Archivum Immunologiae et Therapiae Experimentalis, 2001, **49**, 111–120
PL ISSN 0004-069X

*Review*

# Modeling the Meta-Dynamics of Lymphocyte Repertoires

RAMIT MEHR*

Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel

**Abstract.** The complexity of biological systems, and the explosion of the quantity of biological information which is rapidly becoming available from experimental and clinical studies, necessitate the use of theoretical tools, namely, mathematical and computational modeling. The vertebrate adaptive immune system, with its learning and memory capabilities, is a particularly rich source of modeling challenges. Most difficult within this area is the study of lymphocyte repertoires – the generation of their diversity and the forces that shape the ever-changing dynamics of lymphocyte clones. I review several examples of problems in lymphocyte repertoire modeling, demonstrate the types of solutions employed, and highlight the contribution of these theoretical studies to immunological research.

**Key words:** lymphocyte repertoires; lymphocyte development; mathematical model; computer simulation.

## Introduction

Biological systems are highly complex, not only due to the large number of factors affecting any biological process, but also because their dynamics are almost always non-linear. The human mind, which tends to think in linear patterns, often cannot predict all the possible modes of behavior of a multi-factorial, non-linear system. Mathematical and computational modeling can enhance the human mind as powerful tools for forming increasingly complex "thought experiments", thus aiding the study of biological systems, from the formation of concepts to the design of experiments and therapies. As a result, the closely-related fields of mathematical and computational biology have in recent years experienced tremendous growth, paralleling the explosion of information obtained in all areas of biology. (For a glimpse of these fields, check out the web pages of the Society for Mathematical Biology: www.smb.org, or the International Society for Computational Biology: www.iscb.org.).

One of the most intriguing challenges to theoretical biologists is presented by the adaptive immune system, one of the only two biological systems capable of continuously learning and memorizing its experiences. Theoretical immunology is currently a rapidly growing field of research which is, nonetheless, highly integrated with mainstream experimental and clinical immunology. Studies of the interactions of the immune system with various pathogens, or of modes of regulation within the immune system itself, are increasing in number and usefulness. Within this rich source of theoretical problems, one particularly difficult area is the study of lymphocyte repertoires: the generation of antigen receptor repertoire diversity, the dynamics of lymphocyte development under normal or immune-deficient conditions, the forces of selection and interaction with antigen that shape lymphocyte repertoires. My

Abbreviations used: BCR – B cell receptor, GC – germinal center, MHC – major histocompatibility complex, TCR – T cell receptor.

* Correspondence to: Dr. Ramit Mehr, Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel, tel.: +972 3 531 79 90, fax: +972 3 535 18 24, e-mail: mehrra@mail.biu.ac.il

intention in this review is to describe some of the problems involved in the theoretical modeling of the complex repertoires of B or T lymphocytes, and the solutions found.

The difficulty in modeling lymphocyte repertoires does not result from the astronomic number of lymphocyte clones in the human patient or the experimental animal, even though the actual modeling of a comparable number of clones would require a much higher computing power than is available to the average theoretician. In principle, studies on a smaller number of clones can often give sufficient insight into the behavior of the whole system. The main difficulty stems from the need to formulate a description of the system on several levels: genetic, molecular, cellular and systemic. For example, when studying the selection of developing T lymphocytes in the thymus, there is only so much one can do with a model that is limited to the level of cell population. Before long, the rearrangement of T cell receptor (TCR) genes, and the interaction of their protein product with major histocompatibility (MHC)-peptide complexes, must be taken into account. Hence, it is not only the population dynamics within each clone (i.e. cell division, differentiation and death processes) that must be modeled, but also the meta-dynamics: creation, selection and elimination of whole clones in the population, and the molecular interactions or genetic changes that form the basis for these meta-dynamics. The response to the demand for increasing complexity depends on the problem at hand. No model can integrate all aspects of the real system, nor should it attempt to do so, because models must be simple enough to enable an exact analysis. Hence, it is up to the researcher to decide which aspects of the system to focus on and which aspects to neglect in each part of the study, as demonstrated below.

## Multiclonal Competition Driven by Mutations: the Repertoire Shift Problem

When studying the humoral immune response, with its features of clonal selection of B lymphocytes and affinity maturation of their antigen receptors, it is crucial to take into account the dynamics of a number of competing clones. Hypermutation is a genetic-level process which stands at the base of the population meta-dynamics through the creation of new clones, so that we need at least a two-level model. The intermediate level, that of molecular interactions between the receptor and its nominal antigen, has the important role of determining the fate of B cell clones. For the purpose of the present model, however, receptor-antigen interactions can be modeled using a single parameter, the affinity of the interaction, because affinity is by far the main factor affecting the fate of each B lymphocyte during the response. This is the strategy that was employed in a recent study of the problem of repertoire shift.

