

A Model for TCR Gene Segment Use

Aryeh Warmflash and Aaron R. Dinner

J Immunol 2006; 177:3857-3864; ;

doi: 10.4049/jimmunol.177.6.3857

<http://www.jimmunol.org/content/177/6/3857>

This information is current as
of September 15, 2014.

References This article **cites 31 articles**, 17 of which you can access for free at:
<http://www.jimmunol.org/content/177/6/3857.full#ref-list-1>

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

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

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

A Model for TCR Gene Segment Use¹

Aryeh Warmflash* and Aaron R. Dinner^{2†‡§¶}

The TCR α -chain is assembled by somatic recombination of variable (V) and joining (J) gene segments at the CD4⁺CD8⁺ stage of development. In this study, we present the first analytical model for deletional rearrangement and show that it is consistent with almost all available data on V $_{\alpha}$ J $_{\alpha}$ use in mice and humans. A key feature of the model is that both “local” and “express service” models of rearrangement can be obtained by varying a single parameter that describes the number of gene segments accessible at a time. We find that the window is much larger for V $_{\alpha}$ segments than J $_{\alpha}$ segments, which reconciles seemingly conflicting data for the former. Implications for the properties of the repertoire as a whole and experiments that seek to probe them are discussed. Special considerations for allelic inclusion are treated in the Appendices. *The Journal of Immunology*, 2006, 177: 3857–3864.

To provide protection from pathogens, T lymphocytes must react to an enormous variety of foreign molecules. The specificity of a clone for Ag is determined by the TCRs on its surface, which are typically heterodimers of α - and β -chains. The diversity of the TCR repertoire derives in large part from the fact that both α - and β -chains are generated during intrathymic development by somatic recombination of gene segments encoding the constant and variable portions of these molecules. The latter include variable (V), diversity (D), and joining (J) gene segments in the case of β -chains and V and J gene segments in the case of α -chains.

Obtaining a quantitative understanding of how physical factors impact rearrangement of TCR loci is important for assessing the diversity of specificities in the TCR repertoire. Loci with D segments can only rearrange each chromosome once because all but the used D segment are deleted by the primary rearrangement. In contrast, those without D can rearrange repeatedly; here, we focus on α loci. In mice, there are 104 V $_{\alpha}$ segments and 61 J $_{\alpha}$ segments; in humans, the number of J $_{\alpha}$ segments is about the same, but there are only about half as many V $_{\alpha}$ segments. When a rearrangement brings specific V $_{\alpha}$ and J $_{\alpha}$ segments together, intervening gene segments are deleted rather than inverted due to the orientation of the recombination signal sequences (1).

There is general agreement that J $_{\alpha}$ segments are used in an essentially sequential manner (termed “local service”), starting at the 5′ end (proximal to the V $_{\alpha}$ segments) and proceeding to the 3′ end (2–8). There is no such consensus concerning the V $_{\alpha}$ segments. One group reported preferential use of 3′ (J $_{\alpha}$ proximal) over 5′ (J $_{\alpha}$ distal) V $_{\alpha}$ segments (8), but another argued for nonsequential use (“express service”) of V $_{\alpha}$ segments based on the fact that there is only a loose correlation in V $_{\alpha}$ /V $_{\alpha}$ -pairs in cells with productive rearrangements on both chromosomes (6). In the present paper, we develop a mathematical model that reconciles these seemingly conflicting observations. We estimated the V $_{\alpha}$ and J $_{\alpha}$ window sizes

(W_V and W_J , respectively) from the average separations of segments on different alleles in selected cells and then show that, without further adjustment, these parameters yield good agreement with independent experimental data on V $_{\alpha}$ and J $_{\alpha}$ use in the selected TCR repertoire. We find that the window is much larger for V $_{\alpha}$ segments than J $_{\alpha}$ segments (37 segments compared with 13 segments, respectively), which accounts for quasisquential use of V $_{\alpha}$ gene segments (8) with only loose correlation between alleles (6). Beyond shedding light on mechanisms of rearrangement, the model is useful for extrapolating statistics on the TCR repertoire from data obtained with limited numbers of V $_{\alpha}$ -specific mAbs, as demonstrated in the Appendices.

Materials and Methods

We derived an expression for the number of possible ways each V $_{\alpha}$ J $_{\alpha}$ gene configuration can be generated and then used it to estimate probabilities of observing in the periphery: 1) V $_{\alpha}$ and J $_{\alpha}$ segments irrespective of with which gene they are paired, 2) V $_{\alpha}$ J $_{\alpha}$ pairs, and 3) V $_{\alpha}$ /V $_{\alpha}$ and J $_{\alpha}$ /J $_{\alpha}$ pairs in dual TCR cells.

Counting paths to V $_{\alpha}$ J $_{\alpha}$ gene configurations

As mentioned in the Introduction, we can treat V $_{\alpha}$ and J $_{\alpha}$ rearrangements with the same expressions by encoding the degree to which each set is used sequentially by a parameter (W) that describes the number of accessible gene segments. In other words, we assumed that rearrangements at any time can be made to the W most proximal gene segments remaining; this scheme is discussed in further detail below. A small value of W corresponds to more sequential use, and a large one to more random use.

For clarity, we numbered the V $_{\alpha}$ and J $_{\alpha}$ segments separately according to their initial (prerearrangement) positions starting from the most proximal ones (9) and refer to the ordered list of (V $_{\alpha}$ or J $_{\alpha}$) segments sampled in multiple rearrangements of a chromosome as a “path” (Fig. 1). The specific goal of this section is to determine $g(n, k, W)$, the number of paths of length k (i.e., those corresponding to k rearrangements) that end in segment n subject to the window constraint described above.

To this end, we determined the number of unrestricted such paths and then removed the contribution from those that violate the window constraint. We began by noting that the segment to which the last rearrangement is made is fixed, and the number of ways of choosing the remaining increasing sequence of $k - 1$ segments from the $n - 1$ prior ones was

$$\binom{n-1}{k-1} = \frac{(n-1)!}{(k-1)!(n-k)!} \quad (1)$$

From these unrestricted paths, we subtracted the number of combinations with immediately sequential segments separated by more than W (“forbidden jumps”). For this purpose, we defined the quantity

$$f(n, k, W, m) = \binom{k}{m} \binom{n-mW-1}{k-1} \quad (2)$$

where m is the number of forbidden jumps in a path. The first binomial coefficient gives the number of ways of placing the m forbidden jumps

*Department of Physics, [†]Department of Chemistry, [‡]Committee on Immunology, [§]Institute for Biophysical Dynamics, and [¶]James Franck Institute, University of Chicago, Chicago, IL 60637

Received for publication March 24, 2006. Accepted for publication June 29, 2006.

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.

¹A.W. was supported by a National Science Foundation graduate research fellowship.

