



How is the T-cell repertoire shaped?

Nuno Henriques dos Santos de Sepúlveda

Dissertação de doutoramento em Ciências Biomédicas

2009

Nuno Henriques dos Santos de Sepúlveda

How is the T-cell repertoire shaped?

Dissertação de Candidatura ao grau de
Doutor em Ciências Biomédicas submetida
ao Instituto de Ciências Biomédicas de Abel
Salazar da Universidade do Porto.

Orientador – Doutor Jorge Carneiro

Categoria – Investigador

Afiliação – Instituto Gulbenkian de Ciência

Laboratório Associado de Oeiras

Co-orientador

Categoria – Doutor Manuel Vilanova

Afiliação – Instituto de Ciências Biomédicas
Abel Salazar da Universidade do Porto.

ABSTRACT

In the adaptive immune system of jawed vertebrates, T cells orchestrate immune responses. Collectively, they express a great diversity of antigen-recognition receptors (TCR, T-cell receptor), forming the so-called T-cell repertoire. The genes encoding these TCRs are generated during T-cell ontogeny in the thymus by a random process of assembling different gene segments, known as V(D)J recombination. Due to the randomness of the process and to the TCR diversity it generates, T cells are able to react to any antigen, including the body ones. Therefore, T-cell repertoire must be somehow fine-tuned in order to efficiently respond to harmful pathogens and yet avoiding deleterious immune responses against body components.

The aim of this thesis is to understand how the T-cell repertoire is shaped. Two main threads were followed. First, we investigated the biological constraints in the very early steps of T-cell repertoire formation by studying the mechanisms governing the V(D)J recombination in T cells. We chose to study $\gamma\delta$ T cells because they just recombine two V-J isotypes in adult mice, allowing to develop mathematical models that could trace all possible rearrangement states of a cell throughout the recombination process. Second, motivated by the important role of regulatory T cells in the control of autoimmunity and immunity against pathogens, we assessed the influence of functional diversity in shaping the peripheral repertoire of regulatory and conventional T cells. We also characterized TCR diversity of these two repertoires in the thymus and periphery. Along the way we also gained some insights on the TCR diversity of $CD4^+$ and $CD8^+$ T cell repertoires in limited TCR diversity mice. We tackled these issues using different mathematical and statistical models.

In Chapters 2 and 3, we studied the rearrangement of $TCR\gamma$ locus that encode the $TCR\gamma$ chain of $\gamma\delta$ T cells. Since only two V-J isotypes at the γ locus are essentially recombined in adult mice, it was feasible to enumerate all possible recombination states of a T-cell precursor and develop different Markov chain models that could trace them throughout the recombination process (Chapter 2). By testing them against experimental data, we argued that V-J rearrangement in a precursor cell can only occur in a certain time period, in which a single locus is first opened for the rearrangement, followed by the opening of the second locus at the other chromosome after some time (asynchronous allelic accessibility). Moreover, the occurrence of a productive V-J rearrangement leads to shutting down of the whole rearrangement process in the γ locus (feedback mechanism). Interestingly, the estimated rearrangement rates for the different isotypes suggested that the outcome of $TCR\gamma$ -chain recombination process involves a tradeoff between maximizing the precursor-cell survival and minimizing the chance of that mature cells exhibit two distinct antigen receptors. In Chapter 3, we provided the experimental evidence supporting the suggested mechanism.

In Chapter 4, we studied the effect of functional diversity in shaping peripheral regulatory and conventional T-cell repertoires via the so-called Crossregulation model. In its essence, the model describes a basic interaction network among regulatory, conventional T cells, and antigen-presenting cells (APCs). In this model, T-cell survival and proliferation requires conjugation with APCs; regulatory T cells need additionally to receive a growth factor from conventional T cells in order to proliferate; conventional T-cell proliferation is inhibited by regulatory T cells. We first revisited previous published work done with the model providing an extended review of experimental data compatible with its assumption and predictions. We took the chance to study the peripheral repertoire by extending the model to a simple situation in which every T-cell clone in the repertoire recognizes an exclusive set of APCs, i. e., T-cell clones do not compete for APCs. The most important result of this model is that the peripheral T-cell repertoire might be partitioned in three subsets of T-cell clones depending on the densities of APCs these clones recognize. The first subset is composed of clones that interact with few APCs, and thus cannot be maintained in the periphery, being short lived. The second subset refers to a diverse set of barely autoreactive conventional T-cell clones, whose expansion is limited only by APC availability. Because these clones lack regulatory T cells, this subset is prone to react to foreign antigens. The third subset is a less diverse set of self-reactive T cell clones with both regulatory and conventional T cells that regulate each other's growth. We argued that this subset is the one that avoids deleterious immune responses against self antigens, because of the presence of regulatory T cells.

In Chapter 5, we extended previous Crossregulation model to a more realistic scenario, in which every T cell clone recognizes a random non-exclusive subset of APCs, giving rise to interclonal competition for APCs. Simulations of this model showed that, in general, peripheral selection favors more crossreactive T-cell clones whose frequency in the repertoire increases in time. Moreover, we could demonstrate that regulatory T cells might have an exclusive repertoire, a possibility that was beyond the scope of the previous model by construction. The origin of this exclusive repertoire might be simply due to shared APCs with other T-cell clones that allow regulatory T cells to persist in clones without conventional T cells. In simulations resembling adoptive transfers, we also showed that the above-mentioned partition of the repertoire holds. When a constant thymic export was included in the model, the repertoire structures emerging from the simulations could be classified into three types. One repertoire structure is characterized by an overdominance of conventional T cells, even in clones recognizing high levels of APCs that should be under control of regulatory T cells. This situation is interpreted as the devastating phenotype of *Foxp3*-deficient mice, which have no regulatory T cells. Other repertoire structure is related to an overdominance of regulatory T cells. This structure is obviously a deleterious phenotype, since all immune responses, even those against harmful pathogens, would be under action of regulatory T cells. The third repertoire structure shows a similar partition of the repertoire similar to the one obtained in simulations without thymic export or with independent APC recognition studied in previous chapter. As argued above, this structure should be displayed by healthy individuals. These three repertoire structures are best distinguished in the resident population that persist longer in the repertoire. For each repertoire structure, we provided qualitative predictions for different structural diversity properties, such

as the relationship between regulatory and conventional T-cell diversities and the correlation between the respective clonal size distributions. The "healthy", partitioned repertoire structure emerged only in a small fraction of all simulations. This structure is transient in the presence of constant thymic export, and the repertoire will eventually tend to an overdominance of regulatory T cells. We argued that a way to prevent this is to decrease the thymic export throughout life to maintain the repertoire in its healthy shape. Therefore, in contrast to what is often argued, thymic involution might have a beneficial effect to the organism by delaying generalized regulation of all immune responses beyond its lifespan. Another important result of this chapter is that we could obtain three repertoire classes in simulations with the same parameter set and the relative proportions of each class is under the control of the parameters. Therefore, the stochasticity in the repertoire formation might be one of the explanations for the incomplete penetrance observed in many mouse models that develop spontaneous autoimmune diseases or for the low disease concordance rate in twin studies.

In Chapter 6 we provided statistical methods to assess the above predictions for the structural diversity properties with available experimental data. We first introduced the Poisson abundance models (PAM) to estimate T-cell receptor diversity and the underlying clonal sizes distributions. This class of models is able to describe different patterns of the underlying clonal size distribution, such as the Lognormal or the Gamma distribution. To fit PAM to experimental data, we made available the package PAM for the R language. For illustrative purposes, we analyzed thymic and peripheral CD4⁺ and CD8⁺ TCR data from a mouse model in which TCR diversity is extremely restricted. We showed that peripheral CD4⁺ and CD8⁺ TCR samples are well described by the Poisson-Lognormal distribution, which, if true, implies a Lognormal probability distribution for the clonal size distribution.

In Chapter 7 we applied the Poisson abundance models to the study of the TCR repertoires of regulatory and conventional T cells in the thymus and periphery. The Lognormal-Poisson distribution revealed again to be the best model for all data sets analyzed. Based on this result, this model was extended to the two-sample scenario in order to dissect the relationship between any two overlapping repertoires. The results obtained suggest that conventional thymocytes are more diverse than their regulatory counterparts, in agreement with the hypothesis that regulatory T cells are biased towards self-reactivities. We could not rule out the hypothesis that peripheral regulatory and conventional T-cell repertoires have the same diversity and, by the analysis of their intersection repertoire, indicates that the two repertoires are identical. These two results are compatible with simulated repertoire structure characterized by an overdominance of regulatory T cells. Notwithstanding this interpretation, the experimental data are also consistent with a relatively small fraction of TCR variants being shared between peripheral regulatory and conventional T cells and this in turn is in agreement with the partitioned repertoire structure. Therefore, the available data are compatible, and do not contradict the predictions of Crossregulation model for the peripheral T-cell repertoire of non-autoimmune individuals.

In chapter 8 we discussed the most important aspects of this thesis, namely, the potential structures of peripheral T-cell repertoire, and the mechanisms that generate them. We built on theoretical results to gain insight into the likelihood of an exclusive regulatory T-cell repertoire

and putative explanations for the different repertoire structures observed in animal models with limited TCR diversity. We also discussed the reason for the good fit of the Poisson-Lognormal distribution in most data sets. Using the Crossregulation model to tackle this issue, the underlying clonal size distribution seems a mixture of different distributions associated with different populations (e.g., recent thymic emigrants and different resident populations). With this result in mind, we argued that the general good fit of this distribution is due to either too small sample sizes or low TCR diversities in the mice under analysis. We also put forward a set of precise recommendations for future TCR repertoire studies in general, and for validation of the predictions of the Crossregulation model in particular. In this regard, we advise to increase current sample sizes, the usage of TCR sequencing at single-cell level, and longitudinal T-cell repertoire studies.

We then conclude that mathematical modeling provided clues on how the T-cell repertoire is structured. It is important to say that most of the above conclusions could only be reached because of our constant attempt to be in close contact with experimental data. In the case of the recombination of the TCR genes, we developed mathematical models with the aim of fitting them directly to the data, which allowed us to put forward more realistic predictions for the outcome of the process. In the case of the Crossregulation model, the simulations that we performed seem an important basis for the interpretation of future T-cell repertoire studies. Moreover, the simulations done with this model were fundamental to understand the need of altering current experimental designs for the study of T-cell repertoire, such as the usage of longitudinal T-cell repertoire studies. Notwithstanding, combining our models with experimental data revealed to be a hard task as illustrated in Chapter 7, where definite conclusions could not be drawn due to insufficient sample sizes and, in some cases, the unavailability of the raw data. This shows the difficulties underlying the type of approach followed in this thesis. These difficulties could be partially overcome with a greater cooperation between "wet" and "dry" labs. This seems the only way to understand better complex biological phenomena, such as the shaping of the T-cell repertoire at its different stages.

RESUMO

O sistema imunitário adaptativo dos vertebrados é orquestrado pelos linfócitos T. Estas células apresentam colectivamente um conjunto diverso de receptores (TCR, do inglês *T-cell receptor*) capazes de reconhecerem e desencadearem respostas específicas a antígenos. Este conjunto é denominado por repertório dos linfócitos T. Os genes que codificam os TCR são gerados durante o desenvolvimento dos linfócitos T no timo por um mecanismo de rearranjo aleatório de vários segmentos de genes, o chamado processo de recombinação V(D)J. Devido à aleatoriedade deste processo e à enorme diversidade de receptores que este gera, o sistema imunitário fica habilitado a responder a todo e qualquer antígeno, inclusive aos do próprio corpo. Assim, o repertório dos linfócitos T necessita de estar bem estruturado de forma a assegurar uma boa eficiência de resposta contra patógenos mas evitando, ao mesmo tempo, reacções imunitárias contra componentes do próprio corpo que, ocorrendo, poderão conduzir às chamadas doenças autoimunes.

O objectivo desta tese é dissecar os mecanismos que estruturam o repertório dos linfócitos T. Duas linhas de investigação foram seguidas para atingir este objectivo. A primeira consistiu no estudo das restrições biológicas subjacentes aos primeiros estádios da formação do repertório dessas células no timo. Com este fim, escolheu-se estudar o processo de recombinação do locus γ que codifica a respectiva cadeia do receptor dos linfócitos T $\gamma\delta$, uma vez que apenas dois isótipos V-J recombinaem essencialmente em ratinhos adultos. A segunda foi motivada pelo papel fulcral dos linfócitos T reguladores no controlo de autoimunidade e de imunidade contra diversos agentes infecciosos, tendo como objectivo estudar os mecanismos de selecção periférica que actuam sobre o repertório dessas células. Caracterizou-se, também, a diversidade de TCR no timo e nos órgãos periféricos através da re-análise de dados previamente publicados na literatura. Ao longo da tese ganhou-se também algum conhecimento sobre a diversidade de TCR em linfócitos T $CD4^+$ e $CD8^+$ de ratinhos cuja geração da diversidade desses receptores estava limitada geneticamente. Para atingir os fins propostos, utilizaram-se diferentes modelos matemáticos e estatísticos.

Nos capítulos 2 e 3 estuda-se o processo de recombinação do locus γ nos linfócitos T $\gamma\delta$. Como essencialmente dois isótipos V-J são recombinaados no locus γ de ratinhos adultos, foi possível enumerar todos os estados de recombinação que uma célula precursora poderia ter durante o processo de recombinação. Assim, desenvolveram-se modelos estocásticos formulados em termos de Cadeias de Markov que pudessem explicar a dinâmica de recombinação nessas células (Capítulo 2). Ao confrontar esses modelos com dados experimentais, conseguiu-se inferir que a recombinação do locus γ parece ocorrer durante um período limitado de tempo (janela temporal), onde um dos locos γ abre primeiro para recombinação, seguido da abertura do outro passado algum tempo (acessibilidade alélica assíncrona).

A ocorrência de um rearranjo produtivo V-J parece conduzir ao encerramento quase imediato do processo de recombinação nesse locus (mecanismo de retroacção). Interessa notar que, ao estudar as consequências do processo de recombinação sob as estimativas das taxas de rearranjo de cada isótipo, os resultados apontam para um balanço entre a maximização da probabilidade de sobrevivência de uma célula precursora e a minimização da probabilidade de ocorrerem dois rearranjos produtivos na mesma célula precursora. O capítulo 3 mostra o conjunto de experiências executadas para corroborar o mecanismo acima referido, o que de facto parece acontecer.

O capítulo 4 dedica-se ao estudo do efeito da diversidade funcional na estruturação do repertório das células T CD4⁺ efectoras e reguladoras nos órgãos linfóides (ou periferia), onde as primeiras são os linfócitos T convencionais que montam respostas (auto-)imunes e as últimas são células T que regulam a proliferação das efectoras, evitando autoimunidade. Com esse fim, aplicou-se o modelo de regulação cruzada que descreve uma rede de interacções entre células T reguladoras e efectoras quando conjugadas com as chamadas células apresentadoras de antígeno (APC, do inglês *antigen-presenting cells*). Este modelo pressupõe que as células T reguladoras precisem de receber algum factor de crescimento secretado pelas células T efectoras quando ambos os tipos de células estão conjugados na mesma APC. Neste caso, as células T reguladoras são capazes de se dividir, concomitantemente, inibindo por um mecanismo ainda desconhecido, a proliferação das células T efectoras. Para além disso, descreve-se a dinâmica do repertório das células T pressupondo que cada clone de células T reconhece um conjunto exclusivo de APC. De acordo com o modelo, o repertório pode ser particionado em três subconjuntos de clones de células T. O primeiro subconjunto refere-se a clones que reconhecem um subconjunto de APC em baixa densidade no organismo e, por isso, não se conseguem manter na periferia. O segundo subconjunto consiste em diversos clones de células T efectoras mantidos por subpopulações de APC em densidades intermédias. Como estes clones não têm células T reguladoras, sugere-se que serão estes que reconhecem microorganismos que entram ocasionalmente na periferia. O terceiro subconjunto é composto por alguns clones autoreactivos com a coexistência de células T reguladoras e efectoras a regularem mutuamente o seu crescimento. Argumenta-se que o controlo da autoimunidade está assegurado na periferia por estes clones com células T reguladoras.

No capítulo 5 estende-se o modelo de regulação cruzada a um cenário mais realista em que cada clone de células T reconhece aleatoriamente um conjunto diverso mas não-exclusivo de APC — competição interclonal. Os resultados das simulações demonstram que a selecção periférica tende a favorecer os clones de maior reactividade cruzada, cuja a frequência no repertório aumenta ao longo do tempo. Para além disso, as simulações mostram agora que a selecção periférica de clones é suficiente para gerar um repertório exclusivo das células T reguladoras, o que era impossível de obter com o modelo de clones independentes. Em simulações que se podem assemelhar a experiências de transferência adoptiva, obtém-se a mesma partição referida no capítulo 4. Quando um fluxo tímico constante é introduzido no modelo, três estruturas distintas de repertório surgem nas simulações. Uma estrutura de repertório está associada a uma sobre-representação de células T efectoras, mesmo em clones

que reconhecem APC em alta densidade que deveriam estar sob o controlo de células T reguladoras. Esta estrutura parece, então, indicada para descrever o fenótipo de indivíduos deficientes no factor de transcrição *Foxp3*, essencial para a geração e função das células T reguladoras. Outra estrutura do repertório caracteriza-se por uma sobre-representação de células T reguladoras no repertório periférico. Este tipo de repertório corresponde também a um fenótipo deletério, já que todas as respostas imunes, inclusivé as protectoras contra microorganismos, estariam sob a acção das células T reguladoras. A terceira estrutura de repertório apresenta uma partição semelhante à supracitada no capítulo 4 e, por isso, parece a mais razoável para descrever um repertório eficiente tanto a combater patógenos como a evitar autoimunidade. Demonstra-se também que as três estruturas de repertório são bem evidentes na população residente, que se mantém na periferia por um longo período de tempo. Para cada estrutura de repertório, faz-se uma previsão qualitativa para diferentes propriedades de diversidade estrutural (diversidade de TCR), tal como a relação entre as diversidades de células T reguladoras e efectoras e a correlação entre as distribuições de tamanhos clonais. Interessa realçar que a última estrutura do repertório, a com a partição de acordo com as células T reguladoras, é apenas uma pequena fracção dos repertórios obtidos no conjunto das simulações efectuadas. Isto acontece porque na presença de um fluxo tímico constante, este repertório particionado é transiente, tendendo eventualmente para a estrutura com sobre-representação de células T reguladoras. Sugere-se que uma forma de evitar este fenómeno consiste em diminuir o fluxo tímico a partir do momento em que a periferia estiver colonizada pela sua população residente, tal como é observado em vários estudos experimentais em humanos e ratinhos. Assim, ao contrariamente sugerido na literatura, o fenómeno de involução tímica pode ser então considerado um processo benéfico para o organismo, evitando uma regulação forte e generalizada de todas as respostas imunes. Outro resultado importante deste capítulo é o facto de se observar estas três estruturas de repertórios em simulações com o mesmo conjunto de parâmetros e que as respectivas proporções são controladas por esses parâmetros. Assim, a estocasticidade na formação do repertório pode ser uma das explicações para a penetrância incompleta frequentemente observada em modelos animais que desenvolvem espontaneamente doenças autoimunes, ou para a baixa taxa de concordância destas patologias em gémeos.

O capítulo 6 centra-se na metodologia estatística para inferir as previsões do modelo de regulação cruzada em dados experimentais disponíveis na literatura. Com esse fim, introduzem-se os modelos poissonianos de abundâncias para estimar a diversidade de TCR e a distribuição subjacente de tamanhos clonais. Esta classe de modelos inclui um razoável número de modelos que permitem descrever diferentes padrões de heterogeneidade nas distribuições de tamanhos clonais, tais como as distribuições Lognormal ou Gama. O ajustamento desses modelos é feito pelo método de máxima verosimilhança implementado no pacote *PAM* disponível para o programa informático R. Para fins ilustrativos, analisam-se dados referentes ao repertório das células T $CD4^+$ and $CD8^+$ no timo e órgãos linfóides periféricos em ratinhos com a diversidade de TCR limitada geneticamente. Estes dados parecem particularmente bem descritos pela distribuição Poisson-Lognormal que pressupõe uma distribuição Lognormal para a respectiva distribuição de tamanhos clonais.

