TCRβ Allelic Exclusion in Dynamical Models of V(D)J Recombination Based on Allele Independence

Etienne Farcot, Marie Bonnet, Sébastien Jaeger, Salvatore Spicuglia, Bastien Fernandez and Pierre Ferrier

J Immunol published online 28 June 2010
http://www.jimmunol.org/content/early/2010/06/28/jimmunol.0904182

Supplementary Material
http://www.jimmunol.org/content/suppl/2010/06/28/jimmunol.0904182.DC1

Why The JI?
• Rapid Reviews! 30 days* from submission to initial decision
• No Triage! Every submission reviewed by practicing scientists
• Speedy Publication! 4 weeks from acceptance to publication

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

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

Email Alerts
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
TCRβ Allelic Exclusion in Dynamical Models of V(D)J Recombination Based on Allele Independence

Etienne Faricot,*† Marie Bonnet,‡ Sébastien Jaeger,‡ Salvatore Spicuglia,‡ Bastien Fernandez,* and Pierre Ferrier†

Allelic exclusion represents a major aspect of TCRβ gene assembly by V(D)J recombination in developing T lymphocytes. Despite recent progress, its comprehension remains problematic when confronted with experimental data. Existing models fail short in terms of incorporating into a unique distribution all the cell subsets emerging from the TCRβ assembly process. To revise this issue, we propose dynamical, continuous-time Markov chain-based modeling whereby essential steps in the biological procedure (D-J and V-DJ rearrangements and feedback inhibition) evolve independently on the two TCRβ alleles in every single cell while displaying random modes of initiation and duration. By selecting parameters via fitting procedures, we demonstrate the capacity of the model to offer accurate fractions of all distinct TCRβ genotypes observed in studies using developing and mature T cells from wild-type or mutant mice. Selected parameters in turn afford relative duration for each given step, hence updating TCRβ recombination distinctive timings. Overall, our dynamical modeling integrating allele independence and noise in recombination and feedback-inhibition events illustrates how the combination of these ingredients alone may enforce allelic exclusion at the TCRβ locus. The Journal of Immunology, 2010, 185: 000–000.
led us to consider an alternative scenario that is not tied into a strictly sequential mode of interallelic activation for V(D)J recombination. In this study, we aimed to use a dynamical approach to model TCRβ gene recombination in an effort to comprehend the stochastic and regulated premises of allelic exclusion within a developmental scheme that would integrate all the observed cell subsets at once. Principally, successive D-to-J and V-to-DJ rearrangements at TCRβ alleles of individual T cells and ensuing feedback inhibition are seen throughout as independent, possibly concurrent and mostly not stringently simultaneous biochemical transactions, with fluctuating modes of initiation and duration. Indeed, evidence that developing T cells exhibit uniformed, biallelic transcription of a Vβ gene before recombination (18) makes allele autonomy in the conduct of TCRβ rearrangements a plausible hypothesis. Besides, the variety of rearranging sequences (19) and multiplicity of molecular factors/mechanisms involved in TCRβ gene recombination and feedback signaling (16, 20) strongly favor a widespread noise in, and randomness of, all these events. Furthermore, because none of the shaping events leading up to a signaling launch (from TCRβ gene transcription to protein synthesis and formation of the pre-TCR) nor the resulting feedback block in rearrangement are likely to take place instantly, the notion of feedback control implies a deferred outcome(s) with the possible prolongation of recombination after a VDJ+ has been made. Hence, dynamical aspects are expected to impinge in a general sense on TCRβ gene recombination and, potentially, allelic exclusion.

With this framework in mind and building on a recent study by Sepulveda et al. (21), which illustrated stochastic modeling of TCRγ/δ gene rearrangement, we depicted TCRβ gene recombination as a continuous-time Markov process contingent on adjustable parameters. Using available data from wild-type (WT) and mutant mice, we obtained quantitative values for these parameters by standard fitting procedures. In these settings, stochastic dynamics combined with feedback control thoroughly accounted for the generation, in experimentally compatible ratios, of cell cohorts comprised of a large proportion of allelically excluded cells along with small subsets of VDJ+VDJc cells and those harboring a VDJ+GL genotype. The parameter values further confirmed our estimates for the mean length of the joining and inhibiting intervals, thus providing a scale of relative timings tailored to TCRβ gene recombination. Our modeling reveals concepts awarding a discrete chromosomal system with the property to display robust allelic exclusion at minimal regulatory cost while keeping the opportunity to maximize genetic diversity via dual allele usage.

**Materials and Methods**

**Source of TCRβ experimental distributions**

Experimental distributions of mouse TCRβ+ T cells used were obtained from published studies and compiled in Tables II and III. Numbers of β-selected T lymphocytes well-defined in terms of TCRβ genotype (Table II) were derived from data sets available in studies by Khor and Sleckman (22) and Senou et al. (23), which used peripheral T cell-derived, cloned hybridomas from WT (22, 23) and engineered mutant βpTα− mice. Distributed numbers of early-developing double-negative (DN) thymocytes prior to β-selection (Table III) were derived from data sets in studies by Aifantis et al. (24, 25), which used FACS-sorted intracytoplasmic TCRβ, cell-surface CD25, and CD11b− small DN cells from WT and pTα− or spl-/76-deficient animals (pTα−/− and spl-/76−/−, respectively).

**Simulation of TCRβ+ T cell distributions and analytic expressions**

To modelize the random dynamics of TCRβ gene recombination in early developing T lymphocytes, we generated a Markov chain based on the transition graph shown in Fig. 1, in which \( x^i(\tau) \in [0,1] \) symbolizes the fraction of individual cells harboring the same genomic status \( i \) (among 14 possible statuses as depicted in Fig. 1), at time \( \tau \) in a differentiating T cell population, the recombination time window of which started at \( t_0 \). The Markov transition matrix \( Q = \begin{pmatrix} Q_{0} & Q_{1} \\ Q_{2} & Q_{3} \end{pmatrix} \) (Table I), composed of the probabilities \( 1/\tau_{T_D}, 1/\tau_{V_D}/23/\tau_{T_D}, 1/\tau_{T} \) for transitions through the corresponding cell states (for a definition of the genomic statuses, recombination time window, and transition rates, see Results, Formulation of the model: overview and Formulation of the model: basic features) is upper triangular due to the feed-forward structure of the transition graph. The Markov chain was exploited using the relation

\[
x^i(\tau) = x^i(t_0) \exp((\tau-t_0)Q), \quad t \geq t_0,
\]

in which \( x^i(t) \) represents the vector \( x^i(t) \in [0,1] \), and, likewise, \( x^i(t_0) = (x^i(t_0)) \in [0,1] \) denotes the initial distribution at the window origin.