Repertoire shift is a phenomenon in which antibodies dominating a primary response to antigen are severely depleted, or even absent, in the secondary response. The antibodies that dominate the secondary response use variable (V) region genes that are different from those of the primary response. This shift is the rule rather than the exception in responses to T-dependent antigens, where hypermutation and affinity maturation occur[21]. Repertoire shift stands in apparent contradiction to the classical affinity-maturation paradigm, which predicts that dominant clones in the primary response would be most readily recruited into memory responses.

Why would a B cell clone, which is dominant in the primary response, be missing from the secondary response? When one considers the role of somatic hypermutation, a way to reconcile repertoire shift with the affinity maturation paradigm becomes apparent: the B cell clone expressing the primary-dominant antibody may fail to dominate the memory response because it is more likely than lower-affinity clones to suffer deleterious mutations. The process of affinity maturation is generally thought to improve the affinity of antibodies to antigen through mutation and selection. If, however, a B cell clone begins with a relatively high affinity to the antigen, a random mutation would be more likely to decrease rather than increase its affinity. At the same time, other, previously minor clones, which can improve through mutation, may become dominant clones. This hypothesis can best be visualized by assuming that, in sequence space, the “landscape” of antibody affinity to the antigen is quite rugged[17], with many local “peaks” and “valleys”. The process of affinity maturation at most takes each antibody from where it started on this landscape only to the nearest local peak, which may not be the point with the highest affinity to the antigen in question. The antibody that dominated the primary response may thus end up at a local affinity peak that is lower than the final peaks reached by antibodies that started at different, initially lower points on the landscape. Indication that the affinity landscape is indeed “rugged” comes from findings of single mutations that confer a large (up to 10-fold) increase in affinity[23].

As only the end results of mutation and clonal se-

lection can be observed experimentally, we chose to use mathematical modeling and computer simulations of the dynamics of B cell clones to test the above hypotheses[21]. The model that we used consists of an array of B cell clones which can proliferate, differentiate and mutate in response to antigenic stimulation (Fig. 1). Each clone is assigned a number that represents the affinity of interaction between the cell's antigen receptor and its nominal antigen. Members of each clone may belong to the naive, activated, germinal center (GC), memory, or plasma cell subsets. There is constant input of naive cells from the bone marrow, while further differentiation (emigration of cells to the next compartment) depends on interaction with antigen proportional to the affinity of the cell's receptor to the antigen and to the amount of antigen present. Proliferation is also antigen-dependent and assumed to occur only in the activated and GC compartment. Cell death may occur in all compartments. The transition from GC to memory cells is assumed to be antigen-independent, and memory cells can also be re-activated by antigen. Antigen depletion results from consumption by activated and plasma cells. The rates of all these processes were taken from the literature or from previous models. Mutations were modeled by taking cells from the GC compartment and allowing them to "mutate", that is, randomly obtain new values of their affinity number. The rate of mutation in the simulations was adjusted to fit what is known from experimental systems, taking into account the probabilities that a mutation will be silent, lethal, neutral or advantageous.

For a single clone, a mathematical analysis of the differential equations describing the above model may have been sufficient. However, we were interested in the much more complex dynamics of affinity maturation which emerge when a large number of clones compete for a given antigen. Additionally, we included in our model the genetic-level process of hypermutation, which is discrete and stochastic. Hence, we turned to computer simulations of this model. Simulations which did not take into account the above-described "landscape" hypothesis, assuming instead that production of memory cells is proportional to the clone's initial affinity, generated the picture predicted by the affinity maturation paradigm: clones that dominated the primary response continued to dominate the secondary response, that is, there was no repertoire shift. However, the total numbers of cells at the peak of each response, the timing of the peak and the decline of the response, all agreed with experimental observations[8, 9], confirming that we had chosen a reasonable set of parameter values. Hence, we proceeded to incorporate alternative

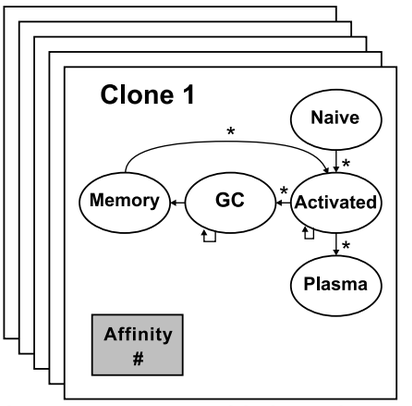


**Fig. 1.** A schematic representation of the humoral response simulation. The simulation is structured as an array of B lymphocyte clones, each associated with a number representing their relative affinity to the antigen. Each clone contains five B cell subsets: naive, activated, plasma, germinal center (GC) and memory. The transitions between the subsets are depicted here as arrows and marked with an asterix if they are antigen-dependent. Activated and GC cells may also proliferate (denoted by small arrows which return to the same compartment they exit from)

hypotheses concerning the cause of repertoire shift into our model.

