

Diversity of the Immune system

Handout for the *Immunobiology* lecture, May 8th, 2018

Rob J. de Boer

Objectives of this exercise:

1. See that immune systems is diverse because there are so many self antigens, and not because there are so many pathogens
2. Learn to use simple probabilistic models to ask evolutionary questions about the size of the repertoire
3. See that number of MHC molecules per host is not optimal, even though several authors claim the opposite

Background

Diversity is a hallmark of the immune system. The repertoires of B cells and of CD4⁺ and CD8⁺ T cells each consist of more than 10⁸ different clonotypes each characterized by a unique receptor Qi *et al.* (2014); Goncalves *et al.* (2017). Each immune response is characterized by a large panel of different cytokines with –partly overlapping– functions. Each individual is characterized by a unique combination of MHC molecules that play an essential role in the selection of peptides presented to the cellular immune system. MHC loci are the most polymorphic genes known for vertebrates, i.e., for most loci several hundreds of alleles have been identified within each population. However, each individual inherits only a limited number of MHC genes from its parents, and expresses about 10 different MHC molecules. We will here address the evolutionary questions why lymphocytes are so diverse within an individual, and why MHC molecules are diverse at the population level, and not diverse within an individual.

The consensus explanation for the enormous diversity of lymphocyte repertoires is the improved recognition of many different pathogens. The consensus explanation for the limited diversity of MHC molecules within an individual is the excessive negative deletion by self tolerance processes when the number of presented self peptides is increased by increasing the diversity of MHC molecules.

Tolerance

We start with a simple toy model revealing some novel expectations for the relationships between lymphocyte specificity p , the number of self epitopes S , and the initial repertoire size R_0 (De Boer and Perelson, 1993). Defining the lymphocyte specificity p as the probability that a lymphocyte responds to a randomly chosen epitope, we have a definition that remains close to the conventional concept of the “precursor frequency” of an epitope. A typical viral epitope activates about one in 10⁵ naive CD8⁺ T cells (Blattman *et al.*, 2002; Su *et al.*, 2013; Kotturi *et al.*, 2008); but see Goncalves *et al.* (2017). This means that the probability that a lymphocyte recognizes a randomly chosen epitope is about $p = 10^{-5}$. It is difficult to estimate the number of self epitopes in general. For the peptides of nine amino acids (9-mers) that are used as epitopes by CD8⁺ T cells, we have made an estimate by enumerating all unique 9-mers in the human genome (Burroughs *et al.*, 2004). Given that there are approximately 10⁷ unique 9-mers in the human self, and that MHC molecules typically present about 1% of these, we would have an estimate of $S = 10^5$ self epitopes per T cell restricted to one particular MHC (Burroughs *et al.*, 2004). Fortunately, for the arguments presented here the precise number of self epitopes turns out to be unimportant, we only need to know that it is large. The diversity of the repertoire before tolerization, R_0 , is also a large number. Because the size of the functional CD4⁺ T repertoire R in man is at least 10⁸ different receptors (Qi *et al.*, 2014), the diversity of the pre-tolerance repertoire is probably several orders of magnitude higher, i.e., $R_0 > 10^{11}$.

Having these concepts at hand we write a simple mathematical model. The diversity of the functional repertoire R is determined by the chance that each clonotype fails to recognize all self epitopes S , i.e.,

$$R = R_0(1 - p)^S . \quad (1)$$

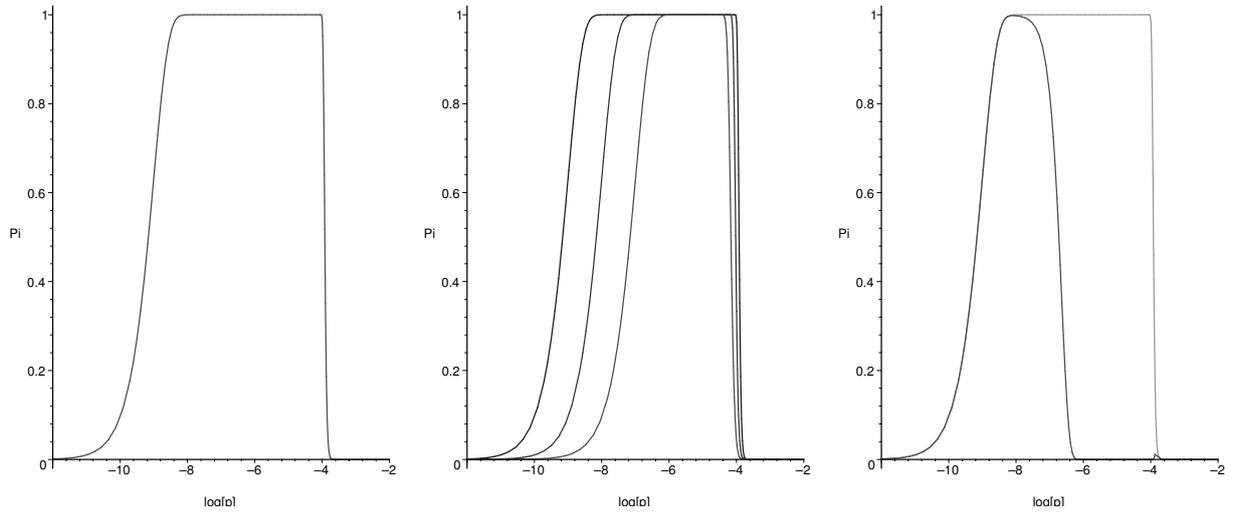


Figure 1: The probability of mounting an immune response P_i from Eq. (2) as a function of the specificity p of the lymphocytes. Parameters $S = 10^5$ and $R_0 = 10^9$. Panel (b) depicts the effect of decreasing the initial repertoire size from $R_0 = 10^9$, $R_0 = 10^8$, to $R_0 = 10^7$. Panel (c) depicts the effect of incomplete tolerance induction, i.e., $f = 1$ and $f = 0.8$ in Eq. (6).

Similarly, the chance that an individual fails to respond to a foreign epitope is the probability that none of its clonotypes in the functional repertoire R recognize the epitope. Expressing one minus the chance of failure as the probability of mounting an immune response to a foreign epitope we obtain

$$P_i = 1 - (1 - p)^R = 1 - (1 - p)^{R_0(1-p)^S} . \quad (2)$$

Plotting P_i as a function of the lymphocyte specificity p gives Fig. 1a which has a very wide region of specificities where the chance of mounting a successful immune response is close to one. If lymphocytes are too specific, i.e., at the left, epitopes remain unrecognized. If they are too cross-reactive, too many clonotypes are deleted by self tolerance processes, and the functional repertoire becomes too small.

Because $(1 - x)^n \simeq e^{-xn}$ whenever $x \ll 1$, we can approximate this model by

$$R \simeq R_0 e^{-pS} \quad \text{and} \quad P_i \simeq 1 - e^{-pR} = 1 - e^{-pR_0 e^{-pS}} . \quad (3)$$

When plotted for the same parameters as those of Fig. 1 the approximation is indistinguishable from the original curve (not shown). The approximation allows us to compute the ‘‘optimal’’ value of P_i by taking the derivative $\partial_p P_i$ of Eq. (3) and solving $\partial_p P_i = 0$ to find that the maximum is at $\hat{p} = 1/S$. This optimum suggests that the lymphocyte specificity is largely determined by the number of self epitopes the immune system has to be tolerant to. Thus, the specificity is not determined by the recognition of pathogens, but by the demand to remain tolerant to a large number of self epitopes. Once lymphocytes are specific, the repertoire has to be sufficiently diverse to guarantee recognition of foreign epitopes (Fig. 1b).

Incomplete tolerance

Although there is promiscuous expression of self antigens in the thymus, it remains unlikely that self tolerance is complete. Healthy individuals do harbor lymphocytes that can recognize self epitopes (Danke *et al.*, 2004; Su *et al.*, 2013; Malhotra *et al.*, 2016). To study how the results change when tolerance is incomplete we define a new parameter f for the fraction of self epitopes that manage to induce tolerance. For $f = 1$ the new model should be identical to the previous one. For a foreign epitope we now require that it is recognized, but that none of the clonotypes recognizing the foreign epitope also recognizes any of the $(1 - f)S$ self epitopes that fail to induce tolerance. Otherwise the clone will be held responsible for auto-immunity. Following Borghans *et al.* (1999) we let α be the fraction of clonotypes recognizing at least one ignored self epitope, i.e.,

$$\alpha = 1 - (1 - p)^{(1-f)S} . \quad (4)$$

The chance that the system remains tolerant when stimulated with a foreign epitope is the probability that none of the clones in the functional repertoire R will respond (with chance p) and is potentially auto-reactive (with chance α), i.e.,

$$P_t = (1 - p\alpha)^R \quad \text{where} \quad R = R_0(1 - p)^{fS}. \quad (5)$$

Now the chance of a “successful immune response” is the probability that the system remains tolerant and responds to the foreign epitope, which is the chance to remain tolerant minus the chance to not respond at all:

$$P_s = P_t - (1 - p)^R, \quad (6)$$

where the functional repertoire R is given in Eq. (5). To study how incomplete tolerance affects the results we plot Eq. (6) for $f = 0.8$ and $f = 1$ in Fig. 1c.

