How many TCR clonotypes does a body maintain?

Grant Lythe a,*, Robin E. Callard b, c, Rollo L. Hoare b, c, Carmen Molina-París a

a Department of Applied Mathematics, School of Mathematics, University of Leeds, Leeds LS2 9JT, UK
b Institute for Child Health, University College London, 30 Guilford Street, London WC1N 1EH, UK
c Centre for Mathematics and Physics in the Life Sciences and Experimental Biology, University College London, Gower Street, London WC1N 1EH, UK

HIGHLIGHTS

• The number of T cells of one clonotype is an integer.
• The history of a clonotype starts with release from the thymus, and ends with extinction.
• Competition and cross-reactivity are included in a natural way.
• The average number of cells per clonotype, in a human body, is only of order 10.

ARTICLE INFO

Article history:
Received 25 May 2015
Received in revised form
13 September 2015
Accepted 7 October 2015
Available online 7 November 2015

Keywords:
T cells
Clonal repertoire
Homeostasis
Competition
Extinction
Stochastic modelling

ABSTRACT

We consider the lifetime of a T cell clonotype, the set of T cells with the same T cell receptor, from its thymic origin to its extinction in a multiclonal repertoire. Using published estimates of total cell numbers and thymic production rates, we calculate the mean number of cells per TCR clonotype, and the total number of clonotypes, in mice and humans. When there is little peripheral division, as in a mouse, the number of cells per clonotype is small and governed by the number of cells with identical TCR that exit the thymus. In humans, peripheral division is important and a clonotype may survive for decades, during which it expands to comprise many cells. We therefore devise and analyse a computational model of homeostasis of a multiclonal population. Each T cell in the model competes for self pMHC stimuli, cells of any one clonotype only recognising a small fraction of the many subsets of stimuli. A constant mean total number of cells is maintained by a balance between cell division and death, and a stable number of clonotypes by a balance between thymic production of new clonotypes and extinction of existing ones. The number of distinct clonotypes in a human body may be smaller than the total number of naive T cells by only one order of magnitude.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Approximately \(4 \times 10^{11}\) T cells circulate in the adult human body (Jenkins et al., 2009), each with multiple T cell receptors (TCR) (Varma, 2008) on its surface. Each T cell is descended from a T cell that left the thymus after selection by binding to self-peptides expressed in association with major histocompatibility complex molecules (self pMHC) (Bains et al., 2009; Van Laethem et al., 2012), or is itself a thymic emigrant. The immune system can respond to very many different pathogens because very many different TCRs are present in the body. However, the TCRs present on the surface of one T cell are usually all identical. The set of cells with the same TCR defines a T cell clonotype, and the set of T cells in the body can be thought of as a repertoire of clonotypes. How many TCR clonotypes are there in humans, mice and other mammals? (Langman and Cohn, 1987; Blattman et al., 2002; Ciupé et al., 2013; Thomas et al., 2014). Estimates of the number of different TCRs that could, in principle, be produced by VDJ gene rearrangement in the thymus,
are about 10^{15} (Sewell, 2012; Nikolic-Zugich et al., 2004; Zarnitsyna et al., 2013; Murugan et al., 2012). However, the human body cannot contain even one T cell of 10^{15} possible types: 10^{15} T cells would weigh about 500 kg (Mason, 1998).

Extracing and reverse transcribing RNA from pools of T cells, and amplifying, by PCR, the gene sequences that encode the TCR β chain, makes it possible to estimate the TCR diversity in a sample of blood. Initial estimates, based on extrapolation from a small fraction of the repertoire (Aristila et al., 1999), and more recent studies that are able to directly count large number of sequences and perform "missing species" analyses (Robins et al., 2009; Warren et al., 2011; Rempala and Seweryn, 2013), yield estimates of 10^8 to 10^9 (Qi et al., 2014). A new field of immunosequencing has emerged with technologies designed to sequence TCRs (Robins, 2013). Millions of TCR sequences can be amplified in a single multiplex PCR reaction, prepared and then read in parallel from a single sample. The distribution of gene usage can be measured with flow cytometry (Salameire et al., 2009; Ciupe et al., 2013), and used to track the dependence on phenotype, age and variation between individuals (Naylor et al., 2005; Britanova et al., 2014; Elhanati et al., 2014; Becattini et al., 2015). Depending on the number of TCRβ chains that each TCRβ chain combines with, the number of distinct clonotypes in one human may be much higher than estimates based on TCRβ alone (Keşmir et al., 2000). The spleen of a mouse has been estimated to contain 2 x 10^6 clones of about 10 cells each (Casrouge et al., 2000). In mice, different T cell types can be compared and the effects of infections and immunization on the repertoire can be tracked (Bouso et al., 1998; Venturi et al., 2008; Bergot et al., 2015; Thomas et al., 2014).

In an adult, the number of recirculating T cells and the number of distinct clonotypes are believed to be held nearly constant for decades by balancing T cell loss with input from the thymus and homeostatic mechanisms controlling cell division and death in the periphery (Cannon, 1932; Tanchot et al., 1997; De Boer and Perelson, 1997; Goldrath and Bevan, 1999; Berzins et al., 2002; Murray et al., 2003; Troy and Shen, 2003; Seddon and Zamoyska, 2003; Takada and Jameson, 2009; Rudd et al., 2011; Germain, 2012). The diversity of the T cell repertoire in the periphery is made possible by the enormous variability of their self pMHC ligands (De Boer and Perelson, 1995; Moses et al., 2003; Blanchfield et al., 2013).