O capítulo 7 tem como objectivo o estudo do repertório das células T reguladoras e efectoras no timo e periferia através do ajustamento dos modelos poissonianos de abundância a dados disponíveis na literatura. Tal como se observou no capítulo 6, o modelo Poisson-Lognormal parece o mais indicado para descrever os dados experimentais. Por este facto, estende-se este modelo à análise de duas amostras de forma a inferir a relação entre dois repertórios de células T que partilham algumas especificidades de TCR (o repertório de sobreposição ou de intersecção). Os dados apontam para uma maior diversidade de timócitos T efectores do que de timócitos T reguladores, o que parece concordar com a hipótese de que as células T reguladoras são geradas no timo com alto grau de autoreactividade. Não se conseguiu rejeitar a hipótese de igual diversidade de células T reguladoras e efectoras na periferia. A análise do respectivo repertório de sobreposição não rejeita a possibilidade de as células T reguladoras e efectoras terem o mesmo repertório. Este resultado é concordante com a estrutura de repertório caracterizada por uma sobrerepresentação de células T reguladoras. Contudo, não se pode rejeitar a hipótese de uma intersecção relativamente baixa entre os repertórios das células T reguladoras e efectoras que, a confirmar, é consistente com a estrutura do repertório tipificada para um indivíduo saudável. Portanto, os dados disponíveis não parecem contradizer as estruturas de repertórios previstas pelo modelo de regulação cruzada para indivíduos que não sofrem de qualquer autoimunidade.

Por fim, no capítulo 8 discute-se os aspectos mais importantes da tese, nomeadamente, as várias estruturas potenciais do repertório de células T e os mecanismos que as geram. Baseados nos resultados teóricos, examina-se a possibilidade da existência de um repertório exclusivo das células T reguladoras, e apresentam-se eventuais explicações para diferentes estruturas de repertório em diversos modelos animais com diversidade limitada de TCR. Discute-se, também, a razão para o bom ajustamento do modelo Poisson-Lognormal nos dados experimentais. Usando o modelo de regulação cruzada, a distribuição de tamanhos clonais parece ser uma mistura de várias distribuições associadas a distintas populações (e.g., imigrantes recentemente vindos do timo e diferentes populações residentes). Assim, a boa adequação do modelo Poisson-Lognormal aos dados pode apenas advir de tamanhos amostrais reduzidos ou de uma baixa diversidade de TCR nos ratinhos analisados. Baseados nos resultados dos Capítulos 6 e 7, propõe-se uma lista de recomendações para futuros estudos de quantificação do repertório em geral, e para a validação das previsões do modelo de regulação cruzada em particular. Dentro deste conjunto de recomendações, destaca-se a necessidade de aumentar significativamente as dimensões amostras, a utilização sistemática de sequenciação do TCR ao nível de uma célula e, se possível, a execução de estudos longitudinais do repertório.

Conclui-se, então, que a modelação matemática produziu um conjunto de pistas para o problema da estruturação do repertório das células T. Importa realçar que estas pistas só foram conseguidas pela constante ligação entre os dados disponíveis e os modelos desenvolvidos. No caso do processo de recombinação do locus γ , os modelos foram concebidos de forma a serem ajustados directamente aos dados experimentais, o que parece implicar previsões mais realistas para diferentes características do processo. No caso do modelo de regulação cruzada, as simulações apresentadas formam uma base interpretativa para futuros

estudos do repertório das células T na periferia. Para além disso, essas simulações foram fundamentais para pesar os prós e contras dos delineamentos experimentais usados nos actuais estudos de repertórios. Contudo, a combinação entre modelação e dados experimentais envolveu um rol de dificuldades, tal como foi ilustrado no Capítulo 7, onde não se puderam retirar conclusões definitivas devido à dimensão insuficiente das amostras e, nalguns casos, a ausência dos dados em bruto. Para evitar estas dificuldades, é necessário uma maior interligação entre experimentação e modelação de forma a satisfazer mutuamente as necessidades de cada abordagem. Só assim parece ser possível obter respostas mais concretas sobre sistemas biológicos complexos, tal como é o caso do repertório das células T.

RÉSUMÉ

Au sein du système immunitaire des gnathostomes, les cellules T orchestrent la réponses immune. Elles expriment collectivement une grande diversité de récepteurs reconnaissant les antigènes et constituant ce que l'on appelle le répertoire lymphocytaire T. Ces récepteurs sont générés lors de l'ontogénie des cellules T dans le thymus, par le biais d'un processus aléatoire d'assemblage de différents segments géniques appelé recombinaison V(D)J. Grâce à la grande diversité des récepteurs ainsi générés, les cellules T sont capables de réagir à toute sorte d'antigènes. Ainsi, le répertoire lymphocytaire T doit être structuré de manière à assurer une réponse efficace contre les pathogènes tout en évitant des réactions indésirables contre les antigènes du soi.

L'objectif de cette thèse est de comprendre comment les mécanismes conduisant à la structuration du répertoire lymphocytaire T. Deux principales directions ont été suivies. D'abord, nous avons étudié les premières étapes de la formation du répertoire en se concentrant sur les mécanismes de la recombinaison V(D)J des cellules T. Nous avons choisi d'étudier les cellules T $\gamma\delta$ car celles-ci ne recombinent que deux isotypes V-J chez la souris adulte, conduisant à la définition d'un modèle mathématique qui permettait de suivre tous les états possibles de réarrangements consécutifs au processus de recombinaison. La deuxième ligne suivie, motivée par le rôle fondamental des cellules T régulatrices dans le contrôle de l'immunité et l'autoimmunité, a consisté à étudier les mécanismes de sélection périphérique structurant le répertoire des cellules T conventionnelles et régulatrices. Nous avons aussi caractérisé la diversité structurelle de ces deux répertoires dans le thymus et en périphérie. Au cours de ce travail, nous avons aussi obtenu une meilleure vue de la diversité structurelle des cellules T CD4 et CD8 chez les souris ayant une diversité limitée du TCR. Nous avons abordé ces questions par le biais de différents modèles mathématiques et statistiques.

Dans les Chapitres 2 et 3, nous avons étudié les réarrangements du locus TCR γ qui encode la chaîne $\gamma\delta$ ces cellules T. Comme seulement deux isotypes V-J sur le locus γ sont recombinés chez la souris adulte, il a été possible d'énumérer tous les états de recombinaison d'une cellule précurseur et de développer des modèles de chaîne de Markov pouvant suivre ces états lors du processus de recombinaison (Chapitre 2). En confrontant ces modèles aux données expérimentales, nous montrons que le réarrangement V-J d'une cellule précurseur ne peut avoir lieu qu'en une période limitée lors de laquelle un locus est ouvert pour le réarrangement, puis le second locus est ouvert après un certain temps (accessibilité allélique asynchrone). De plus, l'occurrence d'un réarrangement productif V-J conduit à l'arrêt de tout le processus de réarrangement sur le locus γ (mécanisme de rétro-contrôle). Il est intéressant de noter que les taux estimés de réarrangement pour les différents isotypes suggèrent un compromis entre maximiser la probabilité de la survie de la cellule précurseur et minimiser la

probabilité d'occurrence de deux réarrangements productifs dans la même cellule précurseur. Dans le Chapitre 3, nous donnons les évidences expérimentales soutenant cette hypothèse.

Dans le Chapitre 4, nous étudions l'effet de la diversité fonctionnelle sur la structuration du répertoire des cellules T effectrices et régulatrices dans la périphérie. Pour cela, nous appliquons le modèle de régulation croisée qui décrit essentiellement un réseau d'interactions entre les cellules effectrices et les cellules présentatrices d'antigènes (APC). Dans ce modèle, la survie des cellules T est favorisée par leur interaction avec des APCs et les cellules régulatrices, pour pouvoir proliférer et exercer leur fonction régulatrice, requièrent la réception de facteur de croissance émis par des cellules T effectrices. Reprenant les publications basées sur ce modèle, nous faisons une revue détaillée des données expérimentales compatibles avec ses hypothèses et prédictions. Nous avons étudié le répertoire périphérique en étendant simplement le modèle en considérant que tout clone de cellule T reconnaît un ensemble exclusif d'APCs (pas de compétition interclonale pour les APCs). Le résultat le plus important de ce modèle est que le répertoire des cellules T en périphérie peut être divisé en trois sous-ensembles de clones de cellules T qui dépendent de la densité des APCs que ces clones reconnaissent. Le premier sous-ensemble est composé de quelques clones à courte durée de vie, interagissant avec quelques APCs, et qui ne peuvent donc pas être maintenus en périphérie. Le deuxième ensemble contient un grand nombre de clones de cellules T effectrices dont l'expansion n'est limitée que par la disponibilité d'APCs. Comme ces clones n'ont pas de cellules régulatrices, ce sous-ensemble est enclin à réagir à des antigènes étrangers. Le troisième sous-ensemble est constitué de quelques clones de cellules T auto-réactives, faisant coexister des cellules effectrices et régulatrices qui régulent mutuellement leur croissance. Nous proposons que c'est ce sous-ensemble qui permet d'éviter, grâce à la présence de cellules régulatrices, des réponses immunitaires délétères contre des antigènes du soi.

Dans le Chapitre 5, nous étendons le modèle précédent de régulation croisée à un scénario plus réaliste, dans lequel chaque clone de cellule T reconnaît un ensemble aléatoire et non-exclusif d'APCs (compétition interclonale pour les APCs). Nos simulations démontrent que, en général, la sélection périphérique favorise les clones de plus grande réactivité croisée dont la fréquence dans le répertoire augmente au cours du temps. De plus, ce qui n'était pas possible avec le modèle précédent, nous avons pu montrer que les cellules régulatrices T pouvaient avoir un répertoire exclusif. Dans des simulations pouvant être assimilées à des expériences de transfert adoptif, nous avons aussi montré que la partition des répertoires décrite précédemment est maintenue. Lorsqu'un flux thymique constant est introduit dans le modèle, trois structures de répertoires distinctes apparaissent dans les simulations. Une structure est associée à une prédominance de cellules T effectrices, même pour des clones qui reconnaissent des APCs en grande densité et qui devraient être sous le contrôle de cellules régulatrices. Cette situation peut être associée au phénotype dévastateur chez les souris présentant une mutation du facteur de transcription *Foxp3*, lequel est essentiel pour la génération et la fonction des cellules régulatrices. Une autre structure du répertoire est caractérisée par la prédominance de cellules régulatrices T. Cette structure devrait être évitée dans l'organisme car toutes les réponses immunes, y compris celles contre les pathogènes nocifs, sont soumises à l'intervention de cellules régulatrices. La troisième structure mon-

tre une partition des répertoires, similaire à celle obtenue dans les simulations du chapitre précédent. Comme précédemment, cette dernière partition paraît adéquate pour décrire un système immunitaire efficace. Cependant, cette structure "saine" ne constitue qu'une fraction réduite de tous les répertoires que nous avons observé dans les simulations. Comme les simulations utilisent un flux thymique constant, elles tendent à une structuration du répertoire présentant une sur-représentation soit des cellules T conventionnelles, soit des cellules T régulatrices. Nous proposons que pour éviter cela, il faut faire décroître le flux thymique au cours du temps, de façon à maintenir le répertoire dans sa forme saine. Ainsi, contrairement aux idées reçues, l'involution thymique pourrait avoir un effet bénéfique sur l'organisme, évitant une auto-immunité ou une régulation très forte de toutes les réponses immunes. Un autre résultat important de ce chapitre concerne l'observation de trois classes de répertoires avec un même ensemble de paramètres. Ainsi, la stochasticité de la formation du répertoire pourrait être une des explications de la pénétrance incomplète fréquemment observée chez les modèles animaux qui développent spontanément des maladies auto-immunes ou encore du taux de concordance réduit pour les maladies auto-immunes chez les jumeaux. Nous montrons aussi que les trois structures de répertoires ci-dessus sont plus clairement différenciées au sein de la population résidente. Pour chaque structure de repertoire, nous énonçons des prédictions qualitatives sur les propriétés de diversité structurelle comme, par exemple, la relation entre les diversités des cellules T conventionnelles et régulatrices, ou encore la corrélation entre les distributions respectives des tailles clonales.

Dans le Chapitre 6, nous établissons des méthodes statistiques pour évaluer les prédictions du modèle de régulation croisée sur des données expérimentales de la littérature. Nous introduisons d'abord les modèles poissonniens de données d'abondance (PAM) pour estimer la diversité des récepteurs de cellules T (TCR) et la distribution sous-jacente des tailles clonales. Cette classe de modèles comprend différentes formes de distributions de la taille clonale, comme la loi log-normale ou la loi Gamma. Pour ajuster le modèle poissonnien aux données expérimentales, nous avons développé une librairie PAM pour le langage R. En guise d'illustration, nous analysons des données thymiques et périphériques de cellules T CD4 and CD8 TCR chez des souris pour lesquelles la diversité TCR est extrêmement réduite. Nos résultats suggèrent que les échantillons de TCR CD4 and CD8 périphériques sont bien décrits par le modèle Poisson-Log normal, ce qui suggère une distribution de probabilité log-normale de la distribution de la taille clonale.

Dans le Chapitre 7 nous appliquons les modèles poissonniens d'abondance pour étudier les répertoires de TCR de cellules T conventionnelles et régulatrices dans le thymus et en périphérie. La distribution log-normale est ici encore la plus adéquate pour toutes les données analysées. C'est pourquoi ce modèle a été étendu au cas de deux échantillons pour analyser les relations entre toutes paires chevauchantes de répertoires. Nos résultats suggèrent que les thymocytes conventionnels sont plus divers que leurs contreparties régulatrices, ce qui est en accord avec l'hypothèse courante que ces dernières sont biaisées vers l'auto-réactivité. Nous n'avons pas pu écarter l'hypothèse d'une diversité similaire de répertoires des cellules T conventionnelles et régulatrices dans la périphérie. Plus important, par l'analyse de l'intersection des répertoire, nous n'avons pu exclure la possibilité que les répertoires des cellules régulatri-

ces et non régulatrices soient identiques. Ces deux résultats sont en accord avec la structure des répertoires liée à une sur-représentation des cellules T régulatrices. Cependant, nous n'avons pas pu non plus rejeter la fraction relativement faible de variants TCR partagés par des cellules régulatrices et non régulatrices en périphérie, ce qui, si tel est le cas, est en accord avec la structure de répertoire saine mentionnée précédemment. Ainsi, les données disponibles ne semblent pas contredire les prédictions du modèle de régulation croisée pour les répertoires des cellules T en périphérie chez des individus non auto-immuns.

Dans le Chapitre 8 nous discutons les points les plus importants de cette thèse, c'est-à-dire de la vraisemblance d'un répertoire exclusif des cellules T régulatrices et des explications possibles pour les différentes structures de répertoires observées chez des modèles animaux aux diversités réduites de TCR. Nous discutons aussi des raisons de la bonne adéquation du modèle Poisson-log normal pour la plupart des données. En utilisant le modèle de régulation croisée pour aborder cette question, la distribution sous-jacente de la taille clonale semble être un mélange de différentes distributions associées à différentes populations (par exemple, cellules récemment émigrées du thymus et différentes populations résidentes). Considérant ce résultat, nous argumentons qu'en général, le bon ajustement d'une distribution est dû soit à une taille réduite d'échantillon, soit à une "petite" diversité de TCR chez la souris analysée. Nous profitons pour établir un ensemble de recommandations pour des études futures des répertoires de TCR, pour mieux vérifier les prédictions sur les répertoires des cellules T en périphérie. Pour cela, nous préconisons une augmentation de la taille des échantillons, l'utilisation du séquençage du TCR au niveau d'une cellule et l'étude de l'évolution des répertoires des cellules T d'un individu au cours du temps (études longitudinales).

Finalement, nous concluons que la modélisation mathématique a fourni des indications sur la façon dont les répertoires de cellules T sont structurés. Il est important de signaler que la plupart des conclusions ci-dessus n'ont pu être émises que par notre préoccupation constante d'être en relation étroite avec les données expérimentales. Dans le cas de la recombinaison du locus $TCR\gamma$, nous avons développé des modèles mathématiques avec l'objectif de les ajuster directement aux données, ce qui nous a permis d'énoncer des prédictions plus réalistes sur les résultats du processus. Dans le cas du modèle de régulation croisée, nos simulations semblent fournir une base importante pour l'interprétation d'études futures sur le répertoire de cellules T. De plus, les simulations réalisées avec ce modèle ont été fondamentales pour comprendre la nécessité de modifier les plans d'expériences actuels pour étudier le répertoire de cellules T, telles que l'emploi d'études longitudinales. Cependant, la combinaison de nos modèles avec des données expérimentales s'est révélée être difficile comme illustré dans le Chapitre 7, quand des tailles d'échantillons insuffisantes conduisent à des résultats peu probants ou que l'absence de données brutes ne permettent pas une analyse statistique détaillée. Cela montre les difficultés inhérentes au type d'approches adoptées dans cette thèse. Ces difficultés pourraient être partiellement surmontées par une plus grande collaboration entre les équipes expérimentales et théoriques. C'est probablement la seule voie pour mieux comprendre des phénomènes biologiques complexes comme la structuration des répertoires des cellules T au cours de ses différentes étapes.

ACKNOWLEDGEMENTS

This work was financially supported by Fundação para a Ciência e Tecnologia (fellowship SFRH/BD2004/19810) and was hosted by the Instituto Gulbenkian de Ciência (IGC) in Oeiras.

I would like to thank:

To Jorge Carneiro, my supervisor, who always believed in my scientific skills more than myself. Almost 8 years have passed since the day I stepped into your office as a fresh statistician and joined the Theoretical Immunology group. So many things I learned since then, scientifically and personally. How can I thank you for that? Words cannot express all my gratitude to you, even when I think about our hard discussions during the lab meetings (it always made part of our *modus operandi*). Now it is time to move on and look for other challenges, as you constantly advised to me during the past year, and this time I will follow your advice. However, I want you to know that the time I spent in your group was so rewarding and fulfilling that it is hard to leave without feeling a big lump of sorrow rising up from my chest. Obrigado por tudo, Jorge!

To Manuel Vilanova for being my always-alert co-supervisor. Thank you as well for being my bridge with ICBAS, and for all the warm welcomes every time I went there. You made me feel at home.

To Carlos Daniel Paulino from Instituto Superior Técnico for all of his support since my graduation in Applied Mathematics. You were always present, always available, always supportive, but always demanding the best of me! Our discussions, our conversations, our outgoing saturday nights and, of course, our laughs had shaped me to what I am today! More than a professor, you are a friend and I hope to keep in this way for many, many years. Valeu, Daniel?!? Grande abraço!

To José Faro from University of Vigo, Spain, for introducing me the problem of T-cell diversity estimation that became an important part of my thesis. Our discussions regarding this topic and the trips to Vigo were fundamental to fully understand the pros and cons of available experimental studies on T-cell repertoires. Unha aperta, José! Besitos para Pilar e Ana!

To Pablo Pereira from Pasteur Institute, France, for the collaboration regarding the $\gamma\delta$ T-cell project. Thanks for believing in the predictions of the mathematical models and trying to test them experimentally.

To Margarida Correia-Neves from University of Minho for being always available to discuss with me the data of LTD mice. Your replies to my emails were fundamental to obtain some results in the thesis.

To Carlos Penha-Gonçalves and Henrique Teotónio for their great job in my PhD thesis committee at the IGC. Your good advices and assertive comments at different stages of my PhD project were priceless. I hope to keep in touch with you guys!