Fig. 1c demonstrates that the effect of incomplete tolerance is enormous. The region of specificity values where the chance of a successful response P_s approaches one is much narrower. Moreover the optimum has shifted leftwards, i.e., towards a specificity much smaller than $p = 1/S$. Thus the $p = 1/S$ estimate (De Boer and Perelson, 1993) is an upper bound for the lymphocyte crossreactivity: when the initial repertoire is sufficiently large the immune system operates even better when lymphocytes are more specific (Borghans *et al.*, 1999). The conclusion remains that lymphocytes are specific to avoid auto-immunity, and not to recognize many pathogens.

Regulatory T cells

For additional documentation you could read the paper by Saeki *et al.* (2015) who extend this type of models with regulatory T cells. It is at least interesting to see how complicated these simple models become.

MHC diversity within the individual

Since individual MHC diversity increases the presentation of pathogens to the immune system, one may wonder why the number of MHC genes is not much higher than it is. The argument that is mostly invoked is that more MHC diversity within the individual would lead to T cell repertoire depletion during self tolerance induction (Nowak *et al.*, 1992). This argument is incomplete, however, because more MHC diversity could also increase the number of clones in the T cell repertoire through positive selection. In order to be rescued in the thymus, lymphocytes need to recognize MHC–self peptide complexes with sufficient avidity. A high MHC diversity thus increases both the number of lymphocyte clones that are positively selected and the number of clones that are negatively selected. To calculate the net effect of these two opposing processes we need a mathematical model (Borghans *et al.*, 2003).

Consider an individual with M different MHC molecules and an initial T lymphocyte repertoire consisting of R_0 different clones. Let p and n denote the (unconditional) chances that a clone is positively selected by a single MHC type, because its avidity is higher than a threshold T_1 , or negatively selected because its avidity exceeds a higher threshold T_2 , respectively (see Fig. 2). By this definition, thymocytes can only be negatively selected by MHC molecules by which they are also positively selected, i.e., $n < p$. Since T cell clones need to be positively selected by *at least one* of the MHC molecules, and avoid negative selection by *all* of the MHC molecules, the number of clones in the functional repertoire R can be expressed as

$$R = R_0 \left((1 - n)^M - (1 - p)^M \right), \quad (7)$$

(Borghans *et al.*, 2003). The functional repertoire R thus contains all T cell clones that are not negatively selected, minus the ones that fail to be positively selected by any of the M different MHC molecules of the host.

Experimental estimates for the parameters of this model are difficult to obtain. In mice, around 3% of the T cells produced in the thymus end up in the mature T cell repertoire Goncalves *et al.* (2017), and at least 50% of all *positively selected* T cells have been shown to undergo negative selection in the thymus (Van Meerwijk *et al.*, 1997; Merckenschlager *et al.*, 1997). Thus, in the absence of negative selection there would have been 6%

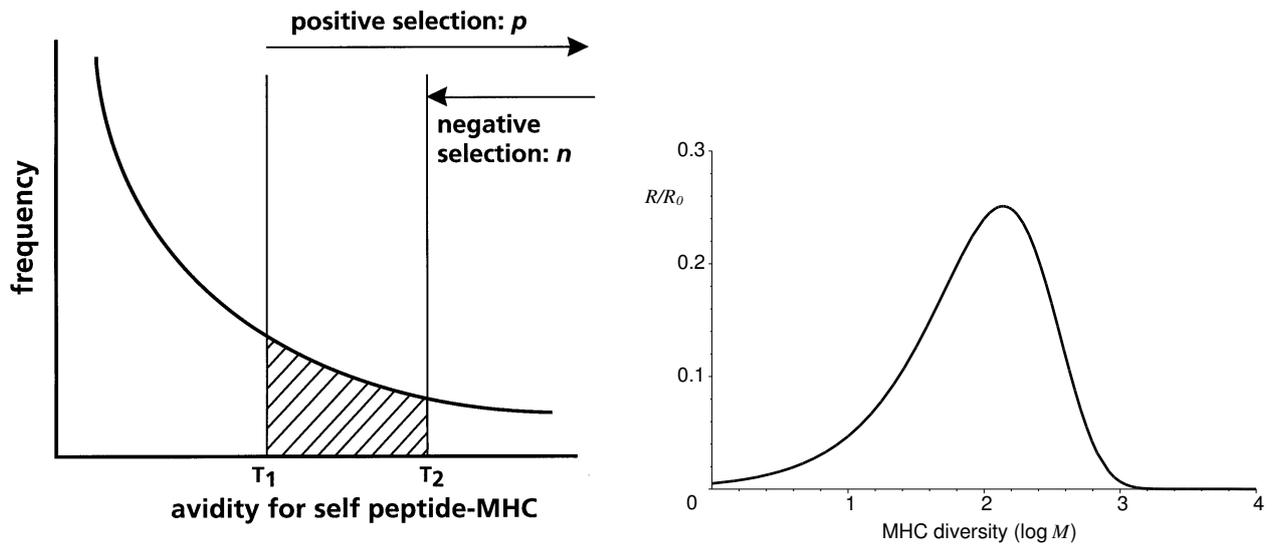


Figure 2: Positive and negative selection according to the avidity model (Janeway and Katz, 1984). The curve in (a) depicts the distribution of thymocyte avidities for self peptide–MHC complexes. In our model, the chance p to be positively selected by a single MHC type is the chance that the avidity between the thymocyte T cell receptor and any of the self peptide–MHC complexes exceeds threshold T_1 . Thymocytes with avidities for self peptide–MHC complexes exceeding the upper threshold T_2 are negatively selected (with chance n per MHC type). Panel (b) depicts the size of the T cell repertoire as a function of MHC diversity as defined by Eq. (7). The number of clones in the functional repertoire R is plotted as a fraction of the total initial lymphocyte repertoire R_0 . Parameters are: $p = 0.01$, and $n = 0.005$.

survival, which means that 94% of the thymic T cells fail to be positively selected by any of the MHC molecules in the host, and that 3% of the thymocytes are negatively selected. These two estimates can be used to calculate the chances p and n of a T cell clone to be positively or negatively selected by a single type of MHC molecule. Taking into account that inbred mice are homozygous and therefore express 3 types of class I MHC and 3 types of class II MHC molecules, one estimates p from $0.94 = (1 - p)^6$ and n from $0.97 = (1 - n)^6$, which yields $p \approx 0.01$ and $n \approx 0.005$.

Using these experimental estimates, the number of clones in the functional T cell repertoire R increases with the number of different MHC molecules M in an individual until $M = 140$ (see Fig. 2b). In other words, the size of the functional T cell repertoire would increase if the MHC diversity M were to exceed its normal value of ten to twenty in heterozygous individuals. The intuitive reason is that only a very small part of the T cell repertoire has sufficient avidity for self peptides presented by a single MHC type to be positively selected by that MHC. As long as additional MHC types positively select hardly overlapping parts of the T cell repertoire, negative selection will only waste T cells that were not even positively selected in the absence of those MHC molecules. A net negative effect of MHC diversity on the size of the functional T cell repertoire is only attained once the individual MHC diversity is so large that thymocytes are selected by multiple MHC types, *i.e.* when $M > 140$. Summarizing, the consensus explanation that the MHC diversity per individual is limited to avoid repertoire depletion is untenable.

Computer Lab Exercises

Today we will plot various functions using “R”, which is a language with which one can easily do statistics, plot functions, and fit mathematical models to data. R is installed on the computers that we use, and the R-code for the functions that you need today is available as the file [Pi.R](#) on the webpage tbb.bio.uu.nl/rdb/practicals/diversity. Open this file in RStudio and spend some time to familiarize yourself with the R-functions in this file before making the exercises. The easiest way to work with RStudio is to highlight some text in the `Pi.R` window, and then hit the Run button (or use the shortcut control Enter) to have it executed in the Console window.

For instance, highlight the `Pi` function in the `Pi.R` file, hit the Run button, and then type `Pi` in the Console window. R then gives the definition of the function. Next click the `curve(Pi(logp=x), from=-10, to=0)` line and Run it. This pops up a graphics window plotting P_i for its default parameters ($R_0 = 10^9, S = 10^5$) as a function of $\log_{10}[p]$. If you want to plot this function for another value of R_0 , just use the up-arrow to retrieve the previous command in the Console, and use the left-arrow to override the R_0 parameter of the function, e.g., insert `r0=1e8` to obtain `curve(Pi(r0=1e8, logp=x), from=-10, to=0)`. Actually, the next line in the `Pi.R` file does this for you, while using the `add=TRUE` option to add the curve to the same graph.

If you decide to write a report on this computer practical, please elaborate the questions on the optimal number of MHC molecules (Questions 3 & 4). Read the papers that we cite, try to find even better and more recent estimates for the probabilities of positive and negative selection, make sure you correctly address all the answers to the questions given below, and try to add some original results.

Exercise 1 Probability of response

Plot Eq. (2) for several values of S and R_0 using R.