To incorporate the "landscape" hypothesis in the model, we have set mutation probabilities such that most mutation-induced affinity changes are small, while large "leaps" in affinity are relatively rare. Selection was applied in our model in the following way: if the post-mutation affinity is smaller than the pre-mutation affinity, the cell fails selection and dies. A cell is positively selected, forming a new sub-clone, only if mutation has improved its affinity by a certain factor, which is a parameter in the simulation. If the improvement is smaller, the mutation is ignored. Simulations with mutation and selection thus implemented give a rich and realistic picture of repertoire shift, evident in all relevant cell subsets, i.e. the activated, plasma, GC and memory B cells. If we extend our simulations to include a third and even a fourth antigenic challenge, the clones that dominate the secondary response continue to dominate subsequent responses, in agreement with experimental observations. This means that the optimal receptor for most antigens can be found by the immune system following the first encounter with the antigen, even though it does not exist in the germline repertoire. Repertoire shift is more pronounced if the distribution of mutations is made narrower, that is, if advantageous mutations are more rare. This supports our key point, that repertoire shift results from the destructive aspect of hypermutation on antibodies that had a high affinity to the antigen prior to mutation. Second,

as expected, a similar effect is observed when increasing the relative advantage conferred by mutations, or the stringency of the selection operating on newly generated mutants.

A related question is whether the average affinity of antibodies to the antigen increases between the primary and the secondary response[23]. Experimental studies[7] concluded that the main effect of somatic hypermutation on the primary antibodies must be destructive, making way for the appearance of new clones, which supports our present conclusion. Studies of the humoral immune response have traditionally focused on the positive aspect of somatic hypermutation. The paradigm of affinity maturation, stating that B cells taking part in the secondary response are "improved versions" of those that dominated the primary response, is often confused with somatic hypermutation. One should keep in mind that somatic hypermutation is only the mechanism generating candidates for affinity improvement, on which selection then operates. One must remember that affinity-improving mutation is more likely to be a rare event in a multitude of deleterious mutations. A thorough understanding of the humoral immune response must acknowledge the destructive aspect of hypermutation.

From the point of view of modeling, it is noteworthy that the rich dynamics of hypermutation, affinity maturation and repertoire shift were obtained in the model using only ten initial B lymphocyte clones. This is comparable to what was found in *in vivo* studies[8, 22]. The key points were the realistic description of B cell subsets within each clone combined with a model of mutation and selection which captured the essential features of the "landscape" hypothesis.

## Combining Genetic and Cellular Processes: Antigen Receptor Gene Rearrangement

The next example is of a problem requiring a simulation of much larger numbers of clones, with many more parameters linked to each clone: the problem of randomness versus sequentiality in antigen receptor gene rearrangement. This series of studies addressed the apparent contradiction between continuous antigen receptor gene rearrangement in lymphocytes and allelic exclusion.

### *The problem*

The long-accepted concept of allelic exclusion states that antigen receptor chain genes in B and T lymphocytes are expressed from only one of the two alleles in each cell, ensuring that each lymphocyte expresses a single antigen receptor specificity on the cell surface. Recent evidence suggests, however, that in the B cell receptor (BCR) light chain[14] and in the TCR$\alpha$ chain[19], rearrangement may not stop after a productive receptor gene has been formed and expressed. Secondary light chain rearrangements occur not only after non-productive rearrangements, but also after productive rearrangements that render a B cell autoreactive ("receptor editing")[20], resulting in one of the mechanisms of central tolerance. However, if the choice of allele for secondary rearrangements is random, it is (at least theoretically) possible that a cell will rearrange and then simultaneously express two different light chains. How then, if at all, is allelic exclusion maintained in the face of continued rearrangement?

For $\alpha\beta$ T cells, the situation is even more complex[19]. Rearrangement and expression of TCR$\alpha$ chain genes does not stop after the expression of the first rearranged $\alpha$ chain, but appears to continue until the cell either is positively selected or dies. Due to the lack of allelic exclusion in TCR$\alpha$, a T cell may not only contain two productively rearranged TCR$\alpha$ alleles, but also simultaneously express the two resulting TCRs. Indeed, the frequency of T cells simultaneously expressing two different V$\alpha$ genes was found[16] to vary between $10^{-4}$–$10^{-3}$. Only the cell surface expression of three V$\alpha$ genes was monitored, which means that the frequency of T cells expressing **any** pair of V$\alpha$ genes may be orders of magnitude higher. Independently, 26% of various T cell clones were found to contain two productive V$\alpha$-J$\alpha$ rearrangements[12, 13]. This is an alarming discovery, because an $\alpha\beta$ T cell that matures in the thymus expressing two different TCRs may be positively selected on one of them, while the other TCR may be autoreactive. Why, then, is allelic exclusion not maintained in TCRs?

The observations of allelic inclusion in TCR$\alpha$ raise several more questions. Can allelic inclusion be fully accounted for by multiple rearrangements alone? Do these rearrangements occur completely at random, or is there some underlying order? What is the role of positive and negative selection in driving, or limiting, the process of TCR gene rearrangement? Several previous models, which attempted to explain experimental observations such as the fractions of cells containing two productive TCR$\alpha$ rearrangements did not sufficiently account for TCR gene organization, which limits secondary rearrangement or for the effects of subsequent thymic selection. The work described in the following was designed to address these questions.

