

**BREAKTHROUGHS TAKE TIME.
ISOLATING CELLS SHOULDN'T.**



Models for Antigen Receptor Gene Rearrangement. III. Heavy and Light Chain Allelic Exclusion

This information is current as of July 15, 2018.

Gil Kalmanovich and Ramit Mehr

J Immunol 2003; 170:182-193; ;
doi: 10.4049/jimmunol.170.1.182
<http://www.jimmunol.org/content/170/1/182>

References This article **cites 68 articles**, 30 of which you can access for free at:
<http://www.jimmunol.org/content/170/1/182.full#ref-list-1>

Why *The JI*? [Submit online.](#)

- **Rapid Reviews! 30 days*** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

**average*

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

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

Email Alerts Receive free email-alerts when new articles cite this article. Sign up at:
<http://jimmunol.org/alerts>

The Journal of Immunology is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852
Copyright © 2003 by The American Association of
Immunologists All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.



Models for Antigen Receptor Gene Rearrangement. III. Heavy and Light Chain Allelic Exclusion¹

Gil Kalmanovich and Ramit Mehr²

The extent of allelic exclusion in Ig genes is very high, although not absolute. Thus far, it has not been clearly established whether rapid selection of the developing B cell as soon as it has achieved the first productively rearranged, functional heavy chain is the only mechanism responsible for allelic exclusion. Our computational models of Ag receptor gene rearrangement in B lymphocytes are hereby extended to calculate the expected fractions of heavy chain allelically included newly generated B cells as a function of the probability of heavy chain pairing with the surrogate light chain, and the probability that the cell would test this pairing immediately after the first rearrangement. The expected fractions for most values of these probabilities significantly exceed the levels of allelic inclusion in peripheral B cells, implying that in most cases productive rearrangement and subsequent cell surface expression of one allele of the heavy chain gene probably leads to prevention of rearrangement completion on the other allele, and that additional mechanisms, such as peripheral selection disfavoring cells with two productively rearranged heavy chain genes, may also play a role. Furthermore, we revisit light chain allelic exclusion by utilizing the first (to our knowledge) computational model which addresses and enumerates B cells maturing with two productively rearranged κ light chain genes. We show that, assuming that there are no selection mechanisms responsible for abolishing cells expressing two light chains, the repertoire of newly generated B lymphocytes exiting the bone marrow must contain a significant fraction of such κ double-productive B cells. *The Journal of Immunology*, 2003, 170: 182–193.

Developing lymphocytes create the genes for their AgRs^{1,2} via gene rearrangement. Elucidation of the mechanisms of AgR gene rearrangement is highly relevant for the understanding of various immunopathologies (1, 2) and is currently under extensive study (3–5). One approach to the study of AgR gene rearrangement is to characterize the rearrangement process through its manifestations in the repertoire of developing and mature lymphocytes. Repertoire characteristics such as the κ : λ ratio among B lymphocytes or the degree of allelic exclusion vs allelic inclusion (the simultaneous expression of two alleles of the same receptor chain gene in the same cell) have received much attention (6–11). Theoretical evaluation of models were suggested as explanations of repertoire characteristics, followed by experimental testing of the predictions of these models. These studies have yielded a detailed picture of how gene rearrangement interacts, causally and temporally, with the processes of receptor-based selection and how all these factors in turn influence the observed population dynamics of developing lymphocytes (12–14). There are, however, many pieces still missing in this fascinating puzzle.

The question of allelic exclusion is one of the most elusive problems related to lymphocyte repertoire development (15). Allelically included cells, in which both receptors are expressed on the cell surface, are potentially dangerous because the lower expres-

sion levels of each of the two receptors may lead to escape of an autoreactive receptor from central tolerance (16). Protein chains of B and T cell receptors vary in the degree of allelic exclusion or inclusion they exhibit. Gene rearrangement is linked to transcription, and chromatic remodeling markers such as DNA methylation and histone acetylation distinguish rearranged from unrearranged alleles (17–19), although the exact role of these markers in induction of allelic exclusion remains to be elucidated (20–22). Expression of the pre-B cell receptor (BCR)³ on the cell surface has been shown to be required for the induction of heavy chain allelic exclusion (23), although neither of the components of the surrogate light chain $\lambda 5$ or VpreB, is required in itself for induction of allelic exclusion (24, 25), and a similar view is held for T cells (26).

A related problem is whether gene segment choice for rearrangement is random or biased, e.g., whether downstream (D-proximal) V segments or upstream J segments are preferably rearranged first. Rearrangement bias may be related to the accessibility of the particular DNA segment to the recombination machinery (27). Data from different systems are conflicting, and largely obscured by differences between the recombination signal sequences of individual gene segments (28–34), which affect the cleavage efficiency of the V elements by the recombination-activating gene complex (35). Stepwise activation of the heavy chain V gene locus may also play a role (36).

Our approach to the complex problem of lymphocyte repertoire development is to formulate mathematical and computational models of the processes of gene rearrangement, receptor-based selection, and the overall population dynamics of developing lymphocytes and to use these models to evaluate the proposed explanations for experimental observations. The first study in this series (37) dealt with the apparent contradiction between continuous rearrangement in the Ig light chain and the concept of allelic exclusion, and showed that the two may be reconciled by a certain

Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel.

Received for publication September 10, 2001; accepted for publication October 23, 2002.

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

¹ This work was supported in part by Israel Science Foundation-The Dorot Science Fellowship Foundation Grant 759/01-1, The Yigal Alon Fellowship, and a Bar-Ilan University internal grant (to R.M.).

² Address correspondence and reprint requests to Dr. Ramit Mehr, Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel. E-mail address: mehrra@mail.biu.ac.il

³ Abbreviations used in this paper: BCR, B cell receptor; DP, double productive; HCDP, heavy chain double productive; κ DP, κ double-positive cell.

degree of order in gene segment choice of rearrangement and by the fact that the number of light chain rearrangement attempts a developing B cell is allowed to perform is limited to two to three attempts. The same characteristics also served to give a full explanation of the observed $\kappa:\lambda$ ratio among mature B lymphocytes. The second study (38) dealt with T lymphocyte development and showed how the observed fraction of T cells containing productive TCR α rearrangements on both alleles can be explained by the relative permissiveness of thymic positive selection, acting on cells with a much higher potential for secondary rearrangements than that of B lymphocytes.

This paper uses similar models to address the following questions (which are explained in more detail in the following paragraphs). (1) Is the feedback model sufficient to explain heavy chain allelic exclusion? (2) Is heavy chain rearrangement simultaneous or ordered (first one allele, then the other)? (3) Is light chain rearrangement ordered (first one allele, then the other), or can successive rearrangements be made on either light chain allele at random? (4) What determines the usage of light chain J segments in rearrangement? And, finally, what is the degree of allelic exclusion in κ light chains? We use the first (to our knowledge) computational model of Ig gene rearrangement that accounts for both heavy and light chain gene rearrangement and for all segments of all Ig genes (39). The model and simulation methodology are described in detail below.

Heavy chain

This paper presents, to our knowledge, the first computational study of BCR heavy chain gene rearrangement, using our model which was first briefly described in Ref. 39. Allelic exclusion in the heavy chain has been considered to be absolute, since very few, if any, mature B lymphocytes express two heavy chains (reviewed in Ref. 40). However, no active mechanism preventing parallel V(D)J recombination on both heavy chain alleles has been found. This has led to the suggestion that heavy chain allelic exclusion is the result of the rapid positive selection of any cell that has managed to productively rearrange one heavy chain gene (i.e., all segments are rearranged in the same reading frame) and express a heavy chain that pairs well with the surrogate light chain (41). This is called the feedback model, evidence for which is reviewed in Ref. 9, and has become the current dogma. Rapid positive selection would result, according to this model, in down-regulation of the recombination-activating genes *RAG-1* and *RAG-2*, before rearrangement on the second allele has been completed. This model can only be valid if rearrangements, at least V to D-J, are not simultaneous. However, nonsimultaneity of heavy chain rearrangements has yet to be proved. In light chains, there is evidence that DNA methylation and histone acetylation, and hence replication status, are not identical in the two alleles (17, 42) before rearrangement; the same may be true in heavy chains. On the other hand, heavy chain D-J rearrangements occur more or less simultaneously. In light of these contrasting types of evidence, we set out to use our model to predict the possible outcomes of the two scenarios, ordered vs simultaneous heavy chain gene rearrangement.

If we denote the probability for a rearrangement to be productive by $P_{product}$ and the probability that a heavy chain pairs well with the surrogate light chain by $P_{pairsurr}$, then the probability of success in the first rearrangement attempt is $P_{product} \times P_{pairsurr}$. Failure in rearrangement of the first allele or failure in the pairing of the resulting chain with the surrogate light chain would preclude this rapid positive selection, enabling the cell to complete the rearrangement on the second allele. (In the heavy chain, only one rearrangement attempt is possible on each allele, because rearrangement deletes all D segments except for the one participating

in the rearrangement process.) This model, which assumes that the two alleles are rearranged one after the other with sufficient time for selection between completion of rearrangement of the first allele to that of the second, can be referred to as the ordered model of heavy chain gene rearrangement (Fig. 1A). It may be contrasted with a synchronous model, in which rearrangement on the two alleles proceeds in parallel (Fig. 1B), or with a more likely hybrid model in which a cell has a certain probability of performing the two rearrangements in parallel. In this case, eventual testing will be based on the status of both alleles, and the probability that a cell passes the test becomes $P(\text{pass due to either allele}) = 1 - P(\text{fail on both alleles}) = 1 - (1 - P_{pairsurr})^2 = 2P_{pairsurr} - P_{pairsurr}^2$.

Under the estimates for the probability for a rearrangement to be productive, $P_{product} = 1/3$ (one of three possible reading frames), and for the probability that a heavy chain pairs well with the surrogate light chain, $P_{pairsurr} = 1/2$ (31, 43), the strictly ordered model predicts that the fraction of mature B cells with complete heavy chain V(D)J rearrangements (productive or nonproductive) on both alleles will be ~45% and that the fraction of mature B cells with productive rearrangements on both heavy chain alleles will be ~9%. Even taking into account the existence of V pseudogenes does not reduce this predicted value by more than a factor of 2. In some of the latter cells, the proper light chain, which will later be rearranged and expressed by the cell, may pair well with both heavy chains, resulting in phenotypic heavy chain allelic exclusion in most (23) but not all cells.