Division of T cells in the periphery is determined by competition among clonotypes, not to add new cells to existing clonotypes. We have previously modelled these processes in a highly cross-reactive pattern of self pMHC recognition, which determines whether they die or survive. The fate of individual clonotypes determines the fate of a clonotype and the fate of individual clonotypes determines the homeostasis of the repertoire. New clonotypes are produced by the thymus but we assume that the thymus does not produce cells of a pre-existing type, and that each individual T cell has only one type of TCR (Molinari-Paris et al., 2011, 2011). In this work, the variables are the integer numbers of cells of each clonotype; a death or division event changes one of the variables by one cell. No clonotype has a pre-established self-limiting size: individual clonotypes compete for resources in a highly cross-reactive pattern of self pMHC recognition, which determines whether they die or survive. The fate of individual clonotypes determines the fate of a clonotype and the fate of individual clonotypes determines the homeostasis of the repertoire.

1.1. Parameter values for humans and mice

The variables of the model are the numbers of cells of each clonotype; the parameters are the mean cell death rate, μ, the rate of cell division caused by stimulus from one self pMHC subset, γ, the rate at which new clonotypes are produced by the thymus, θ, the number of cells in a newly-produced clonotype, n₀, the total number of self pMHC subsets, M, and the probability that any given self pMHC is recognised by a randomly-selected T cell clonotype, p. In order to give a concrete estimate of clonotype numbers, we consider the values of μ, γ, θ, n₀, M and p that give an appropriate description of the naive CD4⁺ T cell repertoire in mice and humans.

The parameter μ sets the overall timescale that permits comparison of computational and physiological dynamics. We take μ
to be constant, based on the simplifications that the T cell population is well mixed and that average availability of the IL-7 resource (Jameson, 2005; Bosco et al., 2005) is constant in time. The mean lifetime of a naive CD4+ T cell in the human body (Michie et al., 1992; Macallan et al., 2004; Borghans and De Boer, 2007) is 1–10 years (De Boer et al., 2012; De Boer and Perelson, 2013; Westera et al., 2013, 2015). Thus, μ should be less than 1 year⁻¹. In a mouse, the corresponding lifetime is about 1 month (Labrecque et al., 2001; Bourgeois et al., 2008; Westera et al., 2013); we take μ = 1 month⁻¹ when considering mice. Next, consider the total number of naive CD4 + T cells. A human body has about 2 × 10¹¹; a mouse has about 10⁷ (Borghans and De Boer, 2007; Jenkins et al., 2009; Bains et al., 2009; den Braber et al., 2012). As we shall see, this number fixes the value of the combination γM/μ in humans, and constrains it in mice.

In our model, the number of subsets of pMHC stimuli is large, and individual TCRs are cross-reactive (Perelson and Oster, 1979; Evavold et al., 1995; Borghans et al., 1999; Newell et al., 2011; Calis and De Boer, 2012; van den Berg and Rand, 2007; Hao et al., 2006; Zarnitsyna et al., 2013; Yates, 2014; Birnbaum et al., 2014); recognising many different self pMHC; the number of subsets of self pMHC present is the parameter M. One estimate, based on the number of distinct 9-mers in the human proteome, is 10²⁷ (Broughs and de Boer, 2004). Another, based on the observation that peptides up to at least 14 amino acids long can be presented by MHC molecules, is as high as 10⁴⁶ (Mason, 1998; Sewell, 2012). It may be more appropriate to count the different possible antigen presentation patterns on the surface of antigen presenting cells (Van Den Berg et al., 2001). The appropriate value of M is thus difficult to establish. Fortunately, our predictions are not very sensitive to the choice of M.

The total number of cells, and the rate of thymic production, differ from one type of mammal to another. On the other hand, the universe of different epitopes, presented as self pMHC to T cells, is a property of the mammal’s environment, lifestyle and physiology. Patterns of cross-reactivity depend on molecular properties of T cell receptors and may depend on HLA class (Košmrlj et al., 2010). We take the value of p, the probability that interaction with a given self pMHC is capable of causing a randomly-selected T cell to divide, to be the same in humans and mice: p = 10⁻⁶ (Rizzuto et al., 2009; Su et al., 2013; Jenkins et al., 2009). Similarily, even though the body of a mouse is smaller than that of a human by close to a factor of 10³, we use the estimate M = 10¹⁰ in both cases. The number of cell divisions per unit time, in the whole body, produced by one self-pMHC subset, γ, is determined by the total amount of the resource available. We expect γ to be proportional to body weight, and thus larger in humans than in mice. The product γM is the total number of cell divisions per unit time in the periphery.

