A dynamical model of TCR β gene recombination: Coupling the initiation of D β -J β rearrangement to TCR β allelic exclusion

Sébastien Jaeger^{1†*}, Ricardo Lima^{2†}, Arnaud Meyroneinc^{3†}, Marie Bonnet¹, Edgardo Ugalde⁴, Pierre Ferrier^{1*}

*For correspondence:

ferrier@ciml.univ-mrs.fr (PF)

[†]These authors contributed equally to this work

¹Centre d'Immunologie de Marseille-Luminy, Aix Marseille Université, CNRS, INSERM, CIML, Marseille, France; ²Centre de Physique Théorique, Aix Marseille Université, Université de Toulon, CNRS, CPT, Marseille, France; ³Departamento de Matemáticas, Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela; ⁴Instituto de Física, Universidad Autónoma de San Luis Potosí, San Luis Potosí, México

---

Abstract One paradigm of random monoallelic gene expression is that of T-cell receptor (TCR) β allelic exclusion in T lymphocytes. However, the dynamics that sustain asymmetric choice in TCR β dual allele usage and the production of TCR β monoallelic expressing T-cells remain poorly understood. Here, we develop a computational model to explore a scheme of TCR β allelic exclusion based on the stochastic initiation of DNA rearrangement [V(D)J recombination] at homologous alleles in T-cell progenitors, and thus account for the genotypic profiles typically associated with allelic exclusion in differentiated T-cells. Disturbances in these dynamics at the level of an individual allele have limited consequences on these profiles, a robust feature of the system that is underscored by our simulations. Our study predicts a biological system in which locus-specific, prime epigenetic allelic activation effects set the stage to both optimize the production of TCR β allelically excluded T-cells and curtail the emergence of their allelically included counterparts.

Introduction

In developing T and B lymphocytes, T-cell receptor (TCR) and immunoglobulin (Ig) genes are assembled from separate variable (V), diversity (D), and joining (J) gene segments in a process called V(D)J recombination (*Schatz and Swanson, 2011; Alt et al., 2013*). This biochemical reaction relies on the lymphoid-specific RAG1 and RAG2 proteins. The RAG1/2 machinery cleaves DNA at asymmetric recombination signal sequences (12- and 23-RSS) that flank each pair of rearrangeable

gene segments. The resultant DNA double-stranded breaks (DSBs) are repaired by components of the non-homologous end joining (NHEJ) pathway to yield coding- and signal-joints (CJs and SJs). Coding joints often display limited deletion and/or addition of nucleotides. Randomization of CJ sequences, while contributing to immune cell diversity, ultimately only yields, on average, one third of productively (in-frame) assembled TCR/Ig genes.

V(D)J recombination obeys genetic programs that integrate signaling cues from generated receptors to sustain cell development in distinct T- and B-cell lineages. For example, during $\alpha\beta$ T-cell development in the mouse thymus, TCR β gene recombination occurs in CD4/CD8 double negative (DN)2 and DN3 evolutionary thymocytes, with D β -to-J β rearrangement preceding V β -to-DJ β joining. Expression of a productively rearranged V β DJ β CJ (hereafter VDJ+) and formation of a primary receptor, namely the pre-TCR, triggers further differentiation into DN4 cells and subsequently CD4/CD8 double positive (DP) cells. This developmental shift is known as β -selection. V(D)J recombination is arrested during this period of cell differentiation. It resumes in DP cells by selectively targeting the TCR α locus to achieve V α -to-J α joining, followed by further selection events involving the completed $\alpha\beta$ TCR. Work done by many groups over the past three decades has revealed multiple molecular levels involved in regulating these dynamics, including RAG1/2 expression, chromatin remodeling, nuclear positioning and 3D-folding of TCR/Ig loci, and occasionally, restricted 12/23-RSS pairing usage (*Schatz and Swanson, 2011; Alt et al., 2013*).

Individual T/B lymphocytes operate under clonal selection, and usually display TCR/Ig receptor components from only one of two chromosomal alleles, a property termed allelic exclusion (AE). For TCR β (as well as *Igh* and *Iglk*/ λ) gene products, AE primarily results from contrasting rearrangements at corresponding alleles (*Khor and Sleckman, 2002*). Thus, the most common TCR β genotypes in $\alpha\beta$ T-cells is a VDJ+ on one allele, with either an intermediate DJ β (DJ) CJ or a non-productive V β DJ β (VDJ-) CJ on the other allele, in a $\sim$ [50%-60%]/[50%-40%] ratio (so-called 60/40 ratio), respectively. The prevailing theory for the interpretation of these profiles relies upon a regulated model wherein following biallelic DJ β joining and the assembly of a primary VDJ+, AE prevents further V β rearrangement *via* inhibitory signaling from the newly formed pre-TCR (*Khor and Sleckman, 2002; Brady et al., 2010a*). This scenario is supported by a wealth of experimental data, including the finding that engineered deficiencies in pre-TCR formation or pre-TCR-mediated signaling in developing T-cells result in the disruption of TCR β AE (*von Boehmer et al., 1998*).

AE completion is thought to consist of two functional phases: an initiation phase of monoallelic activation to prevent V gene rearrangements from occurring simultaneously on both alleles, and a maintenance phase of feedback inhibition to suppress continued rearrangement (*Gorman and Alt, 1998*). Although the latter phase is sufficiently well defined in developing T- and B-cells to rally community consensus, cracking the code of monoallelic activation still represents a topical and controversial subject. Indeed, common interpretations consider no less than three types of distinct scenarios, including: (i) clonal allelic predetermination on a fifty-per-cent chance basis, associated with differential replication timing and suppressive pericentromeric recruitment of the late-replicating allele (*Mostoslavsky et al., 2001; Levin-Klein and Bergman, 2014*); (ii) interchromosomal transmodulation, possibly involving dual allele pairing and ensuing regulation of monoallelic DSB cleavage (*Hewitt et al., 2009; Chaumeil and Skok, 2013*); or (iii) equal biallelic inefficiency in

recombination *via* stochastic transcription or association of both alleles with the nuclear lamina (Liang *et al.*, 2004; Mostoslavsky *et al.*, 2004; Schlimgen *et al.*, 2008; Murre, 2008). In addition, none of these scenarios explicitly meets a number of seemingly opposing specifications that - alongside the completion of $\sim 60/40$ [VDJ+/DJ]/[VDJ-/VDJ+] AE profiles - are distinctive to this unique form of monoallelic gene expression. For instance, it remains unclear how the mechanisms suggested by the proposed scenarios can in their simplest forms proceed differently in the case of V- and D/J-containing *cis*-domains at TCR β /Igh tripartite V-D-J loci (allelically excluded *versus* allelically included, respectively). Likewise, in the context of scenario (i) or (iii) particularly, it is uncertain how the system manages to proceed efficiently with biallelic usage as suggested by the frequent appearance of VDJ-/VDJ+ cells [*i.e.* following the production of a primary VDJ- joint, how could the suppressed, pericentromeric recruited allele be steadily reactivated - (scenario (i)); or, how could the two alleles be used in a statistically significant manner assuming a necessarily low probability p of individual allele usage to efficiently prevent allelic inclusion - (scenario (iii))?]. Finally, none of these scenarios takes into account the real fact that aside from the two prominent VDJ+/DJ and VDJ-/VDJ+ genotypes, less common TCR β profiles, comprising either two distinct VDJ+ or a VDJ+ and a non rearranged germline (GL) allele, are also observed in low but significant fractions of $\alpha\beta$ T-cells (Khor and Sleckman, 2002; Brady *et al.*, 2010a). In short, despite years of intense investigations, the initiation phase of AE remains a puzzling issue [also see Gorman and Alt (1998); Mostoslavsky *et al.* (2004); Vettermann and Schlissel (2010); Outters *et al.* (2015)].

One strategy for the difficult task of investigating lymphoid cell development is to use modeling and simulation [e.g. Claverie and Langman (1984); Shlomchik *et al.* (1998); Piper *et al.* (1999); Sepulveda *et al.* (2005); Volpe and Kepler (2008); Vibert and Thomas-Vaslin (2017); Krueger *et al.* (2017)]. In this way, we have previously used Markov chain-based modeling to address the aforementioned issues, and thus demonstrated that TCR β AE could be enforced by combining three separate concepts - allele independency, probabilistic behaviors and regulated feedback - on a dynamical basis, with all the cell subsets that emerge from the TCR β assembly process now incorporated into a unique distribution (Farcot *et al.*, 2010). Transition parameters implied that V β -to-DJ β joining occurs over a longer time frame in comparison to either the D β -to-J β rearranging step or the feedback closing step (Farcot *et al.*, 2010). Though these findings seem to support the widespread notion that the initiation of TCR β AE is dependent solely and exclusively on V gene recombination, paradoxical results led us to consider the new and unconventional idea that it might instead primarily depend on probabilistic events that lead to the onset of D β -to-J β rearrangement and that are perhaps overlooked by our Markov-chain-based modeling. Specifically, the period of time to achieve D β -to-J β recombination turned out to be overly brief when based on allotted Markovian times (Farcot *et al.*, 2010). However, this initial step is in fact thought to require intricate epigenetic changes during the regional conversion from a repressive to an active chromatin status (Sikes and Oltz, 2012; Jaeger *et al.*, 2013). Based on the compelling evidence that epigenetics contributes towards the modulation of cellular plasticity in development (Pujadas and Feinberg, 2012), we wished to explore the possibility that, in establishing the chromosomal landscape required for the onset of D β -to-J β rearrangement, noisy kinetics significantly contribute towards promoting TCR β AE. Our mathematical simulations presented here provide strong support for this original hypothesis.