Contrary to the latter prediction, however, only about one in 10^3 – 10^4 mature splenic B lymphocytes were found to express two heavy chains in the cytoplasm (23, 44). The same level of allelic exclusion was found even in mice triallelic for the Ig heavy chain

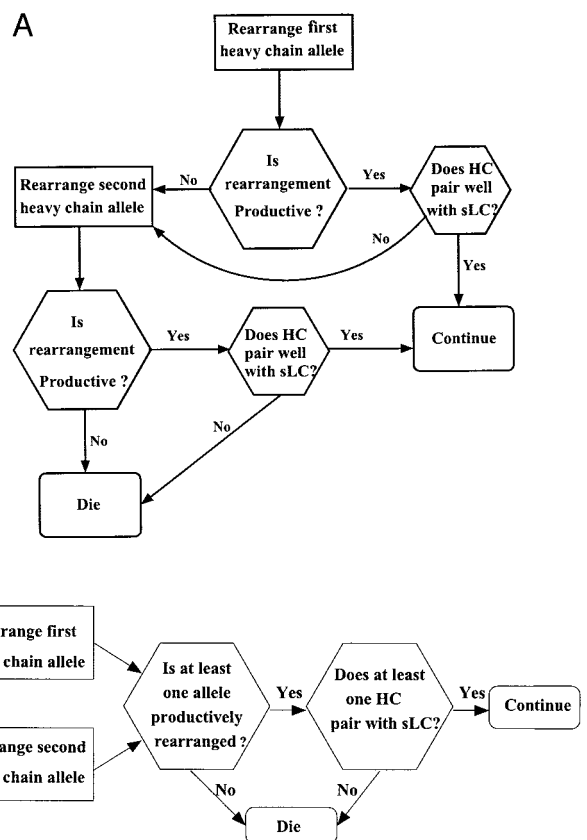


FIGURE 1. A, Ordered model of heavy chain (HC) gene rearrangement. B, Synchronous model of heavy chain gene rearrangement. sLC, Surrogate light chain.

locus (9). Hence, either there is an active mechanism preventing the simultaneous expression of two heavy chain alleles in most cells or there is some additional selection process, possibly in the peripheral lymphoid tissues, favoring cells that express only one heavy chain allele (23), as has been shown for B cells bearing two complete transgenic receptors (45).

Kitamura and Rajewsky (23) have also found that when one of the two heavy chain alleles has a disrupted μ membrane exon, such that it cannot be expressed on the cell membrane and transduce a selecting signal, the fraction of heavy chain double-productives (HCDPs) is larger than expected for this situation. The double-productive (DP) fraction in newly generated B lymphocytes in the bone marrow was expected to be 12% but was much higher, between 20 and 25%. This was explained as an artifact. However, as we show in *Results* this result could be completely accounted for by the inability of the disrupted μ chain to be expressed on the cell surface with the surrogate light chain, which creates a situation equivalent to that of semisynchronous rearrangements. In this regard, it is interesting that $\sim 1\%$ of human T cells express two TCR β chains on their surface (46). Yet, in this case, as with murine peripheral B cells, the effects of peripheral selection cannot be separated from those of putative allelic exclusion mechanisms.

Thus, the following two questions have not as yet been completely answered. 1) Is the rapid selection (feedback) model sufficient to explain heavy chain allelic exclusion, or must other mechanisms, such as competition between two chains for light chain binding (16) or peripheral preference of single-receptor cells, be evoked to explain peripheral B cell allelic exclusion? 2) Does heavy chain rearrangement proceed simultaneously on both alleles or in an ordered manner (i.e., first one allele is rearranged, then the other)? Additionally, the exact value of P_{pairsurr} , the probability that a heavy chain pairs well with the surrogate light chain, has not been sufficiently established. It has been estimated as one-half (31, 43), but this value was based on the pairing of only 14 of 33 heavy chains with the surrogate light chain, which gives about 0.4. We set out in this study to examine the degree of sequentiality or synchronicity in heavy chain gene rearrangement and to make some progress towards the evaluation of P_{pairsurr} . We have dealt with this issue by extending our previous computational model (37) to include all gene rearrangements on the heavy chain (which has not been dealt with at all in our previous studies) as well as the light chain (39). Below, we calculate the predicted fraction of DPs in newly generated B cells (if there is no selection against DPs, and no active mechanism silencing one heavy chain allele), for values of P_{pairsurr} between 0.1 and 1, and values of P_{testpair} , the probability of being tested for heavy chain functionality after the first rearrangement, between zero (a strictly synchronous model) and 1 (a strictly ordered model). Comparisons of the results of our calculations to experimental data enable us to draw several conclusions concerning heavy chain rearrangement order and allelic exclusion.

Light chain

The degree of allelic exclusion in light chains has also been a topic of much debate, especially in light of the high degree of inclusion in TCR α chains (reviewed in Ref. 38). Experimental studies have in many cases used mice transgenic for autoreactive receptors, which shed light on tolerance mechanisms but do not give an estimate of the degree of allelic inclusion in normal situations (see, e.g., Refs. 16 and 47). To approach this issue, one must address two relevant subquestions: 1) Is light chain rearrangement ordered (first one allele, then the other) or can successive rearrangements be made on either light chain allele at random? 2) What determines the usage of light chain J segments in rearrangement? In the

present study, we directly address light chain allelic inclusion by using the first computational model which addresses and enumerates B cells maturing with two productively rearranged light chain genes (39).

An older study (37) dealt only with light chain gene rearrangement, taking into account only J segments because they are the limiting factor in Ig light chains. Results of simulations, incorporating various assumptions on the probabilistic choices cells make during rearrangement, were compared with published data on the $\kappa:\lambda$ ratio in mature B cells, on the fractions of mature B cells with rearrangements on one or two alleles, and on the relative frequencies of J κ segments in these rearrangements. These comparisons led to the following conclusions. First, to explain the observed $\kappa:\lambda$ ratio of $\sim 20:1$ in mature murine B cells, it is necessary to assume both a preference of choosing κ over λ in rearrangement and a certain probability of cell death after each nonproductive rearrangement. This death probability limits the number of rearrangement attempts a cell may go through, which was suggested as an explanation for light chain allelic exclusion. Second, a bias toward using upstream vs downstream J κ segments in rearrangement was indicated. Although this seems to allow the cell to perform a larger number of rearrangement attempts before exhausting J κ segments on both alleles (relative to the random case), the probability of death after each rearrangement failure does not allow the cell to maximize the usage of its pool of eight J κ and four J λ segments. Third, the possibility that the cell prefers to remain on the same allele was examined. The conclusion supporting such a preference was later shown to be incorrect (39, 48) and was not supported by experimental data (49) or by our studies on TCR gene rearrangement (38).

Although the above study evaluated many of the parameters that play a role in determining the fate of a developing B lymphocyte, it was not conclusive on the point of light chain allelic exclusion for several reasons: 1) allelic exclusion was built into the older simulation (cells with two productive light chain rearrangements were killed by default), and hence the degree of allelic inclusion as function of the parameters could not be evaluated; 2) only the two extreme cases of J κ bias were studied, either no bias at all or an almost absolute sequentiality in segment usage, although the actual magnitude of this bias probably lies somewhere between the two extremes; 3) the contribution of V segments was not considered. Although it is obvious that the numbers of J segments is much more limiting than the much higher number of V segments, it remained to be proved that V segment choice does not limit the rearrangement process, and it remained to be seen whether there are biases in V segment choice (28–33).

Given that many new data have recently become available (Ref. 50; reviewed in Refs. 15–24), we set out in our recent and present studies to clarify whether the extent of allelic exclusion in the BCR light chain can or cannot be explained by probabilistic factors alone. Our approach is to simulate gene rearrangement and the subsequent selection of developing B cells under various values of the underlying probabilities and generate the predicted composition of the mature B cell repertoire under each set of probabilities. Wherever experimental data are available, we compare our results with these data and draw conclusions on the underlying mechanisms. Where data are not available, our results point at the parameters that are worth measuring.

Methods: Simulation and Parameters

Our model is a unified model of both rearrangement processes (39). We have chosen to use an individual-based stochastic (Monte Carlo) model, rather than direct probabilistic calculations, for the following reasons. First, a model that follows the structure of the gene loci and the processes of

rearrangement in detail is easier to implement; hence, it is reliable. Second, this type of model would be easier to generalize as more features of the structure of gene loci or the rearrangement process are discovered by experimental study. Thus, this type of simulation has a larger potential for future extension and use. Most important, however, is the relationship between model results and experimental data. Most experimental data published thus far on repertoire characteristics such as the fractions of cells with only one allele rearranged or productively rearranged, or the distribution of $J\kappa$ segment usage in κ -expressing cells, have reached their conclusions based on very small numbers of cells (e.g., Refs. 15–23). In contrast, the simulations in our previous studies have shown that these repertoire characteristics are highly variable between samples of small numbers of cells (37); if any repertoire characteristic is plotted against the number of cells (or heavy chain clones) simulated, an order of 10^4 or even 10^5 cells must be simulated to obtain a reliable result. In our studies, simulation results (e.g., the $\kappa:\lambda$ ratio) were always plotted as function of the number of cells or the number of lineages simulated to find the order of magnitude of cell numbers for which the results have reached a fixed point (other than a small amount of noise due to the inherent stochasticity). Numbers of cells of the order of 10^5 or numbers of clones of the order of 10^3 were usually adequate (Fig. 2).

Thus, a Monte Carlo-type model can, on one hand, generate accurate predictions of relevant repertoire characteristics from a larger number of simulated cells than is experimentally tractable, as we do below for estimating the fractions of κ -expressing cells with productive rearrangements on both κ alleles. On the other hand, this model enables us to evaluate the statistical reliability of small sets of experimental data by simulating small rather than large numbers of cells and calculating the mean and SD of several repeats of the simulation. We can thus deduce the ranges of simulation parameters that give results sufficiently close to the experimental data. The results shown in this paper are, in most cases, means and SDs of 5 simulations of 500 cells each (of which usually between 50 and 100 cells survive and mature, depending on parameter values). These results were usually the same as those obtained from simulations of 10^5 cells; however, their variability was much higher (see below).

The following is a detailed description of our computer simulation of Ig heavy and light chain gene rearrangement and receptor-based selection, first published in Ref. 39. In each run of the simulation, we create many clones of cells and follow their development. A “cell” is an object, which is characterized by the set of probabilities determining what the cell will do next (divide, die, rearrange its heavy or light chain genes and so on). Each cell contains a “genome” composed of six loci (two heavy chain alleles, two κ and two λ light chain alleles). A heavy chain locus is composed of three arrays of gene segments corresponding to the V, D, and J segments. A light chain locus is composed of two arrays of gene segments representing V and J segments. Each segment is characterized by the probability that it will be chosen for rearrangement in the next rearrangement attempt. Deleted segments, or segments currently rearranged, are thus assigned a rearrangement probability of zero, and the probabilities for rearrangement of the remaining segments are normalized accordingly after each rearrangement.