Thymic production rates have been estimated by direct and indirect methods (Scollay et al., 1980; Hazenberg et al., 2003; Thomas-Vaslin et al., 2008; Bains et al., 2009; den Braber et al., 2012). Because we seek to construct a model of homeostasis, we keep the parameter θ constant, although thymic output decreases with age in both mice and humans (Westera et al., 2015). The adult human thymus releases about 10⁷ naive CD4+ cells per day (den Braber et al., 2012), or 4 × 10⁸ cells per year. The thymus of a mouse produces about 4 × 10⁷ cells per day (den Braber et al., 2012), or 10⁷ per month. The last of our parameters, n₀, reflects the number of rounds of cell division in the thymus, after formation and selection of the full TCR but before a clonotype is released to the periphery (Hare et al., 1999, 2000; Sinclair et al., 2013; von Boehmer, 2014). For example, the choice n₀ = 4 corresponds to the assumption that a thymocyte undergoes two rounds of cell division, after expression and selection of the TCR but before exit from the thymus.

2. Results

Numerical realisations update the variables

\[ nᵢ(t) = \text{number of T cells of clonotype } i \text{ at time } t. \]

A numerical realisation starts with N(0) clonotypes, when the index i that labels a clonotype runs from 1 to N(0). We shall be particularly interested in

\[ N(t) = \text{number of surviving clonotypes at time } t. \]

That is, \( N(t) \) is the number of indices \( i \) such that \( nᵢ(t) > 0 \). In any short time interval, a T cell either stays alive, dies, or divides into two cells of the same clonotype. The dynamics are governed by the birth (division) and death rates. For the latter, we adopt the simplest hypothesis: that each T cell, regardless of its clonotype, has probability μ per unit time of dying. In the absence of cell division, the number of cells in a population would decrease following an exponential decay curve (Tanchot et al., 1997). The probability per unit time of a T cell undergoing division into two daughter cells of the same clonotype, is proportional to the stimulus that it receives from the set of self pMHCs. Our assumption for cell division is that the stimulus from each subset is divided equally on average between the set of all T cells that recognise it, a fraction of all T cells that comprises T cells of many different clonotypes.

The rates of cell division are calculated from their interactions with M self pMHCs by generating, at the start of the run, the N(0) × M matrix A with entries (Košmrlj et al., 2010; Molina-París et al., 2011):

\[ Aᵢⱼ = \begin{cases} 1 & \text{if self pMHC } q \text{ stimulates } T \text{ cells of clonotype } i, \\ 0 & \text{otherwise}. \end{cases} \]

If \( Aᵢⱼ = 1 \) then we say “self pMHC q stimulates T cells of clonotype i to divide” or, equivalently, “T cells of clonotype i recognise self pMHC q”. In graph theory, the matrix A is known as the biadjacency matrix of the bipartite graph consisting of the set of T cell clonotypes, the set of self pMHC and the set of connections (Asratian et al., 1998). A small-scale example is depicted in Fig. 1.

The characteristic signature of a T cell clonotype is its pattern of interaction with the set of M self-pMHC stimuli, illustrated in Fig. 1. Each of the self pMHC is recognised with probability p (Borghans et al., 1999) so that the average number of self pMHCs recognised by a TCR (cross-reactivity) is pm. The number of possible patterns is thus at least as large as the number of ways of choosing pm samples from a total of M possibilities. If \( M = 1000 \) and \( p = 0.1 \) then the probability that two clonotypes have the same pattern, given by \((1−p)^M+p^M\), is already less than 10⁻⁸⁵. The recognition pattern of any one clonotype is highly unlikely to ever be repeated and every clonotype, existing in the periphery or emerging from the thymus, may be taken as unique. Similarly, the repertoires of different individual mammals from the same species will be independent samples from the set of possibilities, with little overlap (Bousso et al., 1998).

The number of T cells recognising self pMHC q at time \( t \), \( c_q(t) \), and the division rate of clonotype \( i \), \( Aᵢ(t) \), are given by

\[ c_q(t) = ∑_{i=1}^{N(t)} Aᵢᵢ nᵢ(t) \quad \text{and} \quad Aᵢ(t) = nᵢ(t) ∑_{q=1}^{M} γ_q Aᵢ₁ \cdot \frac{1}{c_q(t)}. \]

where \( γ_q \) is the stimulus rate of self pMHC q. In this paper, we consider the case where all self pMHC deliver stimulus at the same rate, and the case where the rates are drawn from a log-normal distribution (see Methods). As \( Δt → 0 \), \( Aᵢ(t)Δt \) is the probability that \( nᵢ(t+Δt)−nᵢ(t) = 1 \), corresponding to one round of cell division in one of the T cells of clonotype i between \( t \) and \( t+Δt \). The quantity \( nᵢ(t)/c_q(t) \) is the share of the total stimulus from q that is received by clonotype i, because of the assumption that each self pMHC distributes its stimulus, with equal probability, amongst all
number of cells. Let the total number of T cells at time $t$ be
\[ n(t) = \sum_{i=1}^{N(t)} n_i(t) \]
and the mean total number of T cells at time $t$ be $x(t) = E(n(t))$.
Note that $x(t)$, the mean over realisations of the stochastic processes of
the model, is not constrained to be an integer and satisfies the ordinary differential equation
\[ \frac{dx}{dt} = -\mu x + \Phi + \theta n_0, \quad (3) \]
where $\Phi = \sum_{q=1}^{N(t)} \Lambda_q(t)$ is the total rate of production of new cells by
divisional process. As long as each of the $M$ self pMHCs is recognised by at least one cell,
\[ \Phi = \sum_{q=1}^{N} \gamma_q \]
is constant and
\[ x(t) = (\Phi + \theta n_0) \mu^{-1} (1 - e^{-\mu t}) + x(0) e^{-\mu t}, \]
so that $x(t) \to \frac{\Phi + \theta n_0}{\mu}$, after a transient time of duration proportional to $\mu^{-1}$. In the case where $\gamma_q$ is equal to $\gamma$ for every $q$, $\Phi = \gamma M$
and the stationary mean total number of cells is $x^*$, where
\[ x^* = \frac{\gamma M + \theta n_0}{\mu}. \quad (4) \]
Note that this stably-maintained mean number of cells (De Boer and Perelson, 1997) arises as a "global effect of balancing stochastic influences, but is not a unique state in the multi-dimensional space of clonal sizes. Indeed, as we shall see, individual clonotypes continue to follow their dynamics, reacting to changes in competing clonotypes and chance events, which can lead to extinction or successful establishment in the periphery.