In this study, we were first concerned with data from TCRβ+ T cells (in the form of hybridomas) that corresponded to final states in the model (denoted sVDJ+/DJ, sVDJ+/VDJ, see below). Based on relation (1), we made explicit assumptions as follows: the distribution remains constant from the end of the time window onwards (meaning in particular homogeneous proliferation of cells harboring the distinct (sVDJ+/DJ, sVDJ+/VDJ) statuses); i.e., \( x^i(\tau) = x^i(t_0 + 1) \) if \( \tau \geq t_0 + 1 \); and in line with biological evidence that thymus seeding by T cell progenitors (of status denoted GL/GL in the model) proceeds via gated intervals of receptiveness followed by longer refractory periods (26), we postulated that such cells enter the process on a cyclic basis (instead of a constant flow), with every single cell completing its window within the seeding cycle (i.e., for all cells, \( t_0 + 1 \geq \tau \) of the refractory periods). Accordingly, the distribution resulting from any such cycle does not depend formally on \( t_0 \). It reads

\[
x(\tau) = x(0)\exp(\tau Q),
\]

with the initial distribution \( x(0) = x(0)_{14} \) containing GL/GL cells only (i.e., \( x_{GL/GL}(0) = 1 \) and \( x_i = 0 \) for all \( i \neq GL/GL \)). Steady-state recombination takes place during consecutive seeding cycles. Thus, the fractions \( x_i(1) \) for all final states in the model \( i \in \{sVDJ/GL, sVDJ/DJ, sVDJ/VDJ, sVDJ/VDJ^*\} \) (see Results, Formulation of the model: basic features) actually offer a formal expression for the distributions of TCRβ+ T lymphocytes. Given the triangular structure of the matrix \( Q \), analytic formulas for the variables \( x_i(1) \), which, via \( Q \), explicitly depend on the parameters \( \tau_{T_D}, \tau_{V_D}, \tau_i \), may be obtained by solving in a recursive procedure their corresponding linear differential equations. In practice, for simplicity, we used in this study the matrix exponential tool of Mathematica (Wolfram Research, Champaign, IL) to directly deduce these variables from expression (2).

From relation (1), we also computed cell fractions from presorted DNA data in studies by Aifantis et al. (24, 25) using another series of specific statements. In this case, TCRβ+ DN thymocytes prior to β-selection were identified with transient states (VDJ+/...) in the model. Practically, we further presumed that cells ending their recombination time window and not in the (sVDJ+/...) form, including (VDJ+/...), have a limited life span before commencing to apoptosis or diverging toward another lineage. Thus, differing from above where the computations relied on cell fractions integrated over a stationary regimen comprised of multiple cell-seeding cycles, we then had to consider statistical numbers culminated over a limited interval. As an integration period, we chose the nonstationary section of the model corresponding to the timing window \( t = 1 \). Therefore, the sampled fractions (supposedly collected at an instant \( t \) independent of the cell-differentiation course) had to be fitted by the following expression:

\[
X_i(t) = \int_0^t x_i(\tau - 1) d\tau + \int_0^{t_0} x_i(\tau - 1) d\tau,
\]

which integrates the cell populations over the time window that began after \( t-1 \). A simpler expression was achieved by further assuming that experimental sampling is effected once the flood of cells initiating TCRβ recombination has reached a steady-state course (i.e., with the proportions of GL/GL precursors entering the recombination window being constant in time).

Explicitly:

\[
x_i(t_0) = x_i(t - 1) \int_0^{t_0} x_i(\tau - 1 - t_0) d\tau,
\]

implying \( X_i(t) = \int_0^{t_0} x_i(\tau - 1 - t_0) d\tau \)

in which \( x_i(\tau - 1 - t_0) = x_i(t - 1) \exp(\tau Q) \) with, as before, the initial distribution \( x_i(\tau - 1) = x_i(0) \) containing only GL/GL cells. Again, we used the matrix exponential and formal integration tools in Mathematica (Wolfram Research) to obtain an explicit expression of \( X_i(t) = X_i \), which likewise depends on \( \tau_{T_D}, \tau_{V_D}, \tau_i \) by using the \( n \) independent relations

\[
x_i(0) = x(0)_{14} \exp(\tau Q) d\tau + X_i(1),
\]
Least-squares fits and parameter estimation

To fit the analytic predictions of our statistical model to the experimental data sets, we primarily used the least-square method, a standard technique in numerical analysis (27). The method is based on an algorithm designed to tune parameters so as to minimize the Euclidean distance between the predicted distributions and the experimental distributions, relying on the fact that the distance vanishes if (and only if) the two distributions coincide. In this study, given a collection of TCRα* cell numbers \( \{N_i\}_{i \in S} \) (in which \( S \) denotes any subset of cells harboring one potential TCRα* genomic status as depicted in Fig. 1 and in Tables II and III (e.g., sVDJ+/DJ; sVDJ+/VDJ*, sVDJ*/GL; or VDJ*/DJ; VDJ*/VDJ*; VDJ*/VDJ*, respectively), the quantities to minimize would be the distances

\[
\sum_{i \in S} \left( \frac{N_i - x_i(1)}{\sum_{i \in S} x_i} \right)^2
\]

(in fitting data sets for selected T lymphocytes)

or

\[
\sum_{i \in S} \left( \frac{N_i}{\sum_{i \in S} X_i} - X_i \right)^2,
\]

(in fitting data sets for preselected DN lymphocytes)

between the experimental proportions \( \left\{ \frac{N_i}{\sum_{i \in S} N_i} \right\}_{i \in S} \) (in which \( \sum_{i \in S} X_i \) is a normalization factor) retrieved from Tables II or III, respectively, and the corresponding relative fractions \( \left\{ \frac{x_i(1)}{\sum_{i \in S} x_i} \right\}_{i \in S} \) (in which, likewise,

\[
\sum_{i \in S} x_i(1), \text{ or } \left\{ \frac{N_i}{\sum_{i \in S} N_i} \right\}_{i \in S},
\]

computed as described in the previous paragraph.

In practice, numerical calculations consisted of applying built-in procedures and functions (especially function FindMinimum in Mathematica (Wolfram Research) that, starting from arbitrarily chosen values of (\( \tau_{D\gamma}, \tau_{V\gamma\gamma}, \tau_{V\delta} \)), searched for minimizing the above mentioned distances and, from there, directly yielded the matching, resultant parameter triples. Of note, due to finite computation time, a perfectly vanishing distance could not be reached by numerical means. Therefore, to validate the parameter values, we adopted the following criterion: we only retained parameter triples for which the closest integers to the predicted cell numbers \( \left\{ \frac{x_i(1)}{\sum_{i \in S} x_i} \right\}_{i \in S}, \text{ or } \left\{ \frac{N_i}{\sum_{i \in S} N_i} \right\}_{i \in S} \) all coincided with the experimental figures \( \{N_i\}_{i \in S} \).