Results

Modeling the dynamics of TCR β recombination in single developing T-cells

In DN thymocytes, TCR β gene activation for recombination depends on cooperative outputs from *cis*-regulatory elements, including the transcriptional enhancer E β and D β -flanking promoters pD β s (Sikes and Oltz, 2012). At this stage, attributes of relaxed chromatin (Pol-II-generated GL transcripts, CpG demethylation, active histone marks, positioned arrays of nucleosomes) are evident in the genomic regions that encompass the D β and J β gene segments (Morshead et al., 2003; Kondilis-Mangum et al., 2010; Zacarias-Cabeza et al., 2015). These permit the local recruitment of RAG machinery to form a so-called recombination center (Ji et al., 2010b), and also may optimize lineage-specific release of TCR β alleles from suppressive peripheral subnuclear positioning (Schliming et al., 2008; Chan et al., 2013). In mammalian cells, combinatorial epigenetic modulations are a major component of developmental, allele-specific plasticity and nuclear reorganization, and a source of intrinsic noise in gene expression (Pujadas and Feinberg, 2012; Eckersley-Maslin and Spector, 2014; Therizols et al., 2014; Soshnev et al., 2016). For genes whose activation has to reach a trigger threshold level of epigenetic changes, noise accumulated during these processes may lead to substantial interallelic/cell-to-cell variability in the time required to reach the target level (Pujadas and Feinberg, 2012; Eckersley-Maslin and Spector, 2014). Along these lines, we hypothesize that initial cross-regulatory interactions at TCR β alleles likewise proceed *via* threshold kinetics towards establishing the proper chromosomal landscape required for the onset of D β -to-J β rearrangement, and thus, over time, contribute to probabilistic interallelic behaviors and cell-to-cell variability.

Mathematically, the problem of determining the time required for a stochastic process to reach a certain value falls into the category of first-passage time (FPT) problems (van Kampen, 2007). However, although our current assumption primarily concentrates stochasticity on the epigenetic allele-specific activation events that are mandatory to the onset of TCR β recombination, it is likely that all following steps in the developmental process, including D β -to-J β and V β -to-DJ β recombination and feedback inhibition, also obey probabilistic rules, albeit, we surmise, to lower magnitudes. To ensure the simulation of all of the above conditions, we devised a model system based on discrete time modeling (for comprehensive mathematics, *SI Appendix, section S1/Methods*), which altogether defines a rule-based framework for timely progression of V(D)J recombination at dual TCR β alleles in single developing T-cells. In this system, asynchrony that is introduced during the initiation phase is propagated along the whole developmental process until an asymptotic status is achieved. Importantly however, as a corollary of discretization, delay effects [<1 time unit (*t.u.*) in length and which can therefore be overlooked with respect to overall stochasticity as a first approximation] do in fact take place during every single step in the mathematical model [e.g. Coutinho et al. (2006)]. Thus, the selected formalism offers a double advantage: it keeps free-parameterization at a minimum for formal challenging of a dominance of prime activation events at low risk of overfitting, and yet tolerates a lower degree of freedom during each of the subsequent steps that complete the dynamics. Specifically, our system is made up of pairs of three symmetrical, interacting variables - one triplet set for each TCR β allele. In line with our prime hypothesis, the first pair [(ω_1, ω_2) ; equation (1)] are Boolean input variables, which compute the

stochastic transition of GL TCR β alleles from a recombination-incompetent stage ($\omega_i=0$) to a status that is primed for the onset of TCR β gene recombination ($\omega_i=1$), *i.e.* up to the precise time point at which D β -to-J β recombination is initiated (each ω_1 and ω_2 initialized to 0 at $t=0$). At each and every time step, there is a probability p_i for ω_i to switch from 0 to 1. Thus, p represents the one major stochastic parameter of our model with, typically, $p_1=p_2=p$. The next two pairs $[(x_1, x_2$ and $y_1, y_2)$; equations (2) and (3)] are state variables, which represent the current status in the progression of each individual allele's D β -to-J β and subsequent V β -to-DJ β rearrangements, respectively.

Figure 1. Dynamical modeling of gene rearrangement at TCR β alleles in single developing T-cells. **(A)** Schematic overview of one rearrangement scenario leading to the VDJ+/DJ genotype *via* probabilistic initiation of D-to-J recombination up to feedback inhibition. Thin and thick bars correspond to developmental allelic structures refractory to and compulsory for the onset of TCR β gene recombination, respectively. GL, unrearranged alleles; DJ, VDJ, D-to-J- and V-to-DJ-rearranged alleles, respectively. See text for the definition of the indicated time points t (t_0, t_s, t_1, t_2, t_x , and t_y) and related parameter values δ, τ , and ρ . **(B)** All ten possible cell arrangements in terms of TCR β allele configuration (denoted g_0 to g_9) are shown. Bold circles correspond to asymptotic cell states; examples of intermediate (g_1) and terminal (g_3) cell states from the rearrangement scenario depicted in **(A)** are highlighted.

An additional random variable ($\epsilon_{i=1,2}$), incorporated into state variables y_i , features the in-frame *versus* out-of-frame probabilistic outcomes. In the in-frame situation, the dynamics are interrupted after a fixed period of time ($\rho \geq 1$), which corresponds to the interval that is required to achieve feedback inhibition. Conversely, in the alternative, out-of-frame, VDJ- event, progression of the rearrangement sequence continues on the opposite allele. Based on available data on TCR β gene rearrangement (Aifantis *et al.*, 1997; Senoo *et al.*, 2003; Khor and Sleckman, 2005), we considered that it was not essential to take into account either the scenario of a pre-established time window that would prematurely remove cells that had not yet used the opportunities to achieve a VDJ+ (*e.g.* V β DJ β -/DJ β cells) or, separately, the scenario of a significant fraction of VDJ+/VDJ+ cells arising due to the feedback mechanism not being triggered by the first rearrangement. Both scenarios would ultimately result in predicted data, including increased and decreased 60/40 ratios, re-

spectively, that diverge from experimentally determined values (*SI Appendix, section S2*). Likewise, we did not consider extra variables featuring episodes of cellular division, as these mostly take place prior to D β -to-J β , and after V β -to-DJ β recombination (Ciofani *et al.*, 2004; Masuda *et al.*, 2007).

The dynamic rules of our model system are schematically depicted in Fig. 1A. Time t relates to an initial condition $t=t_0$ when, in any given progenitor T-cell, all four state variables display their base values. Further time points of significance are defined as t_{start} (t_s), δ and ρ , respectively [equations

(5) and (6)]. Briefly, t_{start} marks the time at which a D-to-J recombination reaction begins on the earlier-rearranging allele i (as one ω_i switches from 0 to 1 for the first time), while δ defines the time lag between this initial event on the two homologous alleles, specific to the given cell [geometric distribution of δ s directly derived from stochastic parameters p_i , equations (10)-(13); also see Fig. S1 for graphical representations of the theoretical distribution and partition function of δ]. For each allele, it takes t_x and $t_y > t_x$ to achieve D-to-J and V-to-DJ recombination, respectively. Completion of each event corresponds to the point when x_i (resp. y_i) exceeds the threshold T_x (resp. T_y). Thus, the first-rearranging allele (Fig. 1A, *Allele 1*, initiating recombination time t_1) completes D-to-J recombination and V-to-DJ recombination at time $t_s + t_x$ and $t_s + t_y$, respectively. At the remaining allele (*Allele 2*, initiating time t_2), all timings are delayed by the cell-specific interval δ . The parameter $\tau = t_y - t_x$ measures the whole period of time necessary for a newly generated DJ-rearranged allele to complete V-to-DJ joining. If a VDJ+ is produced, the recombination process will stop on the opposite allele at time $t_s + t_y + \rho$, where ρ defines the time interval required for feedback inhibition. Without loss of generality, we alternatively considered that the process also arrests if both TCR β alleles have achieved VDJ- rearrangements. Overall, the above parameters account for all evolutionary paths that a single cell may possibly follow when sequentially switching from one defined TCR β dual allele configuration to the next (out of a total of 10, denoted g_0 - g_9 ; Fig. 1B).