- How do the results explained above depend to the precise values of S and R_0 ? What do you learn from this?
- Phrase in your own words why the immune repertoire should be so diverse.
- What diversity do you expect in other species with a different size of self?
- Suppose the adaptive immune system evolved in a species with about $S = 10^5$ self antigens. How diverse should this first immune system be to provide a reasonable selective advantage? So how do you think the adaptive immune system evolved?
- What is the optimum if one has to survive a 100 pathogens? Does this narrow down the range of “good” immune systems? Note that `curve(Pi(logp=x)^10, from=-10, to=0)` depicts P_i^{10} as a function of $\log_{10}[p]$.
- Since every pathogen consists of a large number of epitopes, one could argue that the host is protected once it mounts an immune response to at least one epitope. If there are n epitopes in a typical pathogen, this probability becomes $P'_i = 1 - (1 - P_i)^n$. What is now the optimum and does this narrow down the range of “good” immune systems?
- Do you think these results depend much on the assumptions of the model?

Exercise 2 Incomplete tolerance

Plot Eq. (6) for several values of f using R.

- How different is this model from that of Eq. (3)?
- Carefully explain what you see when you change f (e.g., look at $f = 0.999$).
- Phrase in your own words how diverse you expect an immune repertoire to be.
- So how do the results of the previous model depend on its assumptions?

Exercise 3 Nowak's optimal #MHC

Nowak *et al.* (1992) also addressed the question of the optimum number of MHC molecules within an individual with a mathematical model. According to their model the functional repertoire is given by

$$R = R_0(1 - (1 - p)^M)(1 - n^*)^M, \quad (8)$$

where n^* is the *conditional* probability that a positively selected clone is negatively selected by a random MHC molecule. Using the $(1 - n^*)^6 = 0.5$ cited above (Van Meerwijk *et al.*, 1997; Merckenschlager *et al.*, 1997), one would estimate that $n^* = 0.11$.

- Discuss the interpretation of each term. What is the chance of positive selection in this model?

- b. What is wrong with this model?
- c. Study the model with R for the same parameters as used in Fig. 2b. Note that both Eq. (7) and Eq. (8) are provided in the R-file [Pi.R](#).

Exercise 4 General optimal #MHC

The curve in Fig. 2 depends on two parameters, p and n , that were estimated from the observation that 3% of the cells survive tolerance induction, and that 50% of the positively selected cells survives negative selection. Actually, there is also something wrong with this model because part of the positive selection is due to the requirement that developing thymocytes should express a functional TCR. Due the fact that only one third of the random re-arrangements leads to an in-frame receptor, and that T cells get a second change to re-arrange the TCR gene segments on the other chromosome when the first re-arrangement is not functional, the probability of a functional re-arrangement is approximately $\phi = 0.33 + (1 - 0.33) \times 0.33 \approx 0.5$. Since this is independent of the MHC, Eq. (7) changes into

$$R = \phi R_0 \left((1 - n)^M - (1 - p)^M \right), \quad (9)$$

where p is now defined as the probability that a clone expressing a functional TCR is positively selected by a single MHC type. Let us now generalize the calculation for calculating p and n by by rewriting this into

$$\rho = R/R_0 = \phi \left((1 - n)^M - (1 - p)^M \right),$$

where $\rho \approx 0.03$ is the fraction of cells surviving thymic selection. Now define α as the fraction of cells that is positively selected, i.e., that have a functional TCR and have sufficient avidity for at least one of the MHC types, and define β as the fraction of positively selected cells that are negatively selected (i.e., above we studied $\alpha = 0.06$ and $\beta = 0.5$). Because $\rho = \alpha(1 - \beta)$ or $\alpha = \frac{\rho}{1 - \beta}$, and we know that $\alpha = \phi \left(1 - (1 - p)^M \right)$ we solve p from

$$(1 - p)^M = 1 - \frac{\rho}{\phi(1 - \beta)}, \quad \text{i.e.,} \quad p = 1 - \sqrt[M]{1 - \frac{\rho}{\phi(1 - \beta)}}.$$

Next we solve n from

$$\rho = \phi \left((1 - n)^M - (1 - p)^M \right) = \phi \left((1 - n)^M - 1 + \frac{\rho}{\phi(1 - \beta)} \right) \quad \text{leading to} \quad n = 1 - \sqrt[M]{1 - \frac{\rho\beta}{\phi(1 - \beta)}}.$$

These two expressions for p and n enable us to estimate the probabilities of positive and negative selection on each MHC type from data providing the fraction of cells surviving thymic selection, ρ , and the fraction of positively selected cells that are negatively selected, β . A function solving p and n from ρ and β is available in the R-file [Pi.R](#), and this can be used to solve the following questions:

- a. What is the effect of including the fraction of functional receptors, ϕ , on the optimal number of MHC types?
- b. Suppose now that selection is less stringent and that 20% of the cells survive tolerance induction. How would this less stringent selection of $\rho = 0.2$ survival change Fig. 2?
- c. The data arguing that 50% of the positively selected cells become negatively selected are based on cell counts in the thymus, and are fairly indirect. Using new methods to identify cells that undergo negative selection, Stritesky *et al.* (2013) recently argued that the impact of negative selection is much stronger. Enumerating positively selected cells only, they estimate that there are 7×10^5 cells dying from negative selection per hour, for every 1.23×10^5 surviving thymocytes surviving per hour. Thus the fraction of positively selected cells dying from negative selection would be $\beta = 7/8.23 = 0.85$. How does this change the results? At least study $\beta = 0.85$ keeping $\rho = 0.03$.
- d. What is your favorite explanation for the fact that the diversity of MHC molecules per host is low, while their degree of polymorphism of the population is high?

Pi.R

```
#Question 1
```

```
Pi <- function(r0=1e9, s=1e5, logp=-5) {
  p <- 10^logp
```

```

  r <- r0*(1-p)^s
  return(1 - (1-p)^r)
}

curve(Pi(logp=x), from=-10, to=0)
curve(Pi(r0=1e8, logp=x), add=TRUE, col="blue")

# n epitopes per pathogen:

Pin <- function(r0=1e9, s=1e5, n=1, logp=-5) {
  p <- 10^logp
  r <- r0*(1-p)^s
  pi <- 1-(1-p)^r
  return(1 - (1-pi)^n)
}

#Question 2

Ps <- function(r0=1e9, s=1e5, f=1, logp=-5) {
  p <- 10^logp
  r <- r0*(1-p)^(f*s)
  a <- 1 - (1-p)^((1-f)*s)
  pt <- (1-p*a)^r
  return(pt - (1-p)^r)
}

curve(Ps(logp=x), from=-10, to=0, add=T, col="red")

#Question 3

# Eq. 7 for the optimal number of MHC molecules:
R <- function(r0=1, p=0.01, n=0.005, m=6) { r0*((1-n)^m-(1-p)^m) }

curve(R(m=x), from=0, to=1000)

# Nowak's MHC model:
RN <- function(r0=1, p=0.01, n=0.11, m=6) { r0*(1-(1-p)^m)*(1-n)^m }

curve(RN(m=x), from=0, to=20)

#Question 4

#Optimal number of MHC molecules:
RF <- function(rho=0.03, beta=0.5, phi=0.5, M=6, m=6) {
  p <- 1 - (1-(rho/(phi*(1-beta))))^(1/M)
  n <- 1 - (1-rho*beta/(phi*(1-beta)))^(1/M)
  return(phi*((1-n)^m-(1-p)^m))
}

curve(RF(m=x), from=0, to=200)

```

Acute Immune Responses to viruses

Handout for the *Immunobiology* lecture, May 15th, 2018

Rob J. de Boer

Objectives of this exercise:

1. Learn to think in a quantitative manner about the immune system
2. Learn to fit mathematical models to experimental data to obtain quantitative parameter estimates.
3. Identify the mechanisms underlying the immunodominance ranking of primary immune reactions.

Background

If you choose to do a project about this you could start to see how in Fig.2 of De Boer and Perelson (2013) four LCMV immune responses are fitted simultaneously their data, in to see how their parameters define the immunodominance ranking. The last part of this practical provides an example enabling your to repeat (and extend) this analysis. More challenging questions to address in a project are: How would the model change when memory cells are already formed during the expansion phase (see Kohler (2007))? Can this explain the data equally well?

Exercise 1 The LCMV primary immune response

Vigorous infections with rapidly replicating pathogenic bacteria or viruses trigger strong adaptive immune responses. A few days after the initial exposure, antigen specific naive T cells will become activated to undergo rapid clonal expansion, until the peak of the response, after which most of the activated cells die by apoptosis. The clonal expansion in mice infected with bacteria typically continues if the antigenic stimulus is removed by treating the mice with antibiotics, but the peak response is somewhat lower (Badovinac *et al.*, 2002). Similarly, the contraction after the peak will take place even if the antigen persists (Badovinac *et al.*, 2002). It has therefore been suggested that after proper antigenic stimulation the acute immune response of T cells is "programmed", and only partly regulated by the current concentration of antigen. Such a program can conveniently be modeled with piece-wise models (De Boer *et al.*, 2001, 2003; De Boer and Perelson, 2013), that can be fitted to the experimental data from acute immune responses to viruses and bacteria. Consider antigen specific activated T cells, A , and memory cells, M , e.g.,

$$\begin{cases} dA/dt = 0 & \text{and} & dM/dt = 0, & \text{if } t \leq T_{\text{on}}, \\ dA/dt = pA & \text{and} & dM/dt = 0, & \text{if } T_{\text{on}} < t \leq T_{\text{off}}, \\ dA/dt = -(d_A + r)A & \text{and} & dM/dt = rA - d_M M, & \text{otherwise,} \end{cases} \quad (10)$$

where $A(0)$ is the initial number of cells specific for the epitope of interest (and $M(0) = 0$). Kotturi *et al.* (2008) determined the affinity and the precursor frequencies of several CD8⁺ T cell responses to LCMV and report that naive B6 mice contain on average 445 precursors for the GP33 epitope, and 118 precursors for NP396, 42 precursors for the GP118, and 57 precursors for the NP205 epitopes, respectively. We will use an R-script [grind.R](#) to fit this model to the data.