This model is obviously a simplified picture of the actual heavy and light chain loci, the structures of which are known in great detail, especially for the murine κ locus (51). First, we neglect the issue of V gene orientation,

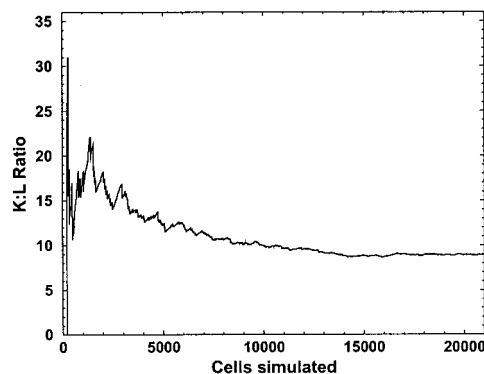


FIGURE 2. Stabilization of the $\kappa:\lambda$ (K:L) ratio among cells leaving the bone marrow, as a function of the total number of cells simulated. It is evident that simulations of $<15,000$ cells would have led to the wrong final result, given that this ratio stabilizes only after 15,000 cells have been simulated.

because the numbers of J segments, rather than of V segments, are the limiting factor, as will be shown below. Second, we neglect the existence of pseudogenes. Taking pseudogenes into account is equivalent to using a value lower than $1/3$ for the probability for a rearrangement to be productive. Reducing this value does not significantly change the results shown here (data not shown).

The simulation flow for each cell is as follows (Fig. 3).

Cell birth

Cells are born into the simulation having all heavy and light chain genes in the unrearranged (germline) configuration. Probabilities for each segment to be chosen for rearrangement in the first attempt are assigned according to the rearrangement bias (e.g., a bias toward rearranging V-proximal J segments first). If there is no bias, then the probabilities are equal for all segments in a given locus.

Heavy chain rearrangement (Fig. 3A)

The cell first selects randomly the heavy chain allele to rearrange, with equal probabilities assigned to both alleles. Then, the cell selects the V, D, and J segments that will be rearranged on this allele. This selection is also probabilistic and uses the rearrangement probabilities assigned to each segment, as described above. The next decision is whether the rearrangement was productive; the probability for a rearrangement to be productive is denoted by P_{product} . If the rearrangement is nonproductive, the cell cannot continue rearranging on the same allele. Hence, the cell attempts to rearrange the other allele. If the rearrangements on both alleles were nonproductive, the cell dies, and a new cell is born.

Synchronicity vs order in heavy chain rearrangement (Fig. 3A)

The parameter P_{testpair} determines the probability that, after the first attempt of heavy chain gene rearrangement and if this rearrangement is productive, the cell will test the resulting heavy chain for pairing with the surrogate light chain. If the cell, rather than performing this test after the first rearrangement, first makes a rearrangement attempt on the second allele, then the two rearrangements can be regarded as synchronous. Thus, $P_{\text{testpair}} = 1$ corresponds to the standard, strictly ordered model, whereas $P_{\text{testpair}} = 0$ corresponds to a fully synchronous model. We have examined intermediate values of this parameter as well (see below).

Pairing test (Fig. 3A)

After a productive heavy chain rearrangement (or two simultaneous rearrangements), the simulation tests the heavy chain(s) for pairing with the surrogate light chain. Success here is, again, a probabilistic event, with a probability of success P_{pairsurr} per heavy chain, as explained above. If the cell passes this test, its probability for further rearrangement and its probability of death are set to zero, and its probabilities for cell divisions or further differentiation are set to positive values that are also parameters of the program.

Cell divisions (Fig. 3A)

If heavy chain rearrangement succeeds, there is a probability that the cell will perform a number of subsequent cell divisions, in which a clone of cells with the same heavy chain rearrangement is created (52). Each cell from this clone is now followed during its subsequent (independent) development. Cell divisions in themselves do not have any effect on the rearrangement process. Nevertheless, we have included cell divisions in our simulation as part of the attempt to create a relatively realistic model. This would enable us to estimate parameters such as the fraction of B cells that actually mature (see below), and the influence of cell divisions, if any, on the variability of the generated repertoire (future work). A cell that has ceased to divide proceeds to the light chain rearrangement stage.

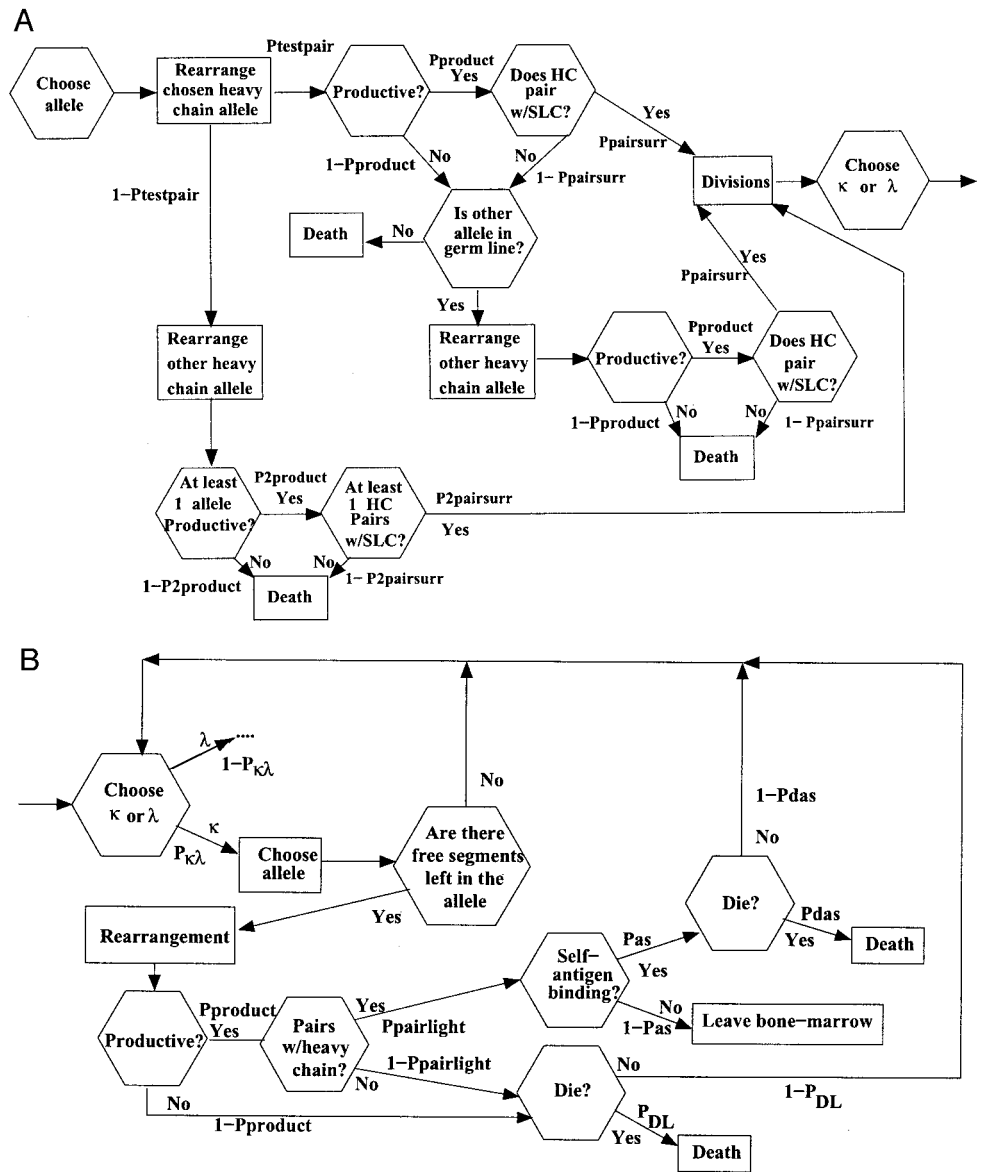
Light chain isotype selection (Fig. 3B)

At each light chain rearrangement step, the isotype to be rearranged, κ or λ , is first selected, according to the selection probability $P_{\kappa\lambda}$, i.e., the cell chooses κ with probability $P_{\kappa\lambda}$ or λ with probability $(1 - P_{\kappa\lambda})$.

Light chain allele selection (Fig. 3B)

After the isotype has been selected, then one or the other allele is randomly selected for rearrangement. The cell may either rearrange the previously rearranged allele or switch to the opposite allele, depending on the probability parameter P_{switch} . If $P_{\text{switch}} = 0$, then the cell rearranges the previously rearranged allele, provided there are J_{κ} (and V_{κ}) segments remaining on it. We call this situation absolute allele preference. If $P_{\text{switch}} = 0.5$, then the choice of the next allele to be rearranged is completely random (no

FIGURE 3. Structure of the simulation algorithm. Each time a new cell is created, it begins with the first heavy chain gene rearrangement. The subsequent development of this cell and all its daughter cells is then followed. **A.** Heavy chain rearrangement in the simulation. The cell selects randomly the first heavy chain allele to rearrange. Two rearrangement attempts are allowed, one for each heavy chain allele, which are performed either sequentially or synchronously. Failure (nonproductive rearrangement or the production of a heavy chain (HC) that does not pair well with (w/) the surrogate light chain (SLC)) in both attempts leads to cell death. Success in at least one rearrangement leads to several cell divisions (52). Following divisions, each cell in the resulting clone proceeds to perform light chain rearrangement, starting with isotype choice (“Choose κ or λ ”); the part of the algorithm that deals with light chain rearrangement is depicted in **B.** **B.** Light chain rearrangement in the simulation. In each rearrangement, first the isotype (κ or λ), then the allele, and then the segments to be rearranged are chosen. After the choice of isotype, the process is similar whether κ or λ was chosen and hence for the sake of simplicity is shown only for κ . The result of each rearrangement is tested for productivity and, if productive, for pairing with the heavy chain of the cell. If the rearrangement passes these tests, meaning that the cell can express the full BCR on its surface, it undergoes negative selection; surviving cells are allowed to mature, whereas failing cells either die or edit their receptors.



allele preference), with equal probabilities to choose either allele, and independent of previous rearrangements. If a cell has run out of J_{κ} segments on one allele, then it automatically switches to the other allele. If a cell is entirely out of J_{κ} segments on both alleles, then it switches to λ rearrangement, and vice versa, until all segments are exhausted. The cell dies if there are no more light chain segments available for rearrangement.