Based on the relative values of δ , τ , and ρ and outcomes from $\epsilon_{i=1,2}$, six timeline scenarios of single-cell histories are anticipated, commensurate with all orders of possible combinations (Fig. 2). Timeline scenarios cluster into three classes, each containing histories endowed with an equal probability of reaching identical dual allele configurations (ψ_1 , ψ_2 , ψ_3 , bordered by unique colors in Fig. 2). The value of δ relative to ρ and $\rho + \tau$ defines the class to which the individual cell belongs: ψ_1 (in cases where $\delta \leq \rho$; pertaining to the VDJ+/VDJ- and VDJ+/VDJ+ cell fates); or ψ_2 ($\rho < \delta \leq \rho + \tau$; including the VDJ+/VDJ- and VDJ+/DJ cell fates); or ψ_3 ($\delta > \rho + \tau$; comprising the VDJ+/VDJ- and VDJ+/GL cell fates). These scenarios of cell histories differentiate our dynamical model from those based on a biphasic view/instructive mode of AE initiation [which is that only following a non-productive V-to-(D)] prime allele rearrangement would the opposite allele be subsequently recombined, e.g. *Chaumeil and Skok (2013)*; *Levin-Klein and Bergman (2014)*], which solely considers a ' ψ_2 -like' distribution of asymptotic outcomes (or ' ψ_3 -like', in case of a one-step V-J junctional event).

Modeling the dynamics of TCR β genotype distributions in developing T-cells

In the model system we describe, individual cells and developmental cell fates are independent. Therefore, addressing the dynamics of TCR β genotypic changes in separate subsets of developing T-cells is straightforward, and is accomplished by summing up the relative fractions of single-cell profiles shown in figure 2. To begin with, we modeled the genotype evolution within a set of progenitors that concurrently engage in TCR β gene recombination. In this case of synchronously developing cells, the distribution of TCR β genotypes at any time t during the recombination process can be viewed as a vector $G_{t,\delta}(t)$ in the space of linear combinations of ten possible g_i subsets ($i = 0, \dots, 9$) [SI Appendix, S1/equations (8)-(14)]. Cells that leave the simulated developmental system independently of TCR β rearrangement outcomes (e.g. towards the $\gamma\delta$ T-cell lineage) should have little effect on these distributions, in so far as relative proportions are considered, as they are

Figure 2. Timeline scenarios of TCR β allelic rearrangements. Relevant histories of single T-cells or, alternatively (see text), of a synchronously developing T-cell subpopulation are shown, depending on relative δ , ρ and τ values and the one-third ‘in frame’ to two-thirds ‘out of frame’ rule of V-to-DJ rearrangement outcomes. The annotated fractions indicate the expected, time-scaled probabilities (or distributions for a cell population) of the corresponding g_0 to g_9 genomic configurations, with the respective probabilities (or distributions) of each final state at asymptotic regimen shown on the right. Lighter lines correspond to cells ending TCR β gene recombination in the VDJ-/VDJ- (g_6) configuration, doomed to disappear from the developmental process. Cell sets sharing the identical final distributions are clustered into the ψ_1 , ψ_2 , and ψ_3 subclasses (gathered by red, green and blue borders, respectively), depending on explicit rules as indicated at the top of the figure. When using model calibration such as defined in the *Results*, the respective distribution between the three sub-classes is: 9% ψ_1 , 87.9% ψ_2 and 3.1% ψ_3 . Note that cell behaviors, as depicted in panel (A) which displays developing single cells with $\delta = 0$, are not necessarily negligible in terms of cell representation since both the few cells displaying no ‘time gap’ between the two rearranging alleles and those with a $\delta < 1$ t.u. are included.

anticipated to be evenly split among the various subsets at each time step.

With the notable exception of early-stage thymic embryogenesis, the DN cell compartment does not typically accommodate a unique cohort of synchronously differentiating T-cells (Ramond *et al.*, 2014). Rather, it is supplied with TCR β -unrearranged cell precursors, and is drained of advanced cells, including TCR β -rearranged (productively or not) thymocytes. Our mathematical model can be applied to these cellular inflow/outflow situations by considering cell-subsets as before, while also having distinct t_0 . To integrate cell departure, we relied on a ‘stage-span’ window σ , so that individual cells will be withdrawn from the pooled system σ t.u. after completion of TCR β recombination. In this context, σ , which is not an additional parameter in the model (whatever setting value, the whole dynamics will be respected), simply prevents cell accumulation. Importantly, in the constant inflow situation, and from the steady state asymptotic regimen onwards, ratios of the various TCR β genotype outcomes are mathematically identical to those in the synchronous case described above [SI Appendix, S1/equations (15) and (16)].

In summary, considering the synchronous or asynchronous case without bias, the probability distributions $\mathbb{P}_i(t_i = t)$, defined along with equation (1), determine (via the spectrum of $\delta = t_2 - t_1$

values) the cellular profiles in terms of TCR β genotype at any time t . In particular, for given values of $[\tau, \rho]$, final cell profiles (of which there are five - g_3, g_5, g_6, g_7 & g_9 - that are divided across the three generic classes ψ_1, ψ_2 , and ψ_3 , as described above) depend only on these two related sets of scattered values (*SI Appendix, S1/Suppl. Table 1*).

Calibration of the model system reveals approved values of TCR β 60/40 ratio

We first proceeded with model calibration, to ascertain whether our model system may be solved such that predictions of AE measurable observables can be obtained. In the situation where $p_1=p_2=p$, equation (13) of the model may be used to delineate confidence intervals for parameter values $[p, \tau, \rho]$, based on experimental data. Previous studies of the TCR β sequences from wildtype (wt) mouse T-cells or T-cell hybridomas yielded quantitative information on cell populations displaying only three out of the five potential TCR β final states, namely $[g_3, g_5$ and $g_9]$ (*Aifantis et al., 1997*) or $[g_3, g_5$ and $g_7]$ (*Senoo et al., 2003; Khor and Sleckman, 2005*) (*SI Appendix, S1/Suppl. Table 2*). In equation (13), F is a strictly monotonic function of p for any fixed a , whereas each of the two g_9 (VDJ+/VDJ+) and g_7 (VDJ+/GL) fractions (which we named u and v , respectively) is found in only the ψ_1 or ψ_3 subclass, respectively. Consequently, provided that ρ is set first (without loss of generality we set $\rho = 1$ t.u.), p can be calculated from u ; then, τ can be calculated from v . From data in (*Aifantis et al., 1997*), a mean estimate for u was 1.8%, corresponding to $p_{wt}=0.0606$. From data in (*Khor and Sleckman, 2005; Senoo et al., 2003*), estimates for v vary from 1% to 2.6%, thereby setting $64 \geq \tau_{wt} \geq 48$ (when $p_{wt}=0.0606$). In using these boundary values for testing simulations, we verified that the resulting [VDJ+/D] (or GL) to [VDJ+/VDJ-] cell ratio falls within the range of experimentally-observed distributions [e.g. 54.6/45.4 when $p_{wt}=0.0606$ ($u=1.8$); $\tau_{wt}=55$ ($v=1.87$); (*Mostoslavsky et al., 2004; Brady et al., 2010a*)]. Importantly, we further verified that moderate variations in the parameters' values still yield genotype outputs compatible with TCR β AE hallmarks, a pledge of stability of the model system (*SI Appendix, section S3*).