Continuum of fitting parameters

As described above, the least-squares approach was a comprehensive and practical way to numerically solve the equations

\[
\frac{x_i(1)}{\sum_{i \in S} x_i} = \frac{N_i}{\sum_{i \in S} N_i} \text{ (or } \frac{X_i(1)}{\sum_{i \in S} X_i} \text{ for the parameters } (\tau_{D\delta}, \tau_{V\delta\delta}, \tau_{V\delta})).
\]

Incorporating in this scheme as many equations as the number of analyzed states in \( S \), there are indeed only \( s-1 \) independent conditions to comply with, because one equation is always ascertained by the normalization \( \sum_{i \in S} x_i = \sum_{i \in S} N_i = 1 \).

For all data sets in Tables II and III, \( s = 3 \) (i.e., states sVDJ*/DJ; sVDJ*/VJD*; sVDJ*/GL; and VDJ*/DJ; VDJ*/VJD*; VDJ*/VDJ*, respectively). Accordingly, in each case, there were only two independent equations to resolve the three parameters \( \tau_{D\gamma}, \tau_{V\gamma\gamma}, \text{ and } \tau_{V\delta} \). Therefore, precisely defining those parameters was possible aside from one degree of freedom, and, consequently, the fitting procedure, when feasible, yielded a continuum of triple solutions in the parameter space.

Additional validation of parameter triples

To further access the effectiveness of parameter evaluations, we sought to confirm the above calculations using additional tools and methods. We thus applied the routine optim in \( R \), another minimizing procedure that relies on the well-known Nelder-Mead (downhill simplex) algorithm (28). To be agreed, the proposed estimations had to yield indistinguishable or, if not, similar (according to the above-defined criterion) triple solutions. In addition, to ascertain that solutions were method-independent, we also verified that the same curves of parameter triples were obtained when applying the maximum likelihood method, which determines the most likely parameters by maximizing the sum

\[
\sum_{i \in S} \left( \frac{x_i(1)}{\sum_{i \in S} x_i} \right)^N \text{ (using data sets for selected T lymphocytes)}
\]

or

\[
\sum_{i \in S} \left( \frac{N_i}{\sum_{i \in S} N_i} \right)^N \text{ (using data sets for preselected DN lymphocytes)}
\]

Least-squares fitting procedure and parameter estimation: worked example

Using data from WT mice in the study by Senoo et al. (23), the distance to work example

\[
\left( \frac{N_{sVDJ+/GD}}{N_{sVDJ+/GD}} \right)^2 \frac{\sum_{i \in S} x_{i}(1)}{\sum_{i \in S} x_{i}} + \left( \frac{N_{sVDJ+/VGD}}{N_{sVDJ+/VGD}} \right)^2 \frac{\sum_{i \in S} x_{i}(1)}{\sum_{i \in S} x_{i}} + \left( \frac{N_{sVDJ+/VDJ}}{N_{sVDJ+/VDJ}} \right)^2 \frac{\sum_{i \in S} x_{i}(1)}{\sum_{i \in S} x_{i}}
\]

in which \( \sum_{i \in S} x_{i}(1) = x_{sVDJ+/GD} + x_{sVDJ+/VGD} + x_{sVDJ+/VDJ} \), (1).

To perform minimization using standard fitting algorithms [including when applying the widely used Neelder-Mead algorithm (28) for parameter validations via the routine “optim” in \( R \)] always first requires assigning arbitrary values to all the parameters and iteratively traversing the parameter space through a succession of small displacements along each parameter axis.

In practice, to explore the predicted continuum of parameters for the distribution \( (N_{sVDJ+/GD}, N_{sVDJ+/VGD}, N_{sVDJ+/VDJ}) = (1, 39, 36) \), we therefore initially ran the least-square algorithm for 500 parameter triples equidistributed on a cubic grid. In applying the validating criterion that we adopted in this study [closest integers for the computed cell numbers \( \left\{ \frac{x_i(1)}{\sum_{i \in S} x_i} \right\}_{i \in S} \) coinciding with experimental figures (1, 39, 36)], the procedure returned >450 solutions (parameter triples) that aligned on a smooth curve (i.e., the red dots depicted in Fig. 2; also see Supplemental Table I).

Results

Formulation of the model: overview

We implemented a stochastic modeling of TCRβ gene recombination based on the Markov process formalism (29). Thus, TCRβ dual-allele rearrangement status in a collection of individual single cells is represented via a time-varying random variable distributed into a finite set of states. Holding to the Markov property, this variable evolves with time from one state to the next at a rate specific to the considered transition, independently of all past statuses. Permitted transitions and associated rates are encoded into the so-called transition matrix (30); the structure of which is determined by: 1) biological constraints on TCRβ gene recombination (D-J occurring first, prior to V-DJ; no direct V-J); and 2) feedback inhibition following a VDJ* assembly (see Ref. 1, Fig. 1, Table I). Transition rates represent adjustable parameters expressed in terms of the mean duration of the given rearrangement step and average time lapse to achieve inhibition (covering the whole period from TCRβ gene expression and protein synthesis to trans-allelic block in recombination), respectively. Collectively, the distribution of cell subpopulations harboring distinct TCRβ gene configurations (β-genomic statuses) evolves in a deterministic way along the linear flow generated by the transition matrix. Provided an initial distribution is known for \( t = 0 \), the fraction of cells included in each subpopulation at any one time \( t > 0 \) can be calculated.
regardless of how the allelic pairs are ordered (e.g., DJ/VDJ+ and

in-frame/out-of-frame readability [i.e., at this later transition, the pos-

b

b

functionally relevant, truncated TCR

i

b

b

in a situation in which there is no evidence that DJ

b

b

is only one potential open reading frame through the various J

may occur at each individual allele). Notice that, at the TCR

The model features a standardized compartmentalization of TCR

partially D-J rearranged (DJ); or 3) completely V-DJ rearranged,

2

3

alleles,

regions. Within the window, successive D–J and V–DJ rearrange-

Discussion

The model features a standardized compartmentalization of TCRβ

gene recombination such that, at the single-cell level, each TCRβ

alleles, regardless of how the allelic pairs are ordered (e.g., DJ/VDJ+ and

VDJ'/DJ are seen as one identical status). In this way, the model

only features general aspects regarding the structural organization and

use of gene segments at the TCRβ locus, without dwelling on

which particular cluster(s) of gene segments is being used for recombin-}

b

b

b

b

FIGURE 1. Transition graph of the Markov process modeling TCRβ

gene recombination and feedback inhibition. The graph also depicts or-
dered rearrangement events at the TCRβ locus (DJβ-to-Jβ joining

occurring first, prior to Vβ gene recombination); arrows represent

authorized transitions in TCRβ gene rearrangement/cell genomic status.

