\text{-}2^{b}$, and a third with receptors for $H\text{-}2^{b}$ and not $H\text{-}2^{a}$. It is the latter set that reject $H\text{-}2^{a/a}$ target cells (GEORGE et al. 1997). The importance of subset-specific expression of Ly-49 receptors was demonstrated by experiments in which a transgene encoding the D^{d}-specific Ly-49A receptor was expressed in all NK cells, as opposed to a subset as observed in normal mice. This manipulation abolished the capacity of NK cells from $H\text{-}2^{b}$ mice to destroy $H\text{-}2^{d}$ cells while not diminishing the capacity to destroy class I deficient target cells (HELD et al. 1996a).

2.1 The Product Rule

Coexpression of Ly-49 receptors is common and results in a complex combinatorial repertoire. Functional studies suggest that NK cells that coexpress a particular pair of Ly-49 receptors can be inhibited independently through either receptor (MASON et al. 1995; YU et al. 1996). Interestingly, the distribution pattern suggests that the expression of one receptor is to some extent independent of the expression of other receptors. Thus the fraction of NK cells reacting with any two anti-Ly-49 antibodies is roughly equal to the product of the fractions of cells reacting with each of the antibodies alone (Table 1). We have called this the "product rule," and it appears to be obeyed, to a first approximation, regardless of the MHC background of the mouse (RAULET et al. 1997). The product rule is consistent with the possibility that a stochastic process underlies the Ly-49 receptor distribution mechanism. However, as discussed at length below, there is evidence that the receptor distribution pattern is not entirely stochastic. In fact it is clear that the representation of different subsets is influenced by host MHC molecules (HELD et al. 1996b). While the distribution pattern may be created by mechanisms that incorporate stochastic components, evidence suggests that it is influenced by an MHC-dependent "education" process.

2.2 Monoallelic Expression of Ly-49 Receptors

Interestingly, Ly-49 receptors, similar to T cell receptors, B cell receptors, and odorant receptors, are expressed in a predominantly monoallelic fashion (HELD et al. 1995). This has been most clearly demonstrated in the case of the Ly-49A locus. The limited Ly-49A sequence differences in the BALB versus B6 strains (three amino acid differences in the extracellular domain) results in discrimination of these two proteins by one of the Ly-49A specific monoclonal antibodies. It was demonstrated that $Ly\text{-}49A^{+}$ NK cells in the $(B6{\times}BALB.B)F_1$ Ly-49A heterozygote consist of an approximately equal number of cells expressing $Ly\text{-}49A^{B6}$ and cells expressing $Ly\text{-}49A^{BALB}$. Monoallelic Ly-49A expression is imposed at the level of mRNA abundance, probably at the level of transcription. Evidence was also presented that the SW-5E6 antigen is expressed in a monoallelic fashion (HELD et al. 1995). Based

on recent studies that have subdivided the SW-5E6$^+$ subset into Ly-49C$^+$ and Ly-49I$^+$ cells (BRENNAN et al. 1996), the earlier results can now be interpreted to suggest that either Ly-49I or Ly-49C is expressed in a monoallelic fashion; the available data do not distinguish between these possibilities (RAULET et al. 1997). It appears likely that all members of the family are expressed in a predominantly monoallelic fashion.

Monoallelic Ly-49 gene expression may arise as a consequence of the mechanism that distributes expression of different Ly-49 genes to overlapping NK cell subsets. This mechanism may treat different alleles at the same locus independently just as it treats different loci independently. Consistent with this possibility, the data suggest that the choice of active allele occurs independently at different Ly-49 loci. A cell can express Ly-49A from one chromosome, and Ly-49I from the other (HELD et al. 1995). Another relevant observation, from analyses of short-term clones of NK cells from Ly-49A heterozygous mice, is that monoallelic expression of Ly-49A genes is incomplete. Approximately 90% of the Ly-49A$^+$ clones expressed only one or the other Ly-49A allele at nearly equal frequency. However, approximately 10% of the Ly-49A$^+$ clones expressed both Ly-49A alleles. Thus approximately 10% of all NK cells expressed one Ly-49A allele, 10% expressed the other, and 1%–2% expressed both Ly-49A alleles (Held et al. 1997b; RAULET et al. 1997). These percentages are in keeping with the product rule and suggest that the two Ly-49A alleles are activated independently. Therefore just as there is predictable overlap in the expression of different Ly-49 family members, there is predictable overlap in the expression of different Ly-49A alleles. Hence both the subset distribution and monoallelic expression of Ly-49 genes could be the result of the same stochastic mechanism in which each allele at each Ly-49 locus is conferred with a fixed probability of expression in each progenitor NK cell. As expected by this model, transgenic expression of Ly-49A on all NK cells does not fully suppress endogenous Ly-49A expression (HELD and RAULET 1997a).

2.3 NK Cells Acquire Self Specificity Somatically

Although a stochastic mechanism may underlie the distribution of Ly-49 receptors to different cells, such a system by itself would inevitably lead to the generation of autoaggressive NK cells. This is because Ly-49 genes are not linked to the MHC, and therefore are not coordinately inherited with class I alleles (BROWN et al. 1997). Furthermore, some of the receptors are expected to be non-self-specific in mice of most if not all MHC types. If the receptors were distributed to different NK cells by a purely stochastic process, some of the resulting clones would fail to express self class I specific receptors. Such clones would be expected to be autoaggressive. However, it has been observed that NK cells, at least those that have not been cultured extensively in IL-2, are generally self tolerant, meaning that they are generally inhibited better by self class I MHC molecules than foreign class I molecules (CHADWICK and MILLER 1992; DORFMAN and RAULET 1996; GEORGE et al. 1997). Experiments demonstrating that NK cell functional specificity can adapt to the presence of a class I transgene (OHLEN et al. 1989) as well as to mutations that confer class I deficiency (BIX et al. 1991) strongly suggest that the self specificity of NK

cells is acquired somatically. This conclusion is further supported by the results of bone marrow chimera experiments (HOGLUND et al. 1991; WU and RAULET 1997), which demonstrate that class I^+ NK cells differentiating in the presence of class I deficient cells are rendered tolerant of the latter cells (WU and RAULET 1997).

Two general theories have been proposed to explain the acquisition of NK cell self specificity. One emphasizes mechanisms that ensure that each functional NK cell expresses at least one self class I specific Ly-49 receptor. This theory is discussed further below. The other theory emphasizes the quantitative effects of expressing different cell surface levels of the Ly-49 receptors. The latter model overlaps to some extent with the first model and is based on the observation that the cell surface levels of Ly-49 receptors vary with the MHC type of the host (KARLHOFER et al. 1994; OLSSON et al. 1995). In H-2^d mice, which express a Ly-49A ligand, Ly-49A^+ NK cells exhibit lower levels of Ly-49A per NK cell than is observed in H-2^b mice or class I deficient mice, which express no known ligand. The magnitude of this effect varies in different studies, from at least twofold to more than tenfold. This phenomenon led to the proposal that the levels of Ly-49 receptors are "calibrated" against the expressed class I molecules, increasing Ly-49 cell surface levels in order to increase sensitivity of the NK cell to weak class I ligands, and vice versa (OLSSON et al. 1995). Depending on the cross-reactivity of different Ly-49 receptors with different class I molecules and their distribution pattern this mechanism could result in each NK cell having at least one productive inhibitory interaction with any set of self class I molecules that the animal happens to express.

2.4 Critique of the Calibration Model

The notion that "calibration" of cell surface Ly-49 levels is in fact responsible for determining the self tolerance of NK cells is not easily reconciled with some recent observations. In fact, these findings appear most consistent with the hypothesis that receptor downregulation is an incidental consequence of ligand-induced receptor internalization or shedding (HELD and RAULET 1997). First, Ly-49A downregulation in normal mice is not accompanied by a decrease in the levels of Ly-49A mRNA per Ly-49A+ cell, indicating that receptor downregulation occurs posttranscriptionally perhaps at the protein level. Accordingly, ligand-induced Ly-49A downregulation occurs even with a transgenic Ly-49A receptor that is driven by heterologous regulatory elements. Moreover, in a Ly-49A transgenic line where the cell surface levels of Ly-49A are low to begin with, the presence of the ligand results in further downmodulation of the receptor. In light of these results it is difficult to argue that the levels are adjusted to a specific level to optimize the sensitivity of the cells to specific class I ligands. Rather it appears that ligand engagement results in receptor downregulation compared to whatever level pertains in the absence of the ligand. These reductions in receptor levels may alter the functional specificity of NK cells for class I molecules, but it does not appear likely that NK cells calibrate receptor levels to a *specific* level dependent on the available class I molecules. To our way of thinking, mechanisms that determine the distribution of Ly-49 receptors to functional NK cell subsets can better account for NK cell self specificity and are thus a focus of this review.