To Ricardo Paiva, Íris Caramalho, Andreia Lino for sharing their preliminary results with me, and for calling my attention to important references on regulatory T cells.

To Claudine Chaouiya and Julie Ribot for the help in the French résumé of the thesis. Merci beaucoup!

To Inês Ferreira for taking over my shoulders the burden of image manipulation and composition when I could not longer stand looking at the thesis. That's what friends are for and you were there. Thanks also for the phen, fan, and fun that we had together this past year. I deeply hope you can find a (fair) way to win Catan!

To IGC for being a great place to do Science. More than an institution is a place of people and, to all of them, a huge thanks. The courses, the seminars, the students' meetings, the journal clubs and, of course, the barbecues, the parties, the beer hours, and the weekly football matches would always be in a very special place of my heart.

And now some words in portuguese...

Quando entrei na escola primária do Sport Algés e Dafundo num ano já longínquo, estava longe de imaginar que a minha vida profissional me levaria a um doutoramento em Ciências Biomédicas. Em criança faziam-me a eterna pergunta "o que queres ser quando fores grande?", ao que eu respondia com um encolher de ombros indeciso. Gostava de fazer tantas coisas que não sabia o que escolher. Tenista? Guitarrista? Informático? Mas nunca me passou pela cabeça algo relacionado com Ciência. Com o decorrer dos anos, escolhas foram-se fazendo, umas vezes com tempo suficientemente para reflectir sobre o futuro, outras tantas sob o efeito do acaso que fez o mudar o rumo da minha vida de um dia para outro. Agora aqui estou eu a escrever estas linhas a comprovar que tudo isto não foi um sonho mas sim realidade.

Todos sabemos que fazer doutoramento é uma empreitada complicada e desgastante e, como tal, precisamos sempre de uma equipa de apoio capaz de combater comigo as inúmeras vicissitudes que acompanham este tipo de trabalho. Ao longo dos anos, sinto que estive rodeado de pessoas fantásticas que me ajudaram a crescer como pessoa e que me tornaram mais forte. Hoje ponho preto no branco todo o meu agradecimento para com elas. Se exclui alguém indevidamente, peço desde já as minhas desculpas.

Quero, então, agradecer:

Aos meus amigos de sempre, especialmente, Nuno Silva, Luís Machado, Filipa Martins, Vitor Conceição, Pedro Silva, Lúcia Belourico e Paulo Balão. "O que interessa é estarmos juntos e sermos felizes!", lembro-me de alguém dizer numa certa passagem de ano. Agora, acrescento "e termos amigos como vocês".

Ao pseudo-suíço José Meninas, que disponibilizou prontamente a sua casa em Lausanne para me concentrar exclusivamente na escrita da tese. Obrigado por me teres mantido focado nas minhas prioridades, especialmente quando os estímulos exteriores eram bem mais interessantes.

Aos Ricardo Perdigão e Elisabete Pereira, os meus colegas de licenciatura com quem fiz a maior parte dos trabalhos de curso. Partilhámos dificuldades, raivas, dúvidas existenciais, indecisões sobre o futuro, e claro, os sucessos tanto escolares como pessoais. Relembro com especial carinho aquelas noites longas de trabalho no IST, em que o nosso sentido de entreajuda fez ultrapassar com um sorriso nos lábios as adversidades que o curso nos colocou inúmeras vezes. O que não nos matou, tornou-nos mais fortes!

À Patrícia Ferreira, com quem passei as tormentas do Mestrado. Juntos vencemos as exigências de uma parte escolar que nos puxava para o isolamento e para atirmos a toalha ao chão. Obrigado, ainda, pela ajuda que me deste neste último ano de tese em que te disponibilizaste, sem qualquer ponta de hesitação, em corrigir quase na totalidade os trabalhos e exames da nossa cadeira de Bioestatística com a querida amiga Cláudia Pascoal (um grande beijo para ti, Margarette!). Desejo do fundo do coração que tu e a Cláudia consigam concluir com sucesso esta árdua tarefa que é fazer um doutoramento. Eu estarei cá para o que der e vier.

A todos os colegas do grupo de Imunologia Teórica que me têm aturado com música barulhenta no gabinete, rosnadelas ao vizinho do lado, boladas na parede (e não só) ao longo destes anos aqui no IGC (Rui Gardner, Tiago Paixão, Kalet Léon, João Sousa, Daniel Ferreira, Tom Weber, Ramiro Magno, Afonso Guerra, Luciana Ferrari, Gonçalo Picado e Thiago Guzella). Obrigado pela vossa enorme disponibilidade em discutir comigo a confusão do meu trabalho, as minhas inesgotáveis dúvidas sobre Biologia, Imunologia e áreas afins. Um beijo muito especial para as lourinhas do grupo, Carline van den Dool e Íris Vilarés, que trouxeram um outro ambiente para o grupo. A vossa amizade foi uma fonte inesgotável de energia positiva durante a tese. Van de ene op de andere dag werd Leiden een illusie. Barcelona wordt mijn nieuwe thuis. Kom me bezoeken! Mijn huis is jullie huis (dankje Dani voor de vertaling).

À Mariana Silva, que soube dar-me o seu sorriso e amizade incondicional neste último ano. Os momentos que passámos juntos foram, sem eu dar conta, o remédio perfeito para curar instantaneamente as insanidades temporárias que surgiram amiúde neste período. Go, surfer girl, go!

Ao Rui Gardner e Tiago Paixão por terem sido o meu braço direito e, às vezes, também o esquerdo durante o meu mestrado e doutoramento. Obrigado por terem contribuído para um ambiente excepcional de trabalho. Sinto-me em dívida para convosco desde a nossa viagem a Praga em que me transformaram num pequeno "monstro"! Os efeitos são ainda visíveis! Pergunto-vos: não ouvem de vez enquanto alguém dentro da vossa cabeça a gritar "Coschutz" e "mass bier, mass bier, mass bier"? Eu oiço!

À Tânia Lourenço, a quem mais devo por estar a escrever estas linhas. Tomei contigo muitas das decisões que me conduziram onde estou hoje. Quis a vida que a escuridão da noite tivesse chegado ao nosso lar e, por isso, não partilharemos juntos a conclusão desta minha etapa de vida. Contudo, quero que saibas que estás presente em cada palavra impressa desta tese.

Por fim, dedico esta tese à minha família, especialmente, a Eurico Sepúlveda, Graça Santos e Teresa Sepúlveda, que nunca vacilaram em apoiar-me, mesmo quando eu merecia receber umas valentes reprimendas. Uma última palavra para a minha avó Deolinda cuja memória perdura em mim até hoje. Ela seria certamente a pessoa mais orgulhosa do mundo neste momento.

LIST OF PUBLICATIONS INCLUDED IN THE THESIS

This thesis is based on the following papers:

Chapter 2

Sepúlveda, N. , Boucontet, L. Pereira, P. and Carneiro, J. (2005). Stochastic Modeling of T Cell Receptor γ Gene Rearrangement. *J. Theor. Biol.*, 234:153-165.

Chapter 3

Boucontet, L. Sepúlveda, N., Carneiro, J. and Pereira, P. (2005). Mechanisms controlling termination of V-J recombination at the TCR γ locus: implications for allelic and isotypic exclusion of TCR γ chains. *J. Immunol.*, 174:3912-3919.

Chapter 4

Carneiro, J., Léon, K., Caramalho, I., van den Dool, C., Gardner, R., Oliveira, V., Bergman, M.-L., Sepúlveda, N., Paixão, T., Faro, J. and Demengeot, J. (2007). When three is not a crowd: A crossregulation model of the dynamics and repertoire selection of regulatory CD4 T cells. *Immunol. Rev.*, 216:48-68.

Chapter 5

Sepúlveda, N., and Carneiro, J. (2009). Repertoire dynamics of peripheral regulatory and effector T cells competing for APCs. Manuscript in preparation.

Chapter 6

Sepúlveda, N., Paulino, C. D. and Carneiro, J. (2009). Estimation of T-cell repertoire diversity and clonal size distribution by Poisson abundance models. Submitted to *J. Immunol. Methods*

Chapter 7

Sepúlveda, N., Paulino, C. D. and Carneiro, J. (2008). Diversity of regulatory and effector T lymphocytes (in portuguese). In *Estatística: da teoria à prática* (M. M. Hill, M. A. Ferreira, J. G. Dias, M. F. Salgueiro, H. Carvalho, P. Vicente and C. Braumann, eds.), SPE Editions, Lisbonp, 513-524.

Sepúlveda, N. and Carneiro, J. (2009). Biodiversity estimation: quantifying the immune repertoire in the body's cellular ecosystem. To appear in the Proceedings of 16th European Young Statistician Meeting (CD-Rom).

LIST OF PUBLICATIONS NOT INCLUDED IN THE THESIS

Campino, S., Bagot, S., Bergman, M.-L., Almeida, P., Sepúlveda, N., Pied, S., Penha Gonçalves, C., Holmberg, D. and Cazenave, P. A. (2005). Genetic control of parasite clearance leads to resistance to *Plasmodium berghei* ANKA infection and confers immunity. *Genes Immun.*, 6:416-421.

Gil, S., Sepúlveda, N., Albina, E., Leitão, A. and Martins, C. (2008). The low-virulent African swine fever virus (ASFV/NH/P68) induces enhanced expression and production of relevant regulatory cytokines (IFN α , TNF α and IL12p40) on porcine macrophages in comparison to the highly virulent ASFV/L60. *Arch. Virol.*, 153:1845-1854.

Ramos, A. C., Martins, M. A., Okorokova-Façanha, A. L., Olivares, F. L., Okorokov, L. A., Sepúlveda, N., Feijó, J. A. and Façanha, A. R. (2009). Arbuscular mycorrhizal fungi induce differential activation of the plasma membrane and vacuolar H⁺-pumps in maize root system. *Mycorrhiza*, 19:69-80.

Sepúlveda, N. (2009). Surfando nos mares da Biologia: peripécias de um jovem estatístico aspirante a biólogo. *Boletim SPE*, Primavera de 2009, 75-81.

Sepúlveda, N. and Carneiro, J. (2007). Que loci controlam as proporções das diversas populações de timócitos? In *Estatística: Ciência Interdisciplinar* (M. E. Ferrão, C. Nunes, and C. A. Braumann, eds), SPE Editions, Lisbon, 729-740.

Sepúlveda, N., Paulino, C. D., Carneiro, J. and Penha-Gonçalves, C. (2007). Allelic penetrance as a tool to model two-locus interaction in complex binary traits. *Heredity*, 99:173-184.

Sepúlveda, N., Paulino, C. D. and Penha-Gonçalves, C. (2005). Modelos de interação genética: uma nova modelação e análise bayesiana. In *Estatística Jubilar* (P. Rodrigues, E. Rebelo, and F. Rosado, eds.), SPE Editions, Lisbon, 735-746.

Sepúlveda, N., Paulino, C. D. and Penha-Gonçalves, C. (2006). Bayesian two-gene interaction models in complex binary traits. In *Statistics in Genomics and Proteomics* (W. Urfer, and M. A. Amaral Turkman, eds), CIM, Lisbon, 91-102.

Sepúlveda, N., Paulino, C. D. and Penha-Gonçalves, C. (2009). Bayesian analysis of allelic penetrance models for complex binary traits. *Comp. Stat. Data Anal.*, 53:1271-1283.

Souto-Carneiro, M. M., Fritsch, R., Sepúlveda, N., Lagareiro, M. J., Morgado, N., Longo, N. S. and Lipsky, P. E. (2008). The NF- κ B canonical pathway is involved in the control of the exonucleolytic processing of coding-ends during V(D)J recombination. *J. Immunol.*, 174:3912-3919.

LIST OF ABBREVIATIONS

APC Antigen-presenting cell

BCR B-cell receptor

CDRs Complementary-determining regions, CDR1, CDR2 and CDR3, located in the variable regions of TCRs. CDR1 and CDR2 interact with MHC molecule per se while CDR3 interacts with the antigen presented by the MHC molecule

CRM Crossregulation model

DN thymocytes Double negative $CD4^{-}CD8^{-}$ thymocytes

DP thymocytes Double positive $CD4^{+}CD8^{+}$ thymocytes

FACAM Model with allelic accessibility and feedback

LN Lymph nodes

LTD mice Limited TCR diversity mice developed by Correia-neves et al. (Immunity, 14, 21-32, 2001)

MHC Major histocompatibility complex

PAM Poisson abundance models

PAM A package in R language to fit the Poisson abundance models

PFM Pure feedback model

pMHC peptide-MHC complexes

RT-PCR Real-time polymerase chain reaction

SP Single positive thymocytes, either $CD4^{+}CD8^{-}$ or $CD4^{-}CD8^{+}$

TCR T-cell receptor

TCR^{mini} Limited TCR diversity mice developed by Pacholczyk et al. (Immunity, 25, 249-259, 2006)

T_E cell Effector T cell ($Foxp3^{-}CD25^{+}CD4^{+}$ or $Foxp3^{-}CD25^{-}CD4^{+}$ T cell)

T_R cell Regulatory T cell ($Foxp3^{+}CD25^{+}CD4^{+}$ T cell)

TWM Time window model

LIST OF FIGURES

1.1	Diversity of the $\alpha\beta$ TCR.	6
1.2	Important features for the characterization of a TCR repertoire: structural and functional diversity	9
1.3	Most important steps in T-cell differentiation in the thymus.	10
1.4	Instructive-selected model versus stochastic-selected model for thymic T_R -cell generation.	14
2.1	Structure of the mouse γ gene locus and the rearrangement possibilities.	34
2.2	Probabilities of some γ loci configurations as a function of the time elapsed since the $\gamma\delta$ precursor cell initiated rearrangement.	44
2.3	Tradeoff between yield and allelic exclusion γ gene rearrangement.	45
3.1	Most $\gamma\delta$ thymocytes bear an unique $\gamma\delta$ TCR at the cell surface.	56
3.2	Rearrangement status of the J1 and J4 regions in progenies of individual $V4^+$ and $V1^+$ thymocytes.	58
3.3	Inhibition of endogenous V4-J1 rearrangements by $\gamma\delta$ TCR transgenes.	60
3.4	The number of $\gamma\delta$ thymocytes correlates with the probability of producing a functional $\gamma\delta$ TCR in progenitor cells.	62
4.1	The Crossregulation Model.	78
4.2	Population dynamics according to the Crossregulation Model.	81
4.3	Crossregulation among regulatory T cells and effector T cells in the presence of APCs in vitro.	82
4.4	Candidate molecular mechanisms for APC-dependent suppression of T_E cells by T_R cells.	86
4.5	Individual cell-oriented simulations of in vitro cultures of regulatory T cells, effector T cells, and APCs.	87
4.6	Equilibrium densities of specific T_R and T_E cell populations as a function of the density of their cognate APCs.	88
4.7	Partition of the repertoire in specificity subsets with and without T_R cells as a function of the density of cognate APCs.	90
4.8	Dependence of the incidence of autoimmunity on the shape of the distribution of clonal specificities over the density of cognate APCs and frequency of regulatory T cells.	93
5.1	The Crossregulation Model.	113

5.2	Steady state analysis of the Crossregulation as a function of APC density per T-cell clone.	116
5.3	Network between T-cell clones and APCs.	119
5.4	Dynamical behavior of some variable of interest in a representative simulation without thymic export.	120
5.5	The partition of the repertoire and some structural properties in simulations without thymic export.	121
5.6	Three distinct classes of repertoire arise in the simulations with thymic export. . .	123
5.7	Three repertoire classes are distinguished by structural properties of the resident populations.	126
5.8	Effect of some parameters on the structure of the T-cell repertoire.	127
6.1	Assessment of the information contained in the samples assuming they were obtained sequentially.	146
6.2	Robustness of TCR diversity estimates according to the Poisson abundance models.	149
6.3	Clonal size distributions of thymocyte populations.	151
6.4	Clonal size distributions of LN T cells.	152
6.5	Large sample sizes are required to discriminate the Lognormal-Poisson, Geometric, and Yule models from each other in the case of LN CD4 ⁺ T cell compartment.	154
7.1	Probability of obtaining a new TCR variant as a function of the respective sample size.	167
7.2	Analysis of TCR diversity estimates based on the Lognormal-Poisson model. . .	168
7.3	Analysis of the intersection repertoire based on mouse 1 data.	170
7.4	Comparison between peripheral T _R and T _E TCR repertoires	172
8.1	Illustration of putative differences between the distribution of APC density recognized by T-cell clones generated in the thymus of LTD and TCR ^{mini} mice. . . .	188
8.2	Clonal size distributions predicted by different Crossregulation models	196

LIST OF TABLES

1.1	The number of different gene segments per mouse locus encoding each TCR chain, and the respective potential and effective receptor diversity.	7
2.1	$V_{\gamma}1J_{\gamma}4$ and $V_{\gamma}4J_{\gamma}1$ rearrangements in $V_{\gamma}1+$ and $V_{\gamma}4+$ cells.	35
2.2	Predicted probabilities of the table entries under the time window mechanism . .	36
2.3	States and their links of the Markov chain describing the pure feedback model. .	37
2.4	Results of goodness-of-fit test for the three rearrangement mechanisms.	41
2.5	Best fit of the three mechanisms using the maximum likelihood methodology. . .	42
2.6	Other estimates derived from the maximum likelihood estimates of model with allelic accessibility and feedback at the end of the time window.	43
3.1	Limited role of the pre-TCR in the termination of V_{γ} to J_{γ} rearrangements.	61
5.1	Summary of parameters/variables related to basic T-cell dynamics, characterization of APC population, and interaction between T cells and APCs.	115
6.1	Number of distinct CDR3 $TCR\alpha$ sequences as a function of their repeats in thymic and peripheral $CD4^{+}$ and $CD8^{+}$ samples taken from a LTD mouse.	145
6.2	Diversity estimates of DP $CD3^{low}$, SP $CD4^{+}$ and $CD8^{+}$ thymocytes, and LN $CD4^{+}$ and $CD8^{+}$ T cells, and the corresponding p-value of Pearson's goodness-of-fit test for different Poisson abundance models when fitted to the whole data. .	148
6.3	Characterization of clonal size distributions for selected Poisson abundance models.	153
6.4	TCR diversity estimates for LN $CD4^{+}$ T cells in three mice and in their pooling data.	155
7.1	Frequency of unique TCR sequences as a function of their repeats in T_R and T_E TCR samples from periphery and thymus of two distinct mice.	164
7.2	Outlier TCR sequences in thymic T_R samples and their number of repeats across samples.	168
7.3	Estimates of Morisita-Horn index between thymic and peripheral T cell compartments.	169
8.1	Main advantages and disadvantages of different experimental procedures and analysis.	192

CONTENTS

<i>Abstract</i>	i
<i>Resumo</i>	v
<i>Resumé</i>	xi
<i>Acknowledgements</i>	xv
<i>List of Publications Included in the Thesis</i>	xix
<i>List of Publications not Included in the Thesis</i>	xxi
<i>List of Abbreviations</i>	xxiii
<i>List of Figures</i>	xxvi
<i>List of Tables</i>	xxvii
 <i>Part I General introduction</i>	 1
1. <i>General Introduction</i>	3
1.1 Basics of the immune system	3
1.2 The generation of diversity	5
1.3 Structural and functional diversity	7
1.4 Central tolerance through thymic selection	10
1.5 Peripheral tolerance by the action of regulatory T cells	12
1.6 Ontogeny of regulatory T cells	13
1.7 The shape of regulatory T-cell repertoire	15
1.8 Some shortcomings in the assessment of TCR structural diversity	17
1.9 In this thesis	18
Bibliography	20
 <i>Part II Recombination mechanism of γ genes encoding the T cell receptor γ chain</i>	 29
2. <i>Stochastic Modeling of T Cell Receptor γ Gene Rearrangement</i>	31
2.1 Introduction	31
2.2 Methods	33