*The simulation*

We first created a stochastic simulation of BCR gene rearrangement[14]. The general structure of the simulation is given in Fig. 2. It simulates the following steps in a developing B cell's life:

– Cell "birth": the simulation starts with all light chain genes in the unrearranged (germline) state;

– Light chain choice: decision whether to rearrange κ or λ next;

– BCR κ rearrangement: choice of which allele to rearrange and, on this allele, which Jκ segment to use. J segments 5' to the chosen segment are deleted;

– BCR λ rearrangement-similar to κ rearrangement, but without deletion;

– Selection: probabilistic decision whether the rearrangement is in-frame, whether the resulting light chain can pair with the existing heavy chain, and whether the resulting BCR is autoreactive (anti-self).

Rearrangement is repeated until a productive, H/L--matched, non-autoreactive BCR is produced, or until the cell dies. Cells containing an anti-self rearrangement are assigned a high death probability. Cells containing only out-of frame V-J joins, H/L-mismatched heavy-light chain pairs, or germline κ, are assigned a moderate death probability. If any of the latter types of cells do not die, they may attempt secondary rearrangements. Cells containing one in frame, H/L-matched, non-autoreactive rearrangement are allowed to mature.

The parameters for the simulation, that is, probabilities of death in each stage, probabilities for isotype or allele choices, etc., were changed in each run within the biologically plausible ranges. For example, 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. On the other hand, the death rate of autoreactive cells is not known and, hence, a whole range of values was studied in the simulations.

The simulation of TCR gene rearrangement was similar, with several extensions: 1) it records all variable region segments, V, D and J, for the four possible chains α, β, γ and δ, 2) it has three types of selection: one at the pre-TCR stage checking that the β chain is functional, and then positive and negative selection of cells expressing an αβ TCR, 3) the simulation allows for cell division and follows all cells in each lineage.

The description of the rearrangement status of each allele of each receptor chain, including the probability for further rearrangement associated with each individual segment, necessitates the use of a large number of parameters associated with each cell. Fortunately, we could reduce the amount of computation involved by reducing the description on the level of cell population dynamics to that of a single subset of cells, its divisions, death and so on. Again, selection was described by a single parameter, neglecting the details of the molecular interactions. Additionally, there was no need for simultaneity, so cells could be simulated one at a time. This was fortunate, because we found that we had to simulate very large numbers of cells (up to $10^5$ in the BCR simulations) in order for the properties of the repertoire to be stable between simulations. In the TCR rearrangement simulations, with high cell division probabilities, an order of $10^4$ individual cells may all belong to a small number of clones. Thus, it was necessary to include a large number of independent clones (each possibly containing many cells) in each simulation, for parameters such as the ratio to stabilize. The number of clones was considered to be sufficient when both inter- and intra-simulation variabilities were small (less than 10% of the initial variability). This was achieved for all quantities measured by generating 4000 independent clones in each simulation (data on variability not shown). The variability criterion thus defined helped to identify the proper division probabilities (probability of division following β-selection

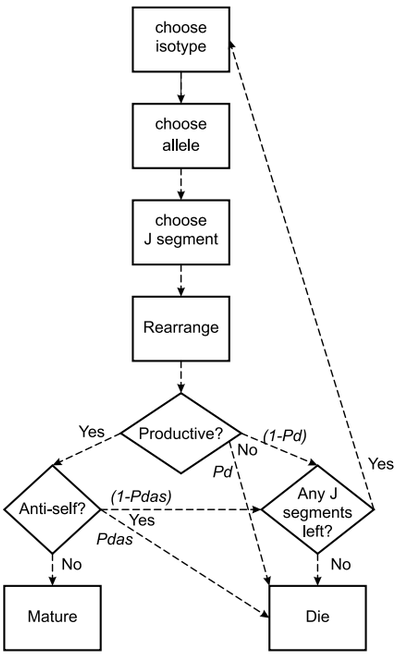


**Fig. 2.** A schematic representation of our simulation of light-chain gene rearrangement. Actions are depicted in rectangles and decisions are depicted in diamonds. *Pd* is the death probability of cells that have not been subjected to negative selection, while *Pdas* is the death probability of "anti-self" cells that have been negatively selected (*Pdas>Pd)*

and following positive selection of αβ T cells) to be used in the simulations.

*Results: BCR gene rearrangement*

The observed ratio between κ light chain- and λ light chain-bearing B cells in murine serum is about 20 : 1; even in immature murine bone marrow cells it is larger than 10 : 1. It is controversial whether the κ : λ ratio can be explained solely on the basis of the higher potential for multiple rearrangements in the κ locus, combined with immature B cell death due to negative selection, without assuming preferential expansion of κ B cells over λ B cells[14].

In addition to explaining the high κ : λ ratio, the postulated ability of a B cell to make secondary rearrangements on a single κ allele could explain why about 70% of mouse splenic κ B cells have only one rearranged κ locus, the other one remaining unrearranged[3]. It would also explain why in mice that have only one functional κ locus, κ B cell production is about 70% (rather than one-half) of that in wild-type mice[2]. Furthermore, it has been suggested[2, 6, 23] that rearrangement may proceed sequentially rather than stochastically, that is, that 5' Jκ segments are used before 3' Jκ segments. Hence, the following question arose: can ordered rearrangement account for the observations on allele bias, Jκ usage and the κ : λ ratio?