²Address correspondence and reprint requests to Dr. Aaron R. Dinner, The University of Chicago, Gordon Center for Integrative Science, 929 E. 57th Street, Chicago, IL 60637. E-mail address: dinner@uchicago.edu

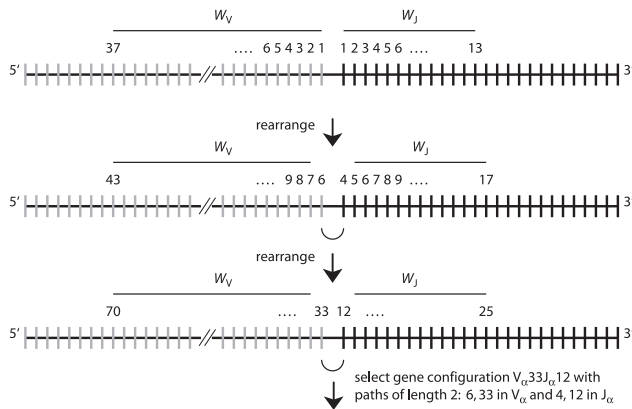


FIGURE 1. Schematic of the model. A path of length of two rearrangements is illustrated. In each rearrangement, intervening DNA is deleted, and the window of accessible gene segments is shifted to begin at the point of the last rearrangement.

among the k rearrangements. The second gives the number of ways of choosing the gene segments for the first $k - 1$ rearrangements given that certain jumps are required to be of length greater than W . In the advent that $n - mW < 1$, we take the second factor on the right side to be 0 here and below.

It is important to note that the quantity $f(n, k, W, m)$ is not in itself the number of paths with m forbidden jumps because it also counts those with more than m forbidden jumps repeatedly. In particular, a path with q ($q \geq m$) forbidden jumps contributes $q!/m!(q - m)!$ times to $f(n, k, W, m)$ because there are that many ways of choosing the minimum of m jumps specified to violate the window constraint. To obtain only the number of allowed paths, begin by considering $f(n, k, W, 0)$. By comparing Eqs. 1 and 2, it can be seen that $f(n, k, W, 0)$ is the number of unrestricted paths. It is thus necessary to subtract $f(n, k, W, 1)$ from $f(n, k, W, 0)$. Then paths with one forbidden jump will no longer be counted. However, doing so overcompensates with regard to the paths with more than one forbidden jump because they are counted $q!/1!(q - 1)! = q$ times in $f(n, k, W, 1)$. Adding back $f(n, k, W, 2)$ overcompensates for paths with more than two forbidden jumps in the opposite direction, and so it is necessary to continue subtracting and adding terms with increasing numbers of forbidden jumps until the maximum possible (k) is reached. Generalizing, in this alternating sum of $f(n, k, W, m)$ over m , the number of times paths with $q \geq 1$ are counted is

$$\sum_{m=0}^{m=q} (-1)^m \binom{q}{m} = (1 - 1)^q = 0 \quad (3)$$

where the first equality derives from the binomial expansion. In other words, in an alternating sum of $f(n, k, W, m)$ over m , contributions from paths with forbidden jumps cancel. Thus, the number of paths that satisfy the window constraint is

$$g(n, k, W) = \sum_{m=0}^{m=k} (-1)^m f(n, k, W, m) \\ = \sum_{m=0}^k (-1)^m \binom{n - mW - 1}{k - 1} \binom{k}{m}. \quad (4)$$

Eq. 4 is the main result of this paper, and all subsequent expressions are derived from it.

Probability distributions

To use Eq. 4 to estimate the probabilities of observing particular V_α and J_α segments in various contexts, we assume that the likelihoods of making productive rearrangements (P_r) and being selected (P_s) are constants. Clearly, the latter does in fact vary for specific gene segments, but here we are concerned with the overall statistics of the repertoire; moreover, we effectively average over different β -chains and CDR sequences. Given P_r and P_s as well as the window sizes, we combine these variables into case-dependent aggregate probabilities for attempting to generate a particular

gene configuration and then being or not being selected (a and b , respectively).

V_α or J_α probability distributions. Here, we consider the probability of observing a single V_α or J_α segment (indexed n), irrespective of with which gene it is paired [$P(n)$]. The likelihood of making a productive rearrangement and then being selected is the product $P_r P_s$, and its complement is $1 - P_r P_s$. We normalize these expressions by the window size for a and b because there is an equal chance of picking any accessible gene segment:

$$a = P_r P_s / W \quad \text{and} \quad b = (1 - P_r P_s) / W. \quad (5)$$

To obtain the probability of interest, we perform a sum over the numbers of rearrangements ($k < N_r$, where N_r is the maximum number of rearrangements, which acts as a surrogate for time in the thymus) weighted according to the number of paths:

$$P(n) = \frac{a}{Q} \sum_{k=1}^{N_r} g(n, k, W) b^{k-1} \quad (6)$$

where Q is a normalization factor determined by summing over all possible values of n . The factor b^{k-1} arises from the fact that to be selected in k rearrangements, a clone must fail to be selected in $k - 1$ previous attempts.

$V_\alpha J_\alpha$ pair distributions. In the case that we want the probability of observing V_α segment n_v with J_α segment n_j [$P(n_v, n_j)$], it is necessary to normalize the aggregate probabilities a and b instead by the number of possible pairs ($W_V W_J$):

$$a = P_r P_s / W_V W_J \quad \text{and} \quad b = \frac{1 - P_r P_s}{W_V W_J} \quad (7)$$

where W_V and W_J are sizes of the windows of accessible V_α and J_α segments. As above, we sum over the number of rearrangements weighted by the numbers of paths to n_v and n_j :

$$P(n_v, n_j) = \frac{a}{Q} \sum_{k=1}^{N_r} g(n_v, k, W_V) g(n_j, k, W_J) b^{k-1} \quad (8)$$

where again Q is a normalization factor, but, in this case, it is computed by considering all possible pairs of V_α and J_α segments.

V_α / V_α and J_α / J_α distributions in dual TCR cells. Statistical data are available for clones with two productively rearranged α loci (6, 10), so it is of interest to determine the likelihood of pairing like types of segments on different alleles. Although the expression for this joint probability is similar to that for a particular $V_\alpha J_\alpha$ pair, a and b must be adjusted for dual TCR cells. Assuming for convenience that both alleles are rearranged simultaneously but independently (see Appendix and Ref. 11 for discussions of this simplification), the probability of productively rearranging both chromosomes and then being positively selected is $P_r^2 P_s$. This product determines a . The aggregate probability b accounts for clones that edit their gene configurations. It is thus necessary to subtract the contributions from both selected dual and single TCR cells: $P_r^2 P_s$ and $P_r(1 - P_r) P_s$, respectively (the possibility of cell death is addressed below). Normalizing by the number of gene segment pairs from the two sets of interest:

$$a = P_r^2 P_s / W^2 \quad \text{and} \quad b = [1 - P_r^2 P_s - 2P_r(1 - P_r) P_s] / W^2. \quad (9)$$

The probability for selecting a cell with segment n_1 from the first chromosome and segment n_2 from the second is

$$P(n_1, n_2) = \frac{a}{Q} \sum_{k=1}^{N_r} g(n_1, k, W) g(n_2, k, W) b^{k-1}, \quad (10)$$

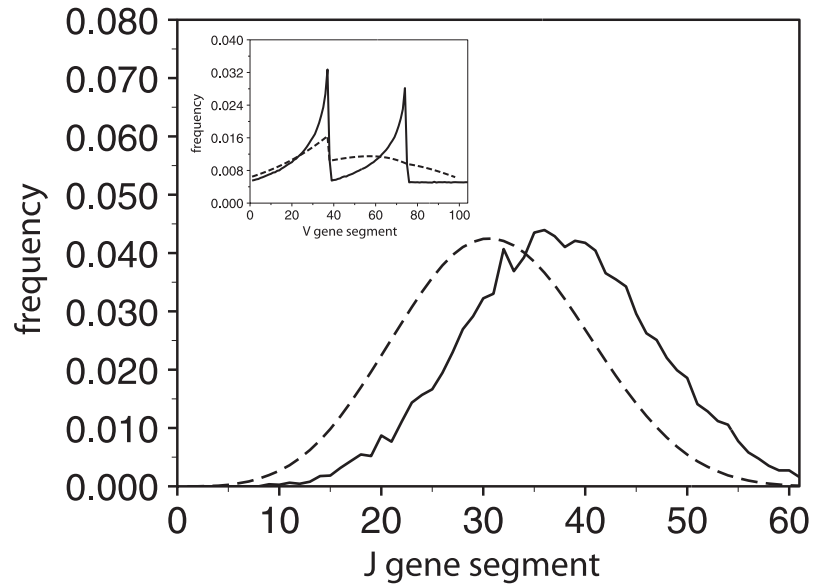
which differs from Eq. 8 in that both counting factors use the same window size, and the normalization factor Q is adjusted accordingly.