The steady-state fraction, $r$, of cells that are thymic emigrants that have not divided in the periphery is given by $r = \theta n_0 / (\gamma M + \theta n_0)$ (den Braber et al., 2012). If division of peripheral cells is the dominant contribution, the mean total number of cells is approximately equal to $\gamma M / \mu$. In adult humans, therefore, if $\mu = 0.5 \text{ year}^{-1}$ and the total number of naive CD4$^+$ T cells is $2 \times 10^{11}$ then $\gamma M = 10^{10} \text{ year}^{-1}$. The observation that thymic production, rather than peripheral division, dominates the dynamics of murine T cell homeostasis implies that $\gamma M = 10^7$, in a mouse. Assuming that the values of $M$ are similar in humans and mice, it is necessary to assume that the ratio $\gamma / \mu$ is smaller in a mouse than in a human, by a factor which reflects the overall size of the pMHC resource available, proportional to body weight (Table 1).

### 2.2. How many TCR clonotypes does a body maintain?

We can make estimates of the stable number of clonotypes, $N^*$, from estimates of the number of cells per clonotype. The dynamics in the "murine" limit is in fact rather simple: clonotypes operate independently, decay stochastically, and comprise few cells. New clonotypes, exported from the thymus, appear in the periphery with rate $\theta$, initially consisting of $n_0$ cells. Each cell has a constant

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Human</th>
<th>Mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.5 year$^{-1}$</td>
<td>1 month$^{-1}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>10 year$^{-1}$</td>
<td>$10^{-4}$ month$^{-1}$</td>
</tr>
<tr>
<td>$\theta$</td>
<td>$10^6$ year$^{-1}$</td>
<td>$2.5 \times 10^6$ month$^{-1}$</td>
</tr>
<tr>
<td>$M$</td>
<td>$10^{10}$</td>
<td>$10^{10}$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$10^{-6}$</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>$n_0$</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
probability per unit time of dying but cell division is very unlikely. If \( n_0 = 1 \), then nearly every clonotype has only one cell. If not, the mean number of cells per clonotype is less than \( n_0 \). With the simple estimate that the mean number of cells per clonotype, averaged over the lifetime of the clonotype, is \( n_0/2 \), we find that the mean number of clonotypes in the body of a mouse is related to the mean number of naive T cells in the mouse, \( x \), as \( t \to \infty \), by

\[
N^* \simeq \frac{2}{n_0} x.
\]

If we choose \( n_0 = 4 \) and \( x \approx 10^7 \), then we find \( N^* \simeq 5 \times 10^6 \).

We introduce the parameter

\[
\alpha = \frac{n_0 \theta}{\gamma M}.
\]

The relation (5) is valid in the weak peripheral division limit, \( \alpha \to 0 \), appropriate to describe T cell homeostasis in a mouse. In the human body, on the other hand, many fewer cells result from thymic production than from peripheral division (Borghans and Tesselaar, 2009; den Braber et al., 2012). The stationary mean number of clonotypes in the periphery is found by multiplying the mean lifetime by the rate of production of new clonotypes, \( \theta \). In Methods, we compute the mean number of clonotypes by calculating the mean lifetime of a clonotype in the limit \( \alpha \to 1 \):

\[
N^* \simeq \frac{\gamma M}{\mu} (1 - \gamma_e - \log (n_0 \alpha)), \quad \text{as } n_0 \to 0,
\]

where \( \gamma_e = 0.577 \ldots \) is the Euler–Mascheroni constant. Recall that \( \gamma M/\mu \) is the mean total number of cells. In Fig. 2, the solid lines use (16) and the dotted line is (7).

Thymic export is estimated to contribute less than 10 percent of the total production of new T cells in adults (den Braber et al., 2012). Let us identify where on the curve in Fig. 2, to locate naive CD4\(^+\) T cell homeostasis in adult humans. We use the estimate of annual production of cells by the thymus of \( 4 \times 10^5 \), and the estimate of annual number of peripheral divisions of \( 10^{11} \), to obtain \( \theta n_0 \gamma /M \approx 0.04 \). Thus, we find that the number of TCR clonotypes in the body is nine percent of the corresponding total number of T cells in the body, close to \( 10^{10} \). In other words, our model of competition produces a mean clonal size, over the lifetime of a clone, that is only of order 10, even though some clonotypes expand to thousands of cells.