We tested the hypothesis of allele bias (preference to rearrange a rearranged allele first) as follows: simulations were performed with either no bias (choice of the allele for the next rearrangement is completely random) or strong bias (the cell continues rearranging the already-rearranged allele until there are no more Jκ segments available for rearrangement on that allele, and only then it may switch to the other allele). The results were compared to experimental observations. Similarly, we tested the hypothesis of sequentiality in Jκ usage, using simulations in which the choice of the next J segment for rearrangement was either random, strictly sequential (first Jκ1, then Jκ2, and so on) or quasi-sequential (an intermediate case with partial bias, that is, preference of Jκ1 and Jκ2 over Jκ4 and Jκ5). We also checked the influence of parameters, such as the probability of choosing κ over λ, and the stringency of negative selection as expressed by death probabilities, on the resulting repertoires.

We have found that BCR gene rearrangement is ordered on three different levels, as follows. First, our simulations support the notion[14] that κ light chain rearrangement precedes λ light chain rearrangement in most, if not all, cases. Second, our results strongly support the hypothesis of allele preference, that is, once a rearrangement exists on one of the κ alleles, the cell is more likely to perform secondary rearrangement, if necessary and possible, on the same allele rather than switching to the other κ allele. Allele preference fully explains the high fractions of κ light chain B cells that contain κ rearrangements on one allele only, even without additional cell divisions. Third, our results show that experimental observations are consistent with the hypothesis of quasi-sequential rearrangement within each κ allele. That is, when Jκ1 and/or Jκ2 are available for rearrangement, the cell is more likely to choose one of these segments over Jκ4 or Jκ5.

The effect of negative selection is determined by the death probability of cells that have not succeeded in producing a productively rearranged, non-autoreactive BCR. We found that the higher the death probability, the larger the κ : λ ratio. Cells are allowed few rearrangement attempts and, hence, are likely to die before exhausting κ and proceeding to λ rearrangements. Our simulations show that, in order for the results to be consistent with experimental observations, we must assume that negative selection limits the number of rearrangement attempts to 2–3 per cell, in agreement with previous results[15]. Therefore, we refer to BCR rearrangement as a negative selection-limited process.

Taken together, the above results reveal our proposed answer to the question of allelic exclusion in B cells. We propose that B cell allelic exclusion results from a high degree of order in gene rearrangement and a stringent process of negative selection. A high degree of order allows a B cell to maximize the number of rearrangement attempts, first on one allele and then on the other. Then, because negative selection limits the number of rearrangement attempts to 2–3 per cell, ordered rearrangement means it is likely that all these attempts will be on a single allele. Thus, it is extremely unlikely for two **productive** rearrangements to exist simultaneously on two light chain alleles. This results in the almost complete absence of cells expressing two different BCRs from the repertoire, i.e. in effective allelic exclusion. Ordered secondary rearrangements thus maximize the cell's ability to make a productive, non-autoreactive rearrangement before exhausting all Jκ segments.

*Results on TCR gene rearrangement*

One of the measurable quantities which has received much attention in the literature is the αβ : γδ ratio in thymocytes and mature T cells, which can be 20 : 1 or even higher, depending on the tissue studied. Theoreti-

cal predictions based on models that do not include multiple rearrangements fall around 2 : 1, which is far from the experimentally observed range of values. The difference was attributed to cell division. Our simulations showed that both multiple rearrangements and cell divisions are required in order to explain this ratio, in contrast to the situation in B cells, where there was no need to invoke cell divisions. The ratio increases with the number of cell divisions after β selection and after thymic selection. Conversely, the ratio decreases when we increase the death probabilities of cells that fail β selection or αβ selection. Additionally, the αβ : γδ ratio increases with the probability that β rearrangement will precede δ rearrangement.

The measure for the extent of allelic exclusion, or rather allelic inclusion, in T cells is the fraction of TCRα "double productives": T cells that carry productive TCRα rearrangements on both alleles. Theoretical models predict that secondary rearrangements are necessary to explain the experimentally observed fraction of up to 26% TCRα "double productives". Our simulations examined the dependence of this quantity on the degree of order in TCRα rearrangement and on selection probabilities. Fractions of TCRα "double-productives" higher than 20%, as observed experimentally, are obtained in our simulations only when we allow multiple TCRα rearrangements, but assume they are **unbiased**. Thus, the results of these simulations cannot exclude the hypothesis that multiple rearrangements in T cells are random, rather than ordered as was found for the B cell light chain. At least, if there is order in TCRα rearrangement, it is masked by the large number of rearrangements per cell (10–12 rearrangements in our simulations). The high fraction of residual rearranged δ alleles in αβ T cells (measured to be as high 80%) would imply, according to our simulations, that TCRα rearrangement is ordered, at least on the level of allele bias. However, most of the observed residual rearranged δ alleles probably exist on extra-chromosomal excised DNA circles[11]. An experimental clasification of the status of these δ rearrangements, combined with our analysis, could lead to a determination of the degree of allele bias in TCRα rearrangement.