The possibility of simultaneous light chain rearrangements on different alleles and/or isotypes is not included in the simulations; evidence suggests that this is a rare event (17, 42), and inclusion of these possibilities in the simulations would complicate them much further. The implications of this choice on the interpretation of our results are discussed below.

Rearrangement of a κ allele

Once a κ allele has been selected, one V_{κ} segment and one J_{κ} segment are chosen for rearrangement, using the probabilities assigned to each segment. Each rearrangement is followed by renormalization of rearrangement probabilities of the allele in question, i.e., a rearranged J_{κ} segment is deleted (as are all unrearranged J_{κ} segments 5' relative to the rearranged segment) and cannot be chosen again by the program. Similar deletions occur for V_{κ} segments. We thus neglect the possibility of inversional rearrangements, in spite of the high fraction of V genes placed in the inversional orientation observed in the murine κ locus (51), under the assumption that the only difference between deletional and inversional rearrangements is that the latter do not delete the intermediate V gene segments. Given that the number of V segments is not limiting (an assumption that we have verified in

the simulations; see below), neglecting inversional rearrangements does not have any significant effect on our results and conclusions.

Rearrangement of a λ allele

If λ is selected, then one of the two λ alleles is chosen at random, with equal probabilities for the two alleles. V_{λ} and J_{λ} segments on the current allele are chosen at random. Failure in λ rearrangement leads to deletion of the rearranged segment only and has no effect on any other segments.

Light chain testing and negative selection

Once a V-J light chain rearrangement is made, the program determines whether the rearrangement is productive (V and J are in the same reading frame), with a probability P_{product} . If the rearrangement is productive, the program determines whether the resulting light chain can pair with the existing heavy chain, with a probability $P_{\text{pairlight}}$. If so, the program determines whether the resulting BCR is autoreactive (anti-self), with a probability P_{as} . Rearrangement is repeated until a productive, heavy-light chain-matched, nonautoreactive BCR is produced, or until the cell dies (see below).

Cell fate after selection

Results of rearrangement are classified into one of three categories:

1. Cells containing an anti-self rearrangement. These cells are assigned a high death probability, denoted by P_{das} . If, however, a cell does not

die after such a rearrangement (an event that occurs with a probability $(1 - P_{das})$), it may try another rearrangement (receptor editing; Refs. 4–55). In reality, the fate of these cells may be affected by the site of antigen encounter (56), the number of such encounters (57), receptor expression levels (58, 59) and the affinity of the receptor for the negatively selecting Ag (47).

2. Cells containing only out of frame V-J joins, heavy-light-mismatched heavy-light chain pairs, or germline light chain alleles. These cells are assigned a moderate death probability, P_{DL} . If such cells do not die, they also continue rearranging their light chain genes.
3. Cells that have arrived at a productive, heavy-light matched, nonautoreactive rearrangement are allowed to mature. The simulation tallies the composition of the repertoire of those cells that have been allowed to mature.

The simulation and a user's manual are available from the authors on request (the current version is a C program, which can run on any computer equipped with a C compiler. We used a Unix computer (Silicon Graphics).

Parameters

Table 1 summarizes the default parameter values used in our simulations. Parameter definitions, and justifications of the choice of default values in case the parameters were not varied in the simulations, are as follows.

1. N : The number of cells created in each run of the simulation. We found that 10^5 cells give a sufficient number of mature cells for stabilization of repertoire properties in most simulations (37, 38), as demonstrated in Fig. 2. Because several cells may belong to the same heavy chain clone, the number of individual clones simulated was also important; in the simulations with a total of 10^5 cells, the number of clones was limited to 4000. Other simulations were done with only 500 cells but were repeated several times (as discussed above).
2. $P_{product}$: The probability for a V-J rearrangement to be productive is one-third, because this is the probability that V will be joined to J in the correct reading frame. Incidentally, $P_{product} = 1/3$ for V-D-J rearrangements as well, because in most cases the D segment can be read in all three reading frames. We neglect here the effect of the existence of pseudogenes (51), which would only slightly decrease $P_{product}$.
3. $P_{testpair}$: The probability to perform a test of pairing with the surrogate light chain after the first heavy chain rearrangement, which is one of the parameters under study here (see below). $P_{testpair} = 0$ corresponds to the fully synchronized model. $P_{testpair} = 1$ corresponds to the fully ordered model. Intermediate values were used as well.
4. $P_{pairsurr}$: The probability of pairing of the heavy chain with the surrogate light chain (31, 43). This parameter is studied below in detail.
5. $P_{pairlight}$: Probability of a light chain to pair with the heavy chain. $P_{pairlight}$ was chosen to be 0.8, because previous studies (60) have shown that this is the minimum value required to give a ratio of $\kappa:\lambda > 2$; we confirmed this choice in our previous study (39), where

this value gave $\kappa^{+/0} \approx 0.68$, confirming past choices (Refs. 37 and 60 and data not shown).

6. P_{DH} : The probability of cell death after an unsuccessful heavy chain gene rearrangement (this parameter is relevant only in the strictly ordered model). Death due to failed rearrangements may result from unresolved double-stranded DNA breaks.
7. P_{DL} : The probability of cell death after an unsuccessful light chain rearrangement or failure in light chain pairing with the heavy chain. The default value was found by comparing simulation results to experimental data on the fraction of cells with only one κ allele rearranged ($\kappa^{+/0}$), which has been found to be $\sim 68\%$ (7, 61). In our previous studies, this fraction was obtained for κ -expressing B cells only with $P_{DL} = 0.2$, with no significant dependence on the values of $P_{\kappa\lambda}$ or $J_{\kappa bias}$ (Ref. 39 and data not shown). Hence, we chose $P_{DL} = 0.2$ as the default value for the rest of the study.
8. P_{div} : The probability of divisions after heavy chain selection. This probability is defined per division; i.e., after a successful heavy chain gene rearrangement or after a cell division, the cell will divide again with a probability P_{div} or stop dividing, differentiate further, and rearrange its light chain genes, with a probability $(1 - P_{div})$. The default value was taken from our previous study (37).
9. $P_{\kappa\lambda}$: The probability of choosing κ over λ light chain for rearrangement. The default value was determined by comparing simulation results to experimental data on the ratio between cells expressing κ and cells expressing λ light chains. The observed minimal $\kappa:\lambda$ ratio in murine B lymphocytes is between 5:1 and 10:1 in immature B cells and 10:1–20:1 in peripheral mature B cells (62, 63). We previously found that high values of this parameter ($P_{\kappa\lambda} \geq 0.95$ are necessary to obtain high $\kappa:\lambda$ ratios in our simulations. In the current simulations, with $P_{DL} = 0.2$, the value of $P_{\kappa\lambda} = 0.975$ (39) gave a $\kappa:\lambda$ ratio of > 10 . This reconfirmed our conclusion that a strong preference for choice of κ over λ for a light chain gene rearrangement is necessary to explain the observed $\kappa:\lambda$ ratios; however, it is not necessary to assume that the order of isotypes to be rearranged is preprogrammed (50, 63, 64).
10. P_{switch} : The probability of switching between light chain κ alleles or λ alleles.
11. $J_{\kappa bias}$: The strength of the J segment bias, that is, the preference of the cell to rearrange V-proximal J segments. The exact definition is:

$$J_{\kappa bias} = \frac{P(J_{\kappa 2})}{P(J_{\kappa 1})} = \frac{P(J_{\kappa 4})}{P(J_{\kappa 2})} = \frac{P(J_{\kappa 5})}{P(J_{\kappa 4})}$$

where $P(J_{\kappa i})$ is the probability to choose $J_{\kappa i}$ for rearrangement (the simulation takes into account the deletion of already used segments and intermediate segments). Hence, the smaller the value of $J_{\kappa bias}$, the stronger is the bias. In the previous study, this parameter was either set to $J_{\kappa bias} = 1$, which corresponds to a completely unbiased gene segment choice, or to $J_{\kappa bias} = 0.001$, which gives a very strong, practically absolute, preference toward choice of upstream segments. Here, we used several intermediate values as well.

12. $V_{\kappa bias}$: The strength of V segment bias, defined in a way similar to that of $J_{\kappa bias}$, with the bias favoring rearrangement of J-proximal V segments.
13. P_{as} : The probability of a cell to be anti-self with its current receptor. The value of P_{as} has been independently estimated to be about $2/3$ by others (65, 66).
14. P_{das} : The probability of cell death if the cell was found to be anti-self. The default value was taken from our previous study (37).

The program follows each new cell, or cells produced by cell division after heavy chain pairing with the surrogate light chain, until it either matures (leaves the bone marrow after surviving all selection stages) or dies, and repeats the process for the predetermined number of cells. The program generates as output the distribution of genotypes among the cells that have matured.

Results

Heavy chain

In this section, we examine the issues of heavy chain allelic exclusion and pairing with the surrogate light chain. The heavy chain composition of the mature B cell repertoire is assumed to be independent of the parameters governing light chain gene rearrangement and selection, for which the default values were used in all the simulations described in this section.

Cells with rearrangements on both heavy chain alleles. We plotted the fraction of newly generated B cells with two heavy chain

Table 1. B cell simulation parameters with our default set of values^a

Parameter	Value Range	Default Value
N	10^2 – 10^7	500 or 10^5
$P_{product}$		0.333
$P_{testpair}$	0–1	1
$P_{pairsurr}$	0–1	0.5
P_{DH}	0–0.5	0
P_{div}		0.2
$P_{\kappa\lambda}$		0.975
P_{switch}	0, 0.5	0.5
$P_{pairlight}$	0.5–1	0.8
P_{DL}	0–0.5	0.2
P_{as}	0.5–0.9	0.67
P_{das}	0.3–0.9	0.7
$J_{\kappa bias}$	0.001–1	1
$V_{\kappa bias}$	2/3–1	1

^a These parameters were used in all simulations unless otherwise noted. See text for justification of the default values. In case a parameter was varied in the simulations described here, the range of values is also indicated.

gene rearrangements (HC-DRs), regardless of whether or not the nonexpressed rearrangement is productive, as a function of $P_{testpair}$, the probability to undergo heavy chain selection after the first heavy chain gene rearrangement, and as function of $P_{pairsurr}$, the probability that a heavy chain pairs well enough with the surrogate light chain to be recognized as functional by the cell (Fig. 4A). The fraction of these “double-rearranged” cells decreases rapidly with $P_{testpair}$, as expected, because the higher the value of $P_{testpair}$, the higher is the probability that the cell will not perform a second rearrangement due to success in the first attempt. However, the fraction of “double-rearranged” cells is almost independent of $P_{pairsurr}$, especially for low values of $P_{testpair}$.