2.2.1	Experimental Data	33
2.2.2	Rearrangement Mechanisms and Models	34
2.3	Results	41
2.3.1	Analysis of γ loci structure in mature thymocytes reveals a rearrangement mechanism involving both feedback and asynchronous chromosome accessibility	41
2.3.2	Further analysis of the rearrangement process according to FACAM	42
2.4	Discussion	44
	Bibliography	48
3.	<i>Mechanisms controlling termination of V-J recombination at the TCRγ locus</i>	53
3.1	Introduction	53
3.2	Materials and Methods	55
3.2.1	Mice	55
3.2.2	Abs, Flow cytometry and cell sorting	56
3.2.3	Cell cultures	56
3.2.4	Single cell and single clone PCRs, cloning and sequencing	56
3.3	Results	57
3.3.1	Most $\gamma\delta$ T cells express one antigen receptor at the cell surface	57
3.3.2	Extent of allelic and isotypic inclusion of the functionally most relevant J1 and J4 gene segments	57
3.3.3	Expression of a functionally rearranged $\gamma\delta$ TCR inhibits rearrangements at the endogenous TCR γ locus	59
3.3.4	$\gamma\delta$ T cell progenitors do not attempt all possible TCR γ rearrangements . .	59
3.3.5	The extent of TCR γ rearrangements found in $\gamma\delta$ T cells is better explained by a model in which initiation of TCR γ rearrangements at both alleles is asymmetric and is terminated upon production of a functional $\gamma\delta$ TCR	60
3.3.6	The number of $\gamma\delta$ thymocytes correlates with the probability of producing a selectable $\gamma\delta$ TCR	62
3.4	Discussion	62
	Bibliography	66
	<i>Part III Peripheral dynamics of regulatory and effector T cell repertoires</i>	71
4.	<i>Repertoire dynamics of peripheral regulatory and effector T cells without interclonal competition</i>	73
4.1	Introduction	73
4.2	Natural tolerance, clonal selection, and regulatory T cells	74
4.3	The Crossregulation Model	76
4.3.1	General Biological Principles	76
4.3.2	Postulates of the Model	77

4.3.3	Minimal Mathematical Formulation of the Model	78
4.3.4	Standard Behavior	80
4.4	Cellular and molecular mechanisms of crossregulation between regulatory and effector T cells	80
4.4.1	Crossregulatory interactions between regulatory and effector T cells in vitro	81
4.4.2	Crossregulatory interactions between regulatory and effector T cells in vivo	83
4.4.3	APC-dependence of interactions among antigen specific T cells	84
4.4.4	Several molecular mechanisms may lead to the same cell population dynamics	85
4.5	Peripheral selection of the preimmune repertoire dependent on crossregulation	87
4.5.1	Dependence on APC density	87
4.5.2	Stationary state in clonally diverse T_E and T_R cell populations	88
4.5.3	Partitioning of the repertoire in subsets with and without regulatory T cells	89
4.5.4	The Basic Ground for Self-nonself Discrimination	91
4.6	Constraints on Thymic Selection and Maturation of Regulatory T cells	91
4.6.1	Constraints on positive and negative selection, and on thymic differentiation of regulatory T cells	92
4.7	Coordination with the Innate Immune System	94
4.8	A Critique of the Crossregulation Model and Other Scenarios	96
4.8.1	Does size matter? Control of Lymphocyte Activation and Effector Function versus the Control of Clonal Expansion	97
4.8.2	Central versus peripheral Generation of Regulatory T cells	97
4.8.3	Specificity of the growth factors regulatory T cell populations	98
4.9	Concluding Remarks	98
	Bibliography	99
5.	<i>Repertoire dynamics of peripheral regulatory and effector T cells with interclonal competition</i>	109
5.1	Introduction	109
5.2	The Crossregulation Model	111
5.2.1	Mathematical formulation	112
5.2.2	Relevant properties of the basic Crossregulation model	114
5.2.3	Modeling the APC population	115
5.2.4	Connecting T cells with APCs	116
5.3	Results	118
5.3.1	Peripheral T-cell repertoire dynamics in the absence of thymic influx	118
5.3.2	Disentangling the T-cell repertoire in the case of a constant thymic influx	122
5.3.3	Effect of some parameters in shaping the repertoires	126
5.4	Discussion	128
	Bibliography	133

6. Poisson abundance models to estimate T cell receptor diversity and clonal size distributions	139
6.1 Introduction	139
6.2 Poisson abundance models	140
6.3 Estimation of Diversity	142
6.4 Estimation of the clonal size distribution	143
6.5 Package PAM	144
6.6 Example	145
6.6.1 Tackling the information contained in the samples	146
6.6.2 Estimating TCR diversity	147
6.6.3 Robustness of TCR diversity estimates	149
6.6.4 Estimating clonal size distributions	150
6.6.5 Large sample sizes are required to discriminate the models	152
6.7 Discussion	155
Bibliography	157
7. Diversity and overlap between regulatory and effector T cell repertoires in mice	161
7.1 Introduction	162
7.2 Materials and Methods	163
7.2.1 Experimental data	163
7.2.2 Estimation of TCR diversity	163
7.2.3 Quantification of the intersection repertoire	164
7.3 Results	166
7.3.1 Similar features among all TCR samples	166
7.3.2 The number of distinct TCR sequences observed in the samples is far from representing TCR diversity at the population level	166
7.3.3 Thymic T_E cells show higher TCR diversity than their thymic T_R counterparts	167
7.3.4 Peripheral T_R and T_E cells exhibit similar TCR diversity	169
7.3.5 Thymic T_R -cell repertoire might share all TCR variants with that of thymic T_E cells	169
7.3.6 Thymic repertoire of the same T cell subset also overlaps greatly its peripheral counterpart	170
7.3.7 T_R and T_E cells might share few or all TCR variants in the periphery	171
7.4 Discussion	172
Bibliography	176

8. General discussion	183
8.1 On the role of allelic exclusion on peripheral tolerance	183
8.2 On the shape of regulatory and effector T-cell repertoires	185
8.2.1 Distinct classes of peripheral T-cell repertoires predicted by the Cross-regulation model	185
8.2.2 Classifying peripheral T-cell repertoires of available limited TCR diversity mice	186
8.2.3 Similar shape for the peripheral repertoire predicted by a model of peripheral T_R -cell generation	188
8.3 On the experimental design and analysis of current TCR repertoires studies . . .	189
8.4 On the good fit of Poisson-Lognormal distribution for experimental data	191
8.4.1 Poisson-Lognormal model should be applied with caution	191
8.4.2 Immunological and ecological communities: the recurrent observation of a Lognormal pattern for species-abundance distributions	193
8.4.3 Can the Crossregulation model predict a Lognormal clonal size distribution?	195
8.5 Concluding remarks	198
Bibliography	198

Part I

GENERAL INTRODUCTION

1. GENERAL INTRODUCTION

The vast diversity of lifeforms is one of the hallmarks of life on Earth. This is particularly true at the microscopic level with an invisible world of microorganisms prompted to fight for their survival. In this battle for survival, multicellular organisms running under a homeostatic regime represent attractive recipients of nutrients. Parasites, bacteria, and viruses might infect them, causing perturbations on the normal behavior of the organism that ultimately lead to disease and death of the host. Nonetheless, multicellular organisms are not sick all the time because, throughout evolution, they acquired mechanisms to fight or, when needed, to tolerate invading pathogens. Such mechanisms are part of what we call the immune system, comprising a diversity of molecules and cellular components that face all the diversity of putative harmful infections that might occur throughout lifetime of an individual. In this line of thought, the most evident and visible function of the immune system is to protect the body against invading and potentially harmful microorganisms. Yet, to fulfill this function, the immune system must be able to discriminate efficiently a harmful pathogen from a body component. Once again, this discrimination seems properly done but by still unclear mechanisms. This thesis aims then to understand how the immune system is structured, or shaped, in order to be highly efficient to fight infections but, at the same time, not attacking the body. We focus our attention on dissecting the repertoire of one of the most important cellular components in the immune system, the T lymphocytes (T cells, for short).

Before entering in the depths of this thesis, it is convenient to introduce some basic background knowledge on the immune system, specially, regarding T-cell immunobiology. This is the purpose of this general introduction.

1.1 Basics of the immune system

As mentioned above, the immune system has a vital role in the maintenance of the healthy status of an organism. Its most evident function is to protect an individual from harmful pathogens. Yet, it is also important in physiology, such as, tissue homeostasis and wound healing (Jame-son and Havran, 2007; Linfert et al., 2009; Park and Barbul, 2004; Schwartz and Moalem, 2001).

Due to its great complexity, the immune system is divided in two main branches, the innate immune system and the adaptive immune system, which communicate with each other via different molecules and cellular components. The innate immune system is present in both invertebrates and vertebrates, while the adaptive immune system is only present in the later. Because of that, the adaptive immune system seems a more recent acquisition in evolutionary terms.

The innate immune system is the first line of defense against invading pathogens. At its disposal, it exhibits a great variety of humoral and cellular components that act upon antigens in a non-specific manner. On the one hand, humoral innate immune responses refer to the complement system that, upon activation of its cascade, leads to a massive amplification of a series of proteins able to clear pathogens via opsonization and cell lysis (Gasque, 2004). On the other hand, the innate immune system also comprises different cell types, such as macrophages, neutrophils, natural killers, eosinophils. Innate immune cells express germ-line encoded pattern recognition receptors (PPRs) that recognize and respond to conserved microbial structures, known as pathogen associated molecular patterns, like lipopolysaccharides and peptidoglycans commonly expressed in bacteria. Some of these PPRs are also able to send signals that activate the adaptive arm of the immune system (Palm and Medzhitov, 2009). In this regard, the family of Toll-like receptors in mammals, drosophila, or even in plants, is of particular interest, as they can recognize a wide variety of microbes, such as bacteria, DNA and RNA viruses, fungi and protozoa (Palm and Medzhitov, 2009; Takeda et al., 2003).

In spite of being highly efficient, as attested by the relatively infrequent disease in invertebrates, the innate immune system is not sufficient to face all different kinds of infections that might occur throughout lifetime of an individual, specially, when thinking in slow evolving vertebrates that are exposed to a fast changing microbial world. For example, some viruses might change drastically their phenotype by a high mutation and, because of these putative phenotypic changes, the innate immune cells might not recognize them. To cope with this fast evolving microbial world, evolution once again found a way to tackle this problem properly by the generation of an adaptive immune system. This branch comprises the army of T and B lymphocytes (also called T and B cells) that would join the fight against invaders when the innate immune system sends signal to do so. B and T lymphocytes exhibit a heterodimer receptor at the cell membrane, generated during B- and T-cell ontogeny by somatic recombination, the so-called V(D)J recombination process that will be discussed in more detail afterwards. The antigen receptor of B cells (BCR) is the immunoglobulin formed by two heavy chains (H and h) and two light chains (λ and κ). BCR is highly focused on native proteins, carbohydrates, and lipids. In its turn, the T cell receptor (TCR) is a heterodimer composed of α - and β -chains in the so-called $\alpha\beta$ T cells and γ - and δ -chains in the $\gamma\delta$ T cells. Since BCR and TCR are generated by somatic recombination, B and T cells exhibit collectively a highly broad repertoire of receptors that virtually recognize and react to all pathogens present in nature. Yet, for a full engagement of T and B cells in an immune response, antigen-presenting cells (APCs) – dendritic cells, macrophages, or even B cells – must show to these cells pathogen-derived antigens complexed with class I or class II major histocompatibility complex (MHC) molecules. Antigen-specific lymphocytes would then recognize and respond to the presented antigen, forming in this way a very specific immune response.

This thesis focuses on T cells and, thus, more emphasis will be given to these cells in this general introduction. A mouse contains around $1 - 2 \times 10^8$ T cells in total (Casrouge et al., 2000), while 10^{12} T cells are present in a human individual (Arstila et al., 1999). The large majority of the T cell pool consists of $\alpha\beta$ T cells, which are yet subdivided in two major lineages: $CD4^+$ and $CD8^+$ T cells that express CD4 and CD8 molecules at the cell surface,

respectively. These two molecules serve as transmembranar co-receptors in the interaction of TCR with MHC complexes, being essentially to signal transduction that ultimately lead to T-cell activation.

CD4⁺ T cells, also called T helper cells, have a fundamental role in the orchestration of the immune system, since they send signals to innate immune cells, B cells and their CD8⁺ counterparts to either start or dampen immune responses against a wide set of antigens, either self or foreign. To this end, they recognize antigen-derived peptides complexed with class II MHC molecules on the surface of APCs. The great importance of this subset in the immune system is particularly evident in HIV-infected individuals, which preferentially attacks these cells (Haase, 1999; McCune, 2001). When the number of CD4⁺ T cells drops below a certain limit, opportunistic infections arise that ultimately lead to death of the individual.

CD8⁺ T cells, also known as cytotoxic T cells, respond to a wide range of infectious agents, such as, viruses, bacteria and protozoa (Wong and Pamer, 2003). As opposed to CD4⁺ T cells, these cells respond to antigens presented by class I MHC molecules on the surface of infected cells. They are able to kill target cells via direct cytolysis mediated by the release of several effector molecules, such as perforin and Fas. In both human and mice, the CD4:CD8 ratio is under genetic control in peripheral organs, and tends to be skewed towards CD4⁺ T cells, ranging from 0.98 to 2.83 in inbred mouse strains (Sim et al., 1998), and from 0.39 to 7.43 in healthy humans (Amadori et al., 1995).

Finally, $\gamma\delta$ T cells represent a small fraction (1-5%) of the T lymphocytes circulating in the blood and peripheral organs of adult mice and humans. They are, however, located in certain tissues, such as, skin, intestine, and reproductive tract, comprising up to 50% of T cells in these organs. Their role in the immune system remains an enigma, because it differs with the structure of their TCRs, the stage of infection, and the tissue these cells occupy (Carding and Egan, 2002). As an example, the epithelia-associated $\gamma\delta$ T cells promote the development of $\alpha\beta$ T-cell responses in the initial stages of infection, while they can also contribute to tissue repair and regeneration after the resolution of the immune response through the production of cytokines and keratinocyte growth factor (Jameson and Havran, 2007).

1.2 The generation of diversity

One of the hallmarks of T-cell immunobiology is the large repertoire of distinct TCRs that these cells collectively express. It is this large TCR diversity that allows T cells to virtually recognize and respond to all microorganisms present in nature. TCR diversity is generated in the thymus during T-cell ontogeny by the V(D)J recombination process, in which non-contiguous gene segments encoding each TCR chain are assembled by genetic recombination. Mature TCR- α and TCR- γ genes are made out of a variable (V) and joining (J) segments (Figures 1.1A), while their β and δ counterparts are the end-product of productive recombinations of V, diversity (D) and J segments (Figure 1.1B). The junction of the J segment to the constant (c) region is made by post-transcriptional splicing. There are many gene segments per TCR-chain locus (Table 1.1), and thus a large diversity of TCR can be obtained not only by combinatorial rearrangement of different gene segments encoding each TCR chain, but also by the pairing of each chain

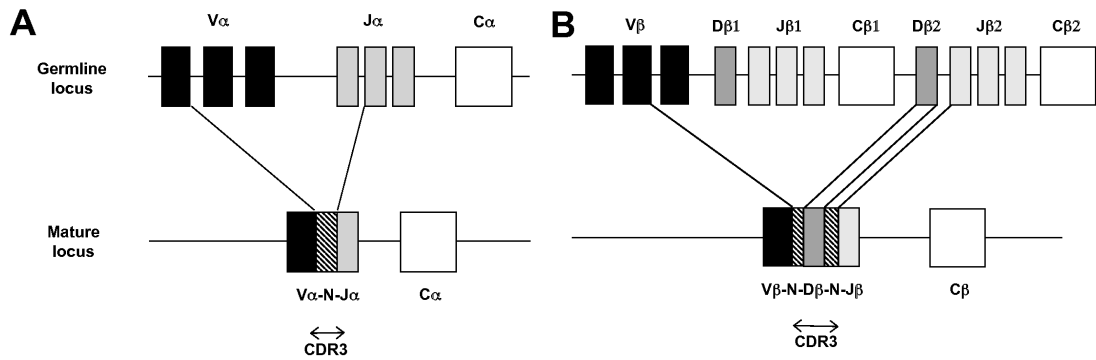


Fig. 1.1: Diversity of the $\alpha\beta$ TCR (adapted from Nikolich-Zugich et al. 2004). Mature and functional genes encoding α (A) and β (B) TCR chains are generated through the recombination of a variable (V), diversity (D), and joining (J) segments present in germline loci. Additional diversity is obtained by addition and deletion of nucleotides (N) (dashed boxes) at the junctions between recombined gene segments. The junction of the J segment to the constant (c) region is made by post-transcriptional splicing. Recombination of the γ and δ locus can be found in Figure 2.1 and in Carding and Egan (2002), respectively.

composing the TCR. Additional diversity is yet achieved by the removal and addition in the interface between V, D, and J gene segments. It is worth noting that similar V(D)J recombination happens in the genes encoding heavy and light chains of BCR during B-cell ontogeny in the bone marrow. B cells show yet other sources of immunoglobulin diversification, such as, gene conversion, somatic hypermutation, and class-switch recombination (Maizels, 2005). Gene conversion occurs mainly in chicken, where a single functional $V\lambda$ gene segment rearranged to a $J\lambda$ gene segment can be further diversified by the transfer of homologous sequences from upstream pseudo $V\lambda$ genes into it. During the so-called germinal center reactions, somatic hypermutation refers to a process of generating random point mutations to variable regions of BCRs in antigen-activated B cells as an attempt to achieve a higher affinity of the BCR for the antigen. Class-switch recombination is a process by which the constant region of a BCR is replaced by another downstream C gene segment to produce a BCR with similar specificity but with a different effector function.

Taking together the number of different gene segments, the imprecise joining between them, and the pairing of each chain, the number of possibilities of generating a distinct $\alpha\beta$ TCR is around 10^{15} in mice (Davis and Bjorkman, 1988). This number reflects then the potential diversity of $\alpha\beta$ TCR repertoire. Since there is $1 - 2 \times 10^8$ T cells in total, only a fraction of this TCR diversity would be present in an individual at any time point, the so-called effective TCR diversity. Currently, effective $\alpha\beta$ TCR diversity has been estimated in around 2×10^6 in mice (Casrouge et al., 2000). In spite of exhibiting a small number of gene segments encoding TCR γ and TCR δ chains, $\gamma\delta$ T cells are more prone to exhibit a higher diversity at V(D)J junctions. Therefore, the size of the potential $\gamma\delta$ TCR repertoire is higher than that of $\alpha\beta$ TCR repertoire, being estimated in 10^{18} (Davis and Bjorkman, 1988; Elliott et al., 1988). Yet, it is still unknown what is the effective $\gamma\delta$ TCR diversity present in an individual. Nonetheless, it is known that the $\gamma\delta$ TCR repertoire is generated as an ordered process, since specific $V\gamma$ segments are

Tab. 1.1: The number of different gene segments per mouse locus encoding each TCR chain and the respective potential TCR diversity (Casrouge et al., 2000; Davis and Bjorkman, 1988).

Gene segments	$\alpha\beta$ TCR		$\gamma\delta$ TCR	
	α	β	γ	δ
Variable	100	25	7	10
Diversity	0	2	0	2
Joining	50	12	2	2
Potential TCR diversity	1×10^{15}		1×10^{18}	

exported from the thymus at defined time periods during embryonic development, while other $V\gamma$ segments, such as $V\gamma 1$ and $V\gamma 4$, only appear in adult mice (Carding and Egan, 2002).