Transition rates (in red) are computed from those at a single allele (e.g., the

first GL→DJ transition occurs at rate 2/τDJ as a DJβ-to-Jβ rearrangement

may occur at each individual allele). Notice that, at the TCRβ locus, there

is only one potential open reading frame through the various Jβ gene

segments and Cβ exons and no stop codon within the Dβ gene segments

(Ref 33; S. Jaeger and P. Ferrier, unpublished observations). Accordingly,

in a situation in which there is no evidence that DJβ joints encode

functionally relevant, truncated TCR β-chains, it is thus the final recom-

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b

b


and subsequent DJ→VDJ+ or DJ→VDJ− transitions occur as independent Poisson processes, with rates equating to, respectively, 1/τVDJ, 1/3τVDJ, and 2/3 τVDJ, in which τDJ and τVDJ represent the average lengths of time of the corresponding transitions, and the V-to-DJ transition alone bears susceptibility to the 1/3 rule of in-frame recombination, as it is thought to be the case at the TCRβ locus (1, 33) (Fig. 1; the figure shows the actual rates for TCRβ dual allele usage). Should a VDJ+ be completed and the resulting pre-TCR produced and made operational, the given cell can then switch to one final state in the model (symbolized sVDJ+/GL, sVDJ+/DJ, sVDJ+/VDJ−, or sVDJ+/VDJ+). The later transitions all arise at rate 1/τi, in which, similar to above, the parameter τi denotes the mean time interval between completion of a VDJ+ and effectiveness of pre-TCR-mediated allelic exclusion. As time runs out, the end of the recombination window eventually puts a stop to any further change in the status of lingering cells (strictly speaking, the latter statement only pertains to the β genomic status; a more elaborate model would be needed to also integrate the outcome—death, commitment to a distinct lineage, etc.—of those cells that did not reach a (sVDJ+/... ) closing stage during their recombination window; see Ref. 34). Given all of the above statements, we relied on standard results from the Markov process theory and on biologically relevant formulations to explicitly compute parameter-dependent estimations of TCRβ+ cell distributions; notably, expressions (2) and (4) proffered analytic predictions on statistical values of developing T cells harboring the distinct β genomic statuses (for details, see Materials and Methods).

### Fitting statistics from β-selected T lymphocytes

To calibrate parameters in the model and, ultimately, assess the statistical relevance of modeling predictions, we relied on available data from experimental analyses of TCRβ+ cell distributions and the application of standard fitting procedures. Briefly, we applied least-squares minimizing methods to fractions ascertained from published numbers of T cell-derived, cloned hybridomas well characterized in terms of TCRβ genotype (for details on the biological approach, see, for example, Ref. 22). The mathematical procedures were implemented using Mathematica (Wolfram Research), function FindMinimum. The parameter triples (τVDJ, τVDJ, τi) obtained in this way were further validated using the maximum likelihood method (see Materials and Methods, Least-squares fits and parameter estimation and Additional validation of parameter triples; all methods also described in Ref. 27).

In practice, studies in the literature yielded quantitative information on cell populations displaying only three out of the four potential TCRβ+ final states outlined above, namely sVDJ+/GL, sVDJ+/DJ, and sVDJ+/VDJ− (22, 23) (Table II). Hence, resolving the three parameters from these fragmentary statistics was possible aside from one degree of freedom (see Materials and Methods, Continuum of fitting parameters). Consequently, for each data set in Table II, the fitting process, if doable, was predicted to yield a continuum of triple solutions situated on a curve in the parameter space. To draft the curves, we ran the minimizing procedures for 500 initial parameter triples, equidistributed along each parameter axis. The validated solutions (full range of numerical values) are displayed in Supplemental Table I and the definitive curves shown in Fig. 2 (for a practical application of the global approach, also see Materials and Methods, Least-squares fitting procedure and parameter estimation: worked example). Thus, fitting the model to quantitative data from hybridomas generated in two studies using WT mice (22, 23) and, in one case, also mice carrying a modified TCRβ locus impinging on some aspects of gene recombination (23, 35) yielded parameter triples that aligned onto a smooth curve in all cases (Fig. 2; each individual triple specifying the model so as to account for the given data set). Even though parameter solutions are depicted for a specific T cell distribution in the particular study (Fig. 2, see legend), fits using different distributions within the limits of those indicated in Table II led to similar solutions with identical properties (data not shown). Notably, the solution curves obtained for WT mice in the two studies investigated in this paper displayed sound consistency (Fig. 2, blue and red dots; slight deviation between the two curves likely reflects statistical variations inherent to relatively small-sized biological samples). Overall, these results thus imply that, provided a proper calibration of the parameters has been performed, a Markov process simulating the basic steps of TCRβ locus assembly in developing T cells reproduces in silico the

<table>
<thead>
<tr>
<th>T Cell Type (Data Source)</th>
<th>NsV DJ/GL</th>
<th>NsV DJ/DJ</th>
<th>NsV DJ/VDJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT (22)</td>
<td>0–200</td>
<td>118–1200</td>
<td>92b</td>
</tr>
<tr>
<td>WT (23)</td>
<td>0–200</td>
<td>38–400</td>
<td>36</td>
</tr>
<tr>
<td>Mutant βLDVLD (23)</td>
<td>14</td>
<td>32</td>
<td>36</td>
</tr>
</tbody>
</table>

The distinct genomic statuses are indicated and were obtained from data in Khor and Sleckman (22) and Senoo et al. (23).

Numerical variability results from formating the experimental data so as to fit the genomic statuses in the model (especially the status of TCRβ alleles carrying an unarranged D1I gene segment is uncertain (i.e., could be GL or DJ depending on the configuration of the cis-linked DJ6-JB2 cluster)).

In this study, the productive versus nonproductive outcome in cells harboring two complete VDJ rearrangements was specified for only 44 out of 92 hybridomas, and all 44 displayed a VDJ+/VDJ− genotype. Accordingly, in the corresponding fitting procedure, we assumed that 92 = NsVDJ+/VDJ = NsVDJ+/VDJ + NsVDJ+/VDJ. We note that, in all cases of fitted parameters, the model did predict at least 44 out of 92 sVDJ+/VDJ cells that harbor 1 VDJ+ only.
statistical data on TCRβ genotype/cell-subset distribution delineated experimentally in independent analyzes of mature T lymphocytes (as discussed in detail below).