A novel quantity defined in this study, for which no observations exist, is the fraction of cells with an auto--reactive receptor among the TCRα "double-productives". This fraction is independent of multiple rearrangements because it depends only on the last rearrangements on the two alleles. However, we found that the fraction of autoreactive "double-productives" is highly sensitive to the death rate of auto-reactive thymocytes. If this rate is low, as it must be to get 26% "double-productives", then the fraction of autoreactive "double-productives" can be as high as 70%. This value is only an upper bound, since it was obtained for the case in which the cell is selected only according to its last rearrangement. Otherwise, this number will be lower, and will also depend on the relative expression levels of the two receptors, which are not addressed by the current model. More experimental data would be beneficial for settling this issue, which may help elucidate instances of escape from central tolerance in T cells.

*TCR vs. BCR gene rearrangement*

In spite of the strong similarities revealed in our studies between the way rearrangement seems to operate in B and T cells, it is worthwhile to note a crucial difference between the development of T cells and that of B cells. While BCR rearrangement seems to be limited by negative selection only, T cell development, on the other hand, seems to be limited by positive rather than by negative selection: developing T cells in the thymus are allowed a much more generous time window for continued TCRα rearrangement, so that multiple rearrangements on **both** alleles become the rule rather than the exception. Furthermore, while B cells generally exhibit allelic and isotypic exclusion due in part to ordered rearrangement, receptor gene rearrangement in T cells is far less ordered. As a result, the probability that a cell will contain more than one productive rearrangement of the TCRα chain, and even express two TCRα chains simultaneously, is far from negligible. Moreover, positive selection may rescue the cell from death and allow it to mature, based on the virtues of only one of its expressed receptors, as long as the other receptor is not so extremely auto-reactive as to cause immediate deletion of the cell. This is a potentially dangerous situation because the second receptor may still be weakly autoreactive or, worse, may be specific, with high affinity, to a self-peptide that is not presented in the thymus. In spite of the existence of peripheral mechanisms of self-tolerance which safeguard against improper activation of T cell, such improper activation does sometimes happen. Thus, understanding TCR gene rearrangement, selection and editing is key to understanding autoimmunity.

## Molecular Interactions and Their Effect on Clonal Dynamics: T Cell Selection

In the previous examples, we focused on the genetic processes (gene rearrangement or mutations) which

drive population meta-dynamics mostly through the creation of new clones. On the other hand, selection processes, which decide fate of the clones – survival or deletion condensed into a single parameter representation. The find example in this paper demonstrates the complementary approach, of focusing in detail on the molecular interactions underlying clonal selection (of thymocytes, in this case) while neglecting the genetic details for the sake of simplicity.

The initial T cell repertoire, after its creation through gene rearrangement, is further shaped in the thymus by two selection processes. Positive selection promotes the differentiation to a further developmental stage of cells bearing receptors with a sufficiently large affinity for peptides presented on molecules of the major histocompatibility complex (MHC) expressed in the thymus. This confers to T cells the property of self MHC restriction: they recognize peptides presented in the groove of host MHC molecules, but ignore them when presented on foreign MHC. Then, negative selection deletes cells whose TCRs bind thymic MHC-peptide complexes with very high affinity, thus preventing the emergence of self reactive T cells. Overall, only about 3% of thymocytes have the intermediate affinity needed to fully mature. In addition to self MHC restriction, the mature repertoire is also characterized by a high alloreactivity. Typically, 1–24% of T cells react against the product of a given foreign MHC allele. This high response frequency is hard to reconcile with the fact that only one T cell in $10^4$–$10^6$ of the naive repertoire recognizes a given pathogen.

What should the quantitative properties of the processes driving TCR generation and selection be in order to produce the experimentally observed levels of self restriction, alloreactivity and antigen response? In a series of studies[4, 5] we have addressed this issue using a computer simulation based on the following model. The features of two proteins that determine their binding can be described with a relatively small number of parameters, such as their geometric shape, charges and hydrophobicity. All these parameters combine to form the protein's "generalized shape", modeled as a string of digits[18] from an alphabet of up to 255 digits. The strength of binding of two proteins is then defined as the degree of complementarity between the digits representing their generalized shapes (Fig. 3). Only the interface between TCRs and MHC-peptide complexes is represented in the model. We define the affinity, K, between two digit-string proteins, as the sum of their individual digit interactions.