Evaluation of simplifications

The model described above makes no reference to underlying molecular details and is consistent with any mechanism that only allows a certain number of sequential gene segments to be accessible at a given time. To make the model mathematically tractable, additional simplifications were made. In particular, we assume that rearrangements can always access the W most proximal remaining segments. However, data for the Ig H chain locus suggest that the windows of accessible gene segments are predetermined (12). In this case, the number of segments to which rearrangements can be made varies because the ends of the windows are restricted to fixed points along the locus.

To determine how varying the mechanism for making gene segments accessible impacts the results, we performed stochastic simulations for models with a sliding window of constant size (as in the mathematical

FIGURE 2. Comparison of different mechanisms for how gene segments become accessible. J_α use for V19, the most J_α -distal V_α gene segment, from the analytical sliding window model (dashed line) and simulation of the fixed block model (solid line). *Inset*, Mouse V_α gene use is plotted as a function of index starting at the proximal end with lines as in the main figure. Calculations of the sliding window model used to derive the analytic expressions were performed with $W_V = 37$ and $W_J = 13$. Simulations of a model in which a new block of segments opens after every other rearrangement were performed with fixed window ends at indices 37, 74, and 104 (assumed to be the last based on Ref. 1).



derivations above) and one with fixed ends but variable size (as mentioned immediately above). The behavior of the latter depends on the number of rearrangements between shifts in the window position and we tried several values. Overall, the results from the two models are similar (Fig. 2). The main difference is that, when the window moves infrequently in the case of predetermined regions of accessible gene segments, segments toward the ends of the windows are used to a somewhat greater degree. Indeed, the abrupt changes in use across the boundaries of the windows are similar in shape to distributions observed for distal J_α segments in cells that are forced to edit rather than die due to transgenic expression of Bcl-x_L (7).

For simplicity, we also explicitly considered only one allele in deriving the first two probability distributions above. However, positive selection following functional rearrangement of a second chromosome terminates rearrangement of the chromosome of interest. Because the alleles are otherwise independent, we can easily account for both by interpreting N_r as the number of rearrangements per allele rather than the total. Computer simulations confirmed that, as long as the normalization is treated consistently (see Appendix B), considering one allele with N_r rearrangements and two alleles with $2N_r$ rearrangements yielded identical results to within simulation error (data not shown).

In the derivations above, cell death (due to autoreactivity or neglect) is not considered explicitly for clarity. Accounting for this phenomenon requires subtracting from the numerators of the aggregate probability b in Eqs. 5, 7, and 9 products of the form $P_r P_d$ where P_d is the probability of cell death. As mentioned with regard to P_s , treating P_d as a constant is expected to be adequate because our focus is on the statistics of the repertoire rather than specific $\alpha\beta$ TCR heterodimers.

Lastly, the T early α promoter situated upstream of the J_α segments appears to target primary rearrangements to the 5' end of that locus, but there is also evidence that a second *cis*-regulatory element initiates rearrangements at a point further downstream (13). This possibility can be incorporated into the model by using in place of $g(n,k,W)$ the weighted average

$$P_u g(n,k,W) + (1-P_u) \frac{W}{W_d} \left[h(n,k) + \sum_{i=1}^{W_d} g(n-n_d-i-1, k-1, W) \right] \quad (11)$$

where P_u is the probability of initiating rearrangement at the upstream targeting element, W_d is the number of gene segments in the downstream window, and n_d is the index of the first gene in the downstream window. The downstream promoter targets rearrangements to $J_{\alpha 49}$ - $J_{\alpha 45}$ (13), which corresponds to $W_d = 5$ and $n_d = 13$. In Eq. 11, the first term, which is weighted by P_u , accounts for the fraction of rearrangements to gene segment n starting from the V-proximal end of the J_α locus as in the original model. The second term, which is weighted by $1 - P_u$, adjusts for the primary rearrangements to the downstream window. The function $h(n,k)$ counts these targeting events; it is 1 if $k = 1$ and n is inside the downstream window and 0 otherwise. The sum counts paths that start in the downstream window and end at gene segment n . The arguments to the function g in this

case can be understood as follows. The first, $n - n_d - i - 1$, is the number of gene segments between the gene segment of interest (n) and the initiation point for rearrangement (the i -th gene segment of the downstream window), which corresponds to shifting the gene segment indices to count from the initiation point. The second argument, $k - 1$, is the number of secondary rearrangements, which reflects the fact that the primary rearrangement is already determined. The third argument is simply the window size, which we take to be the same for rearrangements starting from either *cis*-regulatory element. For this same reason, one factor of W/W_d is necessary to correct for the size of the primary rearrangement window, which enters through the composite probabilities a and b in Eqs. 5–10.

Based on the areas of the peaks in Fig. 3 in Ref. 13, we estimate P_u to be ~ 0.7 . With this choice, we recalculated the curves in Fig. 3 (data not shown) and found that the modified model yields $W_J = 10$ gene segments, which is somewhat smaller than our original estimate (more generally, W_J increases with P_u until the original model is recovered for $P_u = 1$). As a result, secondary rearrangements tend to be shifted upstream slightly; primary rearrangements are shifted downstream slightly due to the second *cis*-regulatory element. Similar agreement with the data is obtained overall. Specifically, the predictions for $V_{\alpha 20S1}$ in Table I and $V_{\alpha 6}$ in Fig. 4a are improved (compare with *inset*), those for $V_{\alpha 19}$ in Table I and Fig. 4b are a bit poorer (data not shown).

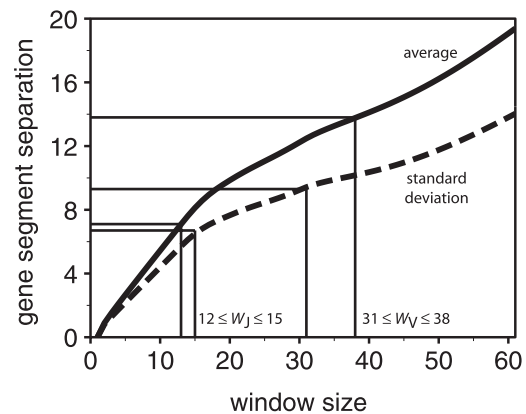


FIGURE 3. Average separation of like segments on different alleles as a function of window size: $\sum_{n_1, n_2} |n_1 - n_2| P(n_1, n_2)$, where $P(n_1, n_2)$ is the probability of obtaining $V_\alpha n_1/V_\alpha n_2$ or $J_\alpha n_1/J_\alpha n_2$ selected cells, computed with Eq. 10 with b as in Eq. 9 and $a = a_1$ in Eq. B2. Because the combinatorial factors involved are often very large, standard computer arithmetic was not sufficient for the accurate computation of these sums; we used the GNU Multiple Precision Arithmetic Library for this and all other numeric evaluations of the analytic expressions.