Parameter estimates and model predictions
Examination of the solution curves in Fig. 2 afforded important information concerning parameter calibration and related predictions relevant to the specific experiments. First and markedly, in those situations in which the number of VDJ+/VDJ+ cells was ignored, the explicit solutions could in all cases be obtained for arbitrary values of \( \tau_1 \) (in fact, unlimited from \( 0 \rightarrow +\infty \)) with, conversely, \( \tau_{VDJ} \) and \( \tau_{vDJ} \) showing variations of small amplitude. In other words, the parameter \( \tau_1 \) entirely epitomized the anticipated degree of freedom. Clearly, this parameter (and corresponding time delays) cannot be inferred from such data sets neglecting the VDJ+/VDJ+ cells. In this context, however, we noticed that the parameter \( \tau_{VDJ} \) in general, as well as the parameter \( \tau_{DJ} \) for the mutant mice (Fig. 2, green dots), displayed either substantial variations or saturation for, respectively, small (\(< 0.1\)) and larger (\(> 0.2\)) \( \tau_1 \) values (Fig. 2, most evident in A and C, respectively), pointing to more robust estimates of recombination periods using this later interval. An interval of \( 0.1 < \tau_1 < 0.2 \) would imply an approximate 0.3–0.5 d (7–12 h) period to achieve pre-TCR synthesis and feedback inhibition assuming that the recombination window is comprised within the 2.5–3 d needed by T cell precursors to transit through the DN2–DN3 compartments (36, 37). Indeed, within the \( 0.1 < \tau_1 < 0.2 \) range, the curves exhibited limited variations enabling a precise determination of \( \tau_{DJ} \) values [0.036 ± 0.006 and 0.038 ± 0.007 for the WT T cells in Khor and Sleckman (22) and Senoo et al. (23), respectively; 0.205 ± 0.020 for the mutant T cells in Senoo et al. (23)] and fairly consistent estimates of \( \tau_{vDJ} \) values (WT T cells: 0.55 ± 0.05 and 0.4 ± 0.03; mutant T cells: 0.33 ± 0.03); as already mentioned, slight variations between numerical figures from WT data sets may reflect statistical fluctuations linked to limited-size sampling. Accordingly, in WT T cells, \( \tau_{DJ} \) would display smaller values (by one order of magnitude) compared with those of \( \tau_{VDJ} \), implying, in this scenario, a prompt completion of D–J assembly and a time window dominated by the V–DJ interval. These results are reminiscent of a handful of experimental data implying an increased intracity of Vβ-to-DJβ rearrangement as compared with that of Dβ-to-Jβ, requiring at least: 1) the partitioning of distant genomic regions via developmentally regulated chromosomal looping (38); and, perhaps, 2) the breaking of Vβ-privileged connection to nuclear repressive compartments (12, 38); 3) overriding the relatively ineffective Vβ recombination signal sequences; and/or 4) supplanting of inter-Vβ antagonisms for productive coupling with a DJβ unit (39–41). In this regard, the \( \tau_{vDJ} \gg \tau_{DJ} \) picture accommodates the assumption generally endorsed in the literature that the control of the initiation phase of allelic exclusion (allelic asynchronism) at the TCRβ locus impinges on the Vβ-to-DJβ rearrangement step (1, 2, 20, 42). In addition, the longer GL→DJ and, conversely, slightly shorter DJ→VDJ transitions observed in parallel between \( \beta^{DKL} \) mutant and WT T cells represent provocative results as the \( \beta^{DKL} \) mutation: 1) specifically enforces the sole usage of the D2–J2 cluster, the segment assembly of which is suspected to proceed less efficiently than that involving the D1–J1 cluster; and 2) greatly reduces the numbers of S′ Vβ gene segments and the intervening distance from the D2–J2 cluster, possibly increasing their recombination rate (23, 39, 43).

Finally, in using the numerical expressions \( x_i(1) \) (see Materials and Methods, Simulation of TCRβ+ T cell distributions and analytic expressions; expressions (1) and (2) and parameter solutions from Fig. 2 and Supplemental Table I, we made predictions concerning the distribution of all TCRβ+ cell populations and, more specifically, of the fraction of sVDJ+/VDJ+ T cells not supplied by the given data sets. Strikingly, all sVDJ+/DJ, sVDJ+/VDJ−, sVDJ+/GL, and sVDJ+/VDJ+ entities were correctly predicted by the model, with the first two approximating the 60:40 ratio (precisely, for \( 0.1 < \tau_1 < 0.2, \) 56 and 39%, respectively; in agreement with exact measuring; e.g., see Ref. 2) and the latter ones restricted to a minority (\(< 5\%\)), as expected (Fig. 3A; similar distributions were obtained using parameters from WT data in Ref. 23; E. Farcoat and B. Fernandez, unpublished observations). Specifically, the fraction of sVDJ+/VDJ+ cells turned out to be fairly low in all cases, varying monotonically with \( \tau_1 \) (Fig. 3B). Depending on the data set, solutions ranged from 3.6–6.4% of total TCRβ+ cells. These estimates are in accordance with assessments on TCRβ allelically included cells based on genomic sequencing of a number of VβDJβ joints (3–10%; even though, at the phenotypic level, dual TCRβ+ cells represent a smaller percentage due to postrearrangement events, such as competition between Vβ gene promoters, TCR β- and α-chain pairing, etc.) (24, 25, 44, 45). In addition, in the case of the \( \beta^{DKL} \) mice, the predicted proportions of sVDJ+/VDJ+ cells are similar to those in WT animals and thus corroborate experimental observations arguing that the specific mutation does not impact on TCRβ allelic exclusion (23). Overall, the ability of the model to integrate predominant and rare subpopulations and to satisfactorily predict their statistical behaviors is testimony to the power of the approach.

Fitting statistics from preselected DN thymocytes
Besides integrating statistical data on TCRβ allelic status in differentiated T cells, potentially accessing such quantitative figures in early developing DN thymocytes prior to β-selection would be an additional advantage. So far, very few experimental studies have addressed this issue. Notably, Aifantis et al. (24, 25)

**FIGURE 3.** Predictions on cell distributions. Productively rearranged cells related to the parameter \( \tau_1 \) were computed via the model in using the solution curves shown in Fig. 2. A. Estimations of the four cell populations of the form (sVDJ+/..) using parameter triples derived from data in Khor and Sleckman (22). B. Estimations of allelically included cells using parameter triples derived from data sets summarized in Table II. The color code is identical to the one used in Fig. 2.
used single-cell PCR in comparative analyses of TCRβ rearrangements in purified (FACS-sorted) intracytoplasmic TCRβ+, preselected DN cells from WT mice versus mutant animals in which pre-TCR setting and/or signaling was perturbed. We thus used their results to further challenge our model by fitting the reported numbers (Table III) with those computed in formal states (VDJ+/DJ, VDJ+/VDJ−, and VDJ+/DJ cells, but not VDJ+/GL cells), a continuum of triplet solutions was similarly predicted. In fact, the solution curve for the WT cells now adopted a distinctive shape, with nonmonomeric variations of τVDJ and τt (the latter limited to a remarkably small amplitude < 5 × 10−3) and the degree of freedom effecting τt (Fig. 4). Although the curve provided evidence for arbitrary solutions of τt (note that the mean length of these rearrangements matters little when VDJ+/GL cells are ignored), it also proffered optimal values for τD-J and τVDJ that were in good agreement with those reported above using data sets from hybridomas (e.g., the triplet (τD-J, τVDJ, τt) = (0.04, 0.52, 0.046) fitted the curve). As for τt, the latter value (0.046) denotes a mean delay for feedback inhibition shorter than that deduced earlier (0.1–0.2). However, the discrepancy may be moderated by the fact that this parameter appeared fairly sensitive to the fraction of the VDJ+/VDJ+ population. For instance, the figures 0.04, 0.67, and 0.13 were found, respectively, for τD-J, τVDJ, and τt in fitting a slightly different distribution (NVDJ+/DJ, NVDJ+/VDJ−, NVDJ+/DJ→VDJ−, NVDJ+/DJ→VDJ+) = (34, 19, 2), workable within the data range (Table III).