2.5 Education Processes Determine NK Cell Specificity

Our favored hypothesis to explain the acquisition of NK cell self specificity emphasizes which receptors are expressed rather than the levels of each receptor and invokes an education process that ensures that each NK cell expresses at least one type of self class I specific receptor (RAULET et al. 1997). This conclusion seems to be the simplest explanation for the results of experiments that investigated the specificity of Ly-49A$^+$ NK cells from H-2^d mice, which express a class I ligand for Ly-49A, compared to those from H-2^b mice, which do not express a known ligand (DORFMAN and RAULET 1996; OLSSON et al. 1995). Neither population lysed H-2^d target cells. However, compared to the H-2^d-derived Ly-49A$^+$ NK cells, the H-2^b derived Ly-49A$^+$ NK cells were also diminished in their capacity to lyse H-2^b target cells. The poor lysis of H-2^b target cells was due to inhibition of the effector cells by H-2^b encoded class I molecules because these effector cells lysed class I deficient lymphoblasts efficiently. Thus these Ly-49A$^+$ NK cells apparently expressed inhibitory receptor(s) for self H-2^b class I molecules. However, although the effector cells expressed Ly-49A, the H-2^b induced inhibition was apparently mediated through distinct receptors, because anti-Ly-49A F(ab')$_2$ fragments failed to block inhibition (DORFMAN and RAULET 1996). As it is well established that individual NK cells can express multiple Ly-49 receptors, it was proposed that the Ly-49A$^+$ NK cells expressed other, H-2^b specific receptors. At least some of the comparable effector cells from H-2^d mice did not express H-2^b specific receptors. It was proposed that acquisition of self class I specificity involves a requirement that each functional cell expresses at least one self-specific receptor, while tolerating expression of irrelevant receptors.

This principle has been difficult to establish by direct analysis of the Ly-49 repertoire because the specificities of at least half of the receptors are unknown, and reagents to detect cells expressing several of the receptors do not exist. Nevertheless, the frequencies of NK cells that express different Ly-49 receptors do vary depending on the MHC class I molecules expressed by the host. The remainder of this review addresses the patterns of MHC-induced changes observed in the Ly-49 repertoire, and whether these trends are consistent with specific education processes that have been proposed.

2.6 The Effect of MHC on the Sizes of Ly-49 Defined Subsets

Considering the hypothesis that NK cells should express self-specific Ly-49 receptors, we predicted that the frequency of cells expressing Ly-49A or Ly-49G2 should be higher in H-2^d mice than in H-2^b mice (Ly-49A and Ly-49G2 react with H-2^d class I molecules and do not react detectably with H-2^b class I molecules). In fact, to our surprise the opposite was true, although the effect was only marginal (HELD et al. 1996b). More substantial effects were observed when receptor overlap was examined: NK cells that express both Ly-49A and Ly-49G2, two H-2^d-specific receptors, were substantially less frequent in H-2^d mice than in H-2^b mice or in class I deficient mice (HELD et al. 1996b). Thus it seems that, on one hand, NK cells are required to express self-specific receptors, and on the other, that expression of at least some particular

self-specific receptors is disfavored. This represents a paradox that any model of Ly-49 repertoire formation must explain.

In terms of NK cell education, the data summarized above could result from a small bias against cells that express either Ly-49A or Ly-49G2, with the more substantial reduction in double-positive Ly-49A$^+$G2$^+$ cells a consequence of the product rule. Alternatively, there may be a stronger though incomplete bias against cells expressing multiple self-specific receptors, for example, Ly-49A$^+$G2$^+$ cells in H-2^d mice, with consequent smaller reductions in the frequencies of cells expressing Ly-49A or Ly-49G2. In the latter case one might expect that the observed frequency of Ly-49A$^+$G2$^+$ cells would be less than that predicted by the product rule. However, because the effects appear to be incomplete, it is difficult to discern whether the existing data are more consistent with one or the other of these schemes (RAULET et al. 1997) (Table 1). As an alternative approach to this question we examined the effects on the repertoire of an Ly-49A transgene that is expressed in all NK cells (HELD and RAULET 1997a). In H-2^d transgenic mice it was observed that the frequency of Ly-49G2$^+$ NK cells (which also express transgenic Ly-49A) was substantially and specifically reduced in the transgenic mice compared to nontransgenic, MHC-matched littermates. The magnitude of the effect was similar to the magnitude of the reduction in Ly-49A$^+$G2$^+$ cells observed in nontransgenic H-2^d mice. The transgene had no effect in class I deficient mice and had little effect in H-2^b mice. Minimally, these results suggest that the education process disfavors NK cells expressing two self-specific receptors more than it disfavors cells expressing only one or the other.

The available data suggest that a central role of the education process is to ensure that each functional NK cell expresses at least one self-specific receptor. A straightforward means to accomplish this would be a "one-step" selection process for cells with self-specific receptors from a "random" preselection repertoire. It should be noted that an identical outcome is to be expected if selection acts *against* cells that do *not* express self-specific receptors. The latter possibility could account for the results of recent bone marrow chimera experiments (WU and RAULET 1997). As these two mechanisms result in the same outcome in normal mice, they are discussed interchangeably with respect to the predicted changes in the repertoire. The important point is that these simple one-step selection models cannot easily account for the available data. Such models invariably predict that cells expressing a given receptor should be more prevalent in ligand-bearing mice, not less so, as was observed for Ly-49A and Ly-49G2.

2.7 Two Models to Account for the Establishment of the Ly-49 Repertoire

Since a single "one-step" selection model is inadequate, we have proposed two other models that can explain the disparate observations that NK cells generally are best inhibited by self class I molecules, while at the same time there is a reduction in the frequencies of cells expressing certain self-specific receptors and especially pairs of these receptors (HELD et al. 1996b; HELD and RAULET 1997; HELD et al. 1995; RAULET et al. 1997). The "selection model" invokes a two-step selection process acting on a

preformed randomly generated repertoire, wherein there is selection for cells expressing at least one self-specific receptor, and an additional (or coordinate) selection step against cells that express "too many" self-specific receptors. It is perhaps unlikely that either selection step is based on actually "counting" the number of self-specific receptors. More likely they would be based on overall Ly-49 dependent signaling, corresponding to some amalgam of the number of different self-specific receptors, their affinity for self class I molecules, and their expression levels.

The second model, the "sequential model," involves a marriage of the mechanisms that activate Ly-49 receptor genes and the education process. It proposes that Ly-49 gene expression occurs in a sequential, cumulative manner, though perhaps in a random order, with ongoing testing of the cells for reactivity against self class I molecules. When the cell achieves the expression of a "sufficient" number and quality of self-specific receptors, Ly-49 mediated signaling would act to prevent expression of any new receptor genes and perhaps induce functional maturation of the cell (though it is not necessary to postulate the latter step). This mechanism would demand that NK cells express some self-specific receptors but would prevent the development of cells with an excess of self-specific receptors. In its purest form such a mechanism is not a selective one as all the cells eventually achieve the desired properties. However, alternate versions involving selection are possible. For example, each cell may be allowed only a limited time period to activate Ly-49 receptors, such that some cells fail to achieve the expression of self-specific receptors. Such cells would be lost, deleted, or silenced in a subsequent or coordinate step.