Model studies predict elements of robustness of the TCR β AE process

To further characterize the properties of our model, we simulated the evolution of TCR β genotypes within synchronous cells displaying distinct sets of p values (Fig. 3), including those in the wt case [row A, ($p_1=p_2=p_{wt}=0.0606$)] or those in situations where one or both p were either decreased or increased relative to the p_{wt} baseline [rows B, ($p_1=p_{wt}, p_2=0.01$); C, ($p_1=p_2=0.01$) and D, ($p_1=p_{wt}, p_2=0.4$); E, ($p_1=p_2=0.4$)]. We performed similar simulations of asynchronously developing cells considering the ($p_1=p_2=p_{wt}$) condition and two types of influx supply dynamics, either constant or periodic (Fig. 4A/B). Thus, we verified the predicted synchronous/asynchronous equivalence for final subpopulations in the case of constant influx (e.g. far right panels in Fig. 3A/4A). Likewise, even though stable ratios are not permanently conserved in the periodic influx situation (Fig. 4B; observed oscillations here reflect on and off periods in flow supply), averaging these quantities over a whole oscillation period *in fine* yields distributions similar to those in the case of constant influx (Fig. 4C). In short, insofar as g_3, g_5, g_7 , and g_9 populations are considered (in practice those that are commonly accessible to the experimentalist), synchronous and asynchronous regimens using the same u - and v -based parameterization of the modeling system lead to identical distributions - and fit those found in *bona*

Figure 3. Dynamics of TCR β gene recombination considering synchronously developing T-cells. The evolutions in the distribution of the g_0 - g_9 TCR β allelic genotypes are shown in the cases of cell populations where (A) $p_1=p_2=p_{wt}$; (B) $p_1=p_{wt}$, $p_2=p^-$; (C) $p_1=p_2=p^-$; (D) $p_1=p_{wt}$, $p_2=p^+$; and (E) $p_1=p_2=p^+$, with $p_{wt}=0.0606$, $p^-=0.01$ and $p^+=0.4$. The remaining parameters are identical in these five situations ($\tau=55$; $t_x=10$; $\rho=1$). The two left panels show distinct representations (in cumulative and separate manners, respectively) of identical dynamics, using a common scale of cell numbers, as indicated on the left. The middle right panel is a magnification of the curves corresponding to the separate VDJ+/VDJ+ and VDJ+/GL genotypes only. The bars on the far right panels indicate the percentage distribution of each final subpopulation g_3 (VDJ+/DJ), g_5 (VDJ-/VDJ+), g_7 (VDJ+/GL) and g_9 (VDJ+/VDJ+) at the end of the depicted dynamics (200 $t.u.$). Arrows on the dynamical curves in (A - C) point to the maximum values of the DJ/GL and DJ/DJ genotypes (dark green arrows), and to the initiation of V-to-DJ rearrangement (orange arrows).

vide biological samples of developed T-cells [SI Appendix, S1/Suppl. Table 2; (Mostoslavsky et al., 2004; Brady et al., 2010a)]. Only in unusual situations where the inflow/outflow may adopt a prolonged period of continuously increasing or decreasing magnitude [e.g. linked to aging (Balomenos et al., 1995)] might the end distributions be modified. However, we note that the synchronous regimen provides a more complete picture of evolutionary dynamics, while the asynchronous regimen may cause information loss in summing up transient cells from different developmental stages.

To reveal the dynamical attributes of the model, we next examined the simulated trajectories depicted in Fig. 3. In the wt situation, TCR β final states compatible with experimental margins are all handled relatively shortly after initiation of the dynamic process, without the need for any additional parameter (Fig. 3A; asymptotic regimens beginning at $t = 90$ -120 $t.u.$). In this setting, the DJ/DJ genotype has mostly reached its maximal value (and consequently few alleles remain in the GL configuration) when the first attempt for V-to-DJ rearrangement takes place (arrows); this accounts, with minimal constraint, for the general view that most developing T-cells have achieved

Figure 4. Dynamics of TCR β gene recombination considering asynchronously developing T-cells. The evolutions in the distribution of cell populations of the g_0 - g_9 TCR β genotypes are shown in the situation of constant (A) or periodic (B) influx supply dynamics; both with $\sigma = 50$ t.u. Parameter values are the same as in Fig. 3A. (C) corresponds to the periodic influx situation [as in (B)], following integration of distribution mean values over a sliding window of 30 t.u. (theoretical oscillation period = 30.4 t.u.). The bars on the far right represent the distributions of the indicated cell subsets, as integrated at the end of the simulation period (300 t.u.; once their 'stage-span' window σ has expired).

D-to-J recombination on the two TCR β alleles before the initiation of V β -to-DJ β rearrangement (Mostoslavsky et al., 2004; Brady et al., 2010a). However, the dynamics and genotype distributions change as p decreases relative to the p_{wt} baseline at one or both allele(s). Then, the distributions - when compared to the wt alleles - display slower kinetics and equivalent or even higher fractions of transient DJ/GL versus DJ/DJ genotypes at the time of a first V-to-DJ rearrangement (Fig. 3B-C, left two panels). In both cases, this was followed by increased representation of VDJ+/GL (g_7) cells and, conversely, by decreased fractions of both VDJ+/DJ (g_3) and VDJ+/VDJ- (g_5) cells and near-complete absence of VDJ+/VDJ+ (g_9) cells (right two panels). Overall, this implies that in these two instances, TCR β AE would be enforced while the system's productivity (in terms of kinetic efficiency to yield comparable amounts of productively rearranged cells) would be reduced. Finally, the whole dynamics appear accelerated as p increases on one or both alleles, with, in either condition, the somewhat earlier implementation of asymptotic profiles (Fig. 3D-E). However, in contrast to the above findings, eventual effects on cell distribution pertain only to the symmetrical ($p_1=p_2$) situation, which solely displays increased VDJ+/VDJ+ (g_9) cells, near-complete absence of VDJ+/GL (g_7) cells, and an inversion of the VDJ+/DJ (g_3) to VDJ+/VDJ- (g_5) ratio. This suggests that an additional, unanticipated function of p in this system is to act as a point of regulation to ensure robustness of the AE process.

To thoroughly explore the effects of p value variation on the yield of allelically included cells, we established the plot $u(p_1, p_2)$. Thus, u values mostly display a growing series of 'L-shaped' isocurves (Fig. 5). Appreciable expansion of u [where VDJ+/VDJ+ cells are more than around 4 - 5% of mature T-cells, up to the theoretical 20% max. when AE is nil] requires that p be increased on both alleles (note that, conversely, AE is ultimately enforced when either one or both p is/are decreased, as predicted). In the common situation of symmetric p incrementation, u reads along the diagonal $p_1=p_2$. In asymmetric cases [$p_1 < p_2$ (resp. $p_2 < p_1$)], however, u displays 'off-diagonal' entries that hold within

values relatively close to the corresponding lower p_i on the diagonal, and maximum values (and therefore extent of allelic inclusion) that are restrained [Fig. 5, one example ($p_1=0.3$) is illustrated].

Figure 5. Plot of u as a function of parameters p_1 and p_2 . White curves delineate constant values of u (= % of allelically included VDJ+/VDJ+ cells), from $u=1\%$ to $u=4\%$. Black curves delimitate the values of p_2 that maximize u as a function of p_1 (top left curve) and, conversely, those of p_1 that maximize u as a function of p_2 (bottom right curve). The diagonal is for $p_1=p_2=p$. The vertical red lines indicates the curves' intersection with the $p_1=p_{wt}=0.0606$ value estimated in this study (after optimization - see *SI Appendix, section 3*), as well as with the $p_1=0.3$ value.

Likewise, when e.g. p_1 (resp. p_2) is kept constant ($p_1=p_{wt}$), and p_2 (resp. p_1) augments progressively, u also shows little variability - initially increasing slightly and then decreasing (Fig. 5, left red line) - so that the fraction of VDJ+/VDJ+ cells is maintained within an acceptable range regarding AE profiles. In the *SI Appendix, section S4*, we propose an explanation for these surprising behaviors of robustness, based on the compensation effect of $\delta s'$ values between cells that first initiate rearrangement of either allele 1 or allele 2. We conclude that the simulations we demonstrate here suggest that our model is a versatile system able - for the first time to our knowledge - to satisfy, with minimal constraints, known features of, and long-lasting questions relating to, AE and its initiation phase at the TCR β locus (see introduction). Furthermore, the model appears to

comprise key elements of stability and robustness, helping to optimally manage the dual objective of promoting independent dual allele usage and uneven development of allelically excluded/included cells. This may be essential, considering a system in which mono-allelic expression primarily depends on imprecise DNA recombination events that frequently results in non-functional CJs.

Model studies predict measurable outcomes from DJ and VDJ knock-in studies

Two studies that analyzed strains of knock-in (KI) mice carrying either a rearranged DJ or a rearranged VDJ gene segment on one or both TCR β alleles (*Carpenter et al., 2009; Brady et al., 2010b*) were of particular interest for further testing our model, in which asymptotic equilibria depend on relative timespans, including those required to achieve D-to-J (t_x) and V-to-DJ ($t_y = t_x + \tau$) rearrangements. Both studies established end distributions of the VDJ/DJ and VDJ/VDJ configurations of TCR β alleles using T-cell hybridomas from KI animals. We relied on these data, in particular those derived from heterozygous T-cells in which wt and mutant alleles can compete simultaneously, to challenge the fit of our model after either (t_x) or ($t_x + \tau$) were set to 0 for one allele, binding the other allele to wt values. All other parameters were maintained as defined above. In addition, the initial cell population was set to a DJ/GL or VDJ+/GL configuration (rather than GL/GL, as in the wt situation).