Knockout deletion of pTα (the invariant protein that pairs with the TCR β-chain to form the pre-TCR) or SLP-76 (a scaffolding factor that plays a critical role in pre-TCR signaling) were each the TCRα−/−T cells (Table III, second and third rows), we simply set up τt = +∞, which comes to eliminate the ultimate states (sVDJ+/DJ . . .) in the model and used a least-squares algorithm with

\[ \Sigma \left( \frac{N_i - \frac{X_i}{X}}{X} \right)^2 \]

[mathematical details in Materials and Methods, Simulation of TCRβ(β)+ T cell distributions and analytic expressions; expressions (3) and (4)].

Because experimental distributions were again related to only three β genotypes (in this study including VDJ+/VDJ+, VDJ+/VDJ−, and VDJ+/DJ cells, but not VDJ+/GL cells), a continuum of triplet solutions was similarly predicted. In fact, the solution curve for the WT cells now adopted a distinctive shape, with nonmonomeric variations of τVDJ and τt (the latter limited to a remarkably small amplitude < 5 × 10−3) and the degree of freedom effecting τt (Fig. 4). Although the curve provided evidence for arbitrary solutions of τt (note that the mean length of these rearrangements matters little when VDJ+/GL cells are ignored), it also proffered optimal values for τD-J and τVDJ that were in good agreement with those reported above using data sets from hybridomas (e.g., the triplet (τD-J, τVDJ, τt) = (0.04, 0.52, 0.046) fitted the curve). As for τt, the latter value (0.046) denotes a mean delay for feedback inhibition shorter than that deduced earlier (0.1–0.2). However, the discrepancy may be moderated by the fact that this parameter appeared fairly sensitive to the fraction of the VDJ+/VDJ+ population. For instance, the figures 0.04, 0.67, and 0.13 were found, respectively, for τD-J, τVDJ, and τt in fitting a slightly different distribution (NVDJ+/DJ, NVDJ+/VDJ−, NVDJ+/DJ→VDJ−, NVDJ+/DJ→VDJ+) = (34, 19, 2), workable within the data range (Table III) (24).

Knockout deletion of pTα (the invariant protein that pairs with the TCR β-chain to form the pre-TCR) or SLP-76 (a scaffolding factor that plays a critical role in pre-TCR signaling) were each reported figures, with best-predicted scores (NVDJ+/DJ, NVDJ+/VDJ−, NVDJ+/DJ→VDJ−) = (17, 15, 4) [instead of (16, 13, 7); pTα−/−T cells] and (NVDJ+/DJ, NVDJ+/VDJ−, NVDJ+/DJ→VDJ−) = (19, 21, 5) [instead of (18, 19, 8); SLP-76−/− T cells] (minimized distributions >0.25, of ~29 in both cases). We can think of two reasons to account for this unique setback. Firstly, we cannot formally eliminate the possibility that the model in its present form does not fit data sets from pre-TCR−deficient T cells. For example, current predictions may underestimate the length of survival of pre-TCR−deficient cells. That such cells may behave oddly is suggested by their unusual high level of expression (compared with DN thymocytes in WT mice) of a surface marker (CD25) associated with DN differentiation (24). However, we favor the alternative explanation that the model indeed matches the biological reality better than the given experimental data. As mentioned in the original articles, the biological numbers/cell distributions obtained experimentally from the pTα−/− and SLP-76−/− situations may be questionable, notably being prone to overestimate of NVDJ+/VDJ−, possibly linked to a sorting procedure-induced bias that would result in preferential sampling of cells harboring a VDJ+ on the first rearranging allele (and so inevitably increasing the proportion of VDJ+/VDJ+ cells; see discussion in Refs. 24, 25). In support of the latter scenario, we found that when the VDJ+/VDJ+ fractions were ignored, each distribution of the remaining cells (NVDJ+/DJ, NVDJ+/VDJ−) was then precisely predicted using a continuum of parameter pairs (τD-J, τVDJ) (Supplemental Fig. 1).

Moreover, for each continuum, the fitted parameters (τD-J, τVDJ) could be chosen arbitrarily close to some of those determined from WT data sets in Fig. 4. Finally, expected scores for VDJ+/VDJ+ cells were then lower (of three and five cells for the pTα−/− and SLP-76−/− situations, respectively) but still of a larger fraction compared with that generally found in pre-TCR−proficient animals (i.e., again validating a disruption of allelic exclusion). These findings corroborate the biological belief that pTα or SLP-76 depletion prevents feedback signaling without affecting prior rearrangement features including the 1/3 in-frame versus 2/3 out-of-frame Vβ recombination outcomes. Indeed, they better fit the expectation of 20% (at most) of TCRβ allelically included cells among those having completed rearrangements on both alleles in this situation (versus 35% as per the data in Refs. 24, 25) and could then emphasize the projecting power of this modeling method.

Alternative model of feedback inhibition targeting the V-J-D recombination step

In the above-described first version of our model (henceforth called M1), we simulated the inhibitory feedback to impact both the D-J and V-DJ recombination steps (Fig. 1). In this way, our estimates implied that the initial step may be substantially shorter than the average time delay for feedback inhibition (τD-J ~ 0.04 << τt ~ 0.1–0.2). In practice, this would suggest that most D-J joining has already been completed long before inhibition can take place, according to current understanding (42). However, to our knowledge, no study has yet been published in which the time course of D9–Jβ recombinase in the very early hours of TCRβ gene assembly has been investigated. Moreover, in a general scheme whereby immature thymocytes employ distinct pre-TCR downstream pathways to signal for allelic exclusion versus

<table>
<thead>
<tr>
<th>T Cell Type (Data Source)</th>
<th>NVDJ+/DJ</th>
<th>NVDJ+/VDJ−</th>
<th>NVDJ+/DJ→VDJ−</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT (24)</td>
<td>34</td>
<td>18–20a</td>
<td>1–3a</td>
</tr>
<tr>
<td>Mutant pTα−/− (24)</td>
<td>16</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Mutant SLP-76−/− (25)</td>
<td>18</td>
<td>18–19a</td>
<td>8–9a</td>
</tr>
</tbody>
</table>

The distinct genomic statuses are indicated and were obtained from data in Aifantis et al. (24, 25).

Productive versus nonproductive outcomes were not provided for, respectively, two cellsa and one cella+ harboring two complete VDJ rearrangements (for details, see Table 4 and Table I in Refs. 24 and 25, respectively).
cell expansion and differentiation (46, 47), feedback inhibition may well depend on a series of non-mutually exclusive processes impinging on the activity of the recombination apparatus and on the epigenetic status of the Vβ versus DJβ chromosomal templates (38, 48–50). Additional molecular mechanism(s), most of which remain incompletely elucidated, likely also act beyond chromosomal accessibility to impair recombination between sites eluding epigenetic silencing at the TCRβ locus (51–53). In such an intricate context, another relatively unexplored question remains of whether feedback signaling primarily (or initially) targets the V-to-DJ rearrangement step. Our modeling framework might help in appraising this issue by revising the behavior of sVDJ+/GL cells in the transition matrix; in practical terms, in permitting a sVDJ+/GL→sVDJ+/DJ switch at rate 1/τDJ, similar to the other final transitions during the recombination window (model M2; notice that this does not amend the clause that all types of recombination, including D-J, are arrested as the window ends).