The `grind.R` script is a wrapper around R-libraries developed by Karline Soetaert and colleagues (Soetaert and Herman, 2009; Soetaert and Petzoldt, 2010; Soetaert *et al.*, 2010; Soetaert, 2009). For this practical you will need three of `grind.R`'s easy-to-use functions:

- `run()` integrates a model numerically and provides a time plot,
- `fit()` fits a model to data by estimating its parameters, and depicts the result in a timeplot,
- `timePlot()` plots a data matrix.

The `run()` function calls `ode()` from the `deSolve` library, and the `fit()` function calls `modFit()` from the `FME` library. Typing `?ode` in the R-console, provides help on the options of the `ode()` library function. The code completion in RStudio will show options available for the `grind.R` functions. The `grind.R` script (and a tutorial) can be downloaded from the webpage tbb.bio.uu.nl/rdb/practicals/grindR. Download `grind.R` into a directory (folder) on your machine, and open it in RStudio. "Source" the `grind.R` file (button in right

hand top corner) to define the `grind.R` functions. If you get error messages you will probably have to install the Soetaert libraries into the R-environment, e.g., `install.packages(c("deSolve", "rootSolve", "FME"))` in the R-console, or `Install Packages` in the Tools menu of RStudio. You may also have to set the working directory to the folder where you stored all the files (Set `working directory` in the Session menu), otherwise RStudio does not read the data files.

On the webpage tbb.bio.uu.nl/rdb/practicals/lcmv you will find the `lcmv.R` example, coding the solution of Eq. (10) in R. Four text files containing the combined data of Homann *et al.* (2001) and Kotturi *et al.* (2008) are available in the files `gp33.txt`, `np396`, `gp118`, and `np205` (i.e., four epitopes). **Download all these files into a directory (folder) on your machine.** Alternatively, download the `lcmv.zip` file and unzip that into a directory (folder) on your machine. After downloading all files open the `lcmv.R`-script in RStudio. Note that the differential equations of Eq. (10) are also available as the script `lcmv_ode.R`, but solving these ODEs numerically is slow, and more cumbersome. Use this script only if you need to change the equations.

Read the first part of the `lcmv.R`-script, and check how the model of Eq. (10) is defined as its solution in the function `model()`. Note the model returns A , M , and the sum of $T = A + M$. The latter will be fitted to the data. Also note that parameters are defined in the vector `p` and the initial state in the vector `s`. First select the function `model()` by highlighting everything up to the closing curly bracket, and execute this by clicking the Run button (or typing `Control Enter`). Then slowly proceed through file by running it line-by-line (using `Control Enter`), and make sure that you understand what is happening. Make notes!

- Run the model for the initial parameter values (`run()`), read the data (`read.table()`), plot the data (`timePlot()`), and define which parameters should be estimated (`free`), using the `fit()` function. Note that we simplified the call to `fit()` by defining a function `myfit()`.
- Interpret the parameter values: how fast do cells divide, when do they start to divide, etcetera. How good do you find this fit?
- Next fit the `np396` data set. Which parameters differ between the `gp33` and `np396` responses?
- Compare the `gp33`, `np396`, `gp118` and `np205` responses. How would you explain their immunodominance ranking?
- Why do you think we fit the logarithm of the data? How do the parameter estimates change if you don't transform the data?

In the second half of the R-script we illustrate how the data of the various epitopes can be fitted simultaneously. Some of the free parameters will be shared (and are defined by the names in `free`) other free parameters can differ between the data sets (and should be defined by the names in `diff`), and finally the initial number of cells, A , is fixed per data set, and is defined as a list of the Kotturi *et al.* (2008) data.

- Fit the data sets simultaneously for various choices of the parameters that differ. Note that you can use the summed squared residuals (SSR) as a measure for the quality of the fit.
- What is the best way to explain the differences between the responses?

Project

It is a nice project to fit the four data sets simultaneously to study which parameter(s) best explain the immunodominance ranking. Because we have direct estimates of the initial number of cells, one can fix the $A(0)$ values to those observed in the data. This is achieved via the list `fixed` in `grind.R` (see the second half of the `lcmv.R` script, and the tutorial on the webpage tbb.bio.uu.nl/rdb/practicals/grindR). Study whether or not we really need to fix the initial condition. Can it also be estimated? Note that `diff` can similarly be defined as a list of initial conditions with a unique initial guess for each data set.

Finally, one could try different models. For instance, memory cells could be formed during the expansion phase (see Kohler (2007)). Because it may be hard to find an analytical solution for other models, it is probably better to use the `lcmv_ode.R` script, that defines model of Eq. (10) as a set of piece-wise differential equations. Solving these ODEs is numerically hard, and unfortunately requires exceptional tolerance, and tricks to arrest the integrator at T_{on} and T_{off} (see the `lcmv_ode.R` script), but changing that model is a lot easier.

lcmv.R

```

# Source grind.R first
# This model provides a solution. Not the derivatives!
# Using the option solution=TRUE one can run() and fit().

model <- function(t,state,parms) {
  with(as.list(c(state,parms)), {
    if (t < Ton) return(c(A,M,A+M))
    else if (t < Toff) {
      At <- A*exp(p*(t-Ton))
      return(c(At,M,At+M))
    }else{
      AToff <- A*exp(p*(Toff-Ton))
      At <- AToff*exp(-(da+r)*(t-Toff))
      Mt <- (r*AToff/(da+r-dm))*(exp(-dm*(t-Toff)) - exp(-(da+r)*(t-Toff)))
      return(c(At,M+Mt,At+Mt))
    }
  })
}

# myfit() calls fit() with optimized parameters:
# lower=0: only positive parameters
# fun=loglp: log(1+x) transformation
# ymin, ymax, log="y", tmax: parameters for plotting result:

myfit <- function(data,w=who,...) {
  return(fit(datas=data, ymin=1, ymax=1e8, log="y", tmax=100, who=w, solution=T,
    lower=0, fun=loglp, ...))
}

# Read gp33 data and plot them. Plot the initial guess of parameters

kotturi <- c(gp33=445,np396=118,gp118=42.5,np205=57)

p <- c(Ton=2,Toff=7,p=2,da=0.5,r=0.05,dm=0.001) # Define parameter values
s <- c(A=445,M=0.001,T=0) # Initial state (A=gp33)

run(tmax=100,ymin=1,ymax=1e8,log="y",solution=T) # Run the model for initial parameters

gp33 <- read.table("gp33.txt",header=TRUE) # Read the data
timePlot(gp33,log="y",draw=points) # Plot them
timePlot(gp33,log="y",tmax=50,draw=points) # Zoom in

who <- c("Ton","Toff","p","da","r","dm") # Select subset of parameters to fit

fit33 <- fit(gp33,who=who,solution=T) # Simple linear fit
fit33$par
fit33 <- fit(gp33,who=who,solution=T,fun=log) # Transform the data
fit33$par

fit33 <- myfit(gp33,who,main="gp33") # Plot on logarithmic axis
summary(fit33) # Note that dm is almost zero

p["dm"] <- 0 # Set dm=0
who <- c("Ton","Toff","p","da","r") # and remove from who
fit33 <- myfit(gp33,who,main="gp33")
summary(fit33)

```

```

p33 <- p # Copy p into p33
p33[who] <- fit33$par # Substitute estimated parameters
run(p=p33, tmax=100, ymin=1, ymax=1e8, log="y", solution=T) # Run for estimated parameters

np396 <- read.table("np396.txt", header=TRUE) # Read np396 data
s["A"] <- kotturi["np396"]; s # Fix A(0)
fit396 <- myfit(np396, who, main="np396")
summary(fit396)

gp118 <- read.table("gp118.txt", header=TRUE)
s["A"] <- kotturi["gp118"]; s
fit118 <- myfit(gp118, who, main="gp118")
summary(gp118)

np205 <- read.table("np205.txt", header=TRUE)
s["A"] <- kotturi["np205"]; s
fit205 <- myfit(np205, who, main="np205")
summary(fit205)

fit33$par; fit396$par; fit118$par; fit205$par # Summary of the four fits

# Questions:
# 1. Interpret the parameter values that you found for the gp33 response.
# How fast do cells divide, when do they start to divide, etcetera.
# 2. Which parameters differ between the gp33 and np396 responses?
# How does that explain the immunodominance of gp33?
# 3. Compare the gp33, np396, gp118 and np205 responses
# How would you explain their immunodominance ranking?
# 4. Why do you think we take the logarithm of the data?
#

#####
# Fit the data together.
# Use fixed to define which fixed parameters differ
# and use differ to define which free parameters differ
# For example:

fixed <- list(A=kotturi)
both<-myfit(list(gp33, np396), who, fixed=fixed, differ=c("Ton", "p"), main=c("gp33", "np396"))
both$par

# Questions (co-fitting):
# 5. Which parameter(s) do you really need to explain the difference
# between the gp33 and np396 responses?
# 6. Which parameters best describe the immunodominance ranking of the four responses?