For splenic B lymphocytes, the fraction of B cells with rearrangements to both J_H fragments was between 80 and 100% (61). However, this does not necessarily mean that V(D)J rearrangement was completed on both alleles, although it indicates that rearrangement may have started as a synchronous process. The fraction of heavy chain “double V(D)J rearranged” cells among pre-B cells was later measured and found to be close to 50% (29–40). This corresponds to values of $P_{testpair} > 0.8$, i.e., supporting an almost strictly sequential model of heavy chain gene rearrangement, in agreement with current models.

The results presented here (and below, unless stated otherwise) are from simulations of 500 cells each. This is the total number of cells simulated, including cells that died during the simulation in addition to those that matured. The fraction of surviving cells depends on the various parameters but usually does not exceed 10–20%. Hence, these simulations are equivalent to experiments in terms of the number of cells sampled. Shown are the mean and SD of the results of five simulations for each parameter set. It is evident that the variability between simulations, which we regard as an estimate for the variability between experimental repetitions, is large, but not so much so as to mask the effects of parameter changes.

Cells with productive rearrangements on both heavy chain alleles. Next, we plotted the fraction of cells with two productive heavy chain gene rearrangements (HCDPs) as a function of $P_{testpair}$ and $P_{pairsurr}$ (Fig. 4B). Experimental estimates of this fraction vary between 0.01 and 5%, depending on the measurement method (9, 23). As our simulation results show, even in a rigorously ordered model, with $P_{pairsurr}$ values ~ 0.5 , the expected fractions of HCDPs are higher than those measured in peripheral B cells. To obtain an HCDP fraction below 5%, $P_{pairsurr}$ must be higher (almost all heavy chains pair with the surrogate light chain ($P_{pairsurr} = 1$). The latter case is unlikely, given that it has already been shown in mice that the pairing probability is lower than or close to one-half (31, 43). Moreover, even in a strictly ordered model, which also includes a probability P_{DH} for cell death after failure in the first heavy chain gene rearrangement, the fractions of HCDPs are higher than measured in peripheral B cells (Fig. 4C).

The fraction of HCDPs in pre-B cells was $\sim 7\%$ (40), higher than most estimates of the same fraction in splenic B cells. Unless this difference is completely attributable to measurement error, it may suggest that some additional selection mechanisms act against cells expressing two heavy chains, even in the cytoplasm, as will be discussed below. On the other hand, HCDP fractions of $\sim 25\%$, as had been measured for heavy chain mutant (μ MT) mice (23), can be obtained in our simulations for values of $P_{testpair}$ below 0.5 and values of $P_{pairsurr}$ of 0.5 or lower. Hence, the measured value is neither an artifact nor an indication of a priority for the targeted allele to be rearranged first, as suggested, but can be explained as follows. The targeted allele is rearranged first only half the time, but because the heavy chain produced from this allele cannot be

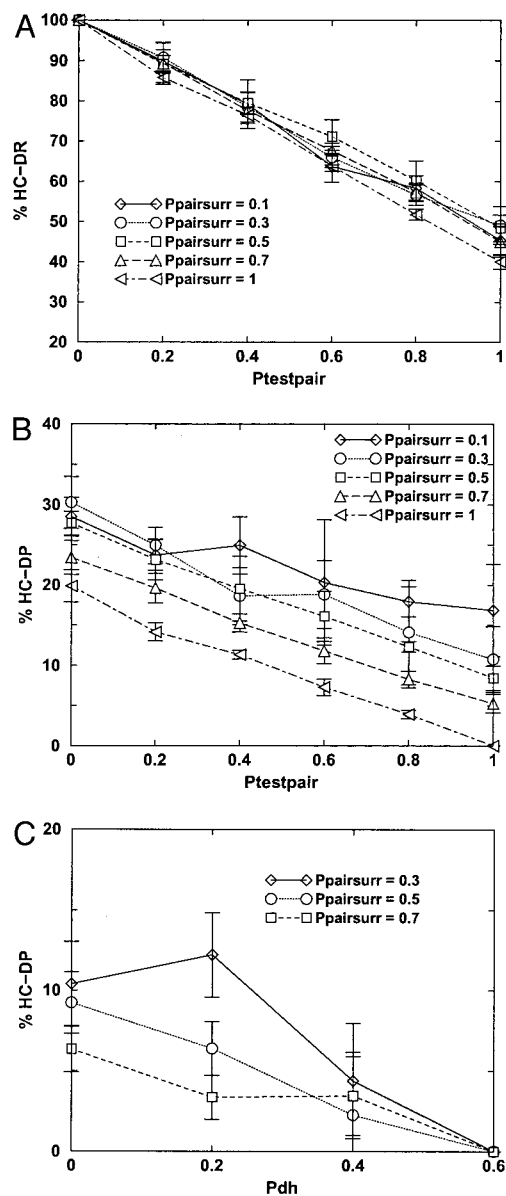


FIGURE 4. A, Fraction of cells with both heavy chain alleles rearranged. This fraction is shown as a function of $P_{testpair}$, the probability of being tested for heavy chain functionality after the first rearrangement, which varied between zero (a strictly synchronous model) and 1 (a strictly ordered model), and of $P_{pairsurr}$, the probability that a heavy chain pairs with the surrogate light chain, which varied between 0.1 and 1. B, Fraction of cells with both heavy chain alleles productively rearranged. This fraction is shown as a function of $P_{testpair}$, which varied between zero (a strictly synchronous model) and 1 (a strictly ordered model), and of $P_{pairsurr}$, which varied between 0.1 and 1. C, Fraction of cells with both heavy chain alleles productively rearranged, in a strictly ordered model ($P_{testpair} = 1$), vs P_{DH} , the probability of death after failure in the first heavy chain rearrangement (HC-DR). Results are plotted for several values of $P_{pairsurr}$.

expressed on the cell surface, in these cases there is no testing of pairing; i.e., in 50% of the cases, the cells behave as if rearrangement of the two alleles was synchronous, which is the meaning of $P_{testpair} = 0.5$.

Light chain

J_κ usage. As observed in our previous study (37), if there is no bias toward upstream J_κ segments ($J_{\kappa bias} = 1$), then the fractions

of κ alleles with rearrangements to downstream J_κ segments actually exceed the fractions of κ alleles with rearrangements to upstream J_κ segments (Fig. 5). This is due to an “accumulation effect”: there are many more pathways leading to cell maturation with a rearrangement to $J_{\kappa 4}$ or $J_{\kappa 5}$ than to $J_{\kappa 1}$ or $J_{\kappa 2}$ (37). Evidence supporting this argument was found in human leukemias and myelomas, in which J_κ usage in λ -expressing cells (i.e., cells in which several κ rearrangements were performed before the cells resorted to rearranging λ genes) was indeed skewed towards $J_{\kappa 4}$ and $J_{\kappa 5}$ (67). On the other hand, the extreme case of $J_{\kappa bias} = 0.001$, which was used in the previous study, leads to an unrealistic distribution of J_κ segment usage among mature cells, with almost 80% of the alleles having rearrangements to $J_{\kappa 1}$. We find that $J_{\kappa bias} = 0.5$ gives results that are most similar to experimental observations (63, 68). Our results thus still support a bias in J_κ segment choice for rearrangement, but this bias is finite, so that J_κ segment choice need not be deterministic (as a bias of 0.001 effectively is).

J_κ usage depended only slightly on P_{DL} (not shown). Obviously, the higher the value of P_{DL} , the smaller is the chance that a cell will perform more than one rearrangement, and hence J_κ usage becomes more skewed toward upstream J_κ segments. However, this effect is marginal compared with that of the bias. Similar results were obtained in simulations of 100,000 cells (not shown).

A cell is allowed, on average, only two rearrangement attempts. Our results show an inverse correlation between the death probability P_{DL} and the average number of rearrangements per cell (Fig. 6). When $P_{DL} = 0.5$, a cell usually manages to perform only one rearrangement, which is unlikely to be the case. However, even for very low P_{DL} values, a cell does not perform, on average, more than four rearrangements. For $P_{DL} = 0.2$, which we consider most plausible, the average is about two rearrangements per cell. Similar results were obtained in simulations of 10^5 cells (39). It is remarkable that even in mice transgenic for autoreactive receptors, the average number of rearrangements per cell was only about three (16, 47). This result is hardly influenced by changes in the values of other parameters, in particular $P_{pairlight}$, P_{as} , and P_{das} (not shown). Hence at least one of the mechanisms responsible for allelic exclusion is the fact that the cell is not given the opportunity to perform too many secondary rearrangements.

This result confirms our previous conclusion that preference to κ rearrangement is only part of the explanation for the $\kappa:\lambda$ ratio. Rearrangement order within κ increases the potential number of κ

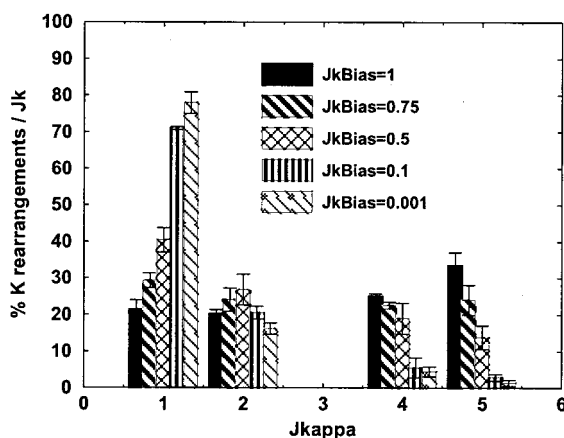


FIGURE 5. J_κ usage (percent of cells expressing a light chain with the indicated J_κ segment) as function of the value of $J_{\kappa bias}$. In these simulations, we used $P_{DL} = 0.2$ and $P_{\kappa\lambda} = 0.95$. The J_κ usage obtained with $J_{\kappa bias} = 0.5$ most resembles experimental results.