Using the same numerical procedure as used above (Materials and Methods, Simulation of TCRβ+ T cell distributions and analytical expressions and Least-squares fits and parameter estimation, simulations of β-selected T lymphocytes), we found that such an M2 alternative also fitted data sets from Khor and Sleckman (22) and Senoo et al. (23), to the same level of accuracy as M1 did (see the solution curves in Figs. 5 and Supplemental Table III; importantly, because the M2 formulation does not affect the VDJ+/DJ, VDJ+/VDJ+, and VDJ+/VDJ− subpopulations, this new version of the model then obviously also fits the DN thymic cell data from Refs. 24, 25). Hence, M2 appears to represent a plausible hypothesis. Notably, the curves in Fig. 5 exhibited all the main characteristics previously found for model M1 (with domain ranges of the parameter τVDJ almost unaffected; e.g., compare A in Figs. 2 and 5), implying that similar conclusions could be drawn in the M2 situation (i.e., indetermination of τC large versus small variations of τVDJ depending on τJ intensities). This similarly applied to predictions on the final distribution of (sVDJ+/… ) cells that were indistinguishable in this situation from those already depicted in Fig. 3. However, one main difference related to the mean length τDJ, which remained at a (roughly) constant value throughout the curve’s evolution and corresponded to a significantly higher grade than previously observed (τDJ = 0.208 ± 0.02 and τDJ = 0.2201, WT T cells in Refs. 22, 23, respectively; τDJ = 0.5086, mutant T cells in Ref. 23). We conclude that the experimental data sets are faithfully reproduced by either the M1 or M2 version of the model with, in the latter case, an extended duration of D-J rearrangement. Importantly, both versions can accommodate the minute fraction of VDJ+/GL cells displayed in Fig. 3A and basic experimental findings that expression of a rearranged TCRβ transgene, although inhibiting Vβ–DJβ recombination, generally does not block DJβ–β rearrangement (1). Hence, within the framework of our model, the hypothesis that allelic exclusion might differentially impinge on the course of Vβ–DJβ versus DJβ–β rearrangements represents an authentic option, even if this has received little attention so far.

Discussion

We used Markov processes to model TCRβ gene recombination, considering cis-rearrangement and trans-inhibition intervals at opposite alleles in individual cells from a dynamical point of view, with all relevant events obeying a probabilistic rule. From fractional data, we proceeded by systematically determining parameter values, firstly by analytical calculations and then by numerical optimization. Remarkably, we thereby reproduced experimental observations from different strains of mice and could proffer exact predictions on the distribution of cells harboring the various TCRβ genomic statuses arising during T cell development. While keeping allele cross talk at a minimum, our model thus comprehensively satisfies the allelic exclusion landmark, suggesting that its basic precepts truly capture the nature of the underlying controls at the TCRβ locus. This work, together with previous studies in which Markov modeling (using Monte Carlo simulations and empirical predictions or, as in the current study, explicit calculations and analytical predictions) was shown to fit the distribution of subsets
of TCR/BCR-rearranged T and B lymphocytes (21, 34, 54, 55), strengthens evidence for the effectiveness of stochastic-based formalisms at simulating developmental milestones in lymphoid cells.

One hallmark of the modeling design used here is to award total independence to the onset and course of GL→DJ and DJ→VDJ stepwise transitions at the two opposite TCRβ alleles, prior to the completion of feedback inhibition. Although an instant shift between two consecutive states is, in principle, conceivable with the selected formalism, such an outcome rarely occurs. Instead, the transition typically occurs on a biological timeline in a stochastic manner in individual cells, and thus entails, on average, positive time slots τDJ or τVDJ. As the independence premise also entails positive time in events committed to autonomous Poisson processes (30), allelic asynchrony generally ensues. In most cases, the time lag separating VDJ assembly at opposite sites will therefore be significantly larger than that of feedback inhibition, a precondition to allelic exclusion. Occasionally, however, the two values may be comparable, thereby permitting the emergence of a few VDJ+/VDJ− cells, an integral part of this developmental process.

A possible criticism to the present analysis is that the model is only as good as the data it uses for testing and refinement, and, arguably, a weakness of the approach could be its reliance on data sets that, due to the technical difficulties, do not have high enough numbers to make definitive statements. However, we note that, in fitting statistics from the small data sets recorded in comparative analyses of TCRβ rearrangements in preselected DN thymocytes from WT mice versus pTα−/− or SPL-76−/− animals (55, 36, and 45 records, respectively; Table III) (24, 25), we not only observed, for the WT situation, parameter values that were in general agreement with those assessed using larger data sets from hybridomas (e.g., 210 records; Table II) (22), but also, for the mutant situations, numerical figures that, compared with the biological data, better fitted the outcomes predicted for a disruption of allelic exclusion. Consistency and predictive ability are two criteria that back up the strength of our modeling approach.

The property of equal autonomy allocated to each individual allele differs from proposals intended to explain the control of allelic exclusion (especially of its initiation phase) based on deterministic models. These proposals argue that opposite alleles are nonequivalent substrates for the VDJ recombinase and that this regulated choice directs the order, on a strictly sequential mode, of subsequent allelic rearrangements in that particular cell (10, 38, 56) (even though reassessment of the data initially presented in Ref. 56 led to a less definite view in this case; see addendum in Ref. 57). However, our calculations showing that experimental cell distributions can be predicted by combining allele autonomy and stochasticity in rearrangement and feedback inhibition events better accommodate scenarios whereby a global decrease in recombination efficiency reduces the likelihood of the two alleles initiating rearrangement simultaneously (11, 12). Overall, they reinforce the plausibility that stochastic events cause asynchronous rearrangements, at least at the TCRβ locus. Yet, the model preserves deterministic facets (ordered rearrangements, feedback control) displaying an oriented and constrained evolution as stated by the triangular form of the transition matrix.