2.8 Mathematical Modeling

Both models discussed above incorporate mechanisms that ensure self class I specificity and yet also limit the number of cells expressing multiple self-specific Ly-49 receptors. In order to provide more specific predictive information, we have worked out mathematical treatments of each model. These treatments provide more direct evidence that each model can account for the MHC and Ly-49 transgene dependent changes in the repertoire that have been observed. The mathematical modeling has been particularly important because the calculations have demonstrated that our intuitions about the behavior of our models were not always reliable. Equally important, the mathematical treatments reveal that the models can account for the data only under specific conditions, in terms of the composition of the Ly-49 repertoire and other variables that we define. Readers not interested in the detailed calculations (Sects. 3.1, 3.2, 4.1 and 4.2) can still obtain insights from the predictions of the models that follow. At present many basic features of the composition and specificity of the Ly-49 repertoire remain unknown. As knowledge of the system grows, however, the predictions of the mathematical models can be tested against observation. It is apparent that the models differ in various predictions, such as the effects of Ly-49 transgenes and knockouts. This information provides a basis for future tests of the models against each other and also against other possible models that can be envisaged. For the sake of clarity we contrast two "extreme" models of repertoire formation. However, it must be emphasized that the models are not in all

respects mutually exclusive, that evidence for one model does not rule out the other, and that there is every reason to suppose that in reality NK cells employ a combination of mechanisms.

To ease the mathematical modeling we have made several simplifying assumptions, some or all of which may turn out to be exaggerations or even incorrect, but which nevertheless allow trends to be predicted: (a) We assume that each receptor gene has an equal initial probability of being activated in an NK cell (the fact that there may be substantially fewer Ly-49A$^+$ than, for example, Ly-49G2$^+$ cells in all strains tested already suggests that this assumption may not be correct). (b) We assume that a given receptor, in a binary fashion, either binds or does not bind to self MHC class I molecules. Receptors are therefore divided cleanly into self and non-self-specific receptors, and we treat all non-self-specific receptors equivalently to each other; similarly, we treat all self-specific receptors equivalently. (c) For reasons stated in Sect. 2.4 we ignore the effect that variations in the *levels* of Ly-49 surface expression may have; here a given receptor is assumed to be either fully expressed or fully repressed. (d) We assume that the underlying mechanisms actually "count" the number of self-specific receptors; in actuality, it is likely that the mechanisms depend on overall Ly-49 dependent signaling, corresponding to some amalgam of the number of different self-specific receptors, their affinity for self class I molecules, and their expression levels. (e) We tentatively ignore the potential role of activating receptors in Ly-49 repertoire development by holding such signals as constant between the two models. (f) For simplicity, when comparing two strains we ignore receptors that cross-react with class I molecules of both strains; this makes little difference for most of the calculations, and cross-reactive receptors can be easily incorporated into the models if desired. (g) We assume that Ly-49 receptor specificity is not modified by somatic hypermutation mechanisms. Lastly, (h) we assume that once formed the repertoire is not biased by the preferential expansion of certain NK subsets.

With these assumptions trends can be predicted. It is more difficult to predict exact frequencies of the various subsets because of several uncertainties. It should be possible eventually to incorporate into the models additional variables such as differing probabilities of Ly-49 gene activation, receptor affinity, and levels. As the information base concerning Ly-49 receptor specificity and expression grows, such modifications will become increasingly relevant.

Initially we consider only a limited number of variables, which are defined in Table 2. Despite the fact that we focus on only very few variables, we demonstrate that both models are able to account for nearly all the existing quantitative data that bear on Ly-49 repertoire development.

3 Mathematical Treatment of the Selection Model

In the selection model as we have formulated it repertoire formation begins by randomly deciding whether each Ly-49 gene is expressed or not. For simplicity we assume that the probability, p, of being expressed is initially the same for all receptors,

Table 2. Definitions of variables

General variables	
n_g	Number of non-self-specific receptor genes
s_g	Number of self-specific receptors genes
t_g	n_g+s_g
n_e	Number of non-self-specific receptors expressed on a given cell
s_e	Number of self-specific receptors expressed on a given cell
t_e	n_e+s_e
R_{s1}, R_{n1}	Denote a given self-specific, or non-self-specific receptor, respectively
Selection model variables	
s_{min}	Minimum number of permitted self-specific receptors per cell
s_{max}	Maximum number of permitted self-specific receptors per NK cell
$f_i[x]$	Fraction of cells in the initial (pre-selection) population that express x
$f_f[x]$	Fraction of cells in the final (post-selection)population that express x
$f_i[s_e]$	Fraction of cells in the initial population that express s_e self-specific receptors
$f_{se}[x]$	Fraction of cells expressing s_e self-specific receptors that express x
p	Initial probability that a given Ly-49 gene is activated
q	1–p
Sequential model variables	
s_t	Target number of self-specific receptors
$f_{ne}[x]$	Fraction of *all* cells that express n_e non-self-specific receptors *and* express x

irrespective of whether they are self- or non-self-specific. (We can also consider the probability, q=1–p, of a receptor not being expressed). Consequently, after an initial phase of Ly-49 gene activation, the total number of receptors expressed by a given NK cell (t_e) might vary anywhere from 0 to the total number of germline receptor genes, with most NK cells distributed between the two extremes. If p is large, most NK cells express a large number of receptors; conversely, if p is small, most NK cells express only a few receptors.

We are most interested in predicting the behavior of variables that are commonly measured experimentally, such as the fraction of NK cells that express any one receptor (non-self-specific or self-specific) or pair of receptors. Our strategy is to calculate the predicted representation of self-specific and non-self-specific receptors separately, since different rules apply to each.

3.1 Frequencies of NK Cells Expressing Non-Self-specific Receptors According to the Selection Model

Calculating the frequencies of cells expressing non-self-specific receptors is quite simple in the selection model. The fraction of cells that express any given receptor in the repertoire before selection, i.e., in the initial repertoire, is equal to p. Since selection acts differentially only on cells expressing different numbers of self-specific receptors, and since the number of non-self-specific receptors on a given cell is assumed to be independent of the number of self-specific receptors, the representation

of a non-self-specific receptor in the final repertoire equals its initial representation. Therefore the fraction of cells expressing a given non-self-specific receptor in the final, or selected, repertoire is predicted to equal p. The non-self-specific receptor is designated R_{n1}, and the frequency of cells expressing it in the final repertoire is designated $f_f[R_{n1}]$. Thus $f_f[R_{n1}]=p$. The fraction of cells predicted to express a specific pair of non-self-specific receptors, $f_f[R_{n1}R_{n2}]$, equals p^2.

3.2 Frequencies of Cells Expressing Self-Specific Receptors in the Selection Model

The calculations for self-specific receptors in the selection model are more involved. The selection model assumes that successful progression of the NK cell to the mature compartment occurs only if it expresses some but not too many self-specific Ly-49 genes. The lower and upper limits are not currently known, and we therefore define these as variables, with s_{min} representing the minimum number of expressed self-specific genes required, and s_{max} representing the maximum.

To calculate the frequencies of cells expressing self-specific receptors it is convenient to consider separately the populations of mature NK cells that express each allowable number of self-specific receptors. From within the initially generated stochastic repertoire we first determine the fraction of the population that expresses s_{min} receptors, the fraction that expresses $s_{min}+1$ receptors etc. until we reach s_{max} receptors. Cells that express more than s_{max} or fewer than s_{min} receptors obviously need not be considered as such cells do not contribute to the final population. For each population that expresses an allowable number of self-specific receptors, we then calculate the fraction of the initially generated population that expresses any *particular* self-specific receptor. We call the self-specific receptor R_{s1}. These subpopulations can be summed and divided by the total number of cells that survive selection. This yields the desired value, the fraction of cells in the final population that express a particular self-specific receptor, $f_f[R_{s1}]$. The denominator, the fraction of all cells that survive selection, is simply the sum of the fraction of all initial cells that express s_{min}, the fraction that expresses $s_{min}+1$, etc., until we reach s_{max}, i.e., the sum of $f_i[s_e]$ for all allowable values of s_e. Using a similar strategy the fraction of cells expressing a particular *pair* of self-specific receptors, $f_f[R_{s1}R_{s2}]$, can be calculated.