In Table 1, we collect our estimated values of the parameters. Based on these estimates, the number of distinct clonotypes in the human body is about ten times smaller than the total number of naive T cells (\( N \approx 10^{10} \)) and the number of distinct clonotypes in a mouse is about two times smaller than the total number of naive T cells (\( N \approx 10^7 \)). These estimates are based on considering the lifetime of the population of cells making up one clonotype, starting with emergence from the thymus and ending in extinction in the periphery. Without peripheral division, the number of cells of one type does not exceed its initial value and the mean number of cells, averaged over the lifetime of the clonotype, is less than the number produced by the thymus, \( n_0 \). When there is peripheral division, the number of cells of a given clonotype may be large at some points in the lifetime of the clone, but the average over the lifetime of the clonotype is less than the maximum.

In Methods, we describe the implementation of the Gillespie algorithm to produce numerical realisations of the model. In Figs. 3, 5 and 6, we display results from small-scale numerical realisations, without thymic input, that illustrate the extinction of clonotypes due to competition. Larger systems, maintaining an invariance from the perspective of any one clonotype, are shown in Fig. 7. The effect of thymic production is considered in Fig. 8. In the numerical realisations summarised in Fig. 9, each self pMHC is assigned a strength \( \gamma_e \) that is nor constant but drawn from a log-normal distribution. In Section 4.3, we consider the extinction time of the ith clonotype as a random variable, and derive an expression for its probability distribution using a diffusion approximation.

### 3. Discussion

Emerging from the thymus with a pattern of recognition of self pMHC, each TCR clonotype is a species that competes in the pool of pre-established ones. We have analysed naive T cell homeostasis by considering the lifetimes of sets of cells making up TCR clonotypes, from generation in the thymus to extinction in the periphery. Our point of view is to consider the stochastic dynamics of the number of T cells of one clonotype as a function of time, which increases or decreases according to the rates of cell division and death. Any one clonotype’s time to extinction is a random variable that depends on its cross-reactivity, on the properties of those clonotypes with which it shares self-pMHC stimuli, and on average properties of the whole naive T cell repertoire. The model presented here has both interclonal and intraclonal competition, in which extinction of a clonotype is irreversible. By considering distributions of times to clonal extinction, we give estimates of clonal sizes, and hence diversity, in adult mice and humans. With the estimate that the ratio of thymic production to peripheral division of naive CD4\(^+\) T cells is four percent, our model predicts that the number of distinct clonotypes in the human body is nine percent of the total number of naive CD4\(^+\) T cells.

Our scheme is consistent with the idea that homeostasis of naive T cell numbers is maintained by availability of cytokines, especially IL-7 (Seddon et al., 2003; Min et al., 2004; Fry and Mackall, 2005; Tan et al., 2001; Pearson et al., 2011; Palmer et al., 2011; Koenen et al., 2013), that are available to T cells of all clonotypes (Ciupé et al., 2009), whereas the diversity of T cells is dependent on the varied self pMHC presented by the body’s antigen presenting cells; each self pMHC being recognised only by T cells from a small fraction of clonotypes (Mahajan et al., 2005; Ciupé et al., 2009; Singh et al., 2012). It is possible to produce deterministic models of competition between clonotypes for different resources, with a biadjacency matrix, based on one ODE for

![Fig. 2. Predicted steady-state number of distinct clonotypes. The value of \( N^* \) is the product of the rate of production of new clonotypes, \( \theta \), and the mean lifetime of a clonotype, \( q_0 \). The dotted lines are (7); valid in the weak-thymus limit. The parameter \( \alpha = \frac{n_0 \theta}{\gamma M} \) measures the strength of thymic production relative to peripheral division. Three values of \( n_0 \) are shown. We use \( \gamma M/\mu = 10^{11} \) cells, approximately equal to the number of naive CD4\(^+\) T cells in a human.](image-url)
each clonotype (De Boer and Perelson, 1997). Stochastic models have the virtue of being able to include chance, as well as systematic competition effects, of being accurate when the numbers of cells of any type are small, and of describing extinction of clonotypes without extra assumptions.

Producing a large-scale competition model with the simplest assumptions and minimum number of parameters comes, of course, at the cost of simplifying the dynamics of homeostasis of the immune system. There are a number of ways that the simple model used here can be changed to more closely resemble reality. One that we have already explored is to relax the assumption that all self-pMHC complexes provide stimuli that produce cell division with the same rate. Instead, we assigned strengths to each drawn from a log-normal distribution. Even when the variance was larger than the mean, the results were similar to the constant-γ case. We believe that it is desirable to include interclonal and intraclonal heterogeneity (Sinclair et al., 2011; Mandl et al., 2013) to allow exploration of post-thymic maturation (Thomas-Vaslin et al., 2008), of the effects of cell-to-cell differences manifested in variation of CD25, IL-7R and CD5 expression levels, and differentiation into distinct phenotypes at the single-cell level (Zehn et al., 2012; Hataye et al., 2006; Jenkins et al., 2008; Pepper and Jenkins, 2011). Changes over time leading to increase or decrease of division rates in some clonotypes (Johnson et al., 2012) may also occur. Our numerical realisations have used the exact Gillespie algorithm, without resort to approximations based on averaging over populations. In order to undertake direct larger-scale simulations, with numbers of cells closer to physiological values, methods that do not update the Λi at every step, or efficient approximations such as “tau-leaping” (Márquez-Lago and Burrage, 2007), will be appropriate.