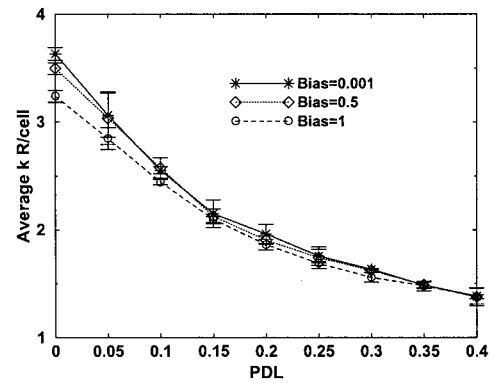


FIGURE 6. Average number of rearrangements ($k R$) per cell as function of P_{DL} and $J_{\kappa bias}$. For $P_{DL} = 0.2$, which we consider most reasonable, a cell performs, on average, two rearrangement attempts.

rearrangements the cell can perform and thus decreases the probability of exhausting the possibilities for κ rearrangement and moving on to λ rearrangement. At the same time, the low number of rearrangements per cell means that few cells survive to the point where the possibilities for κ rearrangement are exhausted and λ rearrangements are initiated. These combined factors contribute to the $\kappa:\lambda$ ratio by decreasing the chance that a cell will arrive at rearranging λ .

Our results do not support allele preference. We used P_{switch} , the probability of switching the allele to be rearranged, to reexamine the issue of allele preference: when $P_{switch} = 0.5$, the choice between the two alleles in each rearrangement attempt is random; when $P_{switch} = 0$, the cell continues to rearrange on the same allele until there are no more κ segments for rearrangement on that allele, and only then does the cell try to rearrange the other κ allele. We ran the simulation with P_{DL} between 0.01 and 0.4, using several values of J_κ bias, and the two values of P_{switch} (Fig. 7). For $P_{switch} = 0$, the fraction of cells that rearranged only one κ allele was much higher than the experimentally observed values of 68% (7, 61). The only exception occurred for very low values of P_{DL} , which we consider very unlikely based on the data presented

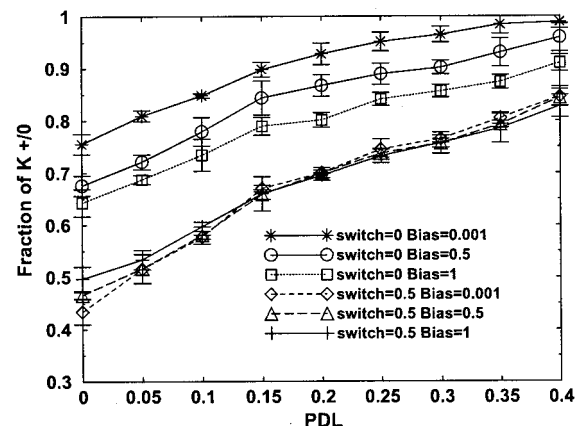


FIGURE 7. Fraction of cells with only one κ allele rearranged (κ^{+0}), as a function of P_{DL} , $J_{\kappa bias}$, and P_{switch} , the probability to switch alleles between rearrangement attempts. The random allele choice case ($P_{switch} = 0.5$) gives values similar to the observed 68% for $P_{DL} = 0.2$ and all $J_{\kappa bias}$ values. In contrast, the allele preference case (the cell continues to rearrange the same allele until exhausted, $P_{switch} = 0$) does not yield results that are similar to observations except for very low values of P_{DL} , which we consider unrealistic due to the observation that there is much cell death at the pro-B and pre-B cell stages.

above. Hence, our results do not support the assumption of allele preference in rearrangement; rather, the cell may rearrange any κ allele it chooses, and the choice is random. These results did not depend on variations of other parameter values (not shown) and were similar to those obtained in simulations of 10^5 cells (39).

The simulations with low values of P_{DL} may be regarded as reflecting a situation where the mechanisms of programmed cell death are defective, such as in Fas-deficient (MRL/*lpr*) mice. In this case, our simulations predict that the fraction of cells with rearrangements on both κ alleles may be as high as 50%.

V segments and V bias. The results reported above on J_κ usage and the average number of rearrangements per cell were qualitatively similar to those obtained in the J segment only study with similar parameter values (37), indicating, as expected, that V segments are not a limiting factor in the rearrangement process. To further verify that putative biases in V segment choice for rearrangement are also inconsequential, we introduced a bias in the choice of V segments toward downstream (J-proximal) V segments, similarly to the bias introduced for J segments, with values between 1.0 and $\frac{2}{3}$ (probability to choose each consecutively more upstream segment is $\frac{2}{3}$ that of the choice probability of the preceding segment). This change did not result in any significant alteration of the above results (data not shown).

Finally, all the results presented up to this point were insensitive to the values of the probability of light chain pairing with the

heavy chain, $P_{pairlight}$, the probability for an expressed BCR to be anti-self, P_{as} , and the probability for an anti-self cell to be deleted, P_{das} , unless otherwise noted.

The fraction of κ double expressors: implications for light chain allelic exclusion/inclusion. Allelic exclusion, or rather to which extent it is broken, can be measured directly in the present model (as in the T cell model (38)) by enumerating the cells that mature with two productively rearranged light chain alleles and counting how many of those cells have a potentially autoreactive BCR. This is in contrast to the previous model of BCR gene rearrangement (37), which did not directly address the question of allelic inclusion in B cells, because in that model, cells with two productive rearrangements were automatically deleted and not allowed to mature. That is, allelic exclusion was “built into” the model. Hence, we set out in the present study, where these limitations were not imposed, to check whether cells with two productively rearranged and expressed light chains can mature and in what frequency. These simulations were done with a large number of cells (10^5) to arrive at more precise estimates of the small fractions of double-productive cells.

Our simulations give a fraction of such κ double-positive cells (κ DPs) of $\sim 4\text{--}5\%$ for most parameter values (Fig. 8). This is consistent with recent experimental measurements (69). As expected, there are more κ DPs with $P_{switch} = 0.5$ than in the case of

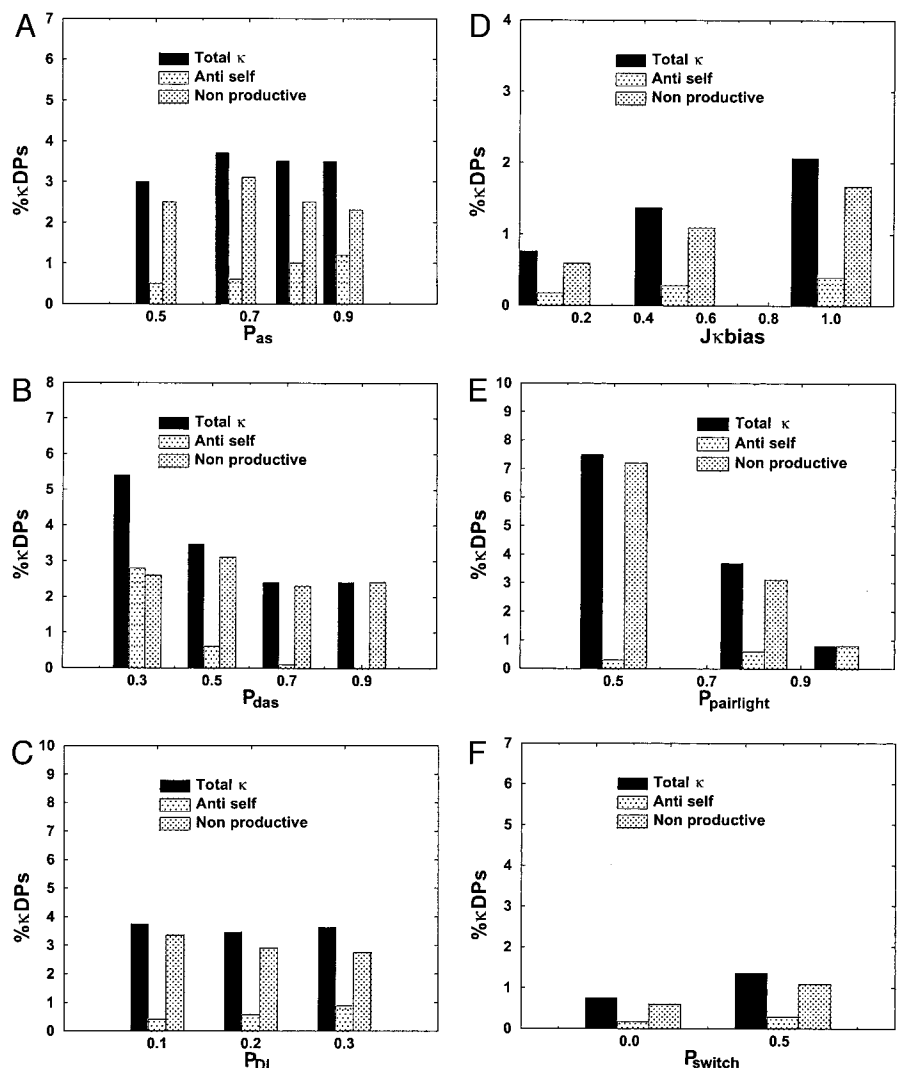


FIGURE 8. Extent of κ allelic inclusion (fraction of newly generated B cells with both κ alleles productively rearranged) as function of P_{as} (A), P_{das} (B), P_{DL} (C), $J_{\kappa bias}$ (D), $P_{pairlight}$ (E), and P_{switch} (F).

$P_{switch} = 0$, because switching means that a productive rearrangement on the previously rearranged allele may be left undeleted. Similarly, there are more κ DPs with random J_κ choice, because it increases the chance that the last J_κ rearrangement on a given allele will be to $J_{\kappa 5}$, which may end up being undeleted (at least in our model, which neglects recombination signal deletion of the whole locus) even if the other allele ends up being rearranged last. Most of these cases are due to a productive light chain rearrangement generating a chain that did not pair well with the heavy chain, but there are always a few cases of “escape” of cells with one anti-self receptor. Obviously, the proportion of these cases increases with the probability for an expressed BCR to be anti-self, P_{as} , and decreases with the probability for an anti-self cell to be deleted, P_{das} , as shown here. Interestingly, the fraction of κ DPs is highly sensitive to $P_{pairlight}$, with the actual fractions of “escaping” anti-self cells increasing when $P_{pairlight}$ is increased. This may be due to the fact that with higher values of $P_{pairlight}$, more cells undergo the negative selection test more than once and have a larger chance of escaping negative selection.