An example will help illustrate the calculations (Table 3). Consider the case in which there are a total of eight Ly-49 genes encoded in the genome, of which four are specific for self MHC class I. Assume, as well that the probability of initially activating any particular Ly-49 gene is 50%, and that the minimum number of self-specific receptors required by the selection process is 1 and the maximum is 2. In this case therefore $t_g=8$, $s_g=4$, $n_g=4$, $p=0.5$, $q=0.5$, $s_{min}=1$, and $s_{max}=2$. Consider first the population of cells expressing any two, and only two, self-specific receptors, i.e., $s_e=2$. The fraction of these cells in the initial population is equal to the fraction of all cells that express any *specific* pair, i.e., 1/16, times the number of possible pairs of receptors, 6; this corresponds to $f_i[s_e=2]=0.375$. By similar reasoning, $f_i[s_e=1]=0.25$. The remaining 0.375 of the cells fail to be selected.

Table 3. Example calculations[a] for the two-step selection model

s_e	$f_i[s_e]$	$f_{se}[R_{s1}]$	$f_i[s_e]\cdot f_{se}[R_{s1}]$	$f_{se}[R_{s1}R_{s2}]$	$f_i[s_e]\cdot f_{se}[R_{s1}R_{s2}]$
Initial repertoire					
1	0.25	0.25	0.063	0	0
2	0.38	0.50	0.18	0.167	0.063
Σ	0.63		0.25		0.063
Final repertoire					
			$f_f[R_{s1}]=0.4$		$f_f[R_{s1}R_{s2}]=0.1$

[a] For the conditions: $t_g=8$; $s_g=4$; $n_g=4$; $p=0.5$; $q=0.5$, $s_{min}=1$ and $s_{max}=2$.

The general formula to calculate the fraction of cells in the initial population expressing s_e receptors can be derived with the aid of the binomial theorem[1]:

$$f_i[s_e] = p^{se}q^{[sg-se]} \cdot \binom{s_g}{s_e} = p^{se}q^{[sg-se]} \cdot \frac{s_g!}{s_e![s_g-s_e]!}$$

To calculate the fraction of the initial repertoire that survives selection the $f_i[s_e]$ values are simply summed. In our example, the result is 0.63 (Table 3).

Now that we know $f_i[s_e]$ for each value of s_e, we wish to calculate the fraction of these cells that express a *particular* self-specific receptor, R_{s1}. This fraction is designated $f_{se}[R_{s1}]$. For each cell expressing two self-specific receptors of four in the genome, the probability that a particular one is expressed is 2/4. Among cells that express one self-specific receptor, 1/4 express R_{s1}. The general expression for the fraction of cells that express R_{s1} among cells expressing s_e receptors is:

$$f_{se}[R_{s1}] = \frac{s_e}{s_g}$$

We also wish to calculate the fraction of cells for each s_e value that express a specific *pair* of self-specific receptors, $f_{se}[R_{s1}R_{s2}]$. This value can be easily calculated as the product of the probabilities of expressing each of them. Considering cells that have expressed one of them, the probability that the second one is expressed equals s_e-1/s_g-1. Therefore:

$$f_{se}[R_{s1}R_{s2}] = f_{se}[R_{s1}] \cdot \frac{s_{e-1}}{s_{g-1}} = \frac{s_e}{s_g} \cdot \frac{s_{e-1}}{s_{g-1}}$$

[1] We make use of the mathematical 'choose' function, whereby the number of ways to choose x items from a pool of y items, i.e., y choose x, is denoted by $\binom{y}{x}$ and equals y!/x! (y–x)!. The exclamation mark denotes the factorial function, i.e. 4!=4×3×2×1

Multiplying $f_{se}[R_{s1}]$ by $f_i[s_e]$ yields the fraction of the initial repertoire that expresses R_{s1}, for cells expressing s_e receptors. Summing all of these values yields the fraction of cells in the initial repertoire that express R_{s1} and can be selected (see Table 3). By dividing this value by the fraction of the initial repertoire that can be successfully selected, we can derive the desired value, $f_f[R_{s1}]$: the fraction of cells expressing R_{s1} in the *final* repertoire.

Similarly, to calculate $f_f[R_{s1}R_{s2}]$, the fraction of selected cells expressing both R_{s1} and R_{s2}, we multiply $f_{se}[R_{s1}R_{s2}]$ by $f_i[s_e]$ for each value of s_e, sum these values, and divide by the fraction of the initial repertoire that can be successfully selected. In our example, $f_f[R_{s1}]=0.25\div0.63=0.4$ and $f_f[R_{s1}R_{s2}]=0.63\div0.63=0.1$.

While Table 2 serves as an aid, the following general formulas can be applied to the problem:

$$f_f[R_{s1}] = \sum_{s_e=s_{min}}^{s_{max}} \left[\frac{s_e}{s_g} \cdot f_i[s_e]\right] \div \sum_{s_e=s_{min}}^{s_{max}} f_i[s_e]$$

and:

$$f_f[R_{s1}R_{s2}] = \sum_{s_e=s_{min}}^{s_{max}} \left[\frac{s_e}{s_g} \cdot \frac{s_{e-1}}{s_{g-1}} \cdot f_i[s_e]\right] \div \sum_{s_e=s_{min}}^{s_{max}} f_i[s_e]$$

3.3 Predictions of the Selection Model

Table 4 depicts the predictions of the selection model for various conditions. For ease of interpretation the table employs a more familiar nomenclature for the receptors and ligands, where Ly-49X1 is a particular H-2^x specific receptor, etc. As anticipated, conditions exist under which the model predicts that cells expressing Ly-49X1 are more frequent in mice that do not express a ligand (H-2^y mice) than in mice that do (H-2^x mice). However, the decrease in the frequencies of cells expressing particular self-specific receptors is not seen for all conditions. Generally three conditions favor the paradoxical decrease: (a) the relative number of self-specific receptors encoded by the genome is large (s_g/t_g is large); (b) the selection process favors cells expressing relatively few self-specific receptors per cell (s_{min} and s_{max} are low); and (c) the probability of initially expressing any given receptor is high (p is relatively large).

As already noted above, the model predicts that the frequency of Ly-49X1$^+$ cells, or Ly-49X1$^+$X2$^+$ cells, in mice that do not express a ligand, is determined only by p. Therefore these frequencies should not be affected by selection, the number of H-2^x specific receptors or the number of H-2^y specific receptors. They should also not be altered by introducing new receptor transgenes into the genome or by interfering with the selection process. In contrast, we demonstrate below that the competing sequential model predicts that usage of non-self-specific receptors should be affected by these manipulations.

Table 4. Predictions of the two-step selection model

						Predicted frequency of cells expressing:			
			Number of germline receptors specific for:			Ly-49X1		Ly-49X1 and Ly-49X2	
p	s_{min}	s_{max}	H-2^x	H-2^y	Other	In H-2^x	In H-2^y	In H-2^x	In H-2^y
0.5	1	2	4	4	0	0.40	0.50	0.10	0.25
0.3	1	2	4	4	0	0.35	0.30	0.070	0.09
0.3	1	2	6	4	0	0.25	0.30	0.035	0.09
0.3	1	2	10	10	0	0.17	0.30	0.015	0.09
0.3	1	3	10	10	0	0.22	0.30	0.040	0.09
0.3	2	3	10	10	0	0.25	0.30	0.046	0.09

4 Mathematical Treatment of the Sequential Model

In the sequential model NK cells are proposed to express Ly-49 receptors in a random sequence until a sufficient number of self-specific receptors have been turned on to ensure self tolerance. The principal feature that distinguishes this model from the selection model is that NK cells *must* be tested frequently or continuously throughout development for the expression of an appropriate number of self-specific receptors, unlike in the selection model, where such a test need occur at one point in time. The model also assumes that all developing NK cells eventually reach a stage where they express an appropriate number of self-specific receptors. No NK cell ever needs to be deleted or anergized since any cell expressing too few self-specific receptors simply continues turning on receptors until a sufficient number is expressed. In all likelihood the extreme version of the sequential model that we present here to bring out its conceptual features will prove to be inaccurate in at least some respects. In particular, it seems probable that the sequential activation of receptors will have to occur within a defined, rather than an unlimited, window of time. After this time has expired, cells that have not achieved sufficient expression of self-specific receptors may be subject to selective forces. Hence it is plausible that NK cell education involves an amalgamation of the sequential and selection models.