The molecular mechanism(s) that might be involved in adapting the dynamics of TCRβ gene assembly and instigating allelic exclusion remains largely unknown. Based on an experimental system consisting of the introduction of the GFP marker into the Igk locus, it has been argued that limiting transcription factors randomly and infrequently activates this locus (<5% of pre-B cells apparently expressed the marker, a value claimed to provide a first statistical insight into this framework; see Ref. 11). Such an effect has also been implied as potentially impacting on TCRβ allelic expression and rearrangement, as both processes may be perturbed by gene dosage and/or deficiency of transcription factors, such as the helix-loop-helix protein E47 (58). However, recent investigations questioning the validity of the GFP system in supplying evidence for probabilistic gene expression at Igk alleles (59, 60) cast doubt on these numerical data. In fact, as with the Vβ locus in pre-T cells (18), GL Igk transcription seems to occur biallelically in single pre-B cells (59–61). Separately, using immunofluorescence in situ hybridization, TCRβ alleles were found to associate at a high frequency and presumably stochastically with repressive nuclear compartments (i.e., the pericentric heterochromatin and nuclear lamina), such that only 5% of nuclei showed two alleles free of both compartments against ~60% still showing potentially repressive dual interactions (12). Although evidence has been provided suggesting that Vβ-DJβ rearrangement occurs more effectively on nonassociated alleles, the analysis did not permit the determination of the extent to which recombination may be inhibited by the interaction, how frequently associated alleles separate from the repressive compartments, or whether dually nonassociated alleles undergo concomitant rearrangement, all figures that would help to validate a probabilistic factor in the initiation of TCRβ allelic exclusion. On the subject, and contrary to common notion, we emphasize in this study that a stochastic-based view of allelic exclusion does not necessarily hinge on poorly efficient, monovalent and sequentially sequential recombination, as long as the inhibiting control becomes effective during the time lag otherwise separating VDJ assembly at opposite sites.

In their modeling of Vγ4Jγ1/Vγ1Jγ4 recombination, Sepulveda et al. (21) opted for a Markov chain whereby recombination rates, feedback inhibition, and differential chromosomal accessibility all contribute to isotypic and allelic exclusion at these loci. Regulated modulations of the chromatin template are known to commonly orchestrate gene expression and recombination in developing thymocytes (62). In this context, it is initially surprising that TCRβ gene recombination can be modeled using only the first two criteria, with maximum fitting to the experimental data at hand. One obvious reason for this discrepancy may relate to the respective numbers of genotypic/isotypic states under consideration (12 in the TCRγ study—see Table IV in Ref. 21—versus at most three in the present study), as an additional parameter, delineating, in the former study, the basic epigenetic changes that might sustain a switch from a densely packed to decondensed chromatin at TCRγ1→γ4 alleles, compulsory to Vγ4Jγ4 recombination, clearly aids the fitting procedures. In any case, this practical consideration does not preclude the potential contribution of an epigenetic-related processes in instigating, still on a stochastic basis, allele asynchrony in TCRβ gene assembly. For example, the extended value of τVDJ may relate to an epigenetic-based recombination hindrance that would bestow on the two alleles differential capabilities for Vβ-to-DJβ assembly. In this context, it is notable that the unrearranged Vβ domains/alleles were found to be overmethylated and to frequently associate with repressive compartments in DN nuclei, even though the Vβ region is biallelically expressed at this stage (12, 18, 63). Of note, as long as the key issue rests on a mechanism capable of generating a time lag in the assembly of homologous alleles, a two-step recombination process should mechanically help to this outcome, not only in mathematical simulation but also in reality at the given locus. Then, one intriguing possibility would be that the mechanisms contributing to allelic asynchrony are not necessarily identical at distinct loci, with those assembled from V, D, and J gene segments (such as TCRβ) relying less on epigenetic cues compared with those (such as TCRγ) utilizing V and J gene segments only.
Parameter values deduced from data obtained from WT T cells in separate studies and further validated using both mature and developing T cell subpopulations, implied that Vβ-ΔDJβ recombination occurs on average over a longer time frame compared with ΔB-Jβ rearrangement (0.5 versus 0.04 time window units (i.e., 30–36 h versus 2.4–3 h according to an estimated 2.5–3-d period of DN2–DN3 cell transit) or 0.5 versus 0.2 time window units in the M2 version (i.e., 30–36 h versus 12–14 h)). As already mentioned, extended τVDJ intervals may be justified by the existence of several structural and biological features particular to the Vβ gene segments/genomic regions, and relatively short τRI timing (as is especially true for the M1 option) is usually accepted. Concerning this shorter interval, the fact that our formalism allows any given cell to proceed with ΔB-Jβ assembly instantly (albeit rarely) following activation of the recombination window suggests it merely spans the period required for these rearrangement events. Even so, when related to the estimated length of the window (a few days), these particular values may seem exceedingly long (a few hours) with regards to the recombination reaction itself. This being the case, the length of the window might have been overestimated. Alternatively, the values may cover additional events including the local epigenetic changes that in effect trigger this first rearrangement step (64, 65); prompt ΔB-Jβ transitions would have no biological reality (with little impact on our data because this option only bears on tiny cell percentages). In the future, thorough examination of ΔB-Jβ recombination in DN cells, under single-cell assay conditions, may be necessary to close this debate.

We stress that the present model is far from complete in so much as the sources of cell-to-cell variation are concerned. Missing aspects relate to the structural organization of the TCRβ locus, including the existence of several V’ β genes, the unique 3’ Vβ14 gene, and the duplication (in human and mouse) into two ΔB-Jβ clusters (66). In this context, Khor and Sleckman (22) have, for example, probed the differential usage of the Jβ1 versus Jβ2 cluster in normal T cells. Notably, these authors emphasized that: 1) utilization of either group of Jβ segments for recombination is equivalent; and 2) all possible VDJβ rearrangements are not completed on one allele before VDJ rearrangement is initiated on the alternate allele. Complementary analysis of their data also shows that the distribution of cells harboring VDJβ rearrangements on both alleles (NvDJβ/VDJβ, NβVβ/VDJβ, NβVDJβ/VDJβ) = (16, 45, 31) (see Table I in Ref. 22) remarkably fits (p = 0.998) a binomial distribution (Np2, 2Np(1−p), N(1−p)2) [total cell number n = 92; fraction p of cells with a VDJβ(+/−) rearrangement involving Jβ1 = 0.42]. This fit in itself confirms the autonomous selection of Jβ segments on separate alleles during recombination (i.e., allele independence). Furthermore, the smaller fraction of VDJβ1-rearranged alleles in cells equipped with two VDJβ rearrangements as opposed to those with just one and bias toward Jβ2 compared with Jβ1 (58 versus 42%) suggests that a VDJ joint using ΔJβ1 has a positive (nonnegligible) probability of being removed by a secondary Vβ–ΔDJβ cis-rearrangement involving a ΔDJ2 unit. This may potentially add to the time lag separating gene assembly at the two alleles, thus contributing even better to allelic exclusion. Along this line, it is also notable that the rate of Vβ to ΔDJβ recombination of discrete Vβ gene segments, including Vβ14, may be reduced due to competing effects from Vβ cluster sequences (39). In the long term, elaborate versions of TCRβ gene rearrangement modeling should benefit from integrating all such recombinational properties.

Acknowledgments

The model was first assembled while the authors attended the workshop “Dynamics of Regulatory Networks” at CIC Cuernavaca, Mexico (July to August 2007). We thank Dr. J. Carneiro (IIGC, Lisbon, Portugal) and Dr. R. Lima (CPT, Marseille, France) for helpful discussions and advice during the early stages of this project.

Disclosures

The authors have no financial conflicts of interest.

References