MHC molecules and peptides are modeled as random strings, $l_m$ and $l_p$ digits long, respectively. TCRs

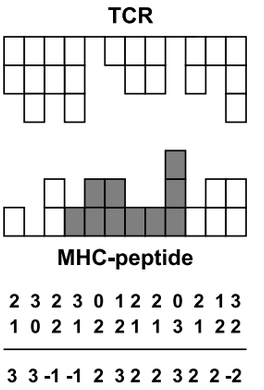


**Fig. 3.** The string representation of MHC-peptide and TCR interactions. The affinity, K, of the interaction between a TCR and an MHC-peptide complex, is the sum of Ki, the interactions between facing digits in the two aligned strings, which represent the complementarity between these digits. In the example shown here, the string length is 12 digits and the affinity K=18

are modeled as random strings, $(l_m + l_p)$ digits long. The number of MHC alleles expressed in an individual is $n_m$. A given MHC allele can present a panel of $n_p$ distinct self peptides. We also assume that, because of allele-specific binding motifs, MHC molecules of different haplotypes present different subsets of self peptides. This is mathematically equivalent to presenting the same peptides in different conformations. Thus, a TCR is selected by a self environment composed of $n_m \times n_p$ MHC-peptide complexes. Selection is implemented by introducing two affinity thresholds for positive and negative selection, $K_+$ and $K_-$ ($K_+ < K_-$). Clones binding at least one self MHC-peptide complex with affinity K larger than $K_+$ survive positive selection. Negative selection deletes clones binding one or more self MHC-peptide complexes with affinity K larger than $K_-$. The values of $K_+$ and $K_-$ are derived from experimental data by considering the fractions of clones surviving the different stages of selection[5]. Furthermore, a mature clone is considered activated by a set of MHC-peptide complexes if the affinity of binding between its TCR and at least one MHC-peptide complex in this set is greater than $K_-$.

All the parameters of the model can be estimated from biological data except the number of possible digits, which controls the discretization of the model and has no effect on it as long as it is chosen large enough[4]. Calculations were done for all parameter combinations that can be obtained from the biologically plausible ranges. The likely operating parameter regime of affinity-driven selection is determined by comparing simulation results to experimental data on antigen response frequency, alloreactivity and self restriction.

Using these comparisons, we found that affinity-driven selection is compatible with experimental estimates of these latter quantities only if the following conditions hold:

– TCRs see more peptide residues than MHC polymorphic residues. We showed that if alloreactivity lies in the range of 1–24%, then one T cell in $10^4$–$10^6$ recognizes a given pathogen, and if the repertoire is self MHC-restricted (by any amount), then the relative contribution of peptide (as opposed to MHC polymorphic) residues to TCR binding should be in the range of 36–85% on average. In other words, if a TCR contacts on average 4 peptide residues, as suggested by several MHC-peptide crystal structures[5], then it should also contact an average of 0.7–6.5 MHC polymorphic residues;

– The majority of positively selected clones are deleted by negative selection, that is, alloreactivity is in the experimental range (1–24%) only when the fraction of positively selected clones that also survive negative selection is less than 37%. Thus the affinity model is not compatible with the upper bound for this fraction, 95%, proposed by LAUFER et al.[10];

– Between 1 and 3.6 clonal divisions occur on average in the thymus after completion of TCR rearrangement;

– Selection is driven by $10^3$–$10^5$ self peptides. Considering together the constraints on alloreactivity and on the response frequency to foreign peptides leads to the conclusion that the number of self peptides that drive thymic selection must be in the range $10^3$–$10^5$.

T cell development has been investigated from a variety of perspectives encompassing, on the one hand, molecular events underlying T cell selection and activation and, on the other hand, dynamics of thymic cell populations. However, the two levels of observation are not independent: molecular processes control population dynamics, and both determine T cell repertoire properties. A thorough understanding of T cell selection cannot be achieved without revealing the quantitative relationships between the these two levels of description of thymic selection. For example, alloreactivity frequency is, by definition, a global property of the T cell repertoire, not the property of any particular TCR/MHC-peptide ternary complex. Our model is a first step in bridging the gap between microscopic molecular events controlling T cell selection and activation and their effect on macroscopic properties of the repertoire. This study revealed how quantitative parameters inferred from data on cellular and molecular level experiments contribute to the shaping of the mature T cell repertoire.

In the work describe here, T cell clones were again dealt with one at a time, rather than simultaneously. Hence, we could not study the effects of competition, or of a changing MHC-peptide "landscape", on the development of the TCR repertoire. These remain open issues for future work.

## Conclusion

In the above I described several examples of theoretical models of lymphocyte repertoire generation and selection. My aim was to point at the difficulties this type of modeling involves and demonstrate various types of solutions. The main message I would like to convey is that theoretical tools, in particular computer simulations, are very helpful for grasping, visualizing and analyzing lymphocyte repertoire meta-dynamics, in spite of their immense complexity. One needs only to bear in mind that no model can, nor should, attempt to be a complete description of reality. The model should be simplified enough to enable a thorough analysis, while focusing in detail on those features of the system in question which are essential for the problem studied. All models described above may potentially be developed further for future study whenever the need arises for complex, quantitative "thought experiments" to accompany experimental investigations.

*Acknowledgment.* The authors gratefully acknowledge collaborations with V. Detours, M. Shannon, A. Perelson, S. Litwin and H. Piper, co-authors of some of the studies reviewed above, and also useful discussions with M. Weigert and P. Seiden.