The sequential model that we have proposed suggests that Ly-49 genes are turned on in a random, rather than defined sequence. This assumption may well turn out to be simplistic, but it fits with much of the available data and serves as a starting point for mathematical modeling. Hence in our modeling we assume that the first receptor to be activated is equally likely to be any of the Ly-49 genes encoded by the genome. The second receptor to be activated is equally likely to be any of the remaining receptors. Whether a particular receptor is self or non-self-specific is also not relevant to the sequence in which the receptor is activated. It should be noted that other versions of the sequential model can be considered, in which the sequence of receptor expression is less random, but we have not addressed such models here.

The mechanism of Ly-49 gene activation is not known. One possibility is that relevant "gene activation factors" are limiting such that there is a defined, relatively low probability of activation of a given gene per unit time over the relevant developmental period. An alternative possibility is that gene activation is somehow tied to a periodic event in cellular physiology, such as DNA replication, such that one Ly-49 gene is activated per period. As an aid in devising a mathematical treatment, the model below incorporates the notion of "periodic" gene activation, but the model nevertheless works for both schemes. Once a gene is initially activated, we assume that it remains activated, and thus that receptor gene activation is cumulative. This assumption is in line with data that we have recently obtained (Dorfman and Raulet, in preparation), and with the observation that NK cells often express non-self-specific Ly-49 receptors. It should be noted that there is now clear evidence for mechanisms by which developmentally regulated gene expression can be maintained permanently in a cell lineage, even after the factors that initially activated gene expression have disappeared from the cell. For instance hypermethylated genes are generally transcriptionally repressed, and the methylation status of a gene is heritable in the daughters of dividing cells (BIRD 1992). As another example, the trithorax and polycomb gene products stably maintain the proper activation/inactivation status of homeotic genes in *Drosophila melanogaster*, even in mature cell lineages (PARO 1995). Thus the process of Ly-49 gene activation could be easily terminated by extinguishing relevant activating factors, while expression of the already activated genes could be maintained.

4.1 Frequencies of Cells Expressing Self-Specific Receptors According to the Sequential Model

The central notion of the sequential model is that the engagement of self-specific receptors by class I molecules terminates the activation of additional receptor genes. As there is evidence that NK cells can express more than one functioning self-specific receptor (Ly-49A$^+$G2$^+$ cells are detectable, though relatively infrequent, in H-2^d mice; Table 1), we presume that in some cases signaling through more than one receptor is necessary to terminate new receptor gene activation. In practice, variations in receptor affinities and cell surface levels might lead to a situation in which signaling through one receptor is sufficient in some cases to terminate new receptor engagement, while signaling through multiple receptors is necessary in others. In order to simplify the mathematical treatment, however, we have assumed that all self-specific receptors are equivalent.

Thus our model makes the assumption that there is a specific target number of self-specific receptors, s_t, that must be activated to terminate the receptor gene activation mechanism. All mature cells therefore express s_t self-specific receptors. If the total number of self-specific receptor genes (s_g) is known, the calculations resemble to those in Sect. 3.2. It is apparent that:

$$f[R_{s1}] = \frac{s_t}{s_g}$$

and:

$$f[R_{s1}R_{s2}] = f[R_{s1}] \cdot \frac{s_{t-1}}{s_{g-1}} = \frac{s_t}{s_g} \cdot \frac{s_{t-1}}{s_{g-1}}$$

Note that, unlike in the selection model, we do not distinguish an initial versus a final repertoire in the sequential model. This is because there is no discrete initial repertoire in the sequential model.

4.2 Frequencies of Cells Expressing Non-Self-specific Receptors According to the Sequential Model

The distribution of non-self-specific receptors is difficult to calculate in the sequential model. It is possible that a mature NK cell would express no non-self-specific receptors, if by chance, it happened to activate only self-specific receptors. It is also possible that a mature NK cell would express *all* of its non-self-specific receptors, as is any combination between these extremes.

To illustrate our approach to calculate the distribution of non-self-specific receptors in the sequential model, consider a more concrete example, in which there are a total of six Ly-49 receptors encoded by the genome, three of which recognize self MHC, and three of which do not. Assume that the expression of two self-specific receptors is sufficient to terminate new receptor gene activation (i.e., s_t=2). A useful way to visualize the sequential model is to use a probability tree diagram (Fig. 1). In this diagram, the cell begins its developmental process with no Ly-49 genes activated. The first receptor to be turned on can be self-specific, in which case the right branch is followed. Alternatively, the first receptor to be activated can be non-self-specific, in which case the left branch is followed. Initially, the probability of turning on a self-specific receptor is 3/6, as is the probability of turning on a non-self-specific receptor. The second receptor to be activated can also either be self-specific or non-self-specific, and, again, the cell follows the right or left branches, respectively. If the first receptor to be activated was self-specific, there are only two self-specific receptors and three non-self-specific receptors remaining that can be activated. Thus in such a case the probability that the second receptor to be activated is self-specific is 2/5; the probability that the second receptor to be activated is non-self-specific is 3/5, etc. All the probabilities are indicated on the tree diagram. The cell stops activating new receptor genes when it has turned on s_t self-specific receptors (i.e., after it has made s_t moves to the right). At this point we consider that it has reached an endpoint and has become an endpoint cell.

The tree diagram illustrates that the process can be divided into sequential *periods*. Since the cells that reach their endpoint in a given period have a number of common features, our approach is to consider each period separately. Clearly, only periods in which cells reach their endpoint are relevant. Each relevant period can be defined by the number of non-self-specific receptors expressed (n_e) by endpoint cells in the period. Thus for each relevant period we determine the fraction of all endpoint cells that reach their endpoint at this period, $f[n_e]$, and the fraction of these cells that express

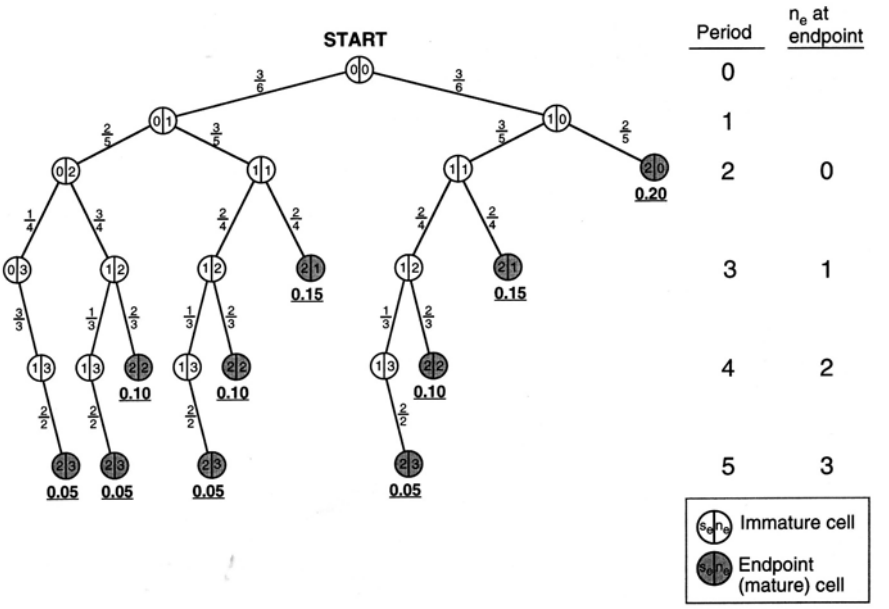

Fig. 1. Probability tree diagram of the sequential model for a system of six receptors, three of which are specific for self MHC and three of which are not. For this example, it is assumed that two self-specific receptors must be activated to terminate new receptor expression. Within each cell is indicated the number of activated self-specific receptors (*left*) and non-self-specific receptors (*right*). The immature cell begins at the top and sequentially activates receptors. Branches heading to the right arise when a cell activates a self-specific receptor; branches to the left arise when a cell activates a non-self-specific receptor. *Numbers on left and right within cells*, the numbers of expressed self and non-self-specific receptors, respectively; *fractions*, the probability of following any particular branch; *shaded circles at the end of branches*, mature enpoint cells; *underlined boldface*, the probability of reaching any particular endpoint (i.e., the endpoint probability, y, calculated by multiplying the probabilities of taking each branch along the route)