Discussion

Our studies address questions regarding the process of lymphocyte AgR gene rearrangement, using detailed computational modeling of the rearrangement process. The present study is, to our knowledge, the first computational study analyzing heavy chain, as well as light chain, gene rearrangement in B lymphocytes, focusing on the explanation of heavy chain allelic exclusion. We examine whether the rapid positive selection (feedback) model is a sufficient explanation. To evaluate this model, the degree of order in heavy chain rearrangement must be elucidated, i.e., whether heavy chain rearrangement is simultaneous or ordered (first one allele, then the other); and the probability of pairing of a heavy chain with a surrogate light chain has to be estimated. We show that, if allelic exclusion in Ig heavy chains is due only to a combination of rapid selection and differentiation of cells with productive rearrangements, then >10% of mature B cells should be HCDPs. Although this is not much higher than the value measured for developing B cells (40), it is higher than most estimates of allelic inclusion in peripheral B cells. Thus, the reduction of the fraction of HCDPs in peripheral B cells relative to its value in newly generated B cells may be due to simple competition between the two heavy chains for light chain binding, which would result in expression of only one of the potential two receptors or to some peripheral selection mechanism, as has been shown for B cells bearing two complete transgenic receptors (45). Otherwise, the conventional light chain of the cell could rescue the expression of a heavy chain which could not pair well with the surrogate light chain (70, 71), resulting in higher proportions of cells actually expressing two different B cell receptors. It remains to be seen whether our predictions will be confirmed by experimental measurements.

In this study, we have also addressed questions concerning order vs stochasticity in light chain rearrangement, namely, whether light chain rearrangement is ordered (first one allele is rearranged, then the other) or can successive rearrangements be made on either light chain allele at random; what determines the usage of light chain J segments in the rearrangement process; and what is the degree of allelic exclusion in κ light chains in normal (rather than transgenic) B lymphocytes. Concerning simultaneous vs serial light chain rearrangements, the evidence suggests that in most cases, light chain genes differ in their DNA methylation and histone acetylation, and hence replication status, before rearrangement, and thus rearrangement may be initiated at different times on the two alleles. However, simultaneous light chain rearrangements on two alleles is a possibility that has not yet been ruled out. This

is obviously related to the question of whether cells switch the rearranged allele between rearrangements. If cells cannot switch ($P_{switch} = 0$), then simultaneous light chain rearrangements cannot occur. If cells can switch, then simultaneous light chain rearrangements might occur, which might increase the fractions of double-rearranged or DP cells in our simulations. Because we have not included this possibility in the current model due to the added complexity this would cause, our estimate for P_{DL} is a minimum estimate (because simultaneous light chain rearrangements will result in lower fractions of κ -expressing cells with only one κ allele rearranged).

Because our results on J_κ usage are not very sensitive to the value of P_{DL} , our conclusions in this regard will be unchanged. Our estimates of the fractions of κ -expressing cells with productive κ rearrangements on both alleles are, obviously, also minimal estimates, given that these fractions would be larger if then simultaneous light chain rearrangements had occurred. This only strengthens the point we make here, i.e., that if allelic exclusion is only the result of the low probability for rearrangement and selection success, then the appearance of κ -expressing cells with productive κ rearrangements on both alleles is inevitable.

Our results show an inverse correlation between the death probability P_{DL} and the average number of rearrangements per cell, with $P_{DL} = 0.2$, the most plausible value, giving about two rearrangements per cell. Hence, at least one of the mechanisms responsible for allelic exclusion is the fact that the cell is not allowed the opportunity to perform too many secondary rearrangements. A probability of death of 0.2 per rearrangement failure is quite high and thus provides the explanation for the considerable loss of cells throughout the process of B lymphocyte development. Additionally, we found that the most likely value probability of a light chain pairing with the heavy chain of the cell, $P_{pairlight}$, is 0.8. The implication is that most light chains can pair well with most heavy chains, although success in pairing is not always guaranteed. It will be interesting to see whether future experimental measurements confirm this predicted value.

The results presented here support a bias in J_κ segment choice for rearrangement, as previously suggested (37); however, this bias is finite, so that J_κ segment choice need not be deterministic as previously proposed. Conversely, our results do not support the assumption of allele preference in rearrangement; rather, the cell may rearrange any κ allele it chooses, and the choice is random. This conclusion is supported by recent experimental studies (49). This might seem to be in conflict with the recent demonstrations that features such as DNA demethylation (17) and histone acetylation (18, 19) mark or possibly even determine the identity of the allele which is rearranged first. However, there is no conflict here, because, although demethylation of the κ locus may indeed determine the location of the first rearrangement, it does not seem to directly activate V(D)J recombination (22), and when the process is activated, both alleles may very quickly become subject to rearrangement.

Experimental studies are divided on whether there is any bias in heavy or light chain V segment choice (28–34). If there is a bias in V segment choice, it is probably masked to a large extent by the much larger differences in rearrangement frequencies between different genes (even within the same gene family (34), as was found for the heavy chain (28, 30, 32, 33). These differences in light or heavy chains have been attributed to variations in the precise sequence of the recombination signal sequence preceding each gene segment (34). However, within large V gene families, there is sometimes a clear bias toward J-proximal segments. Our simulations of light chain gene rearrangement, described in an accompanying paper, show that these biases, although they may affect the

expression levels of individual genes, do not have a significant influence on overall repertoire characteristics, because sufficient numbers of V genes are available for rearrangement. This is probably the case for heavy chain V genes as well.

Thus, our studies of AgR gene rearrangement have delineated the probabilistic parameters governing gene rearrangement and selection, laying the foundation for the construction of comprehensive models of B lymphocyte development in normal and disrupted situations. In this context, our simulations with low values of P_{DL} may be regarded as reflecting a situation where the mechanisms of programmed cell death are defective, such as in Fas-deficient (MRL/lpr) mice. In this case, our simulations predict that the fraction of cells with rearrangements on both κ alleles may be as high as 50%. This prediction exemplifies the potential usefulness of our studies and the directions of research we intend to pursue in the future.

Acknowledgments

We thank Drs. Michele Shannon, Yehudit Bergman, Martin Weigert, and Samuel Litwin for useful discussions and Hanna Edelman for help in running simulations and preparing the drawings.