## References

1. ALLEN D., SIMON T., SABLITZKY F., RAJEWSKY K. and CUMANO, A. (1988): Antibody engineering for the analysis of affinity maturation of an anti-hapten response. EMBO J., **7**, 1995–2001.
2. ARAKAWA H., SHIMIZU T. and TAKEDA S. (1996): Re-evaluation of the probabilities for productive rearrangements on the κ and λ loci. Int. Immunol., **8**, 91–99.
3. COLECLOUGH C., PERRY R. P., KARJALAINEN K. and WEIGERT M. (1981): Aberrant rearrangements contribute significantly to the allelic exclusion of immunoglobulin gene expression. Nature, **290**, 372–378.
4. DETOURS V., MEHR R. and PERELSON A. S. (1999): A quantitative theory of affinity-driven T cell repertoire selection. J. Theor. Biol., **200**, 389–403.
5. DETOURS V., MEHR R. and PERELSON A. S. (2000): Deriving quantitative constraints T-cell selection from data on the mature T-cell repertoire. J. Immunol., **164**, 121–128

6. FOSTER S. J., BREZINSCHEK H. P., BREZINSCHEK R. I. and LIPSKY P. E. (1997): Molecular mechanisms and selective influences that shape the κ gene repertoire of $IgM^+$ B cells. J. Clin. Invest., **99**, 1614–1627.
7. KALINKE U., BUCHER E. M., ERNST B., OXENIUS A., ROOST H.-P., GELEY S., KOFLER R., ZINKERNAGEL R. and HENGARTNER H. (1996): The role of somatic mutation in the generation of the protective humoral immune response against vesicular stomatitis virus. Immunity, **5**, 639–652.
8. KELSOE G. (1996): Life and death in germinal centers (redux). Immunity, **4**, 107–111.
9. KOSCO-VILBOIS M. H., ZENTGRAF H., GERDES J. and BONNEFOY J. (1997): To "B" or not to "B" a germinal center? Immunol. Today, **18**, 225–230.
10. LAUFER T. M., DEKONIG J., MARKOWITZ J. S., LO D. and GLIMCHER L. H. (1996): Unopposed positive selection and autoreactivity in mice expressing class II MHC only on thymic cortex. Nature, **383**, 81–85.
11. LIVAK F., PETRIE H. T., CRISPE I. N. and SCHATZ D. G. (1995): In-frame TCR δ gene rearrangements play a critical role in αβ : γδ T cell lineage decision. Immunity, **2**, 617–622.
12. MALISSEN M., TRUCY J., JOUVIN-MARCHE E., CAZENAVE P.-A., SCOLLAY R. and MALISSEN B. (1992): Regulation of TCR α and β gene allelic exclusion during T-cell development. Immunol. Today, **13**, 315–322.
13. MALLICK C. A., DUDLEY E. C., VINEY J. L., OWEN M. J. and HAYDAY A. C. (1993): Rearrangement and diversity of T cell receptor β chain genes in thymocytes: A critical role for the β chain in development. Cell, **73**, 513–519.
14. MEHR R., SHANNON M. and LITWIN S. (1999): Models for antigen receptor gene rearrangement. I. Biased receptor editing in B cells: implications for allelic exclusion. J. Immunol., **163**, 1793–1798.
15. NEMAZEE D. (1998): Theoretical limits to massive receptor editing in immature B cells. Curr. Top. Microbiol. Immunol., **229**, 163–171.
16. PADOVAN E., CASORATI G., DELLABONA P., MEYER S., BROCKHAUS M. and LANZAVECCHIA A. (1993): Expression of two T cell receptor α chains: Dual receptor T cells. Science, **262**, 422–424.
17. PERELSON A. S. and KAUFFMAN S. A. [eds.] (1991): Molecular evolution on rugged landscapes: proteins, RNA and the immune system. Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley, California, **9**, 73–176.
18. PERELSON A. S. and OSTER G. F. (1979): Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-nonself discrimination. J. Theor. Biol., **81**, 645–670.
19. PIPER H., LITWIN S. and MEHR R. (1999): Models for antigen receptor gene rearrangement. T Cell receptor editing: Allelic exclusion or inclusion? J. Immunol., **163**, 1799–1808.
20. RADIC M. Z. and ZOUALI M. (1996): Receptor editing, immune diversification and self-tolerance. Immunity, **5**, 505–511.
21. SHANNON M. and MEHR R. (1999): Reconciling repertoire shift with affinity maturation: the role of deleterious mutations. J. Immunol., **162**, 3950–3956.
22. TAKAHASHI Y., DUTTA R. P., CERASOLI D. M. and KELSOE G. (1998): *In situ* studies of the primary immune response to (4-hydroxy-3-nitrophenyl) acetyl. V. Affinity maturation develops in two stages of clonal selection. J. Exp. Med., **187**, 885–895.
23. WOOD D. L. and COLECLOUGH C. (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**, 4756–4760.

Received in December 1999
Accepted in January 2000