a particular non-self-specific receptor, $f_{ne}[R_{n1}]$, or pair of non-self-specific receptors, $f_{ne}[R_{n1}R_{n2}]$. Summing these values for all possible periods (i.e., for all relevant values of n_e) yields $f[R_{n1}]$, the fraction of endpoint cells that express R_{n1}, and $f[R_{n1}R_{n2}]$, the fraction of endpoint cells that express both R_{n1} and R_{n2}. The relationships are:

$$f[R_{n1}] = \sum_{n_e=0}^{n_g} f[n_e] \cdot f_{n_e}[R_{n1}]$$

$$f[R_{n1}R_{n2}] = \sum_{n_e=0}^{n_g} f[n_e] \cdot f_{n_e}[R_{n1}R_{n2}]$$

The calculation of $f[n_e]$ depends on the fact that all endpoint cells in a given period arrive at their endpoint with the same probability, regardless of the pathway taken.

Hence the fraction of cells that reach their endpoints at each period, $f[n_e]$, equals the probability of reaching an endpoint at this period (defined as y, the endpoint probability), times the number of endpoints at this period (defined as x).

The derivation of the endpoint probability, y, can be understood by considering the probability of a given pathway to an endpoint in our example (Fig. 1). At period 5 the left-most endpoint arose by expressing receptors in the order R_n, R_n, R_n, R_s, R_s. The probability of this occurring equals the product of the probabilities of each step: $3/6\times2/5\times1/4\times3/3\times2/2$, or $(3\times2\times1\times3\times2)/(6\times5\times4\times3\times2)$. The denominator is seen to equal $t_g!/(t_g-t_e)!$. The numerator can be separated into an expression that concerns non-self-specific receptors and an expression that concerns self-specific receptors: in our example $(3\times2\times1)$ and (3×2). The general formulas can be seen to be $n_g!/(n_g-n_e)!$ and $s_g!/(s_g-s_t)!$, respectively. Thus the endpoint probability can be defined generally as:

$$y = \left[\frac{s_g!}{(s_g - s_t)!} \cdot \frac{n_g!}{(n_g - n_e)!}\right] \div \left[\frac{t_g!}{(t_g - t_e)!}\right]$$

The derivation of the number of endpoints, x, at each period, can also be understood by analysis of the tree diagram. Consider the fifth period. All pathways to endpoints in this or any other period must end with expression of a self-specific receptor. Therefore all the possible pathways to endpoints in this period can be described in the form: $(R_n,R_n,R_n,R_s)R_s$, where the receptors in parentheses can take all possible orders. Thus this problem reduces to determining the number of possible orders of three equivalent R_n, and one R_s, in the previous four periods. This problem can be restated in terms of only the R_n receptors: given four different periods, how many ways are there to put three R_n into them, or more generally, given t_e-1 periods, how many ways are there to put n_e receptors into them (i.e., t_e-1 choose n_e). In the example, there are four endpoints in the fifth period, where $n_e=3$. The general formula for the number of endpoints is:

$$x = \left[\begin{matrix} t_e-1 \\ n_e \end{matrix}\right] = \frac{(t_e-1)!}{(s_t-1)!\, n_e!}$$

As explained above, the product of x and y equals $f[n_e]$. All the calculations for the tree in the example ($s_g=3$, $n_g=3$, $s_t=2$) are summarized in Table 5.

Now we need to determine the fraction of cells that express a *particular* non-self-specific receptor, R_{n1}. These values are determined for each period, defined by n_e. Consider the period where $n_e=2$. The question is, among cells expressing n_e different non-self-specific receptors, what is the fraction that express a particular one, R_n1? As before (Sect. 3.2), it is apparent that this value equals n_e/n_g. Similarly, the fraction of these cells that express a particular pair of receptors, R_{n1} and R_{n2}, equals $(n_e/n_g)\times(n_e-1)/(n_g-1)$. Multiplying these values by $f[n_e]$ yields $f_{ne}[R_{n1}]$, the fraction of *all* cells that express n_e receptors and express R_{n1}:

$$f_{ne}[R_{n1}] = f[n_e] \cdot \frac{n_e}{n_g}$$

Table 5. Calculations for the sequential model[a]

n_e	x	y	$f[n_e]=x \cdot y$	$f_{ne}[R_{n1}]$	$f_{ne}[R_{n1}R_{n2}]$	$n_e \times f[n_e]$
0	1	0.20	0.20	0	0	0
1	2	0.15	0.30	0.10	0	0.3
2	3	0.10	0.30	0.20	0.1	0.6
3	4	0.05	0.20	0.20	0.2	0.6
Total			1.0	$0.5=f[R_{n1}]$	$0.3=f[R_{n1}R_{n2}]$	$1.5=n_{avg}$

[a]for the conditions $s_g=3$; $n_g=3$; $s_t=2$.

Similarly:

$$f_{ne}[R_{n1}R_{n2}] = f[n_e] \cdot \frac{n_e}{n_g} \cdot \frac{n_e-1}{n_g-1}$$

Summing all values of $f_{ne}[R_{n1}]$ or $f_{ne}[R_{n1}R_{n2}]$ yields the desired values $f[R_{n1}]$ and $f[R_{n1}R_{n2}]$, respectively, as depicted in Table 5. Thus on average, given our initial assumptions, we can see that the sequential model predicts that 50% of mature NK cells will express a particular non-self-specific receptor, and 30% will express a particular pair of non-self-specific receptors.

Another potentially useful value that can be calculated is the average number of different non-self-specific receptors expressed on the population of cells, n_{avg}:

$$n_{avg} = \sum_{n_e=0}^{n_g} n_e \cdot f[n_e]$$

Thus in our example, the average NK cell at its endpoint expresses 1.5 non-self-specific receptors.

General expressions for $f[R_{n1}]$ and $f[R_{n1}R_{n2}]$ are:

$$f[R_{n1}] = \sum_{n_e=0}^{n_g} \left[\frac{(t_e-1)!}{(s_t-1)!n_e!}\right] \cdot \left[\frac{s_g!}{(s_g-s_t)!} \cdot \frac{n_g!}{(n_g-n_e)!} \div \frac{t_g!}{(t_g-t_e)!}\right] \cdot \left[\frac{n_e}{n_g}\right]$$

$$f[R_{n1}R_{n2}] = \sum_{n_e=0}^{n_g} \left[\frac{(t_e-1)!}{(s_t-1)!n_e!}\right] \cdot \left[\frac{s_g!}{(s_g-s_t)!} \cdot \frac{n_g!}{(n_g-n_e)!} \div \frac{t_g!}{(t_g-t_e)!}\right] \cdot \left[\frac{n_e}{n_g}\right] \cdot \frac{n_e-1}{n_g-1}$$

Table 6. Predictions of the sequential model when s_t=2

			Predicted frequency of cells expressing:			
Number of germline receptors specific for:			Ly-49X1		Ly-49X1 and Ly-49X2	
$H\text{-}2^x$	$H\text{-}2^y$	Other	In $H\text{-}2^x$	In $H\text{-}2^y$	In $H\text{-}2^x$	In $H\text{-}2^y$
4	4	0	0.50	0.40	0.17	0.20
5	4	0	0.40	0.40	0.10	0.20
6	4	0	0.33	0.40	0.07	0.20
10	4	0	0.20	0.40	0.02	0.20
4	6	0	0.50	0.29	0.17	0.11
10	10	0	0.20	0.18	0.02	0.05
13	10	0	0.15	0.18	0.01	0.05
10	8	4	0.20	0.22	0.02	0.07

4.3 Predictions of the Sequential Model

Using these equations and the equations that govern self-specific receptor distribution (Sect. 4.1), Table 6 can be generated, which depicts the predictions of the sequential model under various conditions. For simplicity, the table addresses predictions for only a single value of s_t, 2.