The generation of extremely diverse $\alpha\beta$ and $\gamma\delta$ TCR repertoires is then one of the fundamental mechanisms of the vertebrate adaptive immune system to keep pace to all fast evolving pathogens present in nature. Yet, among this large collection of TCRs that can be generated during T-cell ontogeny, there is the possibility of producing certain TCRs that might be self-reactive, that is, recognize body components, leading to erroneous immune responses against them. Therefore, it is crucial to understand how the T-cell repertoire is molded to avoid such situations but, at the same time, responding efficiently to harmful microorganisms.

In this regard, it is worth to mention Burnet's clonal selection theory was one of the first attempts to provide a coherent answer to the above problem (Burnet, 1957). Under this theory, autoimmunity was avoided by deletion of autoreactive clones during embryonic life, when all lymphocyte variants were generated. After birth, pathogen-specific clones would expand upon exposure to infection. Lederberg (1959) proposed alternatively that deletion of autoreactive lymphocytes would happen throughout life at immature stages of lymphoid lineage development. Accordingly, autoimmune diseases, such as type I diabetes or rheumatoid arthritis, would be caused by a deficiency in the deletion process that allows autoreactive lymphocyte to escape deletion and enter in circulation. The great impact of this theory in the immunological thinking cannot be overstated. Yet, some shortcomings were pointed out, namely, the presence of autoreactive B and T cells in the periphery of healthy individuals (Avrasmeas, 1991; Coutinho, 1989; Sakaguchi et al., 1995). To overcome the limitations of clonal selection theory, other theories were proposed since then, such as the two-signal model (Bretscher and Cohn, 1970) or the danger theory (Matzinger, 1994); for a comprehensive review on these and other theories, see refs. (Carneiro, 1997; León, 2002). Nowadays, the problem of control of autoimmunity allied to efficiency against infections gained a new aspect with the discovery of the so-called regulatory T cells, as we will see afterwards.

1.3 Structural and functional diversity

TCR plays an important role in antigen recognition and in the activation of the subsequent immune responses. The collection of TCRs expressed by T cells, the TCR repertoire, will determine the immunological status of the organism. We introduce here some important properties required for its characterization in terms of structural and functional diversity (Figure 1.2).

Structural diversity is related to TCR diversity at sequence level (Figure 1.2A). In this regard, a distinct TCR sequence, defined either at the nucleotide or amino-acid level, is a TCR variant, or more generically, a clonotype. TCR diversity is then the number of distinct TCR variant present in the repertoire. The number of cells bearing a certain TCR variant is referred to as the clonal size. The frequency (or probability) distribution of all clonal sizes present in the population is known as the clonal size distribution.

There are different approaches to assess structural diversity. One approach aims to determine the expression or the frequency of T-cell usage of different $V\beta$ or $V\alpha$ segments (see, for example, Matsutani et al. 2006; Takahashi et al. 1998). Yet, this provides solely an overall picture of the repertoire, neither tackling per se the underlying TCR diversity nor clonal size distribution. Other approach is to apply the immunoscope technique, also known as spectratype analysis (reviewed in Boudinot et al. 2008), based on a histogram (the spectratype) related to distribution of CDR3 length of the TCR among a T-cell pool. This technique provides another qualitative insight on the repertoire but gives no information on the underlying TCR diversity and clonal size distributions. The best that one can do with such approach is to detect qualitatively shifts in the repertoire under different experimental conditions or to compare different T-cell subsets (Kepler et al., 2005). Even in this situation, the observation of a difference between spectratype cannot be precisely determined, since the peak on the histogram can be due to an overrepresentation of a single clonotype or owing to many clonotypes exhibiting similar CDR3 length. Therefore, this approach seems a crude approach to assess structural diversity. Other approach is to enumerate each individual TCR sequences expressed in a T-cell sample. Since V(D)J recombination can lead to TCR repertoires with potentially large diversity, this approach requires a prohibited number of TCRs to be sequenced in order to have an accurate snapshot of the repertoires. Because of this, enumerating TCR sequences seems only feasible in experimental settings in which TCR diversity is somehow limited by some genetic trick, as recently done (Correia-Neves et al., 2001; Hsieh et al., 2004; Pacholczyk et al., 2006). In these studies, there are two experimental procedures to determine the TCR sequences. The first procedure is based on a pre-amplified cDNA extract of the TCRs expressed in the samples (Hsieh et al., 2004, 2006). This approach is particularly useful to estimate TCR diversity. Yet, because a cDNA extract coming from a pool of T cells is used, the appearance of several copies of the same TCR sequence cannot be imputed to a single or distinct T cells. Therefore, this procedure is not accurate to dissect the underlying clonal size distribution. The second procedure to obtain TCR sequences is to perform single-cell RT-PCR (Correia-Neves et al., 2001; Pacholczyk et al., 2006, 2007; Wong et al., 2007). This approach seems conceptually equivalent to collect individuals from a community and, therefore, it is appropriate to estimate not only TCR diversity but also the underlying clonal size distribution. However, single-cell RT-PCR does not hitherto work in 100% of cells. In this line of thought, single-cell RT-PCR might exhibit a bias on TCR sequence that fail to be sequenced. Therefore, this procedure still needs to be properly fine-tuned to avoid potential biases in the TCR composition of the samples.

It is known that different TCR sequences can recognize the same antigen and, thus, structural diversity does not capture all relevant information on the T-cell repertoire. Considering the number of antigens recognized by a given TCR, i.e., its cross-reactivity, what might be impor-

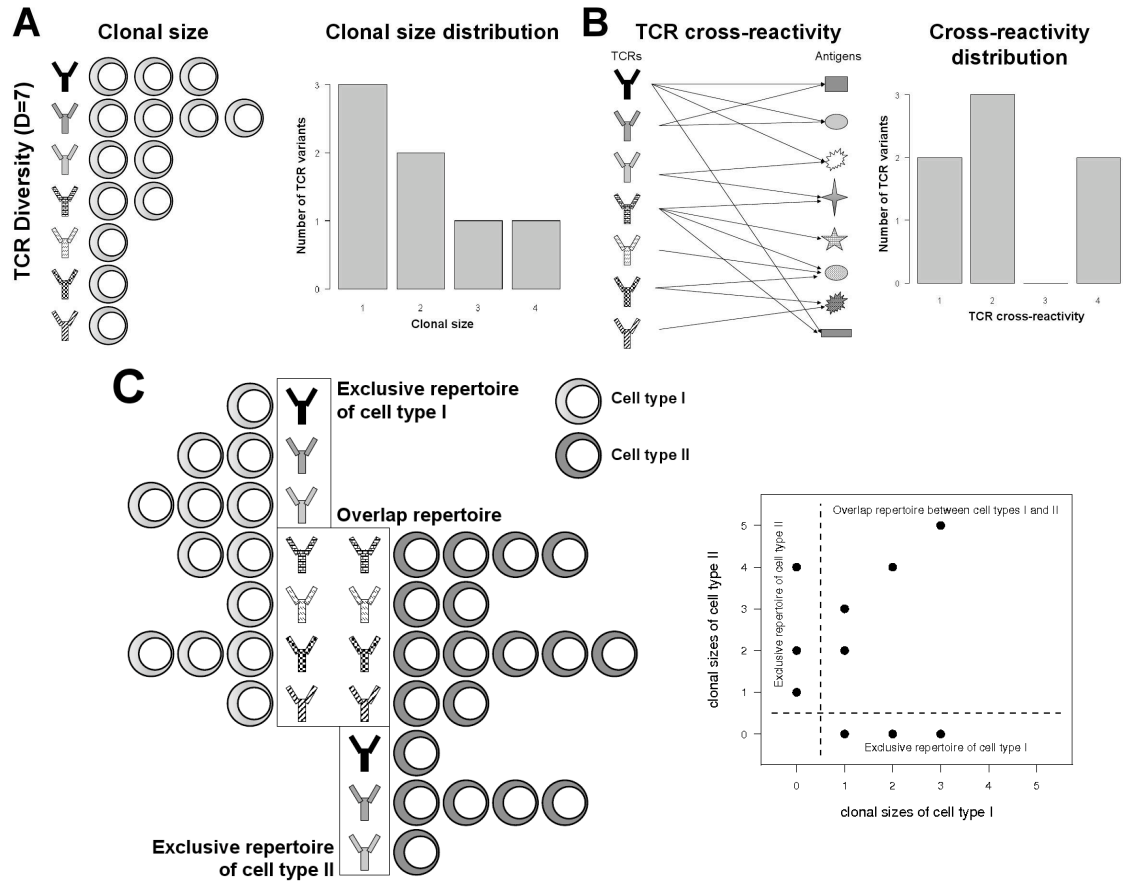


Fig. 1.2: Important features for the characterization of a TCR repertoire. **A.** Structural diversity (TCR diversity, clonal size of a given TCR variant, and the clonal size distribution of the repertoire). **B.** Functional diversity (TCR cross-reactivity and the respective distribution across the repertoire). **C.** The relationship between TCR repertoires of two cell types (exclusive and intersection repertoires, and relationship between clonal size distributions).

tant immunologically is to characterize the functional diversity, which is defined by the number of distinct classes of TCRs according to the antigens they recognize (Figure 1.2B). There are yet other definitions of functional diversity in the literature, namely the number of different functions that activated T cells might exhibit in an immune response, such as proliferation, cytokine secretion, cytotoxicity, migration and homing (Nikolich-Zugich et al., 2004).

It is worth noticing that a particular TCR can react to many different antigens, currently estimated in 10^6 different MHC-associated peptides (Mason, 1998). Moreover, since the number of possible peptides is much larger than the number of T cells in an individual at any given time, it is argued that T cells must be intrinsically cross-reactive (Mason, 1998). On the other hand, a particular antigen can be recognized by many T cells expressing different TCR sequences (Pacholczyk et al., 2006; Wong et al., 2007). The number of T cells that will react vary from antigen to antigen. For example, a study reported that 50-550 CD8⁺ TCR variants can react to two epitopes of influenza A virus (Kedzierska et al., 2006), while other estimated that more than 1000 different CD8⁺ clonotypes might be involved in the immune response to a specific lymphocytic choriomeningitis virus epitope (Pewe et al., 2004). The cross-reactivity distribution is then the frequency of TCR variants present in the repertoire that recognize a certain number of antigens.

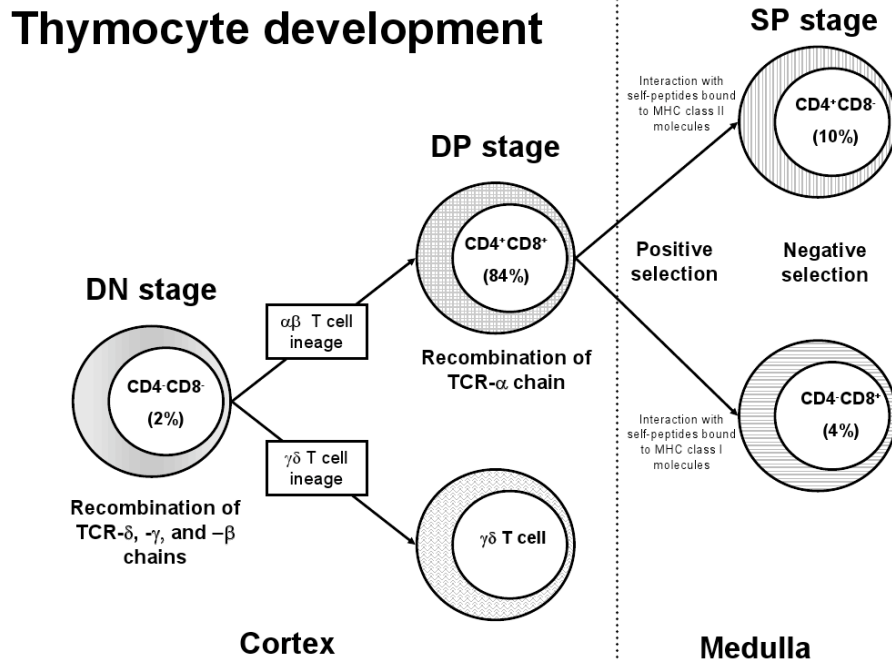


Fig. 1.3: Most important steps in T-cell differentiation in the thymus. Percentages of each thymocyte subset refer to adult F₂ mice derived from an intercross between Balb/C and C57BL/6 mouse strains (Duarte and Penha-Gonçalves, 2001).

As seen before, $\alpha\beta$ T cells are subdivided in two major lineages, CD4⁺ and CD8⁺ T cells. To understand the relationship between these, or any other two, TCR repertoires, it is useful to divide each repertoire in its exclusive and overlapping subsets. The exclusive repertoire of a given cell type is defined by the TCR variants that are only present in the respective repertoire, while the intersection repertoire refers to TCR variants that are present in both cell types (Figure 1.2C). As above, it is important to know the TCR diversity of the exclusive and intersection repertoires and the respective clonal size distributions. Another feature of interest is to infer the relationship between clonal size distributions of TCR variants belonging to both repertoires: uncorrelated or correlated, either negatively or positively (Figure 1.2C).

1.4 Central tolerance through thymic selection

According to modern clonal selection theory, T cells are produced in the thymus through a series of selection steps, which ultimately lead to deletion of most autoreactive clones during thymic T-cell ontogeny. The elimination of these autoreactive clones in the thymus is usually termed central tolerance.

Thymic T-cell ontogeny is usually divided in three main steps, depending on the expression pattern of CD4 and CD8 molecules at the cell surface (Figure 1.3A): the double-negative (DN) stage, in which T-cell precursors neither express the CD4 nor the CD8 molecule, followed by the double-positive (DP) stage, characterized by the expression of both molecules in those cells, ending up in the single-positive (SP) stage in which thymocytes only express one of the molecules, either CD4 or CD8.

The recombination of TCR- β , $-\gamma$, and $-\delta$ chains occur in thymic cortex at the DN stage. During V(D)J recombination of these TCR chains, thymocytes that do not rearrange productively the respective gene segments — that is, T-cell precursors that do not produce a rearrangement in the correct reading-frame — would die by apoptosis, while the remaining ones are allowed to proceed in their developmental program. Therefore, some of the potential TCR diversity is lost due to non-productive rearrangements. This stage seems critical to define whether a precursor will become either a $\alpha\beta$ or a $\gamma\delta$ T cell. In this regard, a productive TCR- β rearrangement in the presence of a surrogate pre-TCR α chain seems enough to commit thymocytes to the $\alpha\beta$ T cell lineage, even when productive γ and δ rearrangements have been achieved in these cells (Dudley et al., 1994; Fehling et al., 1995; von Boehmer et al., 1999). $\gamma\delta$ T cells do not seem to pass through further developmental steps. It is worth noting that DN stage can also be divided in further steps, related to the expression pattern of CD44 and CD25 molecules that seem important to define the further checkpoints in $\alpha\beta:\gamma\delta$ lineage decision, but they will not be discussed here for sake of simplicity; see Hayday and Pennington (2007) for a more detailed view of the DN stage.

$\alpha\beta$ -committed thymocytes proceed then to the DP stage, where they rearrange the TCR- α locus. When a productive rearrangement is obtained, a TCR- α chain is then produced and paired with the TCR- β chain encoded by previous rearranged β genes, forming the heterodimer TCR- $\alpha\beta$. CD4:CD8 lineage decision is thought to be made via interaction of the produced TCR with peptides complexed with MHC molecules (pMHC) presented by cortex thymic epithelial cells. In this regard, the V regions of the TCRs, divided in the three complementary determining regions (CDRs), are particularly important for this interaction. CDR1 and CDR2 are small amino acid sequences inside the variable segments that interact directly with MHC molecule, while CDR3 is a highly diverse amino acid sequence resulting from the imprecise joining between different gene segments that reacts to the peptide presented by the MHC molecule (Figure 1.1). TCR-pMHC interaction is critical to provide survival signals to thymocytes, known as the positive selection step. More precisely, DP thymocytes that do not receive enough signal via TCR will eventually die by apoptosis. It is during positive selection that CD4:CD8 lineage decision seems to be made. Two main models have been proposed for this process (reviewed in Singer et al. 2008). The instructional model postulates that thymocytes interacting with class II pMHC molecules receive instructional signals through TCR to lose their CD8 expression, diverting to the CD4⁺ T cell lineage, while the opposite occurs when the interaction is done with class I pMHC molecules. The instruction signals for the CD4:CD8 lineage might be related to their intensity or duration. In this regard, short or weak TCR signals commit thymocytes to CD8⁺ T cell lineage, while strong or long TCR signals divert T-cell precursors to CD4⁺ T cell lineage. Alternatively, the stochastic model postulates that a first TCR signal randomly terminates the expression of CD4 or CD8 molecule, requiring a second TCR signal for thymocyte survival. Other models can be found in the literature but they are out of the scope of this thesis; for a complete review of this subject, see Singer et al. (2008).

After positive selection, thymocytes migrate to thymic medulla, progressing in the SP stage, where only one of the molecules, CD4 or CD8, is expressed at the cell membrane. Then, thymocytes interact with medullary thymic epithelial cells that would present again present

self-peptides in the context of MHC molecules. At this stage, thymocytes that exhibit TCRs recognizing pMHC complexes with high affinity will be deleted by apoptosis. This process is known as negative selection, and assures that most self-reactive TCR are eliminated in the thymus, before they exert their pathologic effects in the periphery. The surviving thymocytes enter then in the final steps of their developmental program. They must follow migration gradients to reach thymic "gates", wait for their opening, and, finally, enter the periphery (Ladi et al., 2006).

Thymocyte development is then the first big mold of the T-cell repertoire. It has the capacity of generating a large TCR diversity by V(D)J recombination but, at the same time, reducing this diversity by deletion of thymocytes that are either too overreactive to the available pMHC (via negative selection) or unresponsive to them (via positive selection). Interestingly, the proportions of thymocytes in different developmental stages seem to be under genetic control (Duarte and Penha-Gonçalves, 2001; Sepúlveda and Carneiro, 2007). Because of its crucial role in positive and negative selection process, the MHC locus shows the major genetic effect (Duarte and Penha-Gonçalves, 2001). However, reanalyzing available data, we found also minor contributions from two other loci (Sepúlveda and Carneiro, 2007). One of these minor loci includes the terminal deoxynucleotidyl transferase (our unpublished data) that is particularly important for TCR diversity by its essential role in adding nucleotides without template in the end-joining regions of V, J and D segments (Gilfillan et al., 1993; Komori et al., 1993). Therefore, the exact output of thymocyte development process results from an interaction between TCR and MHC diversities. On the one hand, excessive diversity of MHC molecules might lead to a dramatic reduction in T-cell diversity. On the other hand, this effect can be compensated by increasing TCR diversity with a higher activity of terminal deoxynucleotidyl transferase.

Several authors have attempted to understand how thymic selection processes shape structural and functional TCR diversity through probabilistic models (Borghans et al., 1999; de Boer and Perelson, 1993; Mason, 1998). All studies were based on the notion that the output of these processes was optimized by evolution. Mason (1998) proposed that the probability of T cells responding to pathogens is maximized, while Borghans et al. (1999) assumed that what is maximized is the probability of an individual being self-tolerant but at the same time reacting to pathogens. For this general introduction, it is worth noting the work of Mason (1998). This author maximizes the probability of T cells that respond to any pathogen as a function of the proportion of clonal deletion achieved during thymic selection given a mean value for crossreactivity. Interestingly, the optimum was achieved when clonal deletion of self-reactive T cells was incomplete. If thymic selection processes are indeed optimized in such way, some self-reactive T cells would enter the periphery, requiring mechanisms to control their autoimmune potential, the so-called peripheral tolerance to self-antigens. This is in fact the case, as discussed below.