Table I. Comparison of calculated and observed J_α use for $V_\alpha 20S1$ and $V_\alpha 19$ (17) (with nomenclature updated to the IMGT standard)^a

	$J_\alpha 61-49$	$J_\alpha 48-38$	$J_\alpha 37-27$	$J_\alpha 26-13$	$J_\alpha 12-1$
$V_\alpha 20S1$	67 (95)	33 (5)	0 (0)	0 (0)	0 (0)
$V_\alpha 19$	0 (2)	17 (22)	44 (44)	35 (29)	4 (2)

^a The nonparenthetical entry for each J_α range is experimentally measured and the parenthetical entry is the frequency predicted by the model. Data for the $V_\alpha 2$ and $V_\alpha 16$ superfamilies were not included in the comparison because individual members could not be distinguished experimentally.

Choice of parameters

There are five parameters in the model: the V_α and J_α window sizes (W_V and W_J), the probability that rearrangements are productive (P_r), the probability of positive selection following productive rearrangements (P_s), and the maximum number of rearrangements per allele (N_r). To estimate values for W_V and W_J from a relatively small amount of data, we assume values for P_r and P_s . We take P_r to be 0.3; this choice is somewhat less than the one-in-three chance that a rearrangement is in-frame to account for pseudogenes and the possibility of generating stop codons (14). There is little information from experiments to guide the choice of P_s , the probability of selection following productive rearrangement. We expect it to be small because <5% of thymocytes are positively selected (Ref. 15, and references therein). Based on this fact, we take P_s to be 0.03, which yields an overall selection probability of $1 - (1 - 0.3 \times 0.03)^5 \times 100\% = 4.4\%$ given the choices of P_r and N_r (discussed below). Nearly identical results were obtained with P_r, P_s values ranging over an order magnitude.

Data from mice incapable of reintroducing recombination activating gene (RAG)³ at the double-positive stage suggest that ~35% of the T cell repertoire is formed by primary α -chain rearrangements (16) (but see Fig. 5 and associated discussion below). Assuming constant P_r and P_s and that the number of gene segments is not limiting, it is straightforward to show that this percentage is theoretically

$$P_{1\circ} = \frac{P_r P_s}{\sum_{k=0}^{N_r-1} (1 - P_r P_s)^k P_r P_s} \times 100\% \\ = \frac{P_r P_s}{1 - (1 - P_r P_s)^{N_r-1}} \times 100\%. \quad (12)$$

This statistic is not very sensitive to the values of P_r and P_s , but it does depend strongly on N_r , since it is in the exponent. Reasonable values for $P_{1\circ}$ in the range 20–34% are obtained with $4 \leq N_r \leq 6$. We take $N_r = 5$, which is consistent with other estimates (15).

Results

In our model for TCR α rearrangement, the degree to which a set of gene segments (V or J) is used sequentially is encoded in a single parameter that describes the size of the window of accessible gene segments (Fig. 1). Using this idea, we derive in *Materials and Methods* an analytic expression for the number of possible rearrangement “paths” leading to a selected gene configuration (Eq. 4), as well as probabilities for observing specific V_α and J_α gene segments and their combinations on selected cells (Eqs. 5–10). Here, we estimate the V_α and J_α window sizes (W_V and W_J , respectively) from the average separations of segments on different alleles in selected cells and then show that, without further adjustment, these parameters yield good agreement with independent experimental data on V_α and J_α use in the overall TCR repertoire.

Separation of gene segments on different alleles

To estimate the numbers of V_α and J_α segments available for rearrangement at any given time, we calculate the average separation between like types of segments on different alleles in selected cells as a function of window size (Fig. 3). Consistent with intuition, the separation between segments on different chromosomes increases

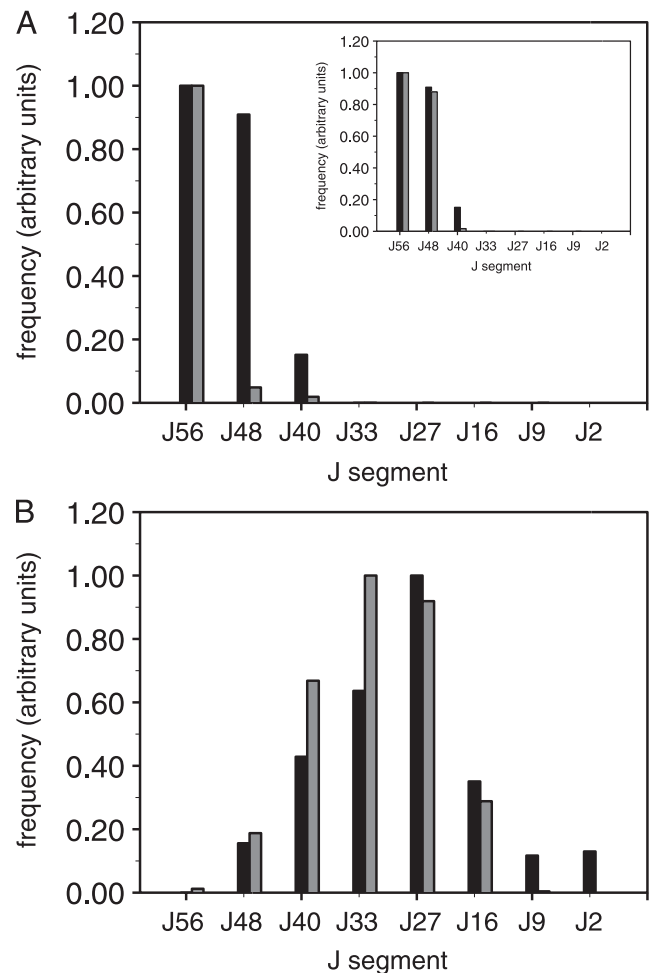


FIGURE 4. Comparison of calculated (gray) and observed (black) J_α use for (A) $V_\alpha 6$ and (B) $V_\alpha 19$. Probabilities of $V_\alpha J_\alpha$ use were computed using Eqs. 7 and 8. Only the relative sizes of bars is important; the scale on the vertical axis is arbitrary. Experimental data is from Ref. 8. *Inset*, $V_\alpha 6$ use calculated with the variation of the model in which primary rearrangement can be to either the V-proximal end of the locus or $J_\alpha 49$ and the four gene segments downstream of it (13).

monotonically with W . In other words, the correlation between alleles decreases as use becomes less sequential.

For the 61 mouse J_α genes, the experimentally measured average separation is 7.1 (SD 6.7), and, for the 58 human V_α genes, it

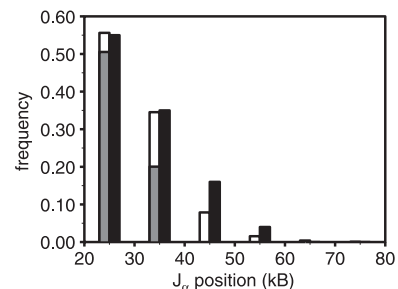


FIGURE 5. Comparison of calculated (gray and white) and observed (black) J_α use in mice incapable of reintroducing RAG at the double-positive stage. We simulated this system by assuming that following each round of rearrangement there is a 70% probability of being incapable of further rearrangements due to insufficient RAG expression; gray contributions are from primary rearrangements, and white contributions are from secondary rearrangements. Experimental data is from Ref. 16.