Although the model is designed to describe the long-term dynamics associated with naïve T cell homeostasis, we envisage including strong but short-term stimuli (Hershberg et al., 2003) in the same framework to understand the effect of foreign antigen, the breadth of the immune response to a mutating virus (van Deutekom et al., 2013), and the creation of populations of memory cells, capable of persisting without TCR signals (Seddon and Zamoycka, 2003; Sanusov et al., 2006; Sprent and Surh, 2011). It is also necessary to consider the circulation of T cells around the body (Sanusov and Auerbach, 2014) and the distribution of different T cell subsets in different organs (Thome et al., 2014; Farber et al., 2014).

Dramatic illustrations of the consequences of perturbing homeostatic processes in the peripheral immune system are found in clinical or experimentally-induced lymphopenic environments (Bosco et al., 2005; Hogan et al., 2013; Paul et al., 2013). We can reproduce the reconstitution of the repertoire by thymic output and peripheral division with the model as presented here, but the constant-death-rate assumption will need to be modified to take into account the abundance of trophic factors during recovery from lymphopenia and heterogeneity in TCR and IL-7R expression levels (Palmer et al., 2011; Singh et al., 2012). Another interesting scenario is that of post-thymic transplant dynamics in infants with DiGeorge anomaly, where thymic production is explicitly a function of time (Freitas and Rocha, 2000; Houston et al., 2011; Ciupé et al., 2009).

4. Methods

4.1. The stochastic algorithm

At each step of the Gillespie algorithm (Renshaw, 2011; Wilkinson, 2006), one of the 2N(t)+1 possible events is chosen. The event is either death of a cell in some clonotype, i, chosen with probability ρm(t,i)/S(t), birth of a cell in clonotype l, chosen with probability Λi(t)/S(t), or thymic production of a new clonotype, chosen with probability θ/S(t). Here, S(t) is the sum of the rates at time t:

\[ S(t) = \sum_{i} \left( \mu_i(t) + \Lambda_i(t) + \theta \right). \]

At the end of the step, time is incremented by an amount sampled from the exponential distribution with mean 1/S(t) (Wilkinson, 2006).

4.2. Competition

In Fig. 3, we display results from one small-scale numerical realisation, with θ = 0 for simplicity. The effect of competition for stimulus is to drive some clonotypes to extinction, reducing the number of surviving clonotypes (Molina-París et al., 2011). The parameters are \( p = 0.1 \) and \( q = 10 \) for every \( q \). With the choice \( \mu = 1 \), one time unit is the mean lifetime of a T cell in the absence of stimulus (Borghans and De Boer, 2007; den Braber et al., 2012; De Boer and Perelson, 2013). In this work, when we consider the cases of T cell homeostasis in humans and mice, we shall use the values \( \mu = 1 \) month\(^{-1} \) (mouse) (Thomas-Vaslin et al., 2008) and \( \mu = 1 \) year\(^{-1} \) (human) (Macallan et al., 2004).

The relationship between the set of clonotypes and the set of self pMHCs is illustrated in Figs. 1 and 4. Consider the following quantities (Stirk et al., 2008, 2010; Molina-París et al., 2011, 2012):

- The set of self pMHCs that are recognised by T cells of clonotype \( i \) is denoted by \( Q_i \), and the number of self pMHC subsets in \( Q_i \) by \( \phi_i \).
- The set of T cells that recognise a self pMHC \( q \) is denoted by \( C_q(t) \), and the number of surviving clonotypes in \( C_q \) by \( |C_q| \).
- The average of \( \phi_i \) over all surviving clonotypes is denoted \( \bar{\phi}(t) \), and the average of \( |C_q| \) over all \( q \) is denoted \( \bar{C}(t) \) (Stirk et al., 2010).

An important exact relationship is (Mason, 1998; Zarnitsyna et al., 2013):

\[ MC(t) = N(t)\bar{\phi}(t). \]

![Fig. 3](image-url) A numerical solution of the competition model without thymic production. The left panel shows the total number of T cells, as a function of time. (The total number of cells continues to fluctuate about the same mean for times later than shown.) The right panel shows the number of surviving clonotypes, \( N(t) \), with a logarithmic time scale. The dotted line is (14). One time unit is the mean lifetime of a T cell in the absence of division stimulus, taken to be one year in a human body. The parameters are \( \mu = 1.0, \gamma = 10, M = 1000 \); T cell clonotype-pMHC connections were assigned randomly with probability \( p = 0.1 \) at the beginning of the run. The initial conditions are \( N(0) = 200 \) and \( n_i(0) = 25 \) for each \( i \).
which holds because the LHS and RHS are, in the representation of Fig. 1, different ways of counting the total number of connections between T cell clonotypes and self pMHCs. In our model, $M$ is constant in time; the remaining three quantities in (8) vary in time but the exact equality is maintained.

There are many ways to assign connections between clonotypes and self pMHCs. A simple algorithm, used in this work, is to assign all connections independently with probability $p$. The values of $\phi_i$ are thus samples from the binomial distribution with mean $pM$, and $\bar{\gamma}(0) = pM$.