Based on the analysis of 247 hybridomas harboring one DJ β -preassembled complex, Carpenter *et al.* reported a distribution of 58% VDJ/DJ and 42% VDJ/VDJ, providing evidence of a two-step recombination process that is dispensable for enforcement of TCR β AE (Carpenter *et al.*, 2009). Remarkably, simulations using our model system yielded comparable values of 57% VDJ/DJ and 43% VDJ/VDJ (Fig. 6A; *SI Appendix, S1/Suppl. Table 3*), showing that our modeling design of AE is consistent with these biological figures. Moreover, in VDJ+/[DJ or GL]-predicted cells, with a ratio of VDJ joints at the (DJ)- and (GL)-initially configured allele of 74% to 26%, respectively, the simulations reproduce the observation that the majority of the corresponding hybridomas contain V β DJ β rearrangements on the KI allele [69% versus 31%, as reported; Carpenter *et al.* (2009)]. In fact, our dynamical view of allele asynchrony from locus-activating events onwards offers an explanation for this phenotypic bias: the mutant allele, when ranking second in starting the recombination process with a disadvantaged δ relative to its wt homologue, nevertheless has the opportunity to achieve a VDJ+ joint first for each $\delta < t_x$, thus adding to the number of cases where it starts and finishes first. This competitive advantage of the mutant allele is reflected in the reduced time needed for VDJ/[DJ or GL] configurations to emerge in the simulations, and by the biphasic dynamics of the VDJ+/GL genotype (Fig. 6A; to be compared with the corresponding data in Fig. 3A).

An analysis of 171 T-cell hybridomas harboring one functionally preassembled V β DJ β complex showed a distribution of 5.3% V β DJ β -rearranged and 26.9% DJ β -rearranged wt alleles, respectively (Brady *et al.*, 2010b). This distribution, showing a significantly reduced DJ fraction, is surprising, as former studies have found that ectopic expression of a VDJ+ exogenous transgene, though inhibiting the endogenous V-to-DJ rearrangement, had no effect on the course of D-to-J recombination [Mostoslavsky *et al.* (2004); Cedar and Bergman (2008); Brady *et al.* (2010a); Vettermann and Schlissel (2010); and thus it has generally been accepted that TCR β AE impinges exclusively on V gene recombination]. However, these conflicting results can be better understood using a dynamical model system like ours, which encompasses all stages of the recombination process (with the reasonable assumption that ectopic versus site-specific genomic integrated VDJ+ inserts behave differently in terms of timing in developmental expression). Indeed, resulting simulations from our revised model ($t_x=0$ / $\tau=0$ at one allele, which is also considered to be functional in all cells, and not just one third of them, throughout the simulation period) retrieved values of 0.94% (V β DJ β) and 28.43% (DJ β) on the wt alleles (Fig. 6B; *SI Appendix, S1/Suppl. Table 3*). It is noteworthy that the relatively low numbers of experimental samples ($\frac{9}{171}$ and $\frac{46}{171}$ hybridomas, respectively) may not be without statistical impact on the confidence intervals for the biological data, especially with respect to the V β DJ β fraction. This may partly account for the variation between values reported in (Brady *et al.*, 2010b) and our simulation results. Furthermore, the KI mice used in this study harbor TCR β alleles with duplicated D β 1-J β 1 and D β 2-J β 2 clusters on the wt allele and a D β 2-J β 2 cluster downstream of the inserted segment on the mutant allele. As parallel analyses implied that DJ β 2 complexes are used in further V-to-DJ rearrangement on the mutant allele in ~3% of homozygous hybridomas (Brady *et al.*, 2010b), one may indeed expect a slightly increased fraction of VDJ+ wt alleles in this case of D-J cluster duplication, *i.e.* following cell rescue as also predicted from our current modeling views (*SI Appendix, section S5*). Overall, we therefore conclude that our model system is not only able to satisfactorily simulate published data, but also able to provide a straightforward

explanation for the observations that KI of a DJ β -rearranged segment does not impact upon AE at the T-cell population level, yet results in an increased frequency of VDJ+ rearrangements at this site (*Carpenter et al., 2009*). Furthermore, the model system also corroborates data showing that KI of a VDJ+ preassembled segment does impact, to varying degrees, on both V-to-DJ and D-to-J rearrangement outcomes at the opposite wt allele (*Brady et al., 2010b*). Considering the overall picture drawn by our simulations, the above statements also suggest that the stochastic events that open the TCR β alleles for recombination and/or expression are independent of the rearrangement status (GL, as in the wt allele; DJ β or V β DJ β , as in the KI allele). We note that one hallmark common to all these alleles, regardless of their configuration, consists of the presence of the enhancer E β , in association with either the promoters pDJ or V β and flanking architectural, long-range tethering elements [e.g. *Majumder et al. (2015)*]. Recent evidence suggests that critical shifts during lymphoid cell development are accompanied by large-scale changes in epigenetic markers associated with enhancers, promoters and architectural elements of cell-lineage-specific genes, including antigen receptor genes (*Benner et al., 2015*). In line with this and our modeling precepts, we propose that chromatin opening of these various alleles proceeds coincidentally (though not necessarily simultaneously) - mostly involving E β -driven probabilistic outputs and functional interactions between these various *cis*-regulatory elements in the 3D nuclear context - to begin the course of TCR β gene expression, recombination and appended AE profiles, as observed in the corresponding T-cells.

Figure 6. Reproducing data from knock-in (KI) DJ-rearranged and VDJ-rearranged TCR β alleles. Evolution in the distribution of the g_1 - g_9 TCR β allelic genotypes considering developing T-cell populations carrying either one wt allele and one KI DJ β -rearranged allele [TCR $\beta^{wt/DJ}$ (**A**)] or one wt allele and one KI functionally-rearranged V β DJ β allele [TCR $\beta^{wt/V\beta DJ+}$ (**B**)]. Legend is as in Fig. 3, except that the middle right panel in (**B**) shows a magnification of the curves corresponding to the separate VDJ+/VDJ- ($g_5=0.63\%$) and VDJ+/VDJ+ ($g_9=0.31\%$) genotypes (total V β DJ $\beta=0.94\%$). (**A**): $p_1=p_2=0.0606$; $t_x=0$ at one allele; (**B**): $p_1=p_2=0.0606$; $t_x=0$ and $\tau=0$ at one allele. All others parameters as defined in the Results. According to model implementation, the TCR $\beta^{wt/DJ}$ and TCR $\beta^{wt/V\beta DJ+}$ distributions lack the GL/GL (g_0) configuration. The biphasic behavior of the VDJ+/GL genotype in (**A**) reflects the early appearance of VDJ+/GL cells because of the dynamical advance of the KI DJ allele (V-DJ rearrangement on this allele), followed by their progressive evolution into VDJ+/DJ cells (D-J rearrangement on the wt allele) and, eventually, by their regular emergence within the kinetic system.

Model studies predict observable TCR β repair foci and system-relevant timings

We next wished to examine whether a better temporal picture of our model could be obtained from experimental results. In the absence of real-time insights into sequential rearrangements at dual TCR β alleles in single DN cells, we considered a recent study that used 3D immuno-fluorescence *in situ* hybridization (FISH) to visualize newly generated RAG-dependent DSBs at TCR loci based on the accumulation of the DNA-repair protein 53BP1 (*Chan et al., 2013*). However, this study reported a low ratio of 53BP1⁺ TCR β foci in DN thymocytes (<10%), which, in light of the recurrent positioning of TCR β alleles at the nuclear periphery (*Schlimgen et al., 2008*), has been taken to suggest that such alleles are mostly inefficient in their ability to undergo V(D)J recombination. Yet, when regarded from a dynamic angle, immuno-FISH data and its significance in terms of recombination frequency must be weighted relative to mean-time values of both cell residency within the system and length of focus detection (for more details, see *SI Appendix, section S6*). On this basis, new calculations again proved the predictive power of our model system in this specific instance and, furthermore, allowed us to predict real-time features of the system. Initially, considering (i) reasonable adjustments regarding lifespan kinetics of developing DN cells (*Vasseur et al., 2001*) as well as of 53BP1⁺ foci at sites of DNA DSBs (*Asaithamby and Chen, 2009*), and (ii) anticipated scenarios of cell developmental histories and their distributions into distinct ψ_1 , ψ_2 , and ψ_3 subsets, as a first approximation we obtained likelihoods of 9.27% and 0.53% for single and dual observable foci, respectively (*SI Appendix, S6A*), which closely correlates with the corresponding 9.4% and 0.4% reported in (*Chan et al., 2013*). Building on these assets, we then took advantage of the latter experimental values to determine - following mathematical refinements aimed at more precisely tracking RAG-dependent DSB kinetics throughout the ψ_1 , ψ_2 , and ψ_3 subsets - realistic estimates for mean residence time and time unit in the model system [RT = 89h; $t.u.$ = 1.18h; *SI Appendix, S6B*, equations (17)-(20) and Fig. S6-2]. This offered the opportunity for a direct assessment of typical time-frames in the system (Table 1; *SI Appendix, S6B*, Additional comments #1 and #2). Importantly, we note that the former two estimates of RT = 89h and $t.u.$ = 1.18h are consistent with experimentally derived approximations of the duration of the DN2-DN3 developmental window (*Vasseur et al., 2001*) and of total timings supposedly integrated by $\rho = 1 t.u.$ (*SI Appendix, section S7*), respectively. We further note that temporal interconnections point to the effectiveness of the system's operation. Notably, median values of initiation time [mIT; 9 $t.u.$ /10.6h] are sufficiently lower, and interallelic time gap [m δ ; 11 $t.u.$ /13h] are sufficiently higher, compared to mean residence time [RT; 75 $t.u.$ /89h] and feedback inhibition [ρ ; 1 $t.u.$ /1.18h], respectively, to bolster both bi-allelic usage and hence cell productivity, and the preponderance of allelically excluded individuals. Finally, it is now obvious that these values better fit the above immuno-FISH data than those derived from previous Markov-chain-based simulations, in which stochasticity mostly pertains to the V β -to-DJ β rearrangement step [(*Farcot et al., 2010*); *SI Appendix, S6B*, Additional comments #3 and #4]. When also considering that, although not further discussed in their article, data reported in (*Chan et al., 2013*) - in particular the low ratio of dual/total 53BP1⁺ foci and corollary of sparse contemporary DSBs - imply that both steps of TCR β recombination are likely to occur asynchronously, we contend that these additional simulation results provide further compelling evidence supporting the consistency of our model.