³ Abbreviation used in this paper: RAG, recombination activating gene.

is 13.8 (SD 9.3) (6). Given these data, we can read the V_α and J_α window sizes off Fig. 3. These averages and SDs correspond to $W_J = 12$ or 15 and $W_V = 38$ or 31 , respectively (Fig. 3); the fact that the averages and SDs yield relatively close estimates for the window sizes suggests that the large SD values observed are inherent to the rearrangement process rather than due to experimental uncertainty.

From the ranges above, we chose $W_J = 13$ and $W_V = 37$ because the mean generally converges more quickly than the second moment of the distribution and the raw data are quite limited. These values confirm the idea that rearrangement of the J_α segments is less random (more sequential) than that of the V_α segments. The value $W_J = 13$ is also consistent with the observation that mice that lack the T early α promoter are unable to use the 10 most proximal J_α segments (5), which suggests a window size of about that number of gene segments.

Due to the limited amount of data, we assume that our values of W_J and W_V are common to mice and humans, which appears to be justified given the remarkably good agreement with experiment that we obtain below. Taking the window sizes to be the same in mice and humans also reconciles seemingly conflicting data on V_α segment use (M. Krangel, unpublished observation). Because 37 gene segments represent roughly 65% of the human but only 35% of the murine V_α segments, the former appear to be used randomly while the latter appear to be used sequentially.

J α distributions for particular V α genes

We now fix the five parameters in the model at the values estimated above ($P_r = 0.3$, $P_s = 0.03$, $N_r = 5$, $W_J = 13$, and $W_V = 37$) and compare calculated $V_\alpha J_\alpha$ pair frequencies with measured ones (8, 17–19). We begin by considering the data from Ref. 8 for J_α segments paired with $V_\alpha 6$ (located at the J-proximal end of the V_α locus) and $V_\alpha 19$ (at the distal end). For these extreme V_α segments, the model and experimentally observed probabilities agree well (Fig. 4). The estimated frequency for $V_\alpha 6$ - $J_\alpha 48$ pairs is very sensitive to whether targeting by the *cis*-regulatory element at $J_\alpha 49$ is considered because almost all of these pairs derive from primary rearrangements and $J_\alpha 48$ falls in the downstream window but not the upstream one. The model predictions compare favorably with the experimental data for $V_\alpha 19$ and $V_\alpha 20S1$ (located at the proximal end of the locus) studied by Huang and Kanagawa (17) as well (Table I). Again, including the downstream initiation site improves the estimates for pair frequencies involving the proximal V_α segment, $V_\alpha 20S1$ (data not shown).

Although the same qualitative trends were observed in Ref. 18, the model cannot reproduce the reported J_α use in detail due to the fact that the distributions are not unimodal. However, it is important to note that these data are putatively for individual members of the $V_\alpha 2$ superfamily, which are difficult to conclusively identify, as noted by Huang and Kanagawa (17). Additional data for this superfamily was obtained recently for mice with only a single functional J_α segment and control animals (19). In this case, reasonable agreement with calculated frequencies for the model in which the ends of the windows are restricted to fixed points along the locus (see Fig. 2), but the model exhibits a greater bias toward J_α -proximal V_α segments than was observed (data not shown).

Regulation of secondary rearrangements

The quasisquential scheme on which our model is based is consistent with the results of knockout experiments directed at elucidating the factors that regulate secondary rearrangements. In mice that are unable to reinduce RAG expression at the double-positive stage because they lack a necessary regulatory element, residual RAG from rearrangement of TCR β at the double-negative stage

catalyzes only limited rearrangement of the TCR α locus. J_α use is restricted to the 5' (proximal) segments in T cells from these mice (16). To simulate these experiments, we performed simulations of our model in which there was a large probability of losing residual RAG and stopping rearrangement following each round of rearrangement. We found that probabilities between 60 and 80% give good agreement with the experimental data (Fig. 5). Although there are data suggesting that the half-life of RAG is short (~ 10 min), they were obtained for the destruction of RAG at the G₁-S phase of the cell cycle (20); that following the formation of the β -chain could be slower. In any event, some RAG must be present in these cells to account for the observed α -chain rearrangements.

Discussion

In this study, we present a model for TCR gene segment use in which the degree that rearrangement is sequential is determined by the size of a window of accessible gene segments. The model is based only on this notion and the fact that intervening gene segments are deleted when V_α and J_α are brought together. We deduced the window sizes from data on the correlations between the two V_α or J_α genes used in selected cells, and showed that without further adjustment, these parameters yield statistics in good agreement between the model and almost all available data on TCR α gene segment use. Although the model cannot reveal the detailed molecular mechanism, it strongly indicates that V_α use is quasisquential, which reconciles seemingly conflicting data for mice and humans.

The model is exactly solvable and provides the first expressions that can be used to extract information directly from data on gene segment use in lymphocytes. Previous theoretical studies were limited to simulations (numeric "experiments") (21, 22) and focused on the use of L chain segments in B cells, in particular J_κ . There are only four functional such gene segments, which are used with a slight bias toward the two more proximal segments. By either assigning probabilities to each of the four (21) or varying the ratio of the likelihoods of choosing each gene segment and the one immediately upstream (22), it was found that quasisquential use best explains the available J_κ data as well.

One consequence of the mechanism identified is that the most distal V_α segments are incapable of pairing with the most proximal J_α segments and vice versa, consistent with Refs. 23 and 24. Although we calculate that 70% of mouse $V_\alpha J_\alpha$ and 58% of human $V_\alpha J_\alpha$ pairs are expressed at frequencies within an order of magnitude of the uniform distribution (with the differences arising from the fact that the V_α locus in humans is roughly twice the size of that in mice), $\sim 4\%$ of $V_\alpha J_\alpha$ pairs are incapable of forming in both species. Thus, the use of specific $V_\alpha J_\alpha$ pairings can vary dramatically depending on their chromosome locations, and care must be used in extrapolating statistics from experiments specific to particular TCR, as discussed in Appendix A.