We noted above that the selection model could explain one counterintuitive feature of the Ly-49 repertoire, namely, that the frequency of cells expressing a particular self-specific receptor can decrease in the presence of its ligand. We can now see that the sequential model is equally able to explain this phenomenon. As in the selection model, this behavior occurs in some but not all hypothetical repertoires and conditions. Consider a repertoire comprised of six $H\text{-}2^x$-specific receptors and four distinct $H\text{-}2^y$-specific receptors. The frequency of cells expressing a given $H\text{-}2^x$ specific receptor, Ly-49X1, is higher in $H\text{-}2^y$ mice (40%) than in $H\text{-}2^x$ mice (33%). An opposite trend is observed when there are four $H\text{-}2^x$ specific receptors and four $H\text{-}2^y$ specific receptors. In common with the selection model, the sequential model predicts that receptor usage will decrease in ligand-expressing mice when a large proportion of all receptors are ligand-specific, and increase in ligand-expressing mice when only a small proportion of all receptors are ligand-specific. This is true over a wide range of s_t values. The frequency of cells expressing a given pair of self-specific receptors (e.g., Ly-49X1 and Ly-49X2) follows the same general trend, as illustrated in Table 6.

5 Comparisons of the Two Models

Having in hand predictions of the two models, one can ask whether either or both models are capable of accounting for the available data concerning the Ly-49 repertoire. If not, other models should be considered. Equally important is to identify the situations in which the models make different predictions, and to ask whether the available data are more consistent with one model than the other.

5.1 Do the Models Fit the Data?

It should first be noted that both models predict deviations from a strict adherence to the product rule. This is expected because both models assume that expression of different receptors is not in fact independent. This assumption is inherent in any education model because education implies a deviation from randomness. The predicted deviations, however, are not large under most conditions that we have modeled and are in line with the observed deviations from the product rule (Table 1).

The two models predict similar trends in terms of the frequencies of cells expressing a given receptor or receptor pair in strains that do or do not express an MHC ligand. Both models predict that the frequency of cells expressing Ly-49X1 is lower in H-2^x mice than in H-2^y mice only under some conditions. In general this occurs in either model only if there exists a relatively large number of H-2^x specific receptors, although other variables affect this outcome (especially p and s_{min} and s_{max} in the selection model). Is this prediction borne out by the data? The reductions in Ly-49A^+ and Ly-49$G2^+$ cells in H-2^d mice would fit this prediction if a relatively large number of the Ly-49 receptors are H-2^d specific. Whether this is so cannot yet be answered because of the limited data available concerning the specificity of most Ly-49 receptors. However, anecdotal evidence suggests that specificity for H-2^d may be common among Ly-49 receptors. Of four inhibitory Ly-49 receptors tested, three – Ly-49A, Ly-49G2, and Ly-49C – are reportedly reactive with D^d and/or L^d. It is also notable that H-$2^{b/d}$ mice reject H-2^d bone marrow only very inefficiently, which might suggest that expression of H-2^d-specific receptors is a common property of NK cells, at least in the case of the H-$2^{b/d}$ host (MURPHY et al. 1990). A final conclusion must await the evaluation of H-2^d reactivity of the remaining Ly-49 receptors.

5.2 The Models as Applied to Class I Deficient Mice

Any model of NK cell repertoire formation must account for the phenotype of NK cells in the class I deficient $\beta_2m^{-/-}$ mouse. These mice contain normal numbers of cells with the NK phenotype (LIAO et al. 1991), yet these cells do not attack $\beta_2m^{-/-}$ normal cells (BIX et al. 1991; HOGLUND et al. 1991; LIAO et al. 1991). The NK cells in these mice are not devoid of function, however, because they do lyse certain tumor cell lines, though with a somewhat reduced efficiency.

The phenotype of NK cells in $\beta_2m^{-/-}$ mice seems initially not to fit easily with either model or indeed with any simple model of NK cell selection by class I molecules. The sequential model in its pure form predicts that all NK cells in mice that fail to express class I molecules should express all Ly-49 receptors. The selection model would predict that such mice have no NK cells. However, it is known that $\beta_2m^{-/-}$ and $TAP^{-/-}$ mice are not completely class I deficient. They express on their cell surfaces low levels of functionally conformed class I molecules. Hence the NK cells in these mice could arise by either postulated education process, depending on interactions with low levels of class I molecules. In the sequential model the higher frequencies of NK cells expressing each tested Ly-49 receptor in $\beta_2m^{-/-}$ mice (Table 1) (HELD et al. 1996b) could result from the requirement for more receptors per cell to terminate new receptor expression when class I levels are low. In the selection model only those clones with more receptors would exhibit the appropriate reactivity with low levels of class I molecules and survive the selection process.

Alternatively, the models may well be too simplified, though some of their basic features may be correct. Perhaps the sequential process operates, but only during a limited time period in the life of a developing NK cell. After this time period, new gene activation could not occur. Cells that had not expressed self-specific receptors might then convert to an "anergized" state where they exhibit poor reactivity to all cells, or at least to normal untransformed cells, regardless of the cells' class I expression. Some or all the NK cells in class I deficient mice might be in this state. The selection model can also be adapted in a similar way. NK cells that fail selection may be induced to enter the putative anergic state. Clearly, further analysis of the NK cells in class I deficient mice is necessary to evaluate these possibilities.

5.3 Differential Predictions of the Models

What are the critical differences in the predictions of the two models? There are several, but two are most relevant in view of data that are currently being generated:

5.3.1 The Expected Effects of Ly-49 Transgenes Expressed in All NK Cells

Both models predict that a Ly-49 transgene (e.g., Ly-49A) expressed by all NK cells, in a strain that expresses a class I ligand (e.g., $H\text{-}2^d$), would result in decreased usage of other self-specific receptors (e.g., Ly-49G2). Such an effect was observed (Sect. 2.5). However, the models make different predictions concerning the effects of the transgene on non-self-specific receptors. In the selection model the transgene should have no effect on the frequencies of cells expressing non-self-specific receptors, since these frequencies are dependent only upon p. However, in the sequential model it can be seen that expression of the transgene early in all NK cells hastens the moment in which cells express s_t receptors and hence decreases the likelihood that any given non-self-specific receptor has time to be activated. A comparison of the

Table 7. Predicted effects of a Ly-49X1 transgene in the sequential versus selection models

Variables			Predicted frequencies of cells expressing			
Number of germline receptors specific for:			Ly-49X2		Ly-49Y1	
$H\text{-}2^x$	$H\text{-}2^y$	Ly-49X1 transgene?	In $H\text{-}2^x$	In $H\text{-}2^y$	In $H\text{-}2^x$	In $H\text{-}2^y$
Sequential model (s_f=2)						
4	4	-	0.50	0.40	0.40	0.50
4	4	+	0.25	0.40	0.20	0.50
6	4	-	0.33	0.40	0.29	0.50
6	4	+	0.17	0.40	0.14	0.50
13	10	-	0.15	0.18	0.14	0.20
13	10	+	0.08	0.18	0.07	0.20
Selection model (p=0.3, s_{min}=1, s_{max}=2)						
4	4	-	0.35	0.30	0.30	0.35
4	4	+	0.16	0.30	0.30	0.35
6	4	-	0.25	0.30	0.30	0.35
6	4	+	0.12	0.30	0.30	0.35
10	10	-	0.17	0.30	0.30	0.17
10	10	+	0.08	0.30	0.30	0.17

predicted effects of a Ly-49X transgene according to the two models under some specific conditions is presented in Table 7.

Does the available data bear on these predictions? It was observed that the Ly-49A transgene caused a modest decrease in the frequency of cells in $H\text{-}2^d$ mice that stained with the Ly-49 specific SW-5E6 mAb. Unfortunately, it is so far difficult to draw a conclusion from this experiment because of the uncertainties concerning the nature and specificity of receptors detected by the SW-5E6 mAb. With these uncertainties the available data are inadequate to distinguish the sequential or selection models. However, at the present pace of research into the specificity of receptors in this family, a clean test should be forthcoming in the near future.