	[RT]	[mIT]	[m δ]	t_x	τ	ρ
Relative Time ($t.u.$)	75	9	11	10	55	1
Absolute time (hours)	89h	10.6h	13h	11.8h	65h	1.18h

Table 1. Relative and estimated absolute time values for mean residence time inside the system [RT]; median value of initiation time [mIT] (which we define as the time when 50% of the cells have initiated a first D-to-J rearrangement, and which therefore reflects the efficiency of the recombination initiation phase); median value of interallelic time gap δ [m δ]; and dwell-time values of D-to-J rearrangement [t_x], V-to-DJ rearrangement [= parameter τ], and feedback inhibition [= parameter ρ].

Discussion

We created a mathematical model in order to test the hypothesis that cell-to-cell interallelic variability during the initiation phase of TCR β gene recombination can account for typical AE profiles at this locus. The model is based on a limited set of parameters - p , τ , and ρ , which define critical time intervals during the recombination process. This work validates our hypothesis in the context of both synchronous and asynchronous flow-dependent behaviors of developing cells. Simulations under diverse allelic conditions, to mimic experimental studies, yielded data that underscore the predictive power of the model.

At V-D-J structured loci, the initiation phase of AE (allele asynchrony) is usually viewed in the context of V-to-DJ assembly, because of observations that dually DJ rearranged *Igh* or TCR β alleles are present in nearly all pre-BCR/pre-TCR selected cells. Since those findings, however, signs of an unforeseen dissociation of D-to-J recombination on opposite alleles have surfaced (*Born et al., 1988; Akamatsu et al., 2003; Dudley et al., 2003; Bonnet et al., 2009*). Our simulations reconcile all these data, at lower regulatory cost, as they support a dynamical cell-system in which allele asynchronicity generated by probabilistic events prior to D β -to-J β rearrangement can account for the observed distribution of TCR β genotypes, while also highlighting that the DJ/DJ prevalence can peak prior to the first V-to-DJ joining attempt. Moreover, the model provides rational explanations of the antipodal observations of V β gene AE, DJ β joining ubiquity, and seeming scarcity of RAG-mediated DSB foci at TCR β alleles (*Chan et al., 2013*). In fact, the model provides these explanations without suggesting that the TCR β alleles are inefficient in their ability to undergo V(D)J recombination (after all, nearly all cells experience dual allele activation/rearrangement at some point in their own development), eventually reconciling stochasticity in recombination with developmental cell productivity [see *e.g. Gorman and Alt (1998)*]. Finally, the model system also presents a broader interpretation of various genetic mutations known to affect AE in T lymphocytes (*SI Appendix, S7B*).

Under this model, asymptotic distributions of TCR β genomic profiles in T-cells split into three subgroups (ψ_1 , ψ_2 , and ψ_3), depending on the relative values between the interallelic time gap δ and parameters τ and ρ , with $\delta > \rho$ as a key precept to enforce AE. From a biological perspective, we believe that δ values in DN cells primarily rely on the asymmetric FPT timing of combinatorial, activating epigenetic changes. These [changes in histone marks from e.g. H3K9me2 to H3K4me3, Pol II-generated short-RNA transcripts, CpG demethylation, mobilization of nucleosomes; *Morshead et al. (2003)*; *Sikes and Oltz (2012)*; *Kondilis-Mangum et al. (2010)*; *Zacarias-Cabeza et al. (2015)*] would be triggered on opposite D β -J β loci by the concerted activity of adjacent *cis*-regulatory elements, whose combinatorial stochastic outputs would thereby be integrated by geometric distributions dependent on parameters p_i . Schematically, primary changes may possibly initiate at E β through the recruitment of a discrete set of master transcription factors and/or co-activators/epigenetic modulators, and subsequently be enforced and regionally spread *via* E β -pD β driven and inter-nucleosome cooperative interactions (*Sikes and Oltz, 2012*; *Zhang et al., 2014*). This in turn would enable the segregation from the suppressive nuclear periphery, the structuring of the TCR β recombination center, and prime D-to-J rearrangement. Longer term epigenetic and structural reorganization, initially insulated within the D-J-C domains, would also involve - in our view, as a ‘canalized’ process (see below) - the V β -containing domains propelling *in fine* their folding into the DN-specific 3D-architecture wherein topological confinement permits V-to-DJ rearrangement (*Majumder et al., 2015*; *Hu et al., 2015*). Accordingly, the δ interallelic hallmarks in individuals cells would grossly be maintained up to this phase. In practice this would be up to the point of ‘prime-encounter’ and ensuing recombination between V and DJ elements (*Lucas et al., 2014*), on first- and second-rearranging alleles, respectively (unless the process is interrupted by ρ timed feedback). Without focused experimental clues on dynamical (epi)genome evolution, at separate TCR β alleles, in statistically significant numbers of single DN developing cells, we acknowledge that this overall scenario remains hypothetical. Still, because our ‘few-parameter’ modeling study demonstrates sufficient robustness under various allelic conditions, it is worth considering this possibility, given the remaining mysteries that surround the initiation phase of AE.

Recombination of genomically distant V and DJ segments requires their juxtaposition within merged CTCF-binding element (CBE)-flanked loop domains (*Majumder et al., 2015*; *Hu et al., 2015*). Underlying principles that direct these events can be accommodated by our dynamical model if considered as part of a hierarchical progression, in which allele- and cell-specific kinetic fluctuations generated during this step (united under parameter τ) remain limited with respect to those responsible for prime asynchronicity (note that we thus do not consider that all these later events are homogeneously invariant at all TCR β alleles). It is now recognized that early developing cells could have a relatively high stochastic variation across the genome compared to more differentiated cells, with epigenetically regulated variability impinging on early developmental plasticity, followed by less variable, canalization periods of development to produce consistent phenotypic outcome(s) (*Pujadas and Feinberg, 2012*). In these contexts, earlier views of autonomous blocks in V-to-DJ recombination, largely from studies involving the *Igh* locus (*Corcoran et al., 1998*) - when examined with the benefit of hindsight - appear to better support a scenario whereby the activation of the two regions is not necessarily uncoupled (*Fuxa et al., 2004*; *Malin et al., 2010*; *Subrahmanyam et al.,*

2012). Overall, we therefore argue that bound τ parameterization (and ρ , see below) does not invalidate the meaning of the model, as their dwell-time $\pm$ standard deviation values plausibly oscillate over shorter time-ranges when compared to those which lead to allelic differential activation.

Parameter ρ , which covers the entire period of time from the expression of a VDJ+ assembled gene to feedback *trans*-inhibition, presumably includes basic cellular processes (transcription and RNA processing, nuclear export and translation of mRNA, and protein folding and delivery), as well as DN cell-specific events (including pre-TCR assembly and ensuing signaling and inhibitory outcomes). Although the precise timelines of these processes are unknown, those steps that have been studied individually (in the case of TCR, Ig, or other gene/protein paradigms) were found to be on the scale of minutes to ≤ 1 hr duration (*SI Appendix, S7A*, for further discussion of timings supposedly integrated by ρ). This suggests that when applied to the 3 - 4.5 day period necessary for DN2-DN3 development (*Vasseur et al., 2001*), these steps normally also have limited impact on the course of the dynamical system process, so that most cells would comply with the $\delta > \rho$ precept.