In general, as rearrangement becomes increasingly sequential, there is a tradeoff between diversity and cell conservation in the thymus. Consequently, it is natural to ask whether the diversity of the TCR repertoire is limited significantly by the quasisquential nature of rearrangement. If rearrangement were totally uniform, every $V_\alpha J_\alpha$ pair would be present in the repertoire with probability $1/N_V N_J$, where N_V and N_J are the total numbers of V_α and J_α genes. To compare different mechanisms, we used our model to calculate the SD of pairing frequencies for all possible values of W_V and W_J (Fig. 6). A lower SD corresponds to a more uniform distribution. Interestingly, the window sizes estimated from the experimental data (Fig. 3) fall close to the bottom of the very shallow basin around the minimum (Fig. 6). Large SDs are obtained for either a perfectly sequential model ($W_V = 1$ and $W_J =$

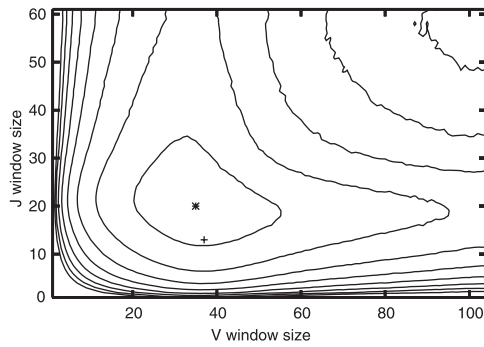


FIGURE 6. SD in pairing frequencies for varying window sizes. Contours indicate lines of constant SD and are spaced by 0.01% with the innermost and outermost lines corresponding to SDs of 0.02 and 0.10%, respectively (compared with a mean probability of $100\%/N_V N_J = 100\%/6344 \approx 0.02\%$). The window sizes corresponding to the minimum SD (*) and estimated from experimental data (+) are indicated. SDs shown assume a V_α locus size of 104 gene segments, which corresponds to mice; similar results are obtained for humans.

1) or a random deletional one ($W_V = 104$ and $W_J = 61$) because gene segments from the proximal and distal ends, respectively, tend to be used disproportionately. These calculations lead us to speculate that the quasisquential rearrangement mechanism that we identified evolved to maximize the diversity of the repertoire.

Appendix A: Corrections for phenotypic allelic inclusion frequencies

Roughly one-quarter of mature $\alpha\beta$ T lymphocytes have productive rearrangements at both their TCR α -chain gene loci (genotypic allelic inclusion) (10), but not all of these cells express two Ag receptors on their surfaces (25–29). Dual TCR cells can thus serve as windows to the mechanisms that regulate T cell surface expression in general and, in turn, how events in the lives of T cells can lead to variations in molecular populations underlying autoimmunity (11). To determine the factors that influence phenotypic allelic exclusion, it is important to be able to quantitate its extent accurately. Unfortunately, only a few mAbs for specific V_α protein segments are available, so it is necessary to extrapolate from limited FACS data to estimate the total fraction of mature T cells that express two TCR on their surface. In this Appendix, we use our model for TCR $V_\alpha J_\alpha$ use together with a brief additional counting argument to improve means for interpreting such experiments.

Correction for overcounting

To determine the extent of phenotypic allelic inclusion in a population of T lymphocytes, FACS is used to count the number of cells that bind reagents specific for two different V_α protein segments. Typically, these data are then used to calculate the fraction

$$f_{ij} = \frac{N_{ij}/N}{(N_i/N)(N_j/N)} \quad (\text{A1})$$

where N_i is the number of cells that are $V_\alpha i^+$, N_{ij} is the number of cells that are $V_\alpha i^+ V_\alpha j^+$, and N is the total number of cells. Although often quoted as such (26–29), f_{ij} is not the frequency of phenotypic allelic inclusion in the total population (F). Rather, it is an approximation for the fraction of $V_\alpha i^+$ and $V_\alpha j^+$ cells which display two receptors at the cell surface (in other words, $f_{ij} \approx f_i$ and f_j , respectively). To relate f_{ij} to F , it is necessary to avoid double-counting cells that express two different V_α protein segments.

The experiments sort cells only according to V_α , which effectively averages over J_α , CDR, and the β -chain. Consequently, it is

not unreasonable to assume $f_{ij} \approx f_i \approx f_j \approx f$ is essentially the same for all V_α (discussed below). Then, denoting the number of $V_\alpha i^+$ cells that express two TCR by d_i , we can write for each V_α protein segment an equation of the form $fN_i = d_i$. Summing over segments,

$$f \sum_i N_i = \sum_i d_i. \quad (\text{A2})$$

The sums above count dual TCR cells with different V_α twice since N_{ij} contributes to N_i , N_j , d_i , and d_j . Denoting the total number of dual TCR cells by d and the number of such cells with the same V_α on both alleles by d_s ,

$$\sum_i d_i = 2d - d_s \quad \text{and} \quad \sum_i N_i = N + d - d_s. \quad (\text{A3})$$

Substituting the expressions in Eq. A3 into Eq. A2 and solving for $F = d/N$, we find

$$F = \frac{f}{2-f} \left[1 + \frac{d_s}{N} \left(\frac{1}{f} - 1 \right) \right] \approx \frac{f}{2-f} \quad (\text{A4})$$

The second (approximate) equality follows from the fact that d_s is expected to be much smaller than N (using the model described in the main text $d_s/N \approx 0.003$), which allows neglect of the term in square brackets. Eq. A4 thus provides a practical means of estimating the fraction of mature T cells that express two TCR on their surfaces from a measurable quantity ($f \approx f_{ij}$, but see below).

Correction for biases in gene rearrangement

V_α gene segments paired in dual TCR cells are weakly correlated (6). As discussed in the main text, such biases in the repertoire come from quasisquential use of the two sets of V_α gene segments at the same rate. In particular, application of Eq. 10 shows that the fraction of dual TCR cells observed depends on the separation of the two V segments detected (Fig. A1). In many cases, f_{ij} as computed from Eq. A1 will be a poor approximation for f .

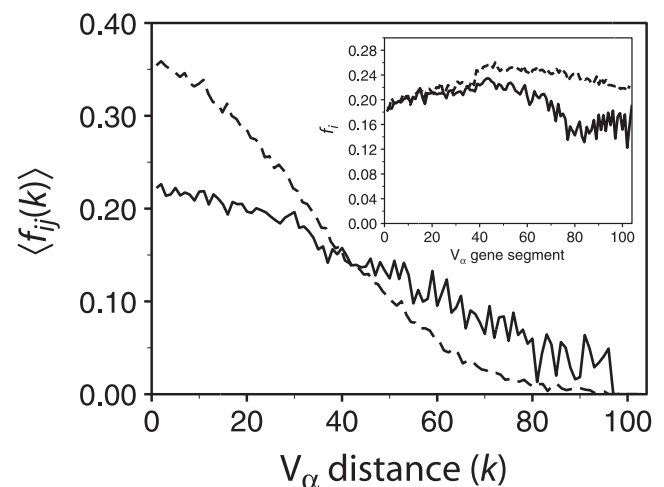


FIGURE A1. Average f as computed by Eq. A1 from simulation data as a function of the distance between the V_α genes studied for $N_r = 5$ (solid line) and $N_r = 10$ (dashed line). The average is performed over all V_α pairs separated by the given distance and is weighted by the numbers of passing single TCR cells which express the V_α genes. In other words, $\langle f_{ij}^{\text{calc}}(k) \rangle = \sum_{|j-i|=k} f_{ij} N_i N_j / \sum_{|j-i|=k} N_i N_j$. Inset, Fraction of cells with $V_\alpha i$ which are dual TCR cells as a function of i calculated from simulation data ($f_i = d_i/N_i$) for $N_r = 5$ (solid line) and $N_r = 10$ (dashed line). The same parameters as in the main text were used.

Here, we show how our model of TCR $V_\alpha J_\alpha$ rearrangement can be used to mitigate the bias that quasisquential deletional rearrangement introduces to interpretation of measured f_{ij} .