```

TRECs and Telomeres

Handout for the *Immunobiology* lecture, June 5th, 2018

José Borghans and Julia Drylewicz (Immunology, UMCU)

Objectives of this exercise:

1. Understand how a simple mathematical model can help the interpretation of experimental data.
2. Understand how a biological phenomenon can be translated into a mathematical model.
3. Learn to translate unexpected modelling results into intuitive understanding.

Exercise 1: Estimating thymic output from T-cell receptor excision circle (TREC) data

Background on TRECs

TRECs are by-products of V(D)J rearrangements that occur in the thymus when the T-cell receptor (TCR) is formed. During this genetic rearrangement process, parts of the TCR genome are excised and form stable DNA circles. These TRECs are not copied when a cell divides; they are simply passed on to one of the daughter cells. TRECs are thus uniquely formed in the thymus. When the average number of TRECs per T cell (the so-called average TREC content) was measured in healthy individuals of different ages, an exponential decline was observed (see Figure 3a). Decreased TRECs per T cell in HIV-infected patients (Figure 3b) were interpreted as evidence for impaired thymic output.

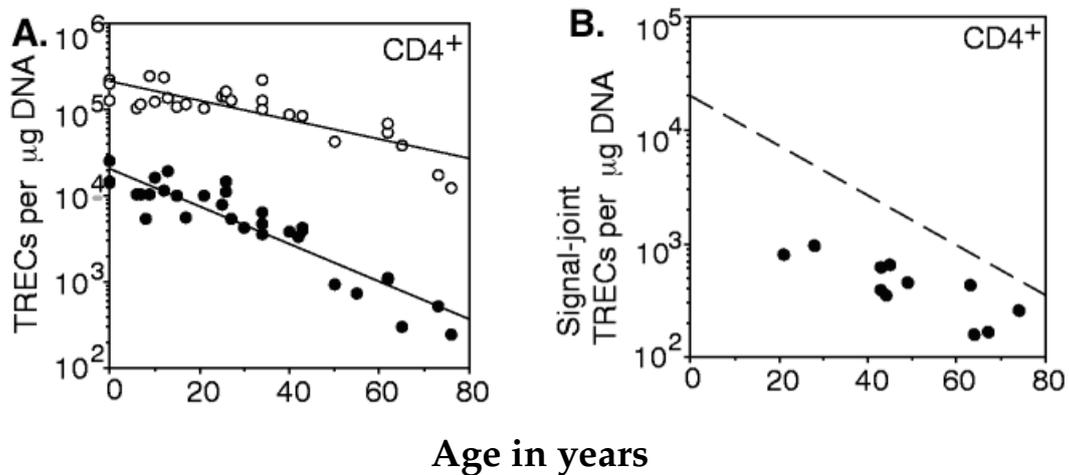


Figure 3: **A.** The average TREC content (in black) of CD4⁺ T cells from healthy individuals as a function of age (in years). From: Douek *et al.* (1998) (Discard the open circles). **B.** The average TREC content (in black dots) of CD4⁺ T cells from HIV-infected individuals of different ages in early stages of infection. The average TREC content in healthy individuals as a function of age is given by the dashed line. From: Douek *et al.* (1998)

A mathematical model for TRECs

Hazenberg *et al.* (2000) developed a mathematical model to investigate how the average TREC content of a cell population depends on thymus output, cell division and cell death rates. The number of naive T cells N was described by:

$$\frac{dN}{dt} = \sigma(t) + pN - dN \quad (11)$$

In these equations, σ represents daily thymic output, p is the average proliferation rate of naive T cells, and d represents the death rate of naive T cells. The dynamics of the total number of TRECs T was described by:

$$\frac{dT}{dt} = c\sigma(t) - dT \quad (12)$$

where c is the average TREC content of a T cell that has just left the thymus. The average TREC content of the naive T cell population (A) can subsequently be calculated from $A = T/N$. Until the age of 30 years, the model is run with the following parameters: $\sigma(t) = 10^9$ cells/day (which for the simplicity assumed to be constant), $p = 0.0004$ per day, $d = 0.0005$ per day, and $c = 0.25$. In this exercise we will use `grind.R` and the file `trec.R` with the following parameters to simulate what happens to an individual who gets HIV infected at the age of 30 years.

The `grind.R` script is a wrapper around R-libraries by developed by Karline Soetaert and colleagues (Soetaert and Herman, 2009; Soetaert and Petzoldt, 2010; Soetaert *et al.*, 2010; Soetaert, 2009). For this practical you will need just a few of `grind.R`'s easy-to-use functions:

- `run()` integrates a model numerically and provides a time plot, or a trajectory in the phase plane,
- `newton()` finds steady states (using the Newton-Raphson method),
- `timePlot()` plots a dataframe.

The `run()` function calls `ode()` from the `deSolve` library, `newton()` calls `steady()` from the `rootSolve` library. Typing `?ode` in the R-console, provides help on the options of the `ode()` library function. The code completion in RStudio will show options available for the `grind.R` functions. The `grind.R` script (and a tutorial) can be downloaded from the webpage tbb.bio.uu.nl/rdb/practicals/grindR. Download `grind.R` into a directory (folder) on your machine, and open it in RStudio. "Source" the `grind.R` file (button in right hand top corner) to define the `grind.R` functions. If you get error messages you will probably have to install the Soetaert libraries into the R-environment, e.g., `install.packages(c("deSolve", "rootSolve", "FME"))` in the R-console, or `Install Packages` in the `Tools` menu of RStudio. You may also have to set the working directory to the folder where you stored all the files (`Set working directory` in the `Session` menu), otherwise RStudio does not read the data files.

On the webpage tbb.bio.uu.nl/immbio/R you will find the `trec.R` example. Read the first part of the R-script, and check how the model is defined in the function `model()`. Also note that parameters are defined in the vector `p` and the initial state in the vector `s`. First select the function `model()` by highlighting everything up to the closing curly bracket, and execute this by clicking the `Run` button (or typing `Control Enter`). Then slowly proceed through file by running it line-by-line (using `Control Enter`), and make sure that you understand what is happening. Make notes!

Question

- Simulate what would happen with the average TREC content and the number of naive T cells in this individual during the first 2 years of HIV infection, if HIV would totally block thymic output from the age of 30 onward. Compare your results to the 10-fold decline in the average TREC content that was observed in HIV patients (Figure 3b). Can you explain intuitively what you see?
- Simulate what would happen if HIV would increase the death rate of T cells from the age of 30 onward. Can you understand this intuitively?
- Simulate what would happen if HIV would increase the rate of T-cell proliferation from the age of 30 onward.
- How would you interpret the decreased TREC content of T cells in HIV infected patients? Is the average TREC content of a T cell population a good indicator of thymic output?

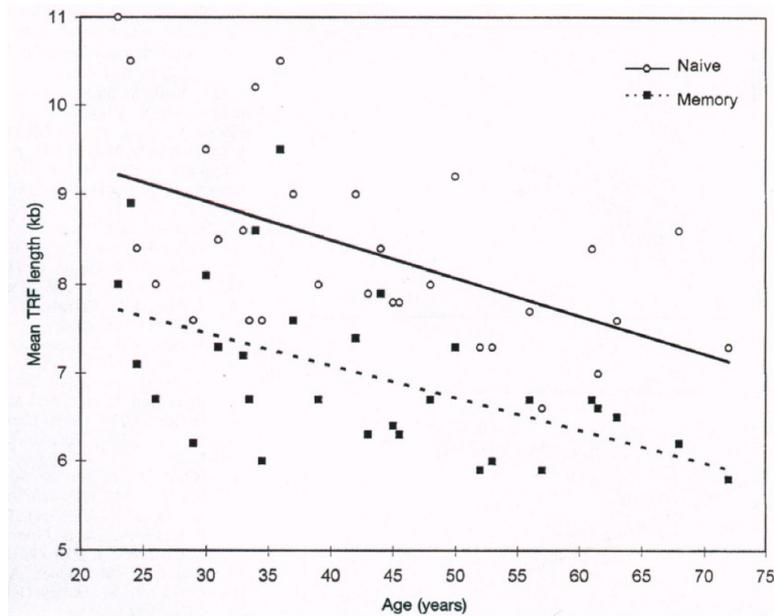


Figure 4: Comparison of the mean telomere restriction fragment (TRF) length of naive and memory T cells. Thirty donors were studied, ranging in age from 24 to 72 years old. Open circles denote the mean TRF length of each individual naive T cells; the solid line represents the regression line through the data; solid squares denote the mean TRF length of each individual memory T cells; the dotted line represents the regression line through the data. From: Weng *et al.* (1995)

Exercise 2: Estimating lymphocyte turnover from telomere data

Background

Telomeres are the unique structures at the end of chromosomes that shorten with each cell division. Each cell division leads to the loss of 50-100 terminal nucleotides from the chromosomes. Weng *et al.* (1995) made use of this phenomenon to study the replication behaviour of naive and memory T lymphocytes. The average telomere lengths of naive and memory T cells were found to decrease with age at very similar rates. The average telomere length of naive T lymphocytes was consistently found to be 1.4 kb longer than that of memory cells (see Fig. 4).