It is tempting to consider that, providing recalibration, our model system could likewise be tailored to the analysis of recombination dynamics at other TCR or Ig genes, and to the analysis of developmental cell fates in distinct lymphoid lineages. This could offer valuable information on mechanistic similarities/disparities among adaptive immune loci subject or otherwise to AE [*Mostoslavsky et al. (2004)*; *Cedar and Bergman (2008)*; *Brady et al. (2010a)*; *Vettermann and Schlissel (2010)*; NB: depending on the locus (and distinct from the TCR β locus), recalibration would firstly involve the adjustment of the probabilistic variable ϵ ($\neq$ from the one-third/two-thirds outcome), *Vettermann and Schlissel (2010)*; and would possibly also require model/parameters' amendment to integrate the cumulative effects of independently controlled chromatin subdomains within larger loci, *e.g.* distal V domains at the *Igh* locus, *Montefiori et al. (2016)*]. Beyond these applications, our modeling hypotheses might also be adapted to the study of discrete cell types that employ random monoallelic gene expression at the transcriptional level, independently of the genetic sequence and parental origin (*i.e.* distinct from classical examples of X-chromosome inactivation and genomic imprinting). A growing body of evidence demonstrates that a significant fraction of mammalian autosomal genes are subject to monoallelic expression in different cell types, in an inheritable manner (*Eckersley-Maslin and Spector, 2014*). Indeed, for some such genes (olfactory receptor genes), a kinetic model assuming a tradeoff between slow activation by step-wise epigenetic events and fast feedback signaling satisfactorily accounts for allelic singularity in expression (*Monahan and Lomvardas, 2015*).

In summary, we provide an original view of TCR β AE as a model system of dynamic distribution, primarily based on noise in the initiation phase of V(D)J recombination. Sufficiently reflective of allelically excluded cells, this unique model is remarkably robust and satisfactorily constrains the generation of allelically included stowaways, with no need for any further sophisticated controls. This property may be of significant value - theoretical frameworks have established the extreme cost of multiplying the sources of signaling events when endeavoring to achieve increased biological accuracy (*Lestas et al., 2010*).

Materials and Methods

Detailed modeling procedures, including mathematical formulations [Equations (1)-(16)], model programming and computational properties, as well as the collection of experimental data used for the calibration of the model and in modeling simulation studies, and associated references, are available in the *SI Appendix, section S1/Methods, section S6/[Equations (17)-(20)], and section S8/References, respectively.*

Acknowledgments

Work in the PF laboratory is supported by institutional grants from 'Institut National de la Santé et de la Recherche Médicale' (Inserm) and 'Centre National de la Recherche Scientifique' (CNRS), and by dedicated grants from the 'Agence Nationale de la Recherche' (ANR), the 'Institut National du Cancer' (INCa), the 'ITMO Cancer Alliance Nationale pour les Sciences de la Vie et de la Santé' (AVIESAN) and the 'Fondation Princesse Grace de la Principauté de Monaco'.

References

- Aifantis I**, Buer J, von Boehmer H, Azogui O. Essential role of the pre-T cell receptor in allelic exclusion of the T cell receptor beta locus. *Immunity*. 1997; 7(5):601–607.
- Akamatsu Y**, Monroe R, Dudley DD, Elkin SK, Gartner F, Talukder SR, Takahama Y, Alt FW, Bassing CH, Oettinger MA. Deletion of the RAG2 C terminus leads to impaired lymphoid development in mice. *Proc Natl Acad Sci USA*. 2003; 100:1209–1214.
- Alt FW**, Zhang Y, Meng FL, Guo C, Schwer B. Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell*. 2013; 152(3):417–429.
- Asaithamby A**, Chen DJ. Cellular responses to DNA double-strand breaks after low-dose gamma-irradiation. *Nucleic Acids Res*. 2009; 37(12):3912–3923.
- Balomenos D**, Balderas RS, Mulvany KP, Kaye J, Kono DH, Theofilopoulos AN. Incomplete T cell receptor V beta allelic exclusion and dual V beta-expressing cells. *J Immunol*. 1995; 155(7):3308–3312.
- Benner C**, Isoda T, Murre C. New roles for DNA cytosine modification, eRNA, anchors, and superanchors in developing B cell progenitors. *Proc Natl Acad Sci USA*. 2015; 112(41):12776–12781.
- von Boehmer H**, Aifantis I, Azogui O, Feinberg J, Saint-Ruf C, Zober C, Garcia C, Buer J. Crucial function of the pre-T-cell receptor (TCR) in TCRb selection, TCRb allelic exclusion and ab versus gd lineage commitment. *Immunol Rev*. 1998; 165:111–119.
- Bonnet M**, Huang F, Benoukraf T, Cabaud O, Verthuy C, Boucher A, Jaeger S, Ferrier P, Spicuglia S. Duality of enhancer functioning mode revealed in a reduced TCR beta gene enhancer knockin mouse model. *J Immunol*. 2009; 183(12):7939–7948.
- Born W**, White J, Kappler J, Marrack P. Rearrangement of IgH genes in normal thymocyte development. *J Immunol*. 1988; 140(9):3228–3232.
- Brady BL**, Steinel NC, Bassing CH. Antigen receptor allelic exclusion: an update and reappraisal. *J Immunol*. 2010; 185(7):3801–3808.
- Brady BL**, Oropallo MA, Yang-lott KS, Serwold T, Hochedlinger K, Jaenisch R, Weissman IL, Bassing CH. Position-dependent silencing of germline V β segments on TCR β alleles containing preassembled V β DJ β C β 1 genes. *J Immunol*. 2010; 185(6):3564–3573.

- Carpenter AC**, Yang-lott KS, Chao LH, Nuskey B, Whitlow S, Alt FW, Bassing CH. Assembled DJ beta complexes influence TCR beta chain selection and peripheral V beta repertoire. *J Immunol.* 2009; 182(9):5586–5595.
- Cedar H**, Bergman Y. Choreography of Ig allelic exclusion. *Curr Opin Immunol.* 2008; 20(3):308–317.
- Chan EA**, Teng G, Corbett E, Choudhury KR, Bassing CH, Schatz DG, Krangel MS. Peripheral subnuclear positioning suppresses Tcrb recombination and segregates Tcrb alleles from RAG2. *Proc Natl Acad Sci USA.* 2013; 110(48):E4628–37.
- Chaumeil J**, Skok JA. A new take on V(D)J recombination: transcription driven nuclear and chromatin reorganization in Rag-mediated cleavage. *Front Immunol.* 2013; 4, 423:1–8.
- Ciofani M**, Schmitt TM, Ciofani A, Michie AM, Cuburu N, Aublin A, Maryanski JL, Zuniga-Pflucker JC. Obligatory role for cooperative signaling by pre-TCR and Notch during thymocyte differentiation. *J Immunol.* 2004; 172(9):5230–5239.
- Claverie JM**, Langman R. Models for the rearrangements of immunoglobulin genes: a computer view. *Trends in Biochemical Sciences.* 1984; 9:293–296.
- Coleclough C**. Chance, necessity and antibody gene dynamics. *Nature.* 1983; 303(5912):23–26.
- Corcoran AE**, Riddell A, Krooshoop D, Venkitaraman AR. Impaired immunoglobulin gene rearrangement in mice lacking the IL-7 receptor. *Nature.* 1998; 391(6670):904–907.
- Coutinho R**, Fernandez B, Lima R, Meyroneinc A. Discrete time piecewise affine models of genetic regulatory networks. *J Math Biol.* 2006; 52(4):524–570.
- Dudley DD**, Sekiguchi J, Zhu C, Sadofsky MJ, Whitlow S, DeVido J, Monroe RJ, Bassing CH, Alt FW. Impaired V(D)J recombination and lymphocyte development in core RAG1-expressing mice. *J Exp Med.* 2003; 198:1439–1450.
- Eckersley-Maslin MA**, Spector DL. Random monoallelic expression: regulating gene expression one allele at a time. *Trends Genet.* 2014; 30(6):237–244.
- Farcot E**, Bonnet M, Jaeger S, Spicuglia S, Fernandez B, Ferrier P. TCR beta allelic exclusion in dynamical models of V(D)J recombination based on allele independence. *J Immunol.* 2010; 185(3):1622–1632.
- Fuxa M**, Skok J, Souabni A, Salvagiotto G, Roldan E, Busslinger M. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev.* 2004; 18(4):411–422.
- Gorman JR**, Alt FW. Regulation of Immunoglobulin light chain isotype expression. *Adv Immunol.* 1998; 69:113–181.
- Hewitt SL**, Yin B, Ji Y, Chaumeil J, Marszalek K, Tenthorey J, Salvagiotto G, Steinell N, Ramsey LB, Ghysdael J, Farrar MA, Sleckman BP, Schatz DG, Busslinger M, Bassing CH, Skok JA. RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. *Nat Immunol.* 2009; 10(6):655–664.
- Hu J**, Zhang Y, Zhao L, Frock RL, Du Z, Meyers RM, Meng FL, Schatz DG, Alt FW. Chromosomal loop domains direct the recombination of antigen receptor genes. *Cell.* 2015; 163(4):947–959.
- Jaeger S**, Fernandez B, Ferrier P. Epigenetic aspects of lymphocyte antigen receptor gene rearrangement or ‘when stochasticity completes randomness’. *Immunology.* 2013; 139(2):141–150.
- Ji Y**, Little AJ, Banerjee JK, Hao B, Oltz EM, Krangel MS, Schatz DG. Promoters, enhancers, and transcription target RAG1 binding during V(D)J recombination. *J Exp Med.* 2010; 207(13):2809–16.
- Ji Y**, Resch W, Corbett E, Yamane A, Casellas R, Schatz DG. The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell.* 2010; 141(3):419–431.