To this end, we use the model with the parameters given in the main text to calculate the fraction of dual TCR cells as a function of V_α gene segment position and the calculated value of f_{ij} :

$$f_i^{\text{calc}} = \sum_{j=i} P(i,j)/P(i) \text{ and } f_{ij}^{\text{calc}} = P(i,j)/P(i)P(j) \quad (\text{A5})$$

where $P(i,j)$ is given by Eq. 10 and $P(i)$ is computed as explained in Appendix B. Note that $P(i)$ includes both single and dual TCR cells. We then scale the measured f_{ij} by the ratio of f_{ij}^{calc} and its average value in the model ($\langle f_i^{\text{calc}} \rangle$, where the average is weighted by the number of $V_\alpha i^+$ cells) to obtain a corrected fraction for use in Eq. A4:

$$f \approx f_{ij} \langle f_i^{\text{calc}} \rangle / f_{ij}^{\text{calc}}. \quad (\text{A6})$$

Basing this correction on our model for TCR $V_\alpha J_\alpha$ rearrangement implicitly assumes that the extent of phenotypic allelic inclusion is directly proportional to the extent of genotypic allelic inclusion. Although this cannot be wholly the case, it is not an unreasonable approximation given, as discussed above, the effective averaging over J_α , CDR, and the β -chain.

One feature of Eq. A6 that might be confusing to the reader is that f_i is a property of one gene segment, but f_{ij} is clearly a property of two. However, f_{ij} as defined by Eq. A1 (26–29) is an estimate for $f_i \approx f_j$. Eq. A6 is used to correct for cases when this is a poor estimate due to the separation between gene segments used in the experiment. It is worth noting that the error in f could thus be reduced by instead staining for one V_α protein segment and CD3 as in Ref. 25. Because the ratio of total TCR to CD3 is roughly constant, cells with a low ratio of CD3 to the V_α protein segment studied express only that V_α , while those with a high ratio express another V_α as well.

The discussion immediately above leads to a related point. Given the dependence of f_{ij} on the separation in position of gene segments $V_\alpha i$ and $V_\alpha j$, it is natural to wonder whether the assumption above that $f_{ij} \approx f$ is essentially the same for all V_α protein segments is appropriate. In deriving Eq. A4, the f used represents the actual fraction of dual TCR cells as a function of V_α gene segment position, not the fraction estimated from a specific pair of $V_\alpha i$ and $V_\alpha j$. Due to averaging over $V_\alpha j$, f_i calculated with Eq. A5 is much flatter than $P(i,j)$ calculated with Eq. 10 (Fig. A1, inset). The assumption that f_i is essentially the same for all V_α is thus reasonable. Moreover, it improves as the number of rearrangements becomes larger, although a kink persists at the transition from primary to secondary rearrangement.

Total fraction of phenotypic dual TCR cells

We use the corrections in Eqs. A4 and A6 to re-evaluate previously published data. The reagents used were for the $V_\alpha 2$, $V_\alpha 8$, and $V_\alpha 11$ gene families. In these cases, $P(i,j)$ is the probability that any gene segments from family i is paired with any member from family j and $P(i)$ is the probability that any member of gene family i is used. Generally speaking, these gene families are distributed throughout the chromosome; the $V_\alpha 2$ family is slightly biased toward the 3' end whereas the $V_\alpha 11$ family is slightly biased toward the 5' end. The $V_\alpha 8$ family is nearly evenly distributed. Thus, it is to be expected that $V_\alpha 2/V_\alpha 8$ and $V_\alpha 8/V_\alpha 11$ experiments need little correction, while $V_\alpha 2/V_\alpha 11$ experiments significantly underestimate the fraction of dual TCR cells. In agreement with this expectation, the calculated correction factors for experiments with $V_\alpha 2/V_\alpha 8$, $V_\alpha 8/V_\alpha 11$, and $V_\alpha 2/V_\alpha 11$ are 1.02, 1.03, and 1.86, re-

Table AI. Estimates for phenotypic allelic inclusion frequencies corrected with Eqs. A4 and A6^a

	Previous Result (%)	Corrected Result (%)	Reference
$V_\alpha 2, V_\alpha 8$	10.8	5.8	26
	14–17, 16	7.6–9.5, 8.9	28
	3.7, 3.9	1.9, 2.0	29
$V_\alpha 2, V_\alpha 11$	6.6, 8.7, 11.0, 13.6	3.5, 5.5, 5.9, 7.4	27
	2.2, 3.1	2.1, 3.0	29
	7–11, 18–21	7.0–11.3, 20.1–24.2	28
$V_\alpha 8, V_\alpha 11$	6.8, 8.0, 31.0	6.8, 8.0, 41.0	27
	16.0	9.0	27

^a Results separated by a comma are independent experiments performed by the same author. Ranges of values result from different methods of background subtraction performed by Heath et al. (28). Data for $V_\alpha 3.2$ from Ref. 26 were not considered because of difficulties involved in determining the total number of $V_\alpha 3.2^+$ cells as noted by the author.

spectively. Because the experiments overestimate the fraction of dual TCR cells by nearly a factor of 2 due to overcounting, the data from $V_\alpha 2/V_\alpha 11$ experiments remains nearly the same after applying both corrections. However, since no genotypic bias was introduced in $V_\alpha 2/V_\alpha 8$ and $V_\alpha 8/V_\alpha 11$ experiments, there was significant overestimation of the fraction of dual TCR cells (Table AI). The separation of $V_\alpha 2$ and $V_\alpha 11$ and the consequent biases in their statistics are likely to account for the fact that the same authors generally found a higher fraction of dual TCR cells in $V_\alpha 2/V_\alpha 8$ and $V_\alpha 8/V_\alpha 11$ experiments with the exception of two aberrant data points (18–21% in Ref. 28; 31.0% in Ref. 27, the latter of which may have been due to gating on CD8). Thus, the corrected statistics suggest that the overall rate of phenotypic allelic inclusion is 2–11%, which supports the idea that posttranslational control mechanisms regulate TCR surface expression (see Refs. 11, 30, and 31 for discussion).

Appendix B: Normalization for dual TCR cells

We explicitly considered only one allele in deriving Eq. 6. However, it is important to treat both chromosomes to normalize the probabilities in Eq. A5 consistently. Specifically, in Eq. A5, for the probability of a cell expressing a receptor which uses the m^{th} gene segment regardless of whether it is a single or dual TCR cell, we have

$$P(m) = \frac{a_1}{Q} \sum_{k=1}^{n_R} g(m,k,W)^2 b^{k-1} + \frac{a_2}{Q} \sum_{k=1}^{n_R} 2g(m,k,W)[W^k - g(m,k,W)]b^{k-1} \quad (\text{B1})$$

The first term represents paths in which both alleles have rearranged to the m^{th} gene segment and at least one of them is in-frame. The second term represents paths in which only one of the two alleles has rearranged to the m^{th} gene segment, in which case that allele must be in-frame and the other can be either in- or out-of-frame. The aggregate probabilities are thus

$$a_1 = [P_r^2 P_s + 2(1 - P_r)P_s]/W^2 \quad (\text{B2})$$

and

$$a_2 = [P_r^2 P_s + (1 - P_r)P_s]/W^2$$

with b given by Eq. 9. Finally, we set the normalization Q such that

$$\sum_i P(i) - \sum_{j<i} P(i,j) = 1. \quad (\text{B3})$$

Acknowledgments

We thank Martin Weigert for helpful discussions and critical reading of the manuscript and Barry Sleckman for providing data on V α 2 use in advance of publication.

Disclosures

The authors have no financial conflict of interest.