For sufficiently large $pM$, the distribution of values of $\phi_i$ is approximately Gaussian with standard deviation $\sqrt{pM}$. Despite its simplicity, the assignment of connections to clonotypes reproduces the basic features of positive and negative thymic selection: all clonotypes recognise some self pMHC, but none are too specific or too promiscuous.

Fig. 5 illustrates the effect of extinction of clonotypes on mean quantities, averaged over the repertoire. Extinction increases $\bar{\gamma}(t)$, which is an average over the surviving clonotypes, but the change is necessarily confined within the distribution of values of $\phi_i$ (mean $pM$, standard deviation $\sqrt{pM}$). On the other hand, the average number of distinct clonotypes recognising a given self pMHC, $\overline{\gamma}(t)$, is reduced because of extinction of clonotypes. As time increases, in the absence of thymic input, $N(t)$ decreases, $\bar{\gamma}(t)$ is nearly constant, and $\overline{\gamma}(t)$ decreases in such a way that the equality (8) is maintained.

### 4.3. Timescales of clonal extinction

Our point of view is to consider the stochastic dynamics of the number of $T$ cells of one clonotype as a function of time. Using the description of the dynamics of a typical clonotype in the periphery, we can find an expression for the distribution of its time to extinction. Let the extinction time of the $i$th clonotype be the random variable $\tau_i$. The number of cells, $n_i(t)$, of the clonotype with label $i$ is an integer, increasing or decreasing by one cell at a time, according to the rates for cell division ("birth") and death; thus $\tau_i$, the smallest value of $t$ such that $n_i(t) = 0$. The death rate is $\mu n_i(t)$, independent of the other clonotypes in the system. In contrast, the division rate, $\Lambda_i(t)$, depends on the number of cells of all clonotypes that compete for any of the self pMHC recognised by clonotype $i$.

If we begin with the expression (2), and replace the sums by their mean values so that $c_i(t) \rightarrow px^e$ and $\sum_i \gamma_i A_{ij} \rightarrow pM$, then

$$A_i(t) = n_i(t)^p M \overline{x}^{e \times \gamma} = \frac{1}{1+\alpha} \mu n_i(t).$$

The most remarkable feature of (9) is that the birth rate $A_i(t)$ takes a form that is similar to the death rate, $\mu n_i(t)$. In the absence of thymic production,

$$A_i(t) \approx \mu n_i(t).$$

Although it is not surprising that death rates and birth rates are in balance for a typical clonotype established in the periphery, it is remarkable that they can take identical functional forms.

To derive an expression for the distribution of $\tau_i$, we use the diffusion approximation, where $n_i(t)$ is replaced by a diffusion process on the real line, $X_i$ (Taylor and Karlin, 1998; Renshaw, 2011). The equation of motion of $X_i$ is a stochastic differential equation (SDE) that is characterised by the mean and mean-square of its increments over a small time interval $\Delta t$ (Gardiner, 2004; Wilkinson, 2006). Thus we must consider the mean change in the size of the population $n_i$ between $t$ and $t+\Delta t$:

$$\Delta n_i = \mu n_i(t + \Delta t) - \mu n_i(t).$$

We require

$$\Delta n_i = \mu n_i(t + \Delta t) - \mu n_i(t) = 2 + \alpha \mu X_i \Delta t.$$}

Hence the stochastic differential equation is

$$dX_i = -\frac{\alpha}{1+\alpha} dt + \frac{2+\alpha}{1+\alpha} \mu X_i dW_i,$$

with initial condition $X_0 = b$.

A simple expression is found when $\alpha = 0$: the distribution of times at which the diffusion process reaches the absorbing state at 0 is Revuz and Yor (2004):

$$\gamma[|X_t = 0| X_0 = b] = \exp \left( -\frac{b}{\mu t} \right).$$

The appropriate initial condition $b$ is the mean total number of cells, $\gamma M / \mu$, divided by the initial number of clonotypes, $N(0)$. That is, the probability that a randomly-chosen clonotype is extinct.
before time $t$, without thymic production, is

$$P[\text{extinction}] = \exp \left( -\frac{\gamma M}{N(0)} \mu t\right). \quad (14)$$

The relation (14) is used to plot the dotted lines in Fig. 3.

Now let us consider $0 < \alpha < 1$. That is, we consider non-zero thymic production, but where more cells are created by peripheral division than by the thymus. Then

$$dX_i = -\alpha \mu X_i dX_i + \sqrt{2\mu X_i} dW_i. \quad (15)$$

The mean lifetime of a clonotype is the mean time until $X_i$ reaches zero, starting at $b$, $T(\alpha, b)$. It satisfies

$$\mu b \frac{d^2}{db^2} T(\alpha, b) - \alpha \mu b \frac{d}{db} T(\alpha, b) = -1.$$  

The solution, with initial number of cells $n_0$, is

$$T(\alpha, n_0) = \frac{1}{\omega_n} \left( e^{\omega_n b} - e^{\omega_n} \right), \quad (16)$$

where $\gamma = 0.577\ldots$ is the Euler–Mascheroni constant. The expression (7) is obtained by using the properties of the exponential integral, $E_n$, for small values of its argument: as $\alpha n_0 \to 0$, $E_n(-\alpha n_0) 
- \gamma + \log(\alpha n_0) - \alpha n_0$ and so $T(\alpha, n_0) \to \frac{1}{\omega_n} (1 - \gamma - \log(\alpha n_0))$.