- van Kampen NG.** Stochastic processes in physics and chemistry. Elsevier Science and Technology Books.; 2007.
- Khor B, Sleckman BP.** Allelic exclusion at the TCRb locus. *Curr Opin Immunol.* 2002; 14(2):230–234.
- Khor B, Sleckman BP.** Intra- and inter-allelic ordering of T cell receptor beta chain gene assembly. *Eur J Immunol.* 2005; 35(3):964–970.
- Kondilis-Mangum HD, Cobb RM, Osipovich O, Srivatsan S, Oltz EM, Krangel MS.** Transcription-dependent mobilization of nucleosomes at accessible TCR gene segments in vivo. *J Immunol.* 2010; 184(12):6970–7.
- Krueger A, Zietara N, Lyszkiewicz M.** T Cell Development by the Numbers. *Trends in immunology.* 2017; 38:128–139.
- Lestas I, Vinnicombe G, Paulsson J.** Fundamental limits on the suppression of molecular fluctuations. *Nature.* 2010; 467(7312):174–178.
- Levin-Klein R, Bergman Y.** Epigenetic regulation of monoallelic rearrangement (allelic exclusion) of antigen receptor genes. *Front Immunol.* 2014; 5:625.
- Liang HE, Hsu LY, Cado D, Schlissel MS.** Variegated transcriptional activation of the immunoglobulin kappa locus in pre-b cells contributes to the allelic exclusion of light-chain expression. *Cell.* 2004; 118(1):19–29.
- Lucas JS, Zhang Y, Dudko OK, Murre C.** 3D trajectories adopted by coding and regulatory DNA elements: First-passage times for genomic interactions. *Cell.* 2014; 158:339–352.
- Majumder K, Koues OI, Chan EAW, Kyle KE, Horowitz JE, Yang-lott K, Bassing CH, Taniuchi I, Krangel MS, Oltz EM.** Lineage-specific compaction of Tcrb requires a chromatin barrier to protect the function of a long-range tethering element. *J Exp Med.* 2015; 212(1):107–120.
- Malin S, McManus S, Cobaleda C, Novatchkova M, Delogu A, Bouillet P, Strasser A, Busslinger M.** Role of STAT5 in controlling cell survival and immunoglobulin gene recombination during pro-B cell development. *Nat Immunol.* 2010; 11(2):171–181.
- Masuda K, Kakugawa K, Nakayama T, Minato N, Katsura Y, Kawamoto H.** T cell lineage determination precedes the initiation of TCRbeta gene rearrangement. *J Immunol.* 2007; 179(6):3699–3706.
- Monahan K, Lomvardas S.** Monoallelic expression of olfactory receptors. *Annu Rev Cell Dev Biol.* 2015; 31:721–740.
- Montefiori L, Wuerffel R, Roqueiro D, Lajoie B, Guo C, Gerasimova T, De S, Wood W, Becker KG, Dekker J, Liang J, Sen R, Kenter AL.** Extremely long-range chromatin loops link topological domains to facilitate a diverse antibody repertoire. *Cell Rep.* 2016; 14(4):896–906.
- Morshead KB, Ciccone DN, Taverna SD, Allis CD, Oettinger MA.** Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks of histone H3 dimethylated at lysine 4. *Proc Natl Acad Sci USA.* 2003; 100(20):11577–11582.
- Mostoslavsky R, Alt FW, Rajewsky K.** The lingering enigma of the allelic exclusion mechanism. *Cell.* 2004; 118(5):539–544.
- Mostoslavsky R, Singh N, Tenzen T, Goldmit M, Gabay C, Elizur S, Qi P, Reubinoff BE, Chess A, Cedar H, Bergman Y.** Asynchronous replication and allelic exclusion in the immune system. *Nature.* 2001; 414(6860):221–5.
- Murre C.** Nuclear geography and allelic exclusion. *Nat Immunol.* 2008; 9(7):718–720.
- Outters P, Jaeger S, Zaarour N, Ferrier P.** Long-range control of V(D)J recombination & allelic exclusion: modeling views. *Adv Immunol.* 2015; 128:363–413.

- Piper H**, Litwin S, Mehr R. Models for antigen receptor gene rearrangement. II. Multiple rearrangement in the TCR: allelic exclusion or inclusion? *J Immunol.* 1999; 163(4):1799–1808.
- Pujadas E**, Feinberg AP. Regulated noise in the epigenetic landscape of development and disease. *Cell.* 2012; 148(6):1123–1131.
- Ramond C**, Berthault C, Burlen-Defranoux O, de Sousa AP, Guy-Grand D, Vieira P, Pereira P, Cumano A. Two waves of distinct hematopoietic progenitor cells colonize the fetal thymus. *Nat Immunol.* 2014; 15(1):27–35.
- Schatz DG**, Swanson PC. V(D)J recombination: mechanisms of initiation. *Annu Rev Genet.* 2011; 45:167–202.
- Schlimgen RJ**, Reddy KL, Singh H, Krangel MS. Initiation of allelic exclusion by stochastic interaction of Tcrb alleles with repressive nuclear compartments. *Nat Immunol.* 2008; 9(7):802–809.
- Senoo M**, Wang L, Suzuki D, Takeda N, Shinkai Y, Habu S. Increase of TCR Vbeta accessibility within Ebeta regulatory region influences its recombination frequency but not allelic exclusion. *J Immunol.* 2003; 171:829–835.
- Sepulveda N**, Boucontet L, Pereira P, Carneiro J. Stochastic modeling of T cell receptor gamma gene rearrangement. *J Theor Biol.* 2005; 234(2):153–165.
- Shlomchik MJ**, Watts P, Weigert MG, Litwin S. Clone: a Monte-Carlo computer simulation of B cell clonal expansion, somatic mutation, and antigen-driven selection. *Curr Top Microbiol Immunol.* 1998; 229:173–197.
- Sikes ML**, Oltz EM. Genetic and epigenetic regulation of Tcrb gene assembly. *Curr Top Microbiol Immunol.* 2012; 356:91–116.
- Soshnev AA**, Josefowicz SZ, Allis CD. Greater than the sum of parts: complexity of the dynamic epigenome. *Molecular Cell.* 2016; 62:681–694.
- Subrahmanyam R**, Du H, Ivanova I, Chakraborty T, Ji Y, Zhang Y, Alt FW, Schatz DG, Sen R. Localized epigenetic changes induced by D_H recombination restricts recombinase to DJ_H junctions. *Nat Immunol.* 2012; 13(12):1205–1212.
- Therizols P**, Illingworth RS, Courilleau C, Boyle S, Wood AJ, Bickmore WA. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science (New York, NY).* 2014; 346:1238–1242.
- Vasseur F**, Campion AL, Penit C. Scheduled kinetics of cell proliferation and phenotypic changes during immature thymocyte generation. *Eur J Immunol.* 2001; 31(10):3038–3047.
- Vettermann C**, Schlissel MS. Allelic exclusion of immunoglobulin genes: models and mechanisms. *Immunol Rev.* 2010; 237(1):22–42.
- Vibert J**, Thomas-Vaslin V. Modelling T cell proliferation: Dynamics heterogeneity depending on cell differentiation, age, and genetic background. *PLoS computational biology.* 2017; 13:e1005417.
- Volpe JM**, Kepler TB. Large-scale analysis of human heavy chain V(D)J recombination patterns. *Immunome Res.* 2008; 4:3.
- Zacarias-Cabeza J**, Belhocine M, Vanhille L, Cauchy P, Koch F, Pekowska A, Fenouil R, Bergon A, Gut M, Gut I, Eick D, Imbert J, Ferrier P, Andrau JC, Spicuglia S. Transcription-dependent generation of a specialized chromatin structure at the TCRbeta locus. *J Immunol.* 2015; 194:3432–3443.
- Zhang H**, Tian XJ, Mukhopadhyay A, Kim KS, Xing J. Statistical mechanics model for the dynamics of collective epigenetic histone modification. *Physical Review Letters.* 2014; 112:068101.