References

- Krangel, M. S., J. Carabana, I. Abbarategui, R. Schlimgen, and A. Hawwari. 2004. Enforcing order within a complex locus: current perspectives on the control of V(D)J recombination at the murine T cell receptor α/δ locus. *Immunol. Rev.* 200: 224–232.
- Thompson, S. D., J. Pelkonen, R. Marja, S. Jacqueline, and J. L. Hurwitz. 1990. Nonrandom rearrangement of T cell receptor $J\alpha$ genes in bone marrow T cell differentiation cultures. *J. Immunol.* 144: 2829–2834.
- Thompson, S. D., J. Pelkonen, and J. L. Hurwitz. 1990. First T cell receptor α gene rearrangements during T cell ontogeny skew to the 5' region of the $J\alpha$ locus. *J. Immunol.* 145: 2347–2352.
- Petrie, H. T., F. Livak, D. Burtrum, and S. Mazel. 1995. T cell receptor gene recombination patterns and mechanisms: cell death, rescue, and T cell production. *J. Exp. Med.* 182: 121–127.
- Villey, I., D. Caillol, F. Selz, P. Ferrier, and J. P. de Villartay. 1996. Defect in rearrangement of the most 5' TCR- $J\alpha$ following targeted deletion of T early α (TEA): implications for TCR accessibility. *Immunity* 5: 331–342.
- Davodeau, F., M. Difilippantonio, E. Roldan, M. Malissen, J. L. Casanova, C. Couedel, J. F. Morcet, M. Merkenschlager, A. Nussenzweig, M. Bonneville, and M. Malissen. 2001. The tight interallelic positional coincidence that distinguishes T-cell receptor $J\alpha$ usage does not result from homologous chromosomal pairing during V α J α rearrangement. *EMBO J.* 20: 4717–4729.
- Guo, J., A. Hawwari, H. Li, Z. Sun, S. K. Mahanta, D. R. Littman, M. S. Krangel, and Y. W. He. 2002. Regulation of the TCR α repertoire by the survival window of CD4⁺CD8⁺ thymocytes. *Nat. Immunol.* 3: 469–476.
- Pasqual, N., M. Gallagher, C. Aude-Garcia, M. Loiodice, F. Thuderoz, J. Demongeot, R. Ceredig, P. N. Marche, and E. Jouvin-Marche. 2002. Quantitative and qualitative changes in V- $J\alpha$ rearrangements during mouse thymocytes differentiation: implication for a limited T cell receptor α chain repertoire. *J. Exp. Med.* 196: 1163–1173.
- Bosc, N., and M. P. Lefranc. 2004. The mouse (*Mus musculus*) T cell receptor α (TRA) and δ (TRD) variable genes. *Dev. Comp. Immunol.* 27: 465–497.
- Malissen, M., J. Trucy, E. Jouvin-Marche, P.-A. Cazenave, R. Scollay, and B. Malissen. 1992. Regulation of TCR α and β gene allelic exclusion during T-cell development. *Immunol. Today* 13: 315–322.
- Warmflash, A., M. Weigert, and A. R. Dinner. 2005. Control of genotypic allelic inclusion through TCR surface expression. *J. Immunol.* 175: 6412–6419.
- Chowdhury, D., and R. Sen. 2001. Stepwise activation of the immunoglobulin mu heavy chain gene locus. *EMBO J.* 20: 6394–6403.
- Hawwari, A., C. Bock, and M. S. Krangel. 2005. Regulation of T cell receptor α gene assembly by a complex hierarchy of germline promoters. *Nat. Immunol.* 6: 481–489.
- Mason, D. 1994. Allelic exclusion of α chains in TCRs. *Int. Immunol.* 6: 881–885.
- Mason, D. 2001. Some quantitative aspects of T-cell repertoire selection: the requirement for regulatory T cells. *Immunol. Rev.* 182: 80–88.
- Yannoutsos, N., P. Wilson, W. Yu, H. T. Chen, A. Nussenzweig, H. Petrie, and M. C. Nussenzweig. 2001. The role of recombination activating gene (RAG) reinduction in thymocyte development in vivo. *J. Exp. Med.* 194: 471–480.
- Huang, C. Y., and O. Kanagawa. 2001. Ordered and coordinated rearrangement of the TCR α locus: role of secondary rearrangement in thymic selection. *J. Immunol.* 166: 2597–2601.
- Aude-Garcia, C., M. Gallagher, P. N. Marche, and E. Jouvin-Marche. 2001. Preferential ADV-AJ association during recombination in the mouse T-cell receptor α/δ locus. *Immunogen* 52: 224–230.
- Huang, C. Y., B. P. Sleckman, and O. Kanagawa. 2005. Revision of T cell receptor α chain genes is required for normal T lymphocyte development. *Proc. Nat. Acad. Sci. USA* 102: 14356–14361.
- Jiang, H., F.-C. Chang, A. E. Ross, J. Lee, K. Nakayama, K. Nakayama, and S. Desiderio. 2005. Ubiquitylation of RAG-2 by Skp2-SCF links destruction of the V(D)J recombinase to the cell cycle. *Mol. Cell* 18: 699–709.
- 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: 1793–1798.
- Louzoun, Y., T. Friedman, E. L. Prak, S. Litwin, and M. Weigert. 2002. Analysis of B cell receptor production and rearrangement part I. Light chain rearrangement. *Semin. Immunol.* 14: 169–190.
- Dave, V. P., M. Larche, S. D. Rencher, B. F. Koop, and J. L. Hurwitz. 1993. Restricted usage of T cell receptor V α sequence and variable-joining pairs after normal T-cell development and bone marrow transplantation. *Hum. Immunol.* 37: 178–184.
- Roth, M. E., P. O. Holman, and D. M. Kranz. 1991. Nonrandom use of $J\alpha$ gene segments: influence of V α and $J\alpha$ location. *J. Immunol.* 147: 1075–1081.
- Padovan, E., G. Casorati, P. Dellabona, S. Meyer, M. Brockhaus, and A. Lanzavecchia. 1993. Expression of two T cell receptor α chains: dual receptor T cells. *Science* 262: 422–424.
- Elliott, J. I., and D. M. Altmann. 1995. Dual T cell receptor α chain T cells in autoimmunity. *J. Exp. Med.* 182: 953–959.
- Elliott, J. I. 1998. Selection of dual V α T cells. *Eur. J. Immunol.* 28: 2115–2123.
- Heath, W. R., F. R. Carbone, P. Bertolino, J. Kelly, S. Cose, and J. F. A. P. Miller. 1995. Expression of two T cell receptor α chains on the surface of normal murine T cells. *Eur. J. Immunol.* 25: 1617–1623.
- Alam, S. M., and N. R. J. Gascoigne. 1998. Posttranslational regulation of TCR V α allelic exclusion during T cell differentiation. *J. Immunol.* 160: 3883–3890.
- Gascoigne, N. R. J., and S. M. Alam. 1999. Allelic exclusion of the T cell receptor α chain: developmental regulation of a post-translational event. *Semin. Immunol.* 11: 337–347.
- Niederberger, N., K. Holmberg, S. M. Alam, W. Sakati, M. Naramura, H. Gu, and N. R. J. Gascoigne. 2003. Allelic exclusion of the TCR α -chain is an active process requiring TCR-mediated signaling and c-Cbl. *J. Immunol.* 170: 4557–4563.