4.4. Selective or stochastic competition

Is the competition selective or stochastic? That is, do some clonotypes survive and other die because they are better suited to the competitive environment, or is it simply a matter of chance? It is possible to answer this question by carrying out a series of independent realisations, each with the same matrix $A$, encoding the pattern of connections between self pMHC and T cell clonotypes. The realisations correspond to independent series of stochastic events occurring to identical in silico individuals. The results of ten such realisations are summarised in Fig. 6. About five per cent of clonotypes survive, which could be achieved in two ways. If survival is determined by chance, any one clonotype will survive any one realisation with the same probability (here, about 0.05). If survival is determined by selection, a small subset of clonotypes will survive in most realisations (De Boer and Perelson, 1997). The presence of several near-complete vertical lines, in the subset of clonotypes shown, indicates that the competition is selective. Note, too, the correlation between higher values of $\phi_i$ and survival frequency.

While it is straightforward to perform computational studies of competition between hundreds or even thousands of T cell clonotypes, it is more challenging to study physiological numbers of clonotypes. It is therefore important to understand how the system can be scaled in such a way that the dynamics of a large system, viewed from any one clonotype, is independent of system size. In the random-connection case, where any one of $N$ T cell clonotypes recognises any one of $M$ self pMHCs with probability $p$, the number recognised by one T cell clonotype has a binomial distribution with parameter $pM$. We generate ever-larger systems, maintaining an invariance from the perspective of any one clonotype, by increasing $M$ but keeping the product $pM$ constant. The result, as illustrated in Fig. 7 with a set of three numerical runs, scales in the sense that the dynamics of the quantity $n(t)$ is preserved. (The product $pM$ is the mean number of self pMHC recognised by a single TCR clonotype; the product $pN$ is the mean number of clonotypes that recognise a single self pMHC.)

4.5. Thymic production

Three numerical realisations are summarised in Fig. 8. As $\theta n_0 < e M$ in all three, thymic production has almost no perceptible effect on the mean total number of cells, $x(t)$. Even so, it is the strength of thymic production that determines the stationary number of surviving clonotypes. With thymic production included in the system, a bona fide steady state is established, where new clonotypes from the thymus balance those that are driven to extinction. The steady state is established on a long timescale, corresponding to decades in humans. The stationary distribution of clonal sizes (that is, the distribution of numbers of cells in surviving clonotypes) and density of times to extinction, are shown on the right for the case $\theta = 10$ year$^{-1}$. In this work we have restricted our attention to homeostasis in adults. In future work we plan to extend the model to consider the full human lifetime, taking into account changes in thymic output and body size during childhood (Bains et al., 2009; Johnson et al., 2012).
In Fig. 9, we display results from a version of the model with an extra element of structural stochasticity. Instead of a constant stimulus rate, each self pMHC is assigned a strength $\gamma_q$ drawn from a log-normal distribution. The parameter $\sigma$ is the standard deviation of the distribution, divided by the mean. Even with the widest distribution of rates, there is little discernible effect on the repertoire dynamics.

**Fig. 8.** Input of new clonotypes from the thymus, while barely affecting the total number of cells, determines the late-time number of surviving clonotypes. For comparison with adult human homeostasis, time is measured in years. The thymic output rates are $\theta = 0.1 \text{ year}^{-1}$ (blue), $\theta = 1 \text{ year}^{-1}$ (green), $\theta = 10 \text{ year}^{-1}$ (red), $\theta = 100 \text{ year}^{-1}$ (black) and $\theta = 400 \text{ year}^{-1}$ (cyan). The remaining parameter values are $M=4000$, $pM=100$, $\gamma=10 \text{ year}^{-1}$, $\mu=1 \text{ year}^{-1}$, $n_\theta=4$ in all cases. The histograms on the right show, for $\theta = 10 \text{ year}^{-1}$, the distribution of survival times of clonotypes from the thymus and the distribution of $n_i(t)$ for surviving clonotypes. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

**Fig. 9.** Total number of cells, and total number of surviving clonotypes, as a function of time. Stimulus rates of the self pMHC, $\gamma_q$, are drawn from a log-normal distribution. The values of $\sigma$ are the ratio of the standard deviation to the mean value of $10 \text{ year}^{-1}$. In the case $\sigma = 1$, the standard deviation is equal to the mean. The remaining parameter values are $M=4000$, $pM=100$, $\theta = 400 \text{ year}^{-1}$, $\mu=1 \text{ year}^{-1}$ and $n_\theta=4$.

In Fig. 9, we display results from a version of the model with an extra element of structural stochasticity. Instead of a constant stimulus rate, each self pMHC is assigned a strength $\gamma_q$ drawn from a log-normal distribution. The parameter $\sigma$ is the standard deviation of the distribution, divided by the mean. Even with the widest distribution of rates, there is little discernible effect on the repertoire dynamics.

**Acknowledgements**

We are grateful for European Union funding FP7-PEOPLE-2012-IRSES 317893 “Mathematics for Health and Disease”, and for the hospitality of the Max Planck Institute for Complex Systems, Dresden, and the International Centre for Mathematical Sciences,