References

- Bassing, C. H., W. Swat, and F. W. Alt. 2002. The mechanism and regulation of chromosomal V(D)J recombination. *Cell* 109(Suppl):S45.
- Hesslein, D. G., and D. G. Schatz. 2001. Factors and forces controlling V(D)J recombination. *Adv. Immunol.* 78:169.
- Oltz, E. M. 2001. Regulation of antigen receptor gene assembly in lymphocytes. *Immunol. Res.* 23:121.
- Kouskoff, V., and D. Nemazee. 2001. Role of receptor editing and revision in shaping the B and T lymphocyte repertoire. *Life Sci.* 69:1105.
- Gellert, M. 2002. V(D)J recombination: rag proteins, repair factors, and regulation. *Annu. Rev. Biochem.* 71:101.
- Coleclough, C. 1992. 47th Forum in Immunology: what determines the κ/λ ratio. *Res. Immunol.* 143:838.
- Arakawa, H., T. Shimizu, and S. Takeda. 1996. Re-evaluation of the probabilities for productive rearrangements on the κ and λ loci. *Int. Immunol.* 8:91.
- Melchers, F., E. ten Boekel, T. Seidl, X. C. Kong, T. Yamagami, K. Onishi, T. Shimizu, A. G. Rolink, and J. Andersson. 2000. Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B cells. *Immunol. Rev.* 175:33.
- Barreto, V., T. Meo, and A. Cumano. 2001. Mice triallelic for the Ig heavy chain locus: implications for V_HDJ_H recombination. *J. Immunol.* 166:5638.
- Khor, B., and B. P. Sleckman. 2002. Allelic exclusion at the TCR β locus. *Curr. Opin. Immunol.* 14:230.
- Tripathi, R., A. Jackson, and M. S. Krangel. 2002. A change in the structure of $V\beta$ chromatin associated with TCR β allelic exclusion. *J. Immunol.* 168:2316.
- Coleclough, C. 1983. Chance, necessity and antibody gene dynamics. *Nature* 303:23.
- Claverie, J.-M. and R. Langman. 1984. Models for the rearrangements of immunoglobulin genes: a computer view. *Trends Immunol. Biol. Sci.* 293.
- Langman, R. E., and M. Cohn. 1995. The proportion of B cell subsets expressing κ and λ light chains changes following antigenic selection. *Immunol. Today* 16:141.
- Melchers, F., E. ten Boekel, T. Yamagami, J. Andersson, and A. Rolink. 1999. The roles of pre-B and B cell receptors in the stepwise allelic exclusion of mouse IgH and L chain gene loci. *Semin. Immunol.* 11:307.
- Li, Y., H. Li, and M. Weigert. 2002. Autoreactive B cells in the marginal zone that express dual receptors. *J. Exp. Med.* 195:181.
- Mostoslavsky, R., N. Singh, A. Kirillov, R. Pelanda, H. Cedar, A. Chess, and Y. Bergman. 1998. κ chain monoallelic demethylation and the establishment of allelic exclusion. *Genes Dev.* 12:1801.
- McBlane, F., and J. Boyes. 2000. Stimulation of V(D)J recombination by histone acetylation. *Curr. Biol.* 10:483.
- Schlissel, M. S. 2000. Transcription: a tail of histone acetylation and DNA recombination. *Science* 287:438.
- Maes, J., L. P. O'Neill, P. Cavellier, B. M. Turner, F. Rougeon, and M. Goodhardt. 2001. Chromatin remodeling at the Ig loci prior to V(D)J recombination. *J. Immunol.* 167:866.
- Agata, Y., T. Katakai, S. K. Ye, M. Sugai, H. Gonda, T. Honjo, K. Ikuta, and A. Shimizu. 2001. Histone acetylation determines the developmentally regulated accessibility for T cell receptor γ gene recombination. *J. Exp. Med.* 193:F27.
- Cherry, S. R., C. Beard, R. Jaenisch, and D. Baltimore. 2000. V(D)J recombination is not activated by demethylation of the κ locus. *Proc. Natl. Acad. Sci. USA* 97:8467.
- Kitamura, D., and K. Rajewsky. 1992. Targeted disruption of μ chain membrane exon causes loss of heavy-chain allelic exclusion. *Nature* 356:154.
- Martensson, I.-L., and R. Ceredig. 2000. Role of the surrogate light chain and the pre-B-cell receptor in mouse B-cell development. *Immunology* 101:435.
- Mundt, C., S. Licence, T. Shimizu, F. Melchers, and I. Martensson. 2001. Loss of precursor B cell expansion but not allelic exclusion in VpreB1/VperB2 double-deficient mice. *J. Exp. Med.* 193:435.
- Mancini, S. J. C., S. M. Candéias, J. P. Di Santo, P. Ferrier, P. N. Marche, and E. Jouvin-Marche. 2001. TCRA gene rearrangement in immature thymocytes in absence of CD3, pre-TCR and TCR signaling. *J. Immunol.* 167:4485.
- Nitzchke, L., J. Kestler, T. Tallone, S. Pelkonen, and J. Pelkonen. 2001. Deletion of the DQ52 element within the Ig heavy chain locus leads to a selective reduction in VDJ recombination and altered D gene usage. *J. Immunol.* 166:2540.
- Pascual, V., L. Verkruyse, M. L. Casey, and J. Capra. 1993. Analysis of Ig H chain gene utilization in human fetal liver. *J. Immunol.* 151:4164.
- Ehlich, A., V. Martin, W. Muller, and K. Rajewsky. 1994. Analysis of the B-cell progenitor compartment at the level of single cells. *Curr. Biol.* 4:573.
- Kraj, P., S. P. Rao, A. M. Glas, R. R. Hardy, E. C. B. Milner, and L. E. Silberstein. 1997. The human heavy chain Ig V region gene repertoire is biased at all stages of B cell ontogeny, including early pre-B cells. *J. Immunol.* 158:5824.
- ten Boekel, E., F. Melchers, and A. Rolink. 1997. Changes in the V_H gene repertoire of developing precursor B lymphocytes in mouse bone marrow mediated by the pre-B cell receptor. *Immunity* 7:357.
- Rao, S. P., J. M. Riggs, D. F. Friedman, M. S. Scully, T. W. LeBien, and L. E. Silberstein. 1999. Biased V-H gene usage in early lineage human B cells: evidence for preferential Ig gene rearrangement in the absence of selection. *J. Immunol.* 163:2732.
- Gallo, M. L., V. E. Ivanov, A. Jakobovitz, and C. G. Davis. 2000. The human immunoglobulin loci introduced into mice: V(D) and J gene segment usage similar to that of adult humans. *Eur. J. Immunol.* 30:534.
- Feeney, A., A. Tang, and K. Ogwaro. 2000. B-cell repertoire formation: role of the recombination signal sequence in non-random V segment utilization. *Immunol. Rev.* 175:59.
- Yu, K. F., A. Taghva, and M. R. Lieber. 2002. The cleavage efficiency of the human immunoglobulin heavy chain V-H elements by the RAG complex: implications for the immune repertoire. *J. Biol. Chem.* 277:5040.
- Chowdhury, D., and R. Sen. 2001. Stepwise activation of the immunoglobulin μ heavy chain gene locus. *EMBO J.* 20:6394.
- Mehr, R., M. Shannon, and S. Litwin. 1999. Models for antigen receptor gene rearrangement. I. Biased receptor editing in B cells: implications for allelic exclusion. *J. Immunol.* 163:1793.
- Piper, H., S. Litwin, and R. Mehr. 1999. Models for antigen receptor gene rearrangement. II. T cell receptor editing: allelic exclusion or inclusion? *J. Immunol.* 163:1799.
- Kalmanovich, G., and R. Mehr. 2000. Secondary antigen receptor gene rearrangements: a violation of allelic exclusion? In *Proceedings of the 14th European Immunology Meeting EFIS2000*, A. Mackiewicz, M. Kurpiz, and J. Zeromski, eds. Monduzzi Editore, Italy, p. 745.
- ten Boekel, E., F. Melchers, and A. Rolink. 1998. Precursor B cells showing H chain allelic inclusion display allelic exclusion at the level of pre-B cell receptor surface expression. *Immunity* 8:199.
- Chang, Y., M. J. Bosma, and G. C. Bosma. 1999. Extended duration of DH-JH rearrangement in immunoglobulin heavy chain transgenic mice: implications for regulation of allelic exclusion. *J. Exp. Med.* 189:1295.
- Mostoslavsky, R., A. Kirillov, Y. H. Ji, M. Goldmit, M. Holzmann, T. Wirth, H. Cedar, and Y. Bergman. 1999. Demethylation and the establishment of κ allelic exclusion. *Cold Spring Harbor Symp. Quant. Biol.* 64:197.
- Melchers, F. 1999. Fit for life in the immune system? Surrogate L chain tests H chains that test L chains. *Proc. Natl. Acad. Sci. USA* 96:2571.
- Barreto, V., and A. Cumano. 2000. Frequency and characterization of phenotypic Ig heavy chain allelically included IgM-expressing B cells in mice. *J. Immunol.* 164:893.
- Rosado, M. M., and A. Freitas. 2000. B cell positive selection by self antigens and counter-selection of dual B cell receptor cells in the peripheral B cell pools. *Eur. J. Immunol.* 30:2181.
- Davodeau, F., M.-A. Peyrat, F. Romagne, A. Necker, M.-M. Hallet, H. Vie, and M. Bonneville. 1995. Dual T cell receptor β chain expression on human T lymphocytes. *J. Exp. Med.* 181:1391.
- Li, H., Y. Yang, E. Luning Prak, M. Radic, and M. Weigert. 2001. Editors and editing of anti-DNA receptors. *Immunity* 15:947.
- Mehr, R., M. Shannon, and S. Litwin. 2000. Models for antigen receptor gene rearrangement. I. Biased receptor editing in B cells: implications for allelic exclusion (correction). *J. Immunol.* 165:3527.
- Lin, X., M. Linden, and B. Van Ness. 2000. Induced κ receptor editing shows no allelic preference in a mouse pre-B cell line. *J. Immunol.* 165:7058.
- Engel, H., A. Rolink, and S. Weiss. 1999. B cells are programmed to activate κ and λ for rearrangement at consecutive developmental stages. *Eur. J. Immunol.* 29:2167.
- Thiede, R., K. F. Schable, A. Bensch, J. Brensing-Kuppers, V. Heim, T. Kirschbaum, H. Mitlohner, M. Ohnrich, S. Pourrajabi, F. Rosenthaler, J. Schwendinger, D. Wichelhaus, I. Zocher, and H. G. Zachau. 1999. The variable genes and gene families of the mouse immunoglobulin κ locus. *Eur. J. Immunol.* 29:2072.
- Hess, J., A. Werner, T. Wirth, F. Melchers, H.-M. Jack, and T. H. Winkler. 2001. Induction of pre-B cell proliferation after de novo synthesis of the pre-B cell receptor. *Proc. Natl. Acad. Sci. USA* 98:1745.
- Radic, M. Z., and M. Zouali. 1996. Receptor editing, immune diversification, and self-tolerance. *Immunity* 5:505.
- Hertz, M., and D. Nemazee. 1998. Receptor editing and commitment in B lymphocytes. *Curr. Opin. Immunol.* 10:208.

55. Fink, P., and C. McMahan. 2000. Lymphocytes rearrange, edit and revise their antigen receptors to be useful yet safe. *Immunol. Today* 21:561.
56. Sandel, P. C., and J. G. Monroe. 1999. Negative selection of immature B cells by receptor editing or deletion is determined by site of antigen encounter. *Immunity* 10:289.
57. Lang, J., and D. Nemazee. 2000. B cell clonal elimination induced by membrane-bound self-antigen may require repeated antigen encounter or cell competition. *Eur. J. Immunol.* 30:689.
58. Braun, U., K. Rajewsky, and R. Pelanda. 2000. Different sensitivity to receptor editing of B cells from mice hemizygous or homozygous for targeted Ig transgenes. *Proc. Natl. Acad. Sci. USA* 97:7429.
59. Watanabe, N., S. Nisitani, K. Ikuta, M. Suzuki, T. Chiba, and T. Honjo. 1999. Expression levels of B cell surface immunoglobulin regulate efficiency of allelic exclusion and size of autoreactive B-1 cell compartment. *J. Exp. Med.* 190:461.
60. Pelavin, P. I. 1996. Why B cells express a single antigen receptor: a probabilistic model. Senior thesis, Department of Molecular Biology, Princeton University, Princeton, New Jersey.
61. Coleclough, C., R. P. Perry, K. Karjalainen, and M. Weigert. 1981. Aberrant rearrangements contribute significantly to the allelic exclusion of immunoglobulin gene expression. *Nature* 290:372.
62. ten Boekel, E., F. Melchers, and A. Rolink. 1995. The status of Ig rearrangements in single cells from different stages of B cell development. *Int. Immunol.* 7:1013.
63. Yamagami, T., E. ten Boekel, J. Andersson, A. Rolink, and F. Melchers. 1999. Frequencies of multiple IgL chain gene rearrangements in single normal or κ L chain-deficient B lineage cells. *Immunity* 11:317.
64. Yamagami, T., E. ten Boekel, C. Schaniel, J. Andersson, A. Rolink, and F. Melchers. 1999. Four of five RAG-expressing $J\kappa\kappa^{-/-}$ small pre-BII cells have no L chain gene rearrangements: detection by high-efficiency single cell PCR. *Immunity* 11:309.
65. DeBoer, R. J., and A. S. Perelson. 1993. How diverse should the immune system be? *Proc. R. Soc. Lond. B* 252:343.
66. Nemazee, D. 1996. Antigen receptor "capacity" and the sensitivity of self-tolerance. *Immunol. Today* 17:25.
67. Brauning, A., T. Goossens, K. Rajewsky, and R. Kuppers. 2001. Regulation of immunoglobulin light chain gene rearrangements during early B cell development in the human. *Eur. J. Immunol.* 31:3631.
68. Wood, D. L., and C. Coleclough. 1984. Different joining region J elements of the murine κ immunoglobulin light chain locus are used at markedly different frequencies. *Proc. Natl. Acad. Sci. USA* 81:4756.
69. Casellas, R., T.-A. Yang Shih, M. Kleinewietfeld, J. Rakonjac, D. Nemazee, K. Rajewsky, and M. C. Nussenzweig. 2001. Contribution of receptor editing to the antibody repertoire. *Science* 291:1541.
70. Novobrantseva, T. I., V. M. Martin, R. Pelanda, W. Muller, K. Rajewsky, and A. Ehlich. 1999. Rearrangement and expression of immunoglobulin light chain genes can precede heavy chain expression during normal B cell development in mice. *J. Exp. Med.* 189:75.
71. Papavasiliou, F., M. Jankovic, and M. C. Nussenzweig. 1996. Surrogate or conventional light chains are required for membrane immunoglobulin μ to activate the precursor B cell transition. *J. Exp. Med.* 184:2025.