A mathematical model for telomere shortening

In order to study how the average telomere length of a population of cells changes with age, De Boer and Noest (1998) developed a mathematical model. The model keeps track of the division history of a population of naive T cells, N , and memory T cells, M , each subdivided in subpopulations of cells that have undergone i cell divisions (see Figure 5). For example, the naive cells in box n_0 have not divided yet, while the memory cells in box m_5 have undergone 5 divisions. The arrows p_N , p_M , d_N and d_M represent proliferation (= cell division) and death rates of naive and memory T cells, respectively. The arrow γC represents the influx of naive cells into the memory pool through activation: Naive T cells are activated during immune responses to foreign antigens at rate γ , and through clonal expansion form C memory T cells. During this activation into the memory population their division index increases by K .

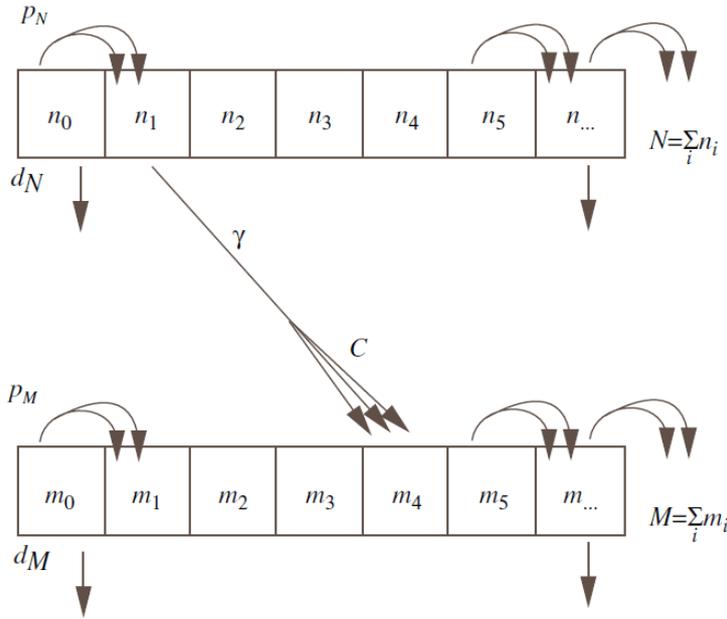


Figure 5: Schematic representation of the mathematical model for telomere shortening of naive (N) and memory (M) T cells. Adapted from: De Boer and Noest (1998).

Using this model it was shown that the average division-index (μ) of naive T cells (N) and memory T cells (M) are given by:

$$\frac{d\mu_N}{dt} = 2p_N \quad (13)$$

and

$$\frac{d\mu_M}{dt} = 2p_M - \gamma C \frac{N}{M} (\mu_M - \mu_N - K) \quad (14)$$

Question

In this exercise we will investigate (using `grind.R` and the file `telomeres.R`) how we should interpret the experimental data of Figure 4. We can use the differential equations for μ_N and μ_M described above to calculate the average telomere lengths of the naive (T_N) and memory (T_M) T-cell population from:

$$T_N = T_{n0} - L\mu_N \quad \text{and} \quad T_M = T_{m0} - L\mu_M, \quad (15)$$

where L is the number of base pairs lost during each cell division, and T_{n0} and T_{m0} are the average telomere lengths of naive and memory T cells at birth, respectively.

- The telomere data of Figure 4 were taken as evidence that naive and memory T cells divide approximately once every 5 years. Test if one division in every five years for both naive and memory T cells would indeed give rise to the data of Figure 4. You can use the following parameter values: $N/M = 1$, $\gamma = 0.001$ per day, $L = 0.1$ kb, $C = 16$ cells, $T_{N0} = 10$ kb, $T_{m0} = 8.6$ kb, $p_N = p_M = 0.0005$ divisions per day and $K = 0$.
- The findings of Figure 4 were remarkable because many other experiments suggested that memory T cells divide a lot more frequently than naive T cells. Yet the fact that telomere lengths of naive and memory T cells declined at equal rates suggested very similar kinetics of naive and memory lymphocytes. Increase the value of p_M to investigate how the telomere data would change if memory cells divide much (e.g. 10 times) more frequently than naive T cells.
- Using the following small steps, investigate why you found the above (perhaps unexpected) result...

- c1. Investigate what happens when you decrease the value of γ (e.g. 100-fold) in the scenario in which memory T cells divide 10-fold more frequently than naive T cells.
- c2. From the equation for μ_M you can tell that the average division index (and hence telomere length) of the memory T-cell population does not directly reflect memory cell division. What is the effect of activation of naive T cells into the memory pool?
- c3. How can we resolve the apparent contradiction between the parallel loss of telomeres in naive and memory T lymphocytes, and the fact that memory T cells are thought to divide more frequently than naive T cells?
- d. Is the average telomere length a good indicator of the division history of a cell population?

trec.R

```
# Source grind.R first

model <- function(t, state, parms){
  with(as.list(c(state,parms)),{
    c <- 0.25
    dN <- sigma - d*N + p*N
    dTr <- c*sigma - d*Tr

    return(list(c(dN, dTr)))
  })
}

myrun <- function(state=s, parms=p, after=NULL, ...) {
  data <- run(tstep=0.1, state=state, parms=p, after=after, table=T, timeplot=F, ...)
  data <- cbind(data, data[,3]/data[,2])
  names(data)[4] <- "A"
  par(mfrow=c(1,2))
  timePlot(data, tmin=28, tmax=32, ymin=1e10, ymax=1e13, log="y", show=c("N", "Tr"), legend=F, xlab="Age (in years)", ylab="Number of naive cells & TRECs", main="(a)")
  timePlot(data, tmin=28, tmax=32, ymin=0.01, ymax=1, log="y", show="A", legend=F, xlab="Age (in years)", ylab="Average TREC content", main="(b)")
  par(mfrow=c(1,1))
}

s <- c(N=1e13, Tr=1e12) # initial state
P <- c(sigma=10^9, p=0.0004, d=0.0005) # healthy parameters (per day!)

## run the model with the parameters of a healthy individual
p <- P*365.25 # convert to yearly rates
run(log="y") # simple grind.R run
myrun() # show age 28-32 and TRECS

##### For questions a-c: fill in your own value for xxx before running the model.

##### Question a: reducing thymic output #####
infect <- "if (t==30) parms[\"sigma\"] <- 365.25*xxx"
myrun(after=infect)

##### Question b: increasing death rate #####
infect <- "if (t==30) parms[\"d\"] <- 365.25*xxx"
myrun(after=infect)

##### Question c: increasing proliferation rate #####
infect <- "if (t==30) parms[\"p\"] <- 365.25*xxx"
myrun(after=infect)
```

telomeres.R

```
# Source grind.R first

model <- function(t, state, parms){
```

```

with(as.list(c(state,parms)),{

  dmuN <- 2*pn
  dmuM <- 2*pm-gamma*C*(muM-muN-K)

  return(list(c(dmuN, dmuM)))

})
}

myrun <- function(state=s, parms=p, ...) {
  nsol <- run(tstep=0.1, state=state, parms=parms, table=T, timeplot=F, ...)
  Tn0 <- 10000; Tm0 <- 8600; L <- 100
  TN <- (Tn0 - L*nsol[,2])/1000
  TM <- (Tm0 - L*nsol[,3])/1000
  data <- as.data.frame(cbind(nsol[,1], TN, TM))
  colnames(data)<-c("time", "TN", "TM")
  timePlot(data, tmin=20, tmax=80, legend=T, xlab="Age in years", ylab="Average telomere
  length")
}

s <- c(muN=0, muM=0) # initial state
P <- c(pn=0.0005, pm=0.0005, gamma=0.001, C=16, K=0) # parameters (first 3 rates per day!)
year <- 365.25
p <- P
p[1:3] <- P[1:3]*year # convert first 3 to per year

## Question a
run() # a simple run depicting muN and muM
myrun() # convert into true telomere lengths

### Question b: increase pm (fill in xxx)
p["pm"] <- xxx*year
myrun()

### Question c: decrease gamma (fill in xx)
p["pm"] <- 0.005*year
p["gamma"] <- xx*year
myrun()

```

T cell turnover

Handout for the *Immunobiology* lecture, June 12th, 2018

José Borghans and Julia Drylewicz (Immunology, UMCU)

Objectives of this exercise:

1. Understand how a simple mathematical model can help the interpretation of experimental data.
2. Understand how to interpret different kinetic and labeling data.

Exercise 1: T-cell dynamics after stem-cell transplantation

During HIV infection, CD4+ T cells show increased levels of T-cell proliferation (as measured by Ki67 expression). Although some have argued that increased T-cell proliferation in HIV infection is a consequence of the low numbers of T cells (the immune system is trying to compensate for the loss of cells), others have argued that increased T-cell proliferation is induced directly by the virus, and is the cause of low CD4+ T-cell numbers. In the article by Hazenberg et al. (Blood 2002) it was found that shortly after stem-cell transplantation, when cell numbers are very low, T cells have increased levels of Ki67 expression.