1.5 Peripheral tolerance by the action of regulatory T cells

An important breakthrough in the understanding of peripheral tolerance was achieved by Sakaguchi et al. (1995); for a historical perspective of this finding, see Sakaguchi et al. (2007). These authors identified a small subset of CD4⁺ T cells, known nowadays as regulatory T

(T_R) cells, through a series of adoptive transfers. This subset was defined in this early report by the expression of the activation molecule CD25. In fact, when $CD25^+CD4^+$ T cells were depleted from the transferred cells, recipient mice developed devastating autoimmune disorders in a great variety of organs. Co-transfer of both $CD25^-CD4^+$ and $CD25^+CD4^+$ T cell populations led to either healthy or sick animals, depending on the relative proportion of each T-cell population in the transfers. These experiments showed then that peripheral tolerance was somehow assured by an active process under influence of T_R cells. In healthy mice, T_R cells comprise 5% to 10% of $CD4^+$ T-cell pool, which demonstrates that, even a small population of $CD4^+$ T cells seems enough to regulate all immune responses against self components. It is worth noting that the CD25 molecule is expressed by all activated T cells and, thus, a better cell surface marker was needed for the identification of T_R cells.

Motivated by the observation that the initial proportions of T_R cells in the transferred cells would determine the phenotype of the animals (either sick or healthy), León et al. (2001) studied a series of mathematical models to determine putative interaction mechanisms between regulatory and nonregulatory T cells in the presence of APCs. Among the models studied, the only ones that could agree with the above observation was those in which T_R cells could inhibit proliferation of nonregulatory T (T_E) cells but in which the growth of T_R -cell population depends on T_E cells. T_R cells may grow dependent on a growth factor secreted by T_E cells or alternatively converting these later cells. These mechanisms are embodied in a mathematical model called as the Crossregulation model. Hitherto, available experimental evidence seem to agree with these mechanisms, as we will seen later in this thesis or Carneiro et al. (2007).

Another important breakthrough was recently made by the discovery of the transcription factor Foxp3 as the master gene for T_R -cell generation and identification (Fontenot et al., 2003; Hori et al., 2003). This discovery was motivated by the observation that mutations on the Foxp3 gene in Scurfy mice resulted in multi-organ autoimmune diseases, as in the above-mentioned adoptive transfer experiments (Brunkow et al., 2001). Moreover, mutations in the human Foxp3 gene were identified as being the cause of the devastating IPEX (immuno dysregulation, polyendocrinopathy, X-linked) syndrome in humans, which is characterized by autoimmune diseases in multiple endocrine organs (Bennett et al., 2001). These two observations prompted researchers to hypothesize that above pathologies were deficient in T_R -cell generation, which was indeed confirmed. In fact, $Foxp3^-CD25^+CD4^+$ T cells comprise essentially activated T cells and not T_R cells. Moreover, Foxp3 expression was predominantly found in $CD25^+CD4^+$ T cells, but not in $CD25^-CD4^+$ T cells in healthy animals, either from the thymus or periphery (Fontenot et al., 2003; Hori et al., 2003). Moreover, the transfer of $Foxp3^+CD4^+$ T cells into Scurfy mice or other autoimmune mouse model, such as SCID, could prevent the animals of developing autoimmune pathologies. Nowadays, Foxp3 is known to be transiently expressed in some $CD4^+$ T cells, but the role of these cells in the maintenance of peripheral tolerance is still unclear (Komatsu et al., 2009)

1.6 Ontogeny of regulatory T cells

T_R cells are mainly generated in the thymus from SP precursors by hitherto not fully understood mechanism. As in the case of $CD4^+$ and $CD8^+$ lineage decision, there are essentially

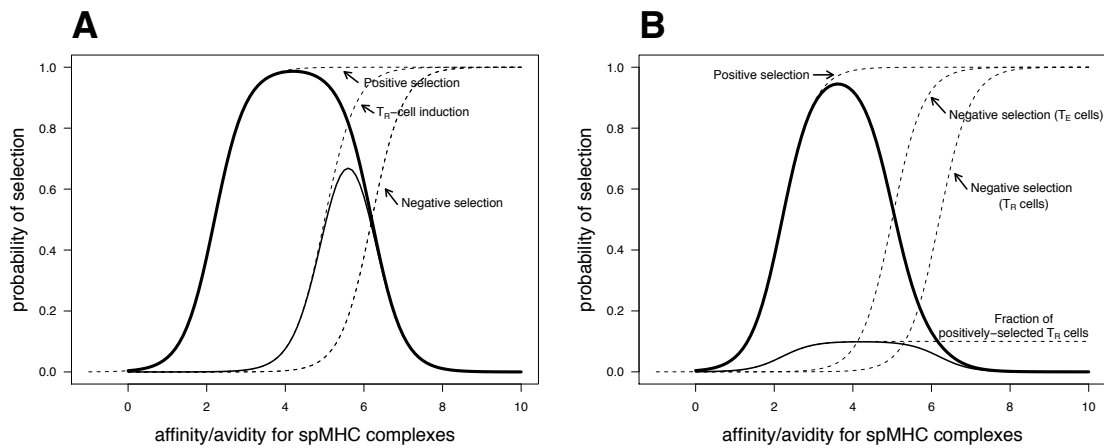


Fig. 1.4: Thymic T cell selection as function of the affinity/avidity of the TCR (in arbitrary units) for pMHC complexes (adapted from Modigliani et al. 1996). **A.** Instructive model in which thymocytes have a probability of being positively and negatively selected, but those with high affinity to pMHC complexes might divert to T_R -cell lineage through integration of instructional signals received from stromal cells in the thymus. **B.** Stochastic-selective model in which there is a probability of generating T_R thymocytes independently of the affinity/avidity for pMHC exhibited by the TCR (in the above case, 10%). In this model, T_R and nonregulatory (T_E) thymocytes have the same probability of being positively selected but T_R precursors are more resistant to negative selection than their T_E counterparts. The probabilities of positive selection, T_R -cell induction, and negative selection are described by appropriate logistic functions (dashed lines) from which the probabilities of selection are derived (solid lines). Thick solid lines refer to the overall probability of selecting a thymocyte with a given affinity/affinity between the respective TCR and pMHC, while thin solid lines refer to the same probability but for T_R thymocytes.

two proposals to thymic T_R -cell generation, the instructive-selective model and the stochastic-selective model. In the instructive-selection model, precursors pass through to positive and negative selection steps, but only those exhibiting high affinity/avidity interactions with pMHC complexes would receive instructional signals to divert to T_R -cell lineage (Figure 1.4A; Apostolou et al. 2002; Bensinger et al. 2001; Modigliani et al. 1996; Ribot et al. 2007, 2006). In the case of stochastic-selective mechanism, a random process not related with TCR affinity would divert precursors to T_R -cell lineage commitment prior to thymic positive and negative selection processes (Bonasio et al., 2006; Liston et al., 2003; van Santen et al., 2004). In this model, an enrichment of high self-reactive specificities in T_R thymocytes is explained by a higher resistance of those cells to negative selection (Figure 1.4B; Bonasio et al. 2006; Liston et al. 2003; van Santen et al. 2004). Recently, a two-step model has been recently proposed where T_R precursors must receive a first TCR-dependent signal, followed by an additional TCR-independent signal via interleukin-2 (IL-2) or IL-15 stimulation to sustain Foxp3 expression in those cells (Lio and Hsieh, 2008).

Recent experimental evidence using different adoptive transfer experiments demonstrated that CD25⁺CD4⁺ T cells can acquire Foxp3 expression in the periphery, exhibiting similar regulatory functions to T_R cells generated in the thymus (Apostolou and von Boehmer, 2004; de Lafaille et al., 2004; Zelenay et al., 2005). Therefore, besides their generation in the thymus,

T_R cells can also be induced in the periphery from conventional $CD4^+$ T cells. This mechanism is known as peripheral conversion, as suggested by Modigliani et al. (1996), but referred to as an "education" process by which naive $CD4^+$ T cells could be educated by T_R cells to gain regulatory functions when conjugated in the same APC. Nonetheless, peripheral conversion does not seem a one way process by solely converting T_E cells into the regulatory T-cell pool. In fact, recent reports showed that the reverse might also happen under lymphopenic or normal conditions (Duarte et al., 2009; Komatsu et al., 2009; Radhakrishnan et al., 2008).

The reasons why these conversion processes occur in the periphery and their exact role in the control of autoimmunity and immunity are currently matters under investigation. Some authors suggest that peripheral $Foxp3^+CD4^+$ T cells are composed of a lineage with $Foxp3^+$ stable expression and a minor subset of $Foxp3^+CD25^-$, which retain a certain degree of phenotypic plasticity, and may be converted into the T_R cell subset by the factors that promote $Foxp3$ expression, such as the $TGF-\beta$ (Komatsu et al., 2009). Other authors claim that TCR specificity is crucial to peripheral conversion due to a small overlap between TCR samples of converted and nonconverted T_R cells when $Foxp3^-CD4^+$ T cells were transferred into lymphopenic hosts (Lathrop et al., 2008). Yet, a word of caution should be put to this interpretation, because a low overlap between samples does not necessarily imply the same interpretation at the level of the whole repertoire, as we will see throughout the thesis. Other authors suggest that peripheral conversion might be driven by homeostatic cues, because T_R -cell numbers can be restored after *in vivo* depletion of $CD25^+CD4^+$ T-cell subset upon treatment with anti- $CD25$ antibody (Zelenay et al., 2005).

Another mathematical model based on peripheral conversion has been recently described (Fouchet and Regoes, 2008). This model is based on a complex interaction between APCs and $CD4^+$ T cells in which naive T cells become either T_R or T_E cells upon contact with resting or activated APCs, respectively. Interestingly, this model features feedback among T_R and T_E cells similar to Crossregulation model but now indirectly via APC. Accordingly, this model shows bistability such that, depending on the initial conditions, the system either goes to a strongly regulated steady state under control of T_R cells or to a weakly regulated steady state with a predominance of T_E cells.

1.7 The shape of regulatory T-cell repertoire

As mentioned above, T_R cells are absolutely necessary for peripheral tolerance against body components. Because of this it was thought that T_R cells were mainly targeted to body components. Yet, this might not be completely true. On the one hand, several lines of evidence suggest an important, but yet not fully understood role of T_R cells in dampening excessive inflammation and immune responses against different microbial agents both in mice and in humans (Cabrera et al., 2008; Demengeot et al., 2006; Franceschini et al., 2009). On the other hand, a recent study showed that T-cell hybridomas with TCRs (loaded) from T_R cells are preferentially activated in presence of pre-stimulated allogeneic APCs rather than in presence of autologous APCs (Pacholczyk et al., 2007). This observation prompted the authors to suggest that TCRs of T_R cells are targetted to nonself-antigens rather than to self-antigens (Pacholczyk

et al., 2007), although this interpretation of the results is questionable. Since T_R cells might be beneficial to the host, avoiding autoimmunity and excessive inflammation, but also beneficial to pathogens, by suppressing protective immune responses against them, it is difficult to decide whether T_R cells are either friends or foes of the host (Mills, 2004).

To understand the exact role of T_R cells in the immune system, it is crucial to characterize T_R and T_E -cell repertoires and their relationship. It is known that peripheral T cells need to compete for limiting resources (APCs, cytokines, antigens, etc) in order to survive and maintain the organism in an equilibrium (or homeostatic) state (Freitas and Rocha, 2000). Because these resources are not evenly distributed throughout the body, several authors suggested that peripheral T-cell repertoire could be compartmentalized in specific T-cell niches, where T cells find their resources (Carneiro et al., 1995; Freitas and Rocha, 2000). Similar reasoning was recently done for the T_R -cell repertoire (Almeida et al., 2005; Carneiro et al., 2007). Since T_R -cell function and maintenance seem critically dependent on the interactions that they make with APCs and T_E cells, a T_R -cell niche might be defined by the co-localization of these three cell types in the different tissues of the body, the APC-dependent foci, as suggested by Carneiro et al. (2007). Anatomically, peripheral T_R cells are found in different locations, such as, spleen, subcutaneous peripheral and mesenteric lymph nodes in the lymphoid system, and lung, skin, and liver in the nonlymphoid organs (Sather et al., 2007). In this regard, the expression of different homing receptors and TCR specificities by T_R cells will determine their broad anatomical distribution throughout the body (Lathrop et al., 2008; Sather et al., 2007; Siegmund et al., 2005).

Several authors had attempted to pinpoint the key features of structural properties of T_R and T_E TCR repertoires. It was suggested that, in wildtype animals, T_R cells do not show any evident bias on the usage of $V\beta$ and $V\alpha$ segments (Takahashi et al., 1998), but this is a very crude analysis of the whole repertoire, as previously mentioned. Recent studies went deeper in the analysis of the repertoires by enumerating TCR sequences present in T-cell samples of mouse models where TCR diversity is genetically limited by a drastic reduction on the number of gene segments that can rearrange during V(D)J recombination (Hsieh et al., 2004, 2006; Lathrop et al., 2008; Pacholczyk et al., 2006, 2007). In spite of this experimental limitation, all studies agree that TCR diversity of each cell type is still large, even when thymic selection processes were under influence of a single peptide (Pacholczyk and Kern, 2008; Wong et al., 2007). Moreover, T_R TCR samples seem more diverse than those of T_E cells, either in the periphery (Hsieh et al., 2004; Pacholczyk et al., 2007; Wong et al., 2007) or in the thymus (Hsieh et al., 2006; Pacholczyk et al., 2006). Samples of T_R cells partially overlap with those of T_E cells according to Morisita-Horn index, but the extent of this overlap varies across different animal models (Hsieh et al., 2004; Pacholczyk et al., 2006; Wong et al., 2007) and anatomical regions (Pacholczyk and Kern, 2008). Similar observation was obtained in human studies (Fazilleau et al., 2007). However, as already mentioned, a small overlap does not necessarily imply the same interpretation for the whole repertoire. Moreover, the above studies applied crude statistical tools to assess the different aspects of structural diversity, as we will discuss below, and thus the experimental characterization of T_R and T_E TCR repertoires remains elusive.

Some clues on the shape of T_R and T_E -cell repertoires have been recently provided by means of mathematical modeling. Carneiro et al. (2007) (and also in this thesis) explored the Crossregulation model to a situation where different T-cell clones recognize their own exclusive set of APCs. In this situation, $CD4^+$ T-cell repertoire can be partitioned into three subsets. The first subset is composed of few clones with short life-span that interact with few APCs, and thus cannot be maintained in the periphery. The second subset refers to a diverse set of barely autoreactive T_E cell clones, whose expansion is limited only by APC availability. The third subset is related to a less diverse set of small self-reactive T cell clones exhibiting both T_R and autoreactive T_E cells that regulate each other's growth. Since this partition suggests that the majority of T_R cells would belong to the third subset, it was expected that T_R cells would be less diverse than T_E cells, which seem at a first glance in disagreement with a higher diversity of T_R in samples taken from the repertoire. However, this disagreement might be just apparent since proper statistical analysis was not done in available TCR repertoire studies, as we will discuss below and throughout the thesis. Moreover, this partition of the repertoire is based on a strong assumption of the model that every T cell clone recognizes an exclusive set of antigens (or more simply, one clonotype - one specificity). Therefore, the model as it stands does not include competition between different T cell clones population for APCs. Therefore, the above-mentioned partition of the repertoire must be confirmed by less restrictive models in terms of functional diversity.

1.8 Some shortcomings in the assessment of TCR structural diversity

Different statistical approaches are available to dissect structural diversity of T-cell repertoires through TCR sequence data. In the past, an estimator and the respective confidence interval were proposed to infer the number of different $V\beta$ gene segments present in the mouse genome (Barth et al., 1985; Behlke et al., 1985). The estimator was derived under the assumption that every $V\beta$ gene segment was equally represented in the population. Later on, this same estimator was applied to obtain the first estimate of $\alpha\beta$ TCR diversity in mouse (Casrouge et al., 2000), and because of that, became a standard tool in the analysis of TCR diversity, the so-called MLE method (Fazilleau et al., 2007; Hsieh et al., 2004, 2006). Interestingly, the available data show often TCR species with a large number of members in the samples, which seem to disagree with the above assumption (Naumov et al., 2003; Pewe et al., 2004). Using proper statistical analysis, we demonstrate that this was indeed the case (in this thesis and Sepúlveda et al. 2008). Therefore, TCR diversity is not hitherto accurately estimated. With this concern in mind, some authors also estimate TCR diversity via abundance- and incidence-coverage estimators (Hsieh et al., 2004, 2006; Pacholczyk et al., 2006, 2007), imported from Ecology (Chao and Lee, 1992), where the problem of biodiversity estimation is in constant debate. Because these estimators are not parametric in their essence, they lead in theory to more robust TCR estimates in a wide range of situations. Nevertheless, they show a major drawback: they give no information on the underlying clonal size distribution. Another approach to TCR diversity is to use the Simpson's diversity index (Venturi et al., 2007) that aims to capture not only TCR diversity but also the clonal size of each TCR variant in the sam-

ples. In spite of being useful to compare TCR diversity among different samples (Venturi et al., 2007), this measure cannot be related to the actual TCR diversity of the underlying populations and to the respective clonal size distributions.

To tackle the intersection between two TCR repertoires, the standard statistical approach is to use some descriptive measures, such as the Jacard and the Morisita-Horn indexes, which attempt to assess the similarity between two TCR samples (Hsieh et al., 2004, 2006; Venturi et al., 2008). The Jacard index is simply the number of TCR variants shared by two samples divided by the total TCR diversity observed in both samples. In its turn, the Morisita-Horn includes the information of the number of TCR variants common to both samples as well as the respective number of repeats of shared TCR variants in the samples; the exact mathematical definition will be given in Chapter 7. Yet, the Jacard and Morisita-Horn indexes are just summary statistics and, thus, they neither can estimate the TCR diversity of intersection and exclusive repertoires nor the respective relationship between clonal size distributions of the intersection repertoire.

We thus conclude that alternative statistical methods are needed to better characterize the structural diversity of TCR repertoires, as also discussed in Boudinot et al. (2008).

1.9 In this thesis

This thesis addresses how the T-cell repertoire is shaped. Two main threads were followed. First, we investigate the very early steps of T cell repertoire formation unravelling the mechanisms underlying the V(D)J recombination in $\gamma\delta$ T cells. In this line of research, we tackled the biological constraints of thymocyte development. Second, motivated by the great interest on the role of T_R cells in the immune system, we assess the influence of functional diversity in shaping the peripheral T_R and T_E cell repertoires. We also characterize the structural diversity of these two repertoires in the thymus and periphery. Along the way we also gained some insights on the structural diversity of $CD4^+$ and $CD8^+$ TCR repertoires in mice with limited TCR diversity. We tackled these issues using different mathematical and statistical models.

In Chapters 2 and 3, we deal with V(D)J recombination in T-cell precursors in adult mice with the aim of assessing the biological constraints shaping the potential T-cell repertoire. We studied the rearrangement of TCR γ locus. Since two isotypes are essentially recombined at the TCR γ locus in adult mice, it was possible to enumerate all possible recombination states of a thymocyte and develop different stochastic models that trace them throughout the recombination process (Chapter 2). By testing them against available experimental data, we argue that V-J rearrangement in each precursor cell can only occur in a certain time period, in which a locus is first opened for the rearrangement, followed by the opening of the other locus after some time (differential allelic accessibility). Moreover, the occurrence of a productive V-J rearrangement leads to a shutting down of the whole rearrangement process (feedback mechanism). Interestingly, the estimated rearrangement rates for the different gene segments suggest that TCR γ -chain rearrangement is somehow "optimized" to minimize the probability of a precursor cell presenting two distinct productively rearranged gene segments, maximizing at the same time the probability of cell survival during the process. In Chapter 3, we provide the experimental evidence supporting the suggested mechanism.