5.3.2 The Effects of "Irrelevant" MHC Expression on the Frequencies of Cells Expressing a Given Ly-49 Receptor

The selection model predicts that the frequency of cells expressing a Ly-49 receptor in mice that do not express a cognate class I molecule should simply equal p. This should be true equally in class I deficient mice and in mice that express noncognate class I molecules. In contrast, the sequential model predicts differences between class I deficient mice and mice that express noncognate class I molecules. In class I deficient mice new receptor gene activation continues for a longer duration in each cell than in class I^+ mice, although the duration may have an upper limit. Hence class I deficient mice might be predicted to harbor a higher frequency of cells expressing

a given Ly-49 receptor than class I^+ mice expressing irrelevant class I molecules. The available data are inconclusive on this point. We have reported that there are more cells expressing Ly-49A or Ly-49G2 or both in class I deficient mice than in H-2^b mice which are thought not to express a ligand for these receptors (HELD et al. 1996b). These data might be seen to support the sequential model. We have thus far refrained from drawing this conclusion because of the possibility that one or both of these receptors reacts weakly with H-2^b encoded class I molecules (HELD and RAULET 1997).

6 Concluding Remarks

The models elaborated here make various predictions concerning the effects of specific manipulations on the NK repertoire. While available data are so far inadequate to distinguish these models or verify them, the current pace of research in this area is dramatic, and it is likely that new reagents and information will be soon forthcoming which will allow rigorous testing of the predictions of each model. We hope that the mathematical treatments described above will serve as an aid to this research.

Acknowledgements. We thank Drs. Montgomery Slatkin and George Oster for valuable advice on the mathematical modeling, and Jeffrey Dorfman for comments on the manuscript. This research was supported by National Institutes of Health grants RO1AI35021 and RO1AI39642. R.E.V. is a predoctoral fellow of the Howard Hughes Medical Institute. Those interested in obtaining a Macintosh-compatible spreadsheet designed to perform calculations for the mathematical models described in this paper should contact us by email: rvance@ocf.berkeley.edu.

References

Bennett M (1987) Biology and genetics of hybrid resistance. Adv Immunol 41:333–443

Bird A (1992) The essentials of DNA methylation. Cell 70:5–8

Bix M, Liao N-S, Zijlstra M, Loring J, Jaenisch R, Raulet D (1991) Rejection of class I MHC-deficient hemopoietic cells by irradiated MHC-matched mice. Nature 349:329–331

Brennan J, Lemieux S, Freeman J, Mager D, Takei F (1996) Heterogeneity among Ly49C NK cells: characterization of highly related receptors with differing functions and expression patterns. J Exp Med 184:2085–2090

Brown MG, Scalzo AA, Matsumoto K, Yokoyama WM (1997) The natural killer gene complex: a genetic basis for understanding natural killer cell function and innate immunity. Immunol Rev 155:53–65

Chadwick BS, Miller RG (1992) Hybrid resistance in vitro. Possible role of both class I MHC and self peptides in determining the level of target cell sensitivity. J Immunol 148:2307–2313

Correa I, Corral L, Raulet DH (1994) Multiple natural killer cell-activating signals are inhibited by major histocompatibility complex class I expression in target cells. Eur J Immunol 24:1323–1331

Cudkowicz G, Stimpfling JH (1964) Induction of immunity and of unresponsiveness to parental marrow grafts in adult F1 hybrid mice. Nature 204:450–453

Dorfman JR, Raulet DH (1996) Major histocompatibility complex genes determine natural killer cell tolerance. Eur J Immunol 26:151–155

George T, Yu YYL, Liu J, Davenport C, Lemieux S, Stoneman E, Mathew PA, Kumar V, Bennett M (1997) Allorecognition by murine natural killer cells: lysis of T-lymphoblasts and rejection of bone marrow grafts. Immunol Rev 155:29–40

Held W, Cado D, Raulet DH (1996a) Transgenic expression of the Ly49A natural killer cell receptor confers class I MHC-specific inhibition and prevents bone marrow allograft rejction. J Exp Med 184:2037–2041

Held W, Dorfman JR, Wu M-F, Raulet DH (1996b) Major histocompatibility complex class I-dependent skewing of the natural killer cell Ly49 receptor repertoire. Eur J Immunol 26:2286–2292

Held W, Raulet DH (1997a) Ly49A transgenic mice provide evidence for an MHC-dependent education process in NK cell development. Exp Med 185:2079–2088

Held W, Raulet DH (1997b) Expression of the Ly49A gene in murine natural killer cell clones is predominantly but not exclusively mono-allelic. Eur J Immunol 27:2876–2884

Held W, Roland J, Raulet DH (1995) Allelic exclusion of Ly49 family genes encoding class I-MHC-specific receptors on NK cells. Nature 376:355–358

Hoglund P, Ohlen C, Carbone E, Franksson L, Ljunggren H, Latour A, Koller B, Karre K (1991) Recognition of β2-microglobulin-negative (β2m-) T-cell blasts by natural killer cells from normal but not from $\beta 2m^-$ mice: nonresponsiveness controlled by $\beta 2m^-$ bone marrow in chimeric mice. Proc Natl Acad Sci USA 88:10332–10336

Karlhofer FM, Hunziker R, Reichlin A, Margulies DH, Yokoyama WM (1994) Host MHC class I molecules modulate in vivo expression of a NK cell receptor. J Immunol 153:2407–2416

Karlhofer FM, Ribaudo RK, Yokoyama WM (1992) MHC class I alloantigen specificity of $Ly\text{-}49^+$ IL-2 activated natural killer cells. Nature 358:66–70

Karre K, Ljunggren HG, Piontek G, Kiessling R (1986) Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defense strategy. Nature 319:675–678

Liao N, Bix M, Zijlstra M, Jaenisch R, Raulet D (1991) MHC class I deficiency: susceptibility to natural killer (NK) cells and impaired NK activity. Science 253:199–202

Ljunggren HG, Karre K (1990) In search of the 'missing self': MHC molecules and NK cell recognition. Immunol Today 11:237–244

Mason LH, Ortaldo JR, Young HA, Kumar V, Bennett M, Anderson SK (1995) Cloning and functional characteristics of murine LGL-1: a member of the Ly-49 gene family (Ly-49G2). J Exp Med 182:293–303

Murphy WJ, Kumar V, Cope JC, Bennett M (1990) An absence of T cells in murine bone marrow allografts leads to an increased susceptibility to rejection by natural killer cells and T cells. J Immunol 144:3305–3311

Ohlen C, Kling G, Hšglund P, Hansson M, Scangos G, Bieberich C, Jay G, Karre K (1989) Prevention of allogeneic bone marrow graft rejection by H-2 transgene in donor mice. Science 246:666–668

Olsson MY, Karre K, Sentman CL (1995) Altered phenotype and function of natural killer cells expressing the major histocompatibility complex receptor Ly-49 in mice transgenic for its ligand. Proc Natl Acad Sci USA 92:1649–1653

Paro R (1995) Propagating memory of transcriptional states. Trends Gen 11:295–297

Raulet DH, Held W, Correa I, Dorfman J, Wu M-F, Corral L (1997) Specificity, tolerance and developmental regulation of natural killer cells defined by expression of class I-specific Ly49 receptors. Immunol Reviews 155:41–52

Roland J, Cazenave P-A (1992) Ly-49 antigen defines an αβ TCR population in i-IEL with an extrathymic maturation. Int Immunol 4:699–706

Ryan JC, Seaman WE (1997) Divergent functions of lectin-like receptors on NK cells. Immunol Rev 155:79–89

Takei F, Brennan J, Mager DL (1997) The Ly49 family: genes proteins and recognition of class I MHC. Immunol Rev 155:67–77

Wu M-F, Raulet DH (1997) Class I-deficient hematopoietic cells and nonhematopoietic cells dominantly induce unresponsiveness of NK cells to class I-deficient bone marrow grafts. J Immunol 158:1628–1633

Yu YY, George T, Dorfman J, Roland J, Kumar V, Bennett M (1996) The role of Ly49A and 5E6 (Ly49C) molecules in hybrid resistance mediated by murine natural killer cells against normal T cell blasts. Immunity 4:67–76