- a. Although these data seem to support the idea that low T-cell numbers (even in the absence of HIV infection) lead to increased levels of T-cell proliferation through a homeostatic mechanism, Hazenberg et al. argue that even after stem-cell transplantation there is no homeostatic response of the immune system. Give two arguments for their interpretation, and point out which figures support their interpretation.
- b. Calculate manually which average TREC content you expect in a healthy individual with $\sigma = 0.2$ cells/ μl blood per day, $p = 0.0005$ per day, $d = 0.001$ per day, and $c = 0.25$. Check with the file [trec2.R](#) whether this is indeed the equilibrium.
- c. Before patients can receive a stem-cell transplantation their lymphocyte numbers are dramatically decreased because of 'conditioning' for the transplantation. Investigate what happens with the average TREC content if the T-cell pool (and the total number of TRECs) in an individual is reduced 10-fold due to conditioning for stem-cell transplantation?
- d. How can you explain intuitively what happens to the TREC content after stem cell transplantation?
- e. There is some dispute in the field as to whether the thymus in healthy adults is still contributing new T cells. What do the findings of Hazenberg's Figure 2A tell you about the thymus in a healthy adult?

trec2.R

```
# Source grind.R first

model <- function(t, state, parms){
  with(as.list(c(state, parms)), {
    c <- 0.25
    dN <- sigma - d*N + p*N
    dTr <- c*sigma - d*Tr

    return(list(c(dN, dTr)))
  })
}

myrun <- function(state=s, parms=p, after=NULL, ...) {
  data <- run(tstep=0.1, state=state, parms=p, after=after, table=T, timeplot=F, ...)
  data <- cbind(data, data[,3]/data[,2])
  names(data)[4] <- "A"
  par(mfrow=c(1,2))
  timePlot(data, tmin=0, tmax=5, ymin=0, ymax=500, show=c("N", "Tr"), legend=F, xlab="Age", ylab="Number
of Naive T cells & TRECs", main="(a)")
  timePlot(data, tmin=0, tmax=5, ymin=0.1, ymax=0.2, log="y", show="A", legend=F, xlab="Age", ylab="Average
TREC content", main="(b)")
}
```

```

par(mfrow=c(1,1))
}

s <- c(N=10000, Tr=10000)          # initial state
P <- c(sigma=0.2, p=0.0005, d=0.001) # healthy parameters (per day!)

##### For questions b-c: fill in your own value for xxx before running the model.

##### Question b: #####
# First calculate by hand the equilibrium value for the average TREC content with the
  given parameters
## Then check the equilibrium by replacing xx with the equilibrium values that you
  calculated for Tr and N
p <- P*365.25                      # convert to yearly rates
s <- c(N=xx, Tr=xx)                # initial state
myrun()

##### Question c: decreasing 10-fold initial state #####
s <- c(N=xx, Tr=xx)                # initial state
myrun()

```

Exercise 2: Quantification of T-cell turnover using BrdU or stable isotope labelling

Background

Hellerstein et al. (2003) studied the increase in deuterium label during 9 weeks of heavy water (D₂O) administration in the DNA of T cells from healthy and HIV-infected subjects (see the following abstract and Figure 6).

Model for stable isotopes

As shown in today's lecture, the fraction of deuterium labelled DNA L during deuterium administration changes according to:

$$\frac{dL}{dt} = p - dL \quad (16)$$

where p is the production rate of the cells (in the thymus or through peripheral T-cell proliferation) and d their death rate. After stop of deuterium administration the fraction of deuterium labelled DNA L changes according to:

$$\frac{dL}{dt} = -dL \quad (17)$$

In this exercise we will use the file [labeling.R](#) to study deuterium and BrdU labelling.

- Hellerstein *et al.* (2003) reported that the accrual of deuterium in the DNA of total and memory/effector CD4 T cells occurs 'bi-phasically'. Based on this it was concluded that the memory/effector CD4 T-cell pool consists of two different subpopulations: one with short-lived and one with long-lived cells. Simulate a deuterium labelling experiment for a cell population with $p = 0.015$ per week and $d = 0.05$ per week when the individual is healthy. Increase the proliferation and death rate of cells 10-fold when the individual is HIV-infected. What is causing the non-linear increase? Was Hellerstein's interpretation correct?

Subpopulations of long-lived and short-lived T cells in advanced HIV-1 infection

See the related Commentary beginning on page 821.

Marc K. Hellerstein,^{1,2} Rebecca A. Hoh,² Mary Beth Hanley,³ Denise Cesar,¹ Daniel Lee,¹ Richard A. Neese,^{1,2} and Joseph M. McCune^{2,3,4}

¹Department of Nutritional Sciences and Toxicology, University of California, Berkeley, Berkeley, California, USA

²Department of Medicine, University of California, San Francisco, and Medical Service, San Francisco General Hospital, University of California, San Francisco, San Francisco, California, USA

³Gladstone Institute of Virology and Immunology, San Francisco General Hospital, University of California, San Francisco, San Francisco, California, USA

⁴Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, California, USA

Antigenic stimulation of T cells gives rise to short-lived effector cells and long-lived memory cells. We used two stable isotope-labeling techniques to identify kinetically distinct subpopulations of T cells and to determine the effect of advanced infection with HIV-1. Long-term deuterated water (²H₂O) incorporation into DNA demonstrated biphasic accrual of total and of memory/effector (m/e)-phenotype but not naive-phenotype T cells, consistent with the presence of short-lived and longer-lived subpopulations within the m/e-phenotype T cell pool. These results were mirrored by biphasic die-away kinetics in m/e- but not naive-phenotype T cells after short-term ²H-glucose labeling. Persistent label retention was observed in a subset of m/e-phenotype T cells (presumably memory T cells), confirming the presence of T cells with very different life spans in humans. In advanced HIV-1 infection, much higher proportions of T cells were short-lived, compared to healthy controls. Effective long-term anti-retroviral therapy restored values to normal. These results provide the first quantitative evidence that long-lived and quiescent T cells do indeed predominate in the T cell pool in humans and determine T cell pool size, as in rodents. The greatest impact of advanced HIV-1 infection is to reduce the generation of long-lived, potential progenitor T cells.

J. Clin. Invest. 112:956-966 (2003). doi:10.1172/JCI200317533.

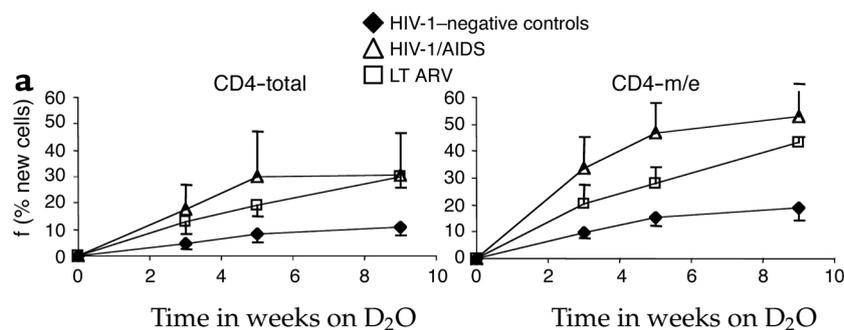


Figure 6: Note that LT ARV stands for patients on long term anti-retroviral therapy, *m/e* stands for memory or effector T cells and *f* (on y-axis) is the percentage of cells that have proliferated (and hence have taken up deuterium)

- b. In many deuterium labelling studies the average turnover rate of cells is determined by finding the maximum level of deuterium incorporation that is ever attained during label administration, and dividing by the number of days it took to reach that maximum.

Based on your simulations of previous question, describe what can go wrong using this method.

How would you design a labelling study to avoid such problems?

How would you design it if you were only allowed to take one blood sample during labeling?

Model for BrdU

As shown in today's lecture, the fraction of BrdU labelled cells L and unlabelled cells U during BrdU administration change according to:

$$\frac{dL}{dt} = \sigma + 2pU + pL - dL \quad (18)$$

$$\frac{dU}{dt} = -pU - dU \quad (19)$$

where σ represents the influx of cells from a labelled source (e.g. the thymus in the case of naive T cells), p is the proliferation rate of the cells and d their death rate. After stop of BrdU administration the fraction of BrdU labelled cells L and unlabelled cells U change according to:

$$\frac{dU}{dt} = \sigma + pU - dU \quad (20)$$

$$\frac{dL}{dt} = pL - dL \quad (21)$$

- c. Consider a mouse T-cell population for which $p = 0.5$ per day and $d = 0.5$ per day, which is labelled for 10 days with both BrdU and deuterium. Simulate how the percentage of BrdU labeled cells and deuterium labeled DNA increases during the labeling period.

Why is the increase in the fraction of BrdU labelled cells faster than the increase in the fraction of deuterium labelled DNA?

- d. Simulate what happens with this cell population when BrdU and deuterium labelling are stopped. What is going on?
 e. What could explain the BrdU loss that is typically observed after label withdrawal?

labeling.R

```
# Source grind.R first

deuterium <- function(t, state, parms){
  with(as.list(c(state,parms)),{

    if (t < Tend) dLd <- p - d*Ld
    else dLd <- -d*Ld

    return(list(c(dLd)))
  })
}

brdu <- function(t, state, parms){
  with(as.list(c(state,parms)),{

    if (t < Tend) dLb <- (p+d)*(1-Lb)
    else dLb <- (p-d)*Lb

    return(list(c(dLb)))
  })
}

model <- function(t, state, parms){
```

```

with(as.list(c(state,parms)),{
  if (t < Tend) {
    dLd <- p - d*Ld
    dLb <- (p+d)*(1-Lb)
  }else {
    dLd <- - d*Ld
    dLb <- (p-d)*Lb
  }
  return(list(c(dLd, dLb)))
})
}

## Question a: 9 week deuterium labeling Hellerstein (replace the values of p and d to
study HIV)
p <- c(p=0.015, d=0.05, Tend=9)
s <- c(Ld=0)
run(tstep=0.1, odes=deuterium, tmax=9, ylab="Fraction labeled")

## Question c: 10 day deuterium and BrdU up-labeling in a mouse; replace xx!
p <- c(p=xx, d=xx, Tend=10)
s <- c(Ld=0, Lb=0)
run(tstep=0.1, odes=model, tmax=10, ylab="Fraction labeled")

## Question d: up- and down-labeling curve of 10 day deuterium and BrdU labeling in a
mouse
run(tstep=0.1, odes=model, tmax=20, ylab="Fraction labeled")

```

References

- Badovinac, V. P., Porter, B. B., and Harty, J. T., 2002. Programmed contraction of CD8⁺ T cells after infection. *Nat. Immunol.* 3:619–626.
- Blattman, J. N., Antia, R., Sourdive, D. J., Wang, X., Kaech, S. M., Murali-Krishna, K., Altman, J. D., and Ahmed, R., 2002. Estimating the precursor frequency of naive antigen-specific CD8 T cells. *J. Exp. Med.* 195:657–664.
- Borghans, J. A. M., Noest, A. J., and De Boer, R. J., 1999. How specific should immunological memory be? *J. Immunol.* 163:569–575.
- Borghans, J. A. M., Noest, A. J., and De Boer, R. J., 2003. Thymic selection does not limit the individual MHC diversity. *Eur. J. Immunol.* 33:3353–3358.
- Burroughs, N. J., De Boer, R. J., and Kesmir, C., 2004. Discriminating self from nonself with short peptides from large proteomes. *Immunogenetics.* 56:311–320.
- Danke, N. A., Koelle, D. M., Yee, C., Beheray, S., and Kwok, W. W., 2004. Autoreactive T cells in healthy individuals. *J. Immunol.* 172:5967–5972.
- De Boer, R. J., Homann, D., and Perelson, A. S., 2003. Different dynamics of CD4⁺ and CD8⁺ T cell responses during and after acute lymphocytic choriomeningitis virus infection. *J. Immunol.* 171:3928–3935.
- De Boer, R. J. and Noest, A. J., 1998. T cell renewal rates, telomerase, and telomere length shortening. *J. Immunol.* 160:5832–5837.
- De Boer, R. J., Oprea, M., Antia, R., Murali-Krishna, K., Ahmed, R., and Perelson, A. S., 2001. Recruitment times, proliferation, and apoptosis rates during the CD8⁺ T-cell response to lymphocytic choriomeningitis virus. *J. Virol.* 75:10663–10669.
- De Boer, R. J. and Perelson, A. S., 1993. How diverse should the immune system be? *Proc. R. Soc. Lond., B, Biol. Sci.* 252:171–175.
- De Boer, R. J. and Perelson, A. S., 2013. Quantifying T lymphocyte turnover. *J. theor. Biol.* 327:45–87.
- Douek, D. C., McFarland, R. D., Keiser, P. H., Gage, E. A., Massey, J. M., Haynes, B. F., Polis, M. A., Haase, A. T., Feinberg, M. B., Sullivan, J. L., Jamieson, B. D., Zack, J. A., Picker, L. J., and Koup, R. A., 1998. Changes in thymic function with age and during the treatment of HIV infection. *Nature* 396:690–695.

- Goncalves, P., Ferrarini, M., Molina-Paris, C., Lythe, G., Vasseur, F., Lim, A., Rocha, B., and Azogui, O., 2017. A new mechanism shapes the naive CD8⁺ T cell repertoire: the selection for full diversity. *Mol. Immunol.* **85**:66–80.
- Hazenberg, M. D., Otto, S. A., Stuart, J. W., Verschuren, M. C., Borleffs, J. C., Boucher, C. A., Coutinho, R. A., Lange, J. M., De Wit, T. F., Tsegaye, A., Van Dongen, J. J., Hamann, D., De Boer, R. J., and Miedema, E., 2000. Increased cell division but not thymic dysfunction rapidly affects the T-cell receptor excision circle content of the naive T cell population in HIV-1 infection. *Nat. Med.* **6**:1036–1042.
- Hellerstein, M. K., Hoh, R. A., Hanley, M. B., Cesar, D., Lee, D., Neese, R. A., and McCune, J. M., 2003. Subpopulations of long-lived and short-lived T cells in advanced HIV-1 infection. *J. Clin. Invest.* **112**:956–966.
- Homann, D., Teyton, L., and Oldstone, M. B., 2001. Differential regulation of antiviral T-cell immunity results in stable CD8⁺ but declining CD4⁺ T-cell memory. *Nat. Med.* **7**:913–919.
- Janeway, Jr, C. A. and Katz, M. E., 1984. Self Ia-recognizing T cells undergo an ordered series of interactions with Ia-bearing substrate cells of defined function during their development: a model. *Surv. Immunol. Res.* **3**:45–54.
- Kohler, B., 2007. Mathematically modeling dynamics of T cell responses: predictions concerning the generation of memory cells. *J. theor. Biol.* **245**:669–676.
- Kotturi, M. F., Scott, I., Wolfe, T., Peters, B., Sidney, J., Cheroutre, H., Von Herrath, M. G., Buchmeier, M. J., Grey, H., and Sette, A., 2008. Naive precursor frequencies and MHC binding rather than the degree of epitope diversity shape CD8⁺ T cell immunodominance. *J. Immunol.* **181**:2124–2133.
- Malhotra, D., Linehan, J. L., Dileepan, T., Lee, Y. J., Purtha, W. E., Lu, J. V., Nelson, R. W., Fife, B. T., Orr, H. T., Anderson, M. S., Hogquist, K. A., and Jenkins, M. K., 2016. Tolerance is established in polyclonal CD4⁺ T cells by distinct mechanisms, according to self-peptide expression patterns. *Nat. Immunol.* **17**:187–195.
- Merkenschlager, M., Graf, D., Lovatt, M., Bommhardt, U., Zamoyska, R., and Fisher, A. G., 1997. How many thymocytes audition for selection? *J. Exp. Med.* **186**:1149–1158.
- Nowak, M. A., Tarczy-Hornoch, K., and Austyn, J. M., 1992. The optimal number of major histocompatibility complex molecules in an individual. *Proc. Natl. Acad. Sci. U.S.A.* **89**:10896–10899.
- Qi, Q., Liu, Y., Cheng, Y., Glanville, J., Zhang, D., Lee, J. Y., Olshen, R. A., Weyand, C. M., Boyd, S. D., and Goronzy, J. J., 2014. Diversity and clonal selection in the human T-cell repertoire. *Proc. Natl. Acad. Sci. U.S.A.* **111**:13139–13144.
- Saeki, K., Doekes, H. M., and Boer, R. J. D., 2015. Optimal t cell cross-reactivity and the role of regulatory t cells. *Journal of Theoretical Biology* **375**:4 – 12. *Theories and Modeling of Autoimmunity.*
- Soetaert, K., 2009. rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations. R package 1.6.
- Soetaert, K. and Herman, P. M., 2009. A Practical Guide to Ecological Modelling. Using R as a Simulation Platform. Springer. ISBN 978-1-4020-8623-6.
- Soetaert, K. and Petzoldt, T., 2010. Inverse modelling, sensitivity and monte carlo analysis in R using package FME. *Journal of Statistical Software* **33**:1–28.
- Soetaert, K., Petzoldt, T., and Setzer, R. W., 2010. Solving differential equations in R: Package desolve. *Journal of Statistical Software* **33**:1–25.
- Stritesky, G. L., Xing, Y., Erickson, J. R., Kalekar, L. A., Wang, X., Mueller, D. L., Jameson, S. C., and Hogquist, K. A., 2013. Murine thymic selection quantified using a unique method to capture deleted T cells. *Proc. Natl. Acad. Sci. U.S.A.* **110**:4679–4684.
- Su, L. F., Kidd, B. A., Han, A., Kotzin, J. J., and Davis, M. M., 2013. Virus-specific CD4⁺ memory-phenotype T cells are abundant in unexposed adults. *Immunity* **38**:373–383.
- Van Meerwijk, J. P., Marguerat, S., Lees, R. K., Germain, R. N., Fowlkes, B. J., and MacDonald, H. R., 1997. Quantitative impact of thymic clonal deletion on the T cell repertoire. *J. Exp. Med.* **185**:377–383.
- Weng, N. P., Levine, B. L., June, C. H., and Hodes, R. J., 1995. Human naive and memory T lymphocytes differ in telomeric length and replicative potential. *Proc. Natl. Acad. Sci. U.S.A.* **92**:11091–11094.