In Chapters 4 and 5, we study the effect of functional diversity in shaping peripheral T_R and T_E cell repertoires via Crossregulation model. As mentioned above, the model describes the dynamics of a basic interaction of T_R and T_E cells competing for APCs. We first revisit previous work done with the model, but also extending it to a population of T cell clones not sharing APCs among them. Some of the predictions were already been described throughout this general introduction chapter. The most important result of the model is that peripheral $CD4^+$ T-cell repertoires might be partitioned in three subsets of T-cell clones depending on the densities of APCs they recognize. Then, we extent previous model to a situation where a heterogeneous APC population is shared by a diverse T_R and T_E -cell population. In simulations resembling adoptive transfers, we showed that the same partition of the repertoire seems to hold. When constant thymic export was contemplated in the model, three distinct repertoire structures emerged from simulations, even for the same parameter step. One repertoire structure is associated with an overdominance of T_E cells, even in clones recognizing high levels of APCs that should be under control of T_R cells. This situation resembles the devastating phenotype of Foxp3-knockout mice. Other repertoire structure is related to an overdominance of T_R cells. This structure is inadequate as it would correspond to a situation where all immune responses, even those against harmful pathogens, would be prevented by T_R cells. The third repertoire structure shows a similar partition of the repertoire as in simulations without thymic export or with independent APC recognition. As argued by Carneiro et al. (2007), this last partition seems adequate to describe an efficient immune system. We also show that the above repertoire structures are best distinguished by the resident population. For each repertoire, we provided a qualitative prediction for different structural diversity properties, such as the relationship between T_R and T_E -cell diversities and the correlation between clonal size distributions.

With these predictions in mind, we re-analyze previously published TCR data in Chapter 6 and 7. To this end, we first introduce the class of Poisson abundance models (PAM) to estimate TCR diversity and the underlying clonal sizes distributions (Chapter 6). This class of models includes many distinct models that describe different patterns for the underlying clonal size distribution, such as the Lognormal or the Gamma distribution. To fit PAM to experimental data, we made available the package PAM for R language. For illustrative purposes, we analyze thymic and peripheral $CD4^+$ and $CD8^+$ TCR data from a mouse model where TCR diversity is extremely restricted (Chapter 6). Our results show that peripheral $CD4^+$ and $CD8^+$ populations are well described by the Poisson-Lognormal model, which predicts a Lognormal probability distribution for the clonal size distribution.

We then apply the Poisson abundance models to study T_R and T_E TCR repertoires in the thymus and periphery (Chapter 7). The Lognormal-Poisson model revealed again the best model for all data sets analyzed. Because of that, this model was extended to the two-sample case in order to dissect the relationship between any two overlapping repertoires. Our results suggest that T_E -cell precursors exhibit higher TCR diversity than their regulatory counterparts, in agreement with the hypothesis that T_R cells are targeted to self-reactivities. More importantly, the correlation between clonal size distribution of TCRs belonging to the respective intersection repertoire is positive, suggesting that thymic T_R -cell differentiation is somehow

dependent on the abundance of each TCR sequence. We also demonstrate that thymic and peripheral T_R cell repertoires might overlap completely, and their respective TCR abundance distributions might be positively correlated. We argue that this complete overlap might be mainly due to mature T_R cells re-entering the thymus. Peripheral T_R and T_E cells exhibit similar TCR diversity, and their repertoires might share 5% to 100% TCR variants. The respective clonal size distributions for TCRs belonging to the intersection repertoire might be either uncorrelated or positive correlated. To reach a more clearer conclusion for the intersection of these two repertoires, we recommend to determine the relative contribution of several populations (recent thymic emigrants, peripheral converted T_R cells, and long-lived T cell clones) to peripheral repertoires. We then conclude that this type of analysis improved the current knowledge on the structural diversity of T_R and T_E cell repertoires in the thymus and periphery.

In chapter 8 we discussed the most important aspects of this thesis, namely, the potential structures of peripheral T-cell repertoire, and the mechanisms that generate them. We built on theoretical results to gain insight into the likelihood of an exclusive regulatory T-cell repertoire and putative explanations for the different repertoire structures observed in animal models with limited TCR diversity. We also discussed the reason for the good fit of the Poisson-Lognormal distribution in most data sets. Using the Crossregulation model to tackle this issue, the underlying clonal size distribution seems a mixture of different distributions associated with different populations (e.g., recent thymic emigrants and different resident populations). With this result in mind, we argued that the general good fit of this distribution is due to either too small sample sizes or low TCR diversities in the mice under analysis. We also put forward a set of precise recommendations for future TCR repertoire studies in general, and for validation of the predictions of the Crossregulation model in particular. In this regard, we advise to increase current sample sizes, the usage of TCR sequencing at single-cell level, and longitudinal T-cell repertoire studies.

Bibliography

- Almeida, A. R., Rocha, B., Freitas, A. A., and Tanchot, C. (2005). Homeostasis of T cell numbers: from thymus production to peripheral compartmentalization and the indexation of regulatory T cells. *Sem. Immunol.*, 17:239–249.
- Amadori, A., Zamarchi, R., De Silvestro, G., Forza, G., Cavatton, G., Danieli, G. A., Clementi, M., and Chieco-Bianchi, L. (1995). Genetic control of the CD4/CD8 T-cell ratio in humans. *Nat. Med.*, 1:1279–1283.
- Apostolou, I., Sarukhan, A., Klein, L., and von Boehmer, H. (2002). Origin of regulatory T cells with known specificity for antigen. *Nat. Immunol.*, 3:756–763.
- Apostolou, I. and von Boehmer, H. (2004). In vivo instruction of suppressor commitment in naive T cells. *J. Exp. Med.*, 199:1401–1408.
- Arstila, T. P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J., and Kourilsky, P. (1999). A direct estimate of the human $\alpha\beta$ T cell receptor diversity. *Science*, 286:958–961.

- Avrasmeas, S. (1991). Natural autoantibodies: from "horror autotoxicus" to "gnothi seauton". *Immunol. Today*, 12:154–159.
- Barth, R. K., Kim, B. S., Lan, N. C., Hunkapiller, T., Sobieck, N., Winoto, A., Gershenfeld, H., Okada, C., Hansburg, D., Weissman, I. L., and Hood, L. (1985). The murine T-cell receptor uses limited repertoire of expressed V_{β} gene segments. *Nature*, 316:517–523.
- Behlke, M. A., Spinella, D. G., Chou, H. S., Sha, W., Hartl, D. L., and Loh, D. Y. (1985). T-cell receptor β -chain expression: dependence on relatively few variable region genes. *Science*, 229:566–570.
- Bennett, C. L., Christie, J., Ramsdell, F., Brunkow, M. E., Ferguson, P. J., Whitesell, L., Kelly, T. E., Saulsbury, F. T., Chance, P. F., and Ochs, H. D. (2001). The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat. Genet.*, 27:20–21.
- Bensinger, S. J., Bandeira, A., Jordan, M. S., Caton, A. J., and Laufer, T. M. (2001). Major histocompatibility complex class II-positive cortical epithelium mediates the selection of CD4⁺25⁺ immunoregulatory T cells. *J. Exp. Med.*, 194:427–438.
- Bonasio, R., Scimone, M. L., Schaerli, P., Grabie, N., Lichtman, A. H., and von Andrian, U. H. (2006). Clonal deletion of thymocytes by circulating dendritic cells homing to the thymus. *Nat. Immunol.*, 7:1092–1100.
- Borghans, J. A., Noest, A. J., and de Boer, R. J. (1999). How specific should immunological memory be? *J. Immunol.*, 163:569–575.
- Boudinot, P., Marriotti-Ferrandiz, M. E., Du Pasquier, L., Benmansour, A., Cazenave, P.-A., and Six, A. (2008). New perspectives for large-scale repertoire analysis of immune receptors. *Mol. Immunol.*, 45:2437–2445.
- Bretscher, P. and Cohn, M. (1970). A theory of self-nonself discrimination. *Science*, 169:1042–1049.
- Brunkow, M. E., Jeffery, E. W., Hjerrild, K. A., Paeper, B., Clark, L. B., Yasayko, S. A., Wilkinson, J. E., Galas, D., Ziegler, S. F., and Ramsdell, F. (2001). Disruption of a new forkhead/winged-helix protein, scurf, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nat. Genet.*, 27:68–73.
- Burnet, F. M. (1957). A modification of Jerne's theory of antibody production using the concept of clonal selection. *Aust. J. Sci.*, 20:67–77.
- Cabrera, G., Burzyn, D., Mundiano, J., Courreges, M. C., Camicia, G., Lorenzo, D., Costa, H., Ross, S. R., Nepomnaschy, I., and Piazzon, I. (2008). Early increases in superantigen-specific Foxp3⁺ regulatory T cells during mouse mammary tumor virus infection. *J. Virol.*, 82:7422–7431.

- Carding, S. R. and Egan, P. J. (2002). $\gamma\delta$ T cells: functional plasticity and heterogeneity. *Nat. Rev. Immunol.*, 2:336–345.
- Carneiro, J. (1997). *Towards a comprehensive view of the immune system*. PhD thesis, University of Porto, Porto.
- Carneiro, J., Léon, K., Caramalho, I., van den Dool, C., Gardner, R., Oliveira, V., Bergman, M.-L., Sepúlveda, N., Paixão, T., Faro, J., and Demengeot, J. (2007). When three is not a crowd: a Crossregulation model of the dynamics and repertoire selection of regulatory CD4+ T cells. *Immunol. Rev.*, 216:48–68.
- Carneiro, J., Stewart, J., Coutinho, A., and Coutinho, G. (1995). The ontogeny of class-regulation of CD4+ T lymphocyte populations. *Int. Immunol.*, 7:1265–1277.
- Casrouge, A., Beaudoin, E., Dalle, S., Pannetier, C., Kanellopoulos, J., and Kourilsky, P. (2000). Size of the $\alpha\beta$ TCR repertoire of naive mouse splenocytes. *J. Immunol.*, 164:5782–5787.
- Chao, A. and Lee, S.-M. (1992). Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.*, 87:210–217.
- Correia-Neves, M., Waltzinger, C., Mathis, D., and Benoist, C. (2001). The shaping of the T cell repertoire. *Immunity*, 14:21–32.
- Coutinho, A. (1989). Beyond clonal selection and network. *Immunol. Rev.*, 110:63–88.
- Davis, M. M. and Bjorkman, P. J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature*, 334:395–402.
- de Boer, R. J. and Perelson, A. S. (1993). How diverse should the immune system be? *Proc. Biol. Sci.*, 252:171–175.
- de Lafaille, M. A. C., Lino, A. C., Kutchukhidze, N., and Lafaille, J. J. (2004). CD25- T cells generate CD25+Foxp3+ regulatory T cells by peripheral expansion. *J. Immunol.*, 173:7259–7268.
- Demengeot, J., Zelenay, S., Moraes-Fontes, M. A., Caramalho, I., and Coutinho, A. (2006). Regulatory T cells in microbial infection. *Springer Semin. Immun.*, 28:41–50.
- Duarte, J. H., Zelenay, S., Bergman, M.-L., Martins, A. C., and Demengeot, J. (2009). Natural Treg cells spontaneously differentiate into pathogenic helper cells in lymphopenic conditions. *Eur. J. Immunol.*, 39:948–955.
- Duarte, N. and Penha-Gonçalves, C. (2001). The MHC locus controls size variations in the CD4 compartment of the mouse thymus. *Immunogenetics*, 53:662–668.
- Dudley, E. C., Petrie, H. T., Shah, L. M., Owen, M. J., and Hayday, A. C. (1994). T cell receptor β chain gene rearrangement and selection during thymocyte development in adult mice. *Immunity*, 1:83–93.

- Elliott, J. F., Rock, E. P., Patten, P. A., Davis, M. M., and Chien, Y. H. (1988). The adult t-cell receptor delta-chain is diverse and distinct from that of fetal thymocytes. *Nature*, 331:627–631.
- Fazilleau, N., Bachelez, H., Gougeon, M. L., and Viguier, M. (2007). Cutting edge: size and diversity of CD4⁺CD25^{high} Foxp3⁺ regulatory T cell repertoire in humans: evidence for similarities and partial overlapping with CD4⁺CD25⁺ T cells. *J. Immunol.*, 179:3412–3416.
- Fehling, H. J., Krotkova, A., Saint-Ruf, C., and von Boehmer, H. (1995). Crucial role of the pre-T cell receptor $\alpha\beta$ gene in the development of $\alpha\beta$ but not $\gamma\delta$ T cells. *Nature*, 375:795–798.
- Fontenot, J. D., Gavin, M. A., and Rudensky, A. Y. (2003). Foxp3 programs the development and function of CD4⁺CD25⁺ regulatory T cells. *Nat. Immunol.*, 4:330–336.
- Fouchet, D. and Regoes, R. (2008). A population dynamics analysis of the interaction between adaptive regulatory T cells and antigen presenting cells. *PLoS One*, 3:e2306.
- Franceschini, D., Paroli, M., Francavilla, V., Videtta, M., Morrone, S., Labbadia, G., Cerino, A., Mondelli, M. U., and Barnaba, V. (2009). PD-L1 negatively regulates CD4⁺CD25⁺Foxp3⁺ Tregs by limiting STAT-5 phosphorylation in patients chronically infected with HCV. *J. Clin. Invest.*, 119:551–564.
- Freitas, A. A. and Rocha, B. (2000). Population biology of lymphocytes: the flight for survival. *Annu. Rev. Immunol.*, 18:83–111.
- Gasque, P. (2004). Complement: a unique innate immune sensor for danger signals. *Mol. Immunol.*, 41:1089–1098.
- Gilfillan, S., Dierich, A., LeMeur, M., Benoist, C., and Mathis, D. (1993). Mice lacking Tdt: mature animals with an immature lymphocyte repertoire. *Science*, 261:1175–1178.
- Haase, A. T. (1999). Population biology of HIV-1 infection: viral and CD4⁺ T cell demographics and dynamics in lymphatic tissues. *Annu. Rev. Immunol.*, 17:625–656.
- Hayday, A. C. and Pennington, D. J. (2007). Key factors in the organized chaos of early T cell development. *Nat. Immunol.*, 8:137–144.
- Hori, S., Nomura, T., and Sakaguchi, S. (2003). Control of regulatory T cell development by the transcription factor Foxp3. *Science*, 299:1057–1061.
- Hsieh, C.-S., Liang, Y., Tyznik, A. J., Self, S. G., Liggitt, D., and Rudensky, A. Y. (2004). Recognition of the peripheral self by naturally arising CD25⁺ CD4⁺ T cell receptors. *Immunity*, 21:267–277.
- Hsieh, C.-S., Zheng, Y., Liang, Y., Fontenot, J. D., and Rudensky, A. Y. (2006). An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nat. Immunol.*, 7:401–410.

- Jameson, J. and Havran, W. L. (2007). Skin $\gamma\delta$ T-cell functions in homeostasis and wound healing. *Immunol. Rev.*, 215:114–122.
- Kedzierska, K., Day, E. B., Pi, J., Heard, S. B., Doherty, P. C., Turner, S. J., and Perlman, S. (2006). Quantification of repertoire diversity of influenza-specific epitopes with predominant public or private TCR usage. *J. Immunol.*, 177:6705–6712.
- Kepler, T. B., He, M., Tomfohr, J. K., Devlin, B. H., Sarzotti, M., and Markert, M. L. (2005). Statistical analysis of antigen receptor spectratype data. *Bioinformatics*, 21:3394–3400.
- Komatsu, N., Mariotti-Ferrandiz, M. E., Wang, Y., Malissen, B., Waldmann, H., and Hori, S. (2009). Heterogeneity of natural Foxp3⁺ T cells: A committed regulatory T-cell lineage and an uncommitted minor population retaining plasticity. *Proc. Natl. Acad. Sci. USA*, 106:1903–1908.
- Komori, T., Okada, A., Stewart, V., and Alt, F. W. (1993). Lack of N regions in antigen receptor variable region genes of Tdt-deficient lymphocytes. *Science*, 261:1171–1175.
- Ladi, E., Yin, X., Chtanova, T., and Robey, E. A. (2006). Thymic microenvironments for T cell differentiation and selection. *Nat. Immunol.*, 7:338–343.
- Lathrop, S. K., Santacruz, N. A., Pham, D., Luo, J., and Hsieh, C.-S. (2008). Antigen-specific peripheral shaping of the natural regulatory T cell population. *J. Exp. Med.*, 205:3105–3117.
- Lederberg, J. (1959). Genes and antibodies. *Science*, 129:1649–1653.
- León, K. (2002). *A quantitative approach to dominant tolerance*. PhD thesis, University of Porto, Porto.
- León, K., Pérez, R., Lage, A., and Carneiro, J. (2001). Three-cell interactions in T cell-mediated suppression? A mathematical analysis of its quantitative implications. *J. Immunol.*, 166:5356–5365.
- Linfert, D., Chowdhry, T., and Rabb, H. (2009). Lymphocytes and ischemia-reperfusion injury. *Transplant. Rev. (Orlando)*, 23:1–10.
- Lio, C.-W. J. and Hsieh, C.-S. (2008). A two-step process for thymic regulatory T cell development. *Immunity*, 28:100–111.
- Liston, A., Lesage, S., Wilson, J., Peltonen, L., and Goodnow, C. C. (2003). Aire regulates negative selection of organ-specific T cells. *Nat. Immunol.*, 4:350–354.
- Maizels, N. (2005). Immunoglobulin gene diversification. *Annu. Rev. Immunol.*, 39:23–46.
- Mason, D. (1998). A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol. Today*, 19:395–404.
- Matsutani, T., Ohmori, T., Ogata, M., Soga, H., Yoshioka, T., Suzuki, R., and Itoh, T. (2006). Alteration of T-cell receptor repertoires during thymic T-cell development. *Scand. J. Immunol.*, 64:53–60.

- Matzinger, P. (1994). Tolerance, danger, and the extended family. *Annu. Rev. Immunol.*, 12:991–1045.
- McCune, J. M. (2001). The dynamics of CD4⁺ T-cell depletion in HIV disease. *Nature*, 410:974–979.
- Mills, K. H. G. (2004). Regulatory T cells: friend or foe in immunity to infection? *Nat. Rev. Immunol.*, 4:841–855.
- Modigliani, Y., Bandeira, A., and Coutinho, A. (1996). A model for developmentally acquired thymus-dependent tolerance to central and peripheral antigens. *Immunol. Rev.*, 149:155–200.
- Naumov, Y. N., Naumova, E. N., Hogan, K. T., Selin, L. K., and Gorski, J. (2003). A fractal clonotype distribution in the CD8⁺ memory T cell repertoire could optimize potential for immune responses. *J. Immunol.*, 170:3994–4001.
- Nikolich-Zugich, J., Slifka, M. K., and Messaoudi, I. (2004). The many important facets of T-cell repertoire diversity. *Nat. Rev. Immunol.*, 4:123–132.
- Pacholczyk, R., Ignatowicz, H., Kraj, P., and Ignatowicz, L. (2006). Origin and T cell receptor diversity of Foxp3⁺CD4⁺CD25⁺ T cells. *Immunity*, 25:249–259.
- Pacholczyk, R. and Kern, J. (2008). The T-cell receptor repertoire of regulatory T cells. *Immunology*, 125:450–458.
- Pacholczyk, R., Kern, J., Singh, N., Iwashima, M., Kraj, P., and Ignatowicz, L. (2007). Nonself-antigens are the cognate specificities of Foxp3⁺ regulatory T cells. *Immunity*, 27:493–504.
- Palm, N. W. and Medzhitov, R. (2009). Pattern recognition receptors and control of adaptive immunity. *Immunol. Rev.*, 227:221–233.
- Park, J. E. and Barbul, A. (2004). Understanding the role of immune regulation in wound healing. *Am. J. Surg.*, 187:11S–16S.
- Pewe, L. L., Netland, J. M., Heard, S. B., and Perlman, S. (2004). Very diverse CD8 T cell clonotypic responses after virus infections. *J. Immunol.*, 172:3151–3156.
- Radhakrishnan, S., Cabrera, R., Schenk, E. L., Nava-Parada, P., Bell, M. P. van Keulen, V. P., Marler, R. J., Felts, S. J., and Pease, L. R. (2008). Reprogrammed Foxp3⁺ T regulatory cells become IL-17⁺ antigen-specific autoimmune effectors in vivo and in vitro. *J. Immunol.*, 181:3137–3147.
- Ribot, J., Enault, G., Pilipenko, S., Huchenq, A., Calise, M., Hudrisier, D., Romagnoli, P., and van Meerwijk, J. P. M. (2007). Shaping of the autoreactive regulatory T cell repertoire by thymic cortical positive selection. *J. Immunol.*, 179(10):6741–6748.
- Ribot, J., Romagnoli, P., and van Meerwijk, J. P. M. (2006). Agonist ligands expressed by thymic epithelium enhance positive selection of regulatory T lymphocytes from precursors with a normally diverse TCR repertoire. *J. Immunol.*, 177:1101–1107.

- Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M., and Toda, M. (1995). Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol.*, 155:1151–1164.
- Sakaguchi, S., Wing, K., and Miyara, M. (2007). Regulatory T cells - a brief history and perspective. *Eur. J. Immunol.*, 37 Suppl 1:S116–S123.
- Sather, B. D., Treuting, P., Perdue, N., Miazgowiec, M., Fontenot, J. D., Rudensky, A. Y., and Campbell, D. J. (2007). Altering the distribution of Foxp3(+) regulatory T cells results in tissue-specific inflammatory disease. *J. Exp. Med.*, 204:1335–1347.
- Schwartz, M. and Moalem, G. (2001). Beneficial immune activity after CNS injury: prospects for vaccination. *J. Neuroimmunol.*, 113:185–192.
- Sepúlveda, N., Paulino, C. D., and Carneiro, J. (2008). Diversity of regulatory and effector T lymphocytes (in portuguese). In Hill, M. M., Ferreira, M. A., Dias, J. G., Salgueiro, M. F., Carvalho, H., Vicente, P., and Braumann, C. A., editors, *Estatística: da Teoria à Prática*, pages 513–523, SPE editions, Lisboa.
- Sepúlveda, N. and Carneiro, J. (2007). What loci control the proportions of different thymocyte populations? (in portuguese). In Ferro, M. E., Nunes, C., and Braumann, C. A., editors, *Estatística: Ciência Interdisciplinar*, pages 729–740, SPE editions, Lisboa.
- Siegmund, K., Feuerer, M., Siewert, C., Ghani, S., Haubold, U., Dankof, A., Krenn, V., Schn, M. P., Scheffold, A., Lowe, J. B., Hamann, A., Syrbe, U., and Huehn, J. (2005). Migration matters: regulatory T-cell compartmentalization determines suppressive activity in vivo. *Blood*, 106:3097–3104.
- Sim, B. C., Aftahi, N., Reilly, C., Bogen, B., Schwartz, R. H., Gascoigne, N. R., and Lo, D. (1998). Thymic skewing of the CD4/CD8 ratio maps with the T-cell receptor alpha-chain locus. *Curr. Biol.*, 8:701–704.
- Singer, A., Adoro, S., and Park, J.-H. (2008). Lineage fate and intense debate: myths, models and mechanisms of cd4- versus cd8-lineage choice. *Nat. Rev. Immunol.*, 8:788–801.
- Takahashi, T., Kuniyasu, Y., Toda, M., Sakaguchi, N., Itoh, M., Iwata, M., Shimizu, J., and Sakaguchi, S. (1998). Immunologic self-tolerance maintained by CD25⁺CD4⁺ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state. *Int. Immunol.*, 10:1969–1980.
- Takeda, K., Kaisho, T., and Akira, S. (2003). Toll-like receptors. *Annu. Rev. Immunol.*, 21:335–376.
- van Santen, H.-M., Benoist, C., and Mathis, D. (2004). Number of T reg cells that differentiate does not increase upon encounter of agonist ligand on thymic epithelial cells. *J. Exp. Med.*, 200:1221–1230.

- Venturi, V., Kedzierska, K., Tanaka, M. M., Turner, S. J., Doherty, P. C., and Davenport, M. P. (2008). Method for assessing the similarity between subsets of the T cell receptor repertoire. *J. Immunol. Methods*, 329:67–80.
- Venturi, V., Kedzierska, K., Turner, S. J., Doherty, P. C., and Davenport, M. P. (2007). Methods for comparing the diversity of samples of the T cell receptor repertoire. *J. Immunol. Methods*, 321:182–195.
- von Boehmer, H., Aifantis, I., Feinberg, J., Lechner, O., Saint-Ruf, C., Walter, U., Buer, J., and Azogui, O. (1999). Pleiotropic changes controlled by the pre-T-cell receptor. *Curr. Opin. Immunol.*, 11:135–142.
- Wong, J., Obst, R., Correia-Neves, M., Losyev, G., Mathis, D., and Benoist, C. (2007). Adaptation of TCR repertoires to self-peptides in regulatory and nonregulatory CD4⁺ T cells. *J. Immunol.*, 178:7032–7041.
- Wong, P. and Pamer, E. G. (2003). CD8 T cell responses to infectious pathogens. *Annu. Rev. Immunol.*, 21:29–70.
- Zelenay, S., Lopes-Carvalho, T., Caramalho, I., Moraes-Fontes, M. F., Rebelo, M., and Demengeot, J. (2005). Foxp3⁺ CD25⁻ CD4 T cells constitute a reservoir of committed regulatory cells that regain CD25 expression upon homeostatic expansion. *Proc. Natl. Acad. Sci. U.S.A.*, 102:4091–4096.

Part II

RECOMBINATION MECHANISM OF γ GENES ENCODING THE T CELL RECEPTOR γ CHAIN

2. STOCHASTIC MODELING OF T CELL RECEPTOR γ GENE REARRANGEMENT

Sepúlveda, N. Boucontet, L., Pereira, P. and Carneiro, J. (2005). Stochastic Modeling of T Cell Receptor γ Gene Rearrangement. *J. Theor. Biol.*, 234:153-165.

Abstract

The mechanisms controlling the recombination process of the γ genes that encode the γ chain of the antigen receptor of the $\gamma\delta$ T lymphocytes are unclear. Based on experimental data on the recombination status of the two major TCR γ genes expressed in $V_{\gamma}4+$ and $V_{\gamma}1+$ thymocytes, we tested the plausibility of three possible rearrangement mechanisms: 1) a time window mechanism according to which the two chromosomes are accessible to the recombination machinery during a defined period of time; 2) a feedback mechanism in which recombination stops shortly after the first in-frame rearrangement event anywhere in both chromosomes; and 3) a feedback mechanism with asynchronous chromosome accessibility, in which there is a first period when only one chromosome is accessible for recombination, followed by a second period when both chromosomes are accessible; shortly after the first-in frame rearrangement event, during any of these two periods, recombination will definitively stop. We model the time window mechanism using a pure probabilistic approach and the two feedback mechanisms using a continuous-time Markov chain formalism. We used maximum likelihood methodology to infer the goodness-of-fit of the models showing evidence for the last model, which best fits the data. Further analysis of this model suggests an evolutionary tradeoff between allelic and isotypic exclusion and the probability that a precursor differentiates into a mature $\gamma\delta$ T lymphocyte.

2.1 Introduction

The immune system of vertebrates has a remarkable capacity to recognize and respond to diverse and evolving pathogens. This ability is mainly due to the large diversity of antigen-receptors collectively expressed by B and T lymphocytes (Burnet, 1957; Jerne, 1955). The antigen receptor of B cells is the immunoglobulin, which recognizes native proteins, carbohydrates, and lipids, and is composed by two heavy (h) chains and two light chains (either κ or λ). The antigen receptor of T cells (TCR) is a heterodimer made of an α and a β chain in $\alpha\beta$ T cells and γ and δ chains in $\gamma\delta$ T cells. The $\alpha\beta$ TCRs recognize antigenic peptides whereas the ligands of the $\gamma\delta$ TCR are still poorly defined.

The hallmark of Ig and TCR genes is that genes encoding the receptor chains are somatically generated in lymphocyte precursors (Tonegawa, 1993). Different gene segments — V (variable), D (diversity) (for some chains) and J (joining) — are randomly assembled by a process called V(D)J recombination giving rise to the gene that will encode each receptor chain.

Since there are many variants of V, D, and J gene segments and some imprecision in the joining a large receptor repertoire can be built by combinatorial assortment.

The basic steps of the V(D)J recombination reaction have been identified. The reaction is initiated by RAG1 and RAG2 proteins (Oettinger et al., 1990) that recognize conserved recombination signal sequences (RSSs) flanking the gene segments, and introduce a double-strand break at the signal sequence boundary (Gilfillan et al., 1993). The rearrangement is then completed by ubiquitous DNA repair machinery that joins the gene segments (reviewed in Rooney et al. 2004). While the repair is not finalized, specific enzymes chop off or add nucleotides to the coding sequences in a template independent manner (Komori et al., 1993). Due to the randomness in stitching the gene segments the assembled gene product may not conserve the translation reading frame. Hence, some rearrangements will be in-frame such that the coding sequence can be transcribed and translated to a functional protein, while others will be out-of-frame and therefore non-productive (Coleclough, 1983). The frequency of potential in-frame rearrangements is at maximum of 1/3 (Coleclough, 1983), which may be further decreased if a stop codon is introduced into the reading frame aborting translation of the complete chain (Gu et al., 1991).

Overall control of the V(D)J recombination process is mainly due to the developmental regulation of the expression of RAG genes that are transiently upregulated in lymphocyte precursors (Wilson et al., 1994). Targeting specific loci for recombination is mediated by cis-regulatory elements such as enhancers, promoters and locus control regions that contribute to the local opening of the chromatin structure, and make RSSs accessible to RAG proteins (Stanhope-Baker et al. 1996; Yancopoulos and Alt 1985; reviewed in Bergman et al. 2003). The efficiency of the recombination is also fine-tuned by the RSSs structure of the different gene segments (reviewed in Feeney et al. 2004).

The fact that the overwhelming majority of lymphocytes expresses receptor chain proteins from one single allele — a phenomenon known as allelic exclusion — indicates that the mechanisms controlling rearrangement must be very well tuned both individually and collectively. The predominant rationale for the efficiency of allelic exclusion is that rearrangement will be attempted in one gene segment and it will only proceed to the remaining segments if the first attempt was non productive. Preventing more than one rearrangements is partially mediated by a feedback mechanism whereby the first productive rearrangement shuts down the recombination machinery, likely by repressing the expression of the RAG proteins (Aifantis et al., 1997; Alt et al., 1981; Kitamura and Rajewsky, 1992; Malissen et al., 1992; Uematsu et al., 1988). But efficient allelic exclusion will always require some asynchrony in allele rearrangement that can be ensured in two non-mutually exclusive ways: by forcing the rates of rearrangement to be low such that simultaneous rearrangements are improbable (Cohn and Langman, 1990) and/or by making the two chromosomes become differentially accessible to the recombination machinery (Khor and Sleckman, 2002; Mostoslavsky et al., 2001). Hitherto, it is unclear how these three processes act in a concerted fashion to ensure that enough mature lymphocytes, which are allelically excluded, are produced from precursors.

This paper focuses on the rearrangement mechanisms of the mouse TCR γ locus, which contains more than one γ gene isotype (Figure 2.1). Therefore, $\gamma\delta$ T cell precursors could

rearrange productively two isotypes in the same chromosome. The comparison of the extents of allelic and isotypic inclusion in the γ loci (we report elsewhere: Boucontet et al. 2005) can therefore be used to assess the relative importance of chromosome accessibility and feedback mechanisms in controlling rearrangement. Hence, this paper aims to assess how recombination rates, feedback, and loci accessibility come into play to ensure isotypic and allelic exclusion and an adequate yield (i.e. frequency at which a precursor will make at least one functional rearrangement and thus differentiate into a mature cell). Based on experimental data of the recombination status of the two major expressed γ genes in $V_{\gamma}4+$ and $V_{\gamma}1+$ thymocytes, we investigate three possible rearrangement mechanisms: time window, pure feedback and feedback with asynchronous chromosome accessibility. The first one has both loci accessible in the two chromosomes and the recombination machinery upregulated during a fixed period of time (usually known as time window). The pure feedback mechanism postulates also a time window, but, in addition, when the first in-frame rearrangement event takes place, the recombination machinery is downregulated and/or the loci become non-accessible to this machinery. There is a delay between the in-frame rearrangement and the complete cessation of rearrangements in other loci. The last mechanism extends the previous one, in the sense that the two chromosomes do not become simultaneously accessible. So, there is a first period when only one chromosome is accessible for recombination, and a second period where both chromosomes are accessible. Again, upon the first in-frame rearrangement the rearrangement process is downregulated in the remaining gene segments in both chromosomes with some delay. In terms of stochastic modeling, we used the Markov chain formalisms in continuous time for the feedback mechanisms and a pure probabilistic approach for the time window mechanism.

The paper is organized in the following manner: section 2 describes the experimental data, the three rearrangement models, and the methods used for model fitting; section 3 presents the fitting of the models to the data and a detailed analysis of the predictions of the best-fit model; finally, section 4 discusses the results.

2.2 Methods

2.2.1 Experimental Data

The experimental data are relative to two flow cytometry-sorted samples of $V_{\gamma}1+$ and $V_{\gamma}4+$ mature $\gamma\delta$ thymocytes (cells expressing membrane TCRs encoded by $V_{\gamma}1J_{\gamma}4$ and $V_{\gamma}4J_{\gamma}1$ genes, respectively). The experimental details are described elsewhere (Boucontet et al., 2005; Pereira and Boucontet, 2005). In summary, the rearrangement status of the two genes in each chromosome was assessed by PCR. The observed data are summarized in Table 2.1, where 0 denotes "without rearrangement", 1 "with rearrangement" and 1' an "in-frame rearrangement". The pair of rearrangement status does not respect any specific allele order. Another experimental observation that is not explicit in the table is that mature thymocytes containing both $V_{\gamma}1J_{\gamma}4$ and $V_{\gamma}4J_{\gamma}1$ genes rearranged in-frame express predominantly the receptor chain encoded by former gene (Boucontet et al., 2005), perhaps due to pairing biases of the $V_{\gamma}4J_{\gamma}1$ chains with some δ chains.

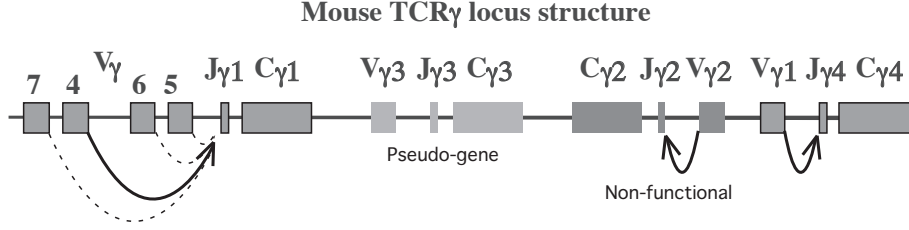


Fig. 2.1: Structure of the mouse γ gene locus and the rearrangement possibilities. The TCR γ chain locus has four J_γ gene segments. Three of them are associated to a single V_γ segment, whereas the other ($J_\gamma 1$) is associated with four different V_γ segments: $V_\gamma 4$, $V_\gamma 5$, $V_\gamma 6$ and $V_\gamma 7$. The V segments rearrange mainly to the nearest J_γ segment as indicated by the arrows. The solid lines indicate predominant rearrangements in adult mice, in which 80% of mature thymocytes express either $V_\gamma 4 J_\gamma 1$ or $V_\gamma 1 J_\gamma 4$ genes. The $V_\gamma 5$ and $V_\gamma 6$ rearrange to $J_\gamma 1$ only during fetal life, and $V_\gamma 7$ rearrangement is much less probable than that of $V_\gamma 4$. $V_\gamma 3$ is a pseudogene and $V_\gamma 2$ despite rearranging often in perinatal cells rarely gives rise to functional chains. A specific isotype or allele rearrangement can happen at most once and further rearrangements can happen elsewhere.

We are dealing with a tridimensional contingency table $2 \times 2 \times 3$: the first variable in the table refers to cell type ($V_\gamma 1+$ or $V_\gamma 4+$), the second is the rearrangement status of the expressed gene ($V_\gamma 1 J_\gamma 4$ for the $V_\gamma 1+$ and $V_\gamma 4 J_\gamma 1$ for the $V_\gamma 4+$), and the third is the rearrangement status of the non-expressed gene ($V_\gamma 4 J_\gamma 1$ for the $V_\gamma 1+$ and $V_\gamma 1 J_\gamma 4$ for the $V_\gamma 4+$). Note that from Table 2.1 we know that $V_\gamma 1+$ thymocytes ($V_\gamma 4+$, respectively) have at least one in-frame $V_\gamma 1 J_\gamma 4$ rearrangement ($V_\gamma 4 J_\gamma 1$, respectively), but we do not know how many in-frame rearrangements are there. For sake of simplicity, we assume that data were generated from a Multinomial-product distribution (one Multinomial distribution for each cell type), i.e.,

$$f(\{n_{ijk}\} | \{n_{i..}\}, \theta) = \prod_{i=1}^2 n_{i..}! \prod_{j=1}^2 \prod_{k=1}^3 \frac{\theta_{(i)jk}^{n_{ijk}}}{n_{ijk}!}, \quad (2.1)$$

where n_{ijk} is the frequency of cells of the i -th type with the j -th rearrangement in the expressed gene and k -th rearrangement in the non-expressed gene, $n_{i..} = \sum_{j,k} n_{ijk}$ and $\theta = (\theta_{(i)jk})$ for $i, j = 1, 2, k = 1, 2, 3$. The parenthesis between the index i of $\theta_{(i)jk}$ indicates a conditional probability of classification in j -th row category and k -th column category given the i -th cell type ($\theta_{(i)jk} = \theta_{ijk}/\theta_{i..}$, where θ_{ijk} is the joint probability of a cell of the i -th type having the j -th rearrangement in the expressed gene, and the k -th rearrangement in the non-expressed gene, with $\theta_{i..} = \sum_{j,k} \theta_{ijk}$).

2.2.2 Rearrangement Mechanisms and Models

It is well established that rearrangement of γ genes can only take place during a limited period of time when the recombination machinery is upregulated, generally known as a time window (see, for example, Livák et al. 1999). In the end of the time window, the $\gamma\delta$ T cell precursors with an in-frame γ rearrangement will continue developing into mature $\gamma\delta$ cells, whereas the rest will not. We are assuming that a functional δ chain is already available in the precursors, and that

Tab. 2.1: $V_\gamma 1 J_\gamma 4$ and $V_\gamma 4 J_\gamma 1$ rearrangement detected by PCR in $V_\gamma 1+$ and $V_\gamma 4+$ cells. In parenthesis is given the usual categorical data analysis notation (n_{ijk} is the frequency of classification in (i, j, k) cell table and its conditional probability $\theta_{(i)jk}$ given cell type i).

$V_\gamma 1+$ Cells		$V_\gamma 4 J_\gamma 1$		
$V_\gamma 1 J_\gamma 4$		(0, 0)	(1, 0)	(1, 1)
(1', 0)	17	$(n_{111}, \theta_{(1)11})$	50	$(n_{112}, \theta_{(1)12})$
(1', 1)	0	$(n_{121}, \theta_{(1)21})$	4	$(n_{122}, \theta_{(1)22})$
				54
				$(n_{113}, \theta_{(1)13})$
				$(n_{123}, \theta_{(1)23})$
$V_\gamma 4+$ Cells		$V_\gamma 1 J_\gamma 4$		
$V_\gamma 4 J_\gamma 1$		(0, 0)	(1*, 0)	(1*, 1*)
(1', 0)	49	$(n_{211}, \theta_{(2)11})$	4	$(n_{212}, \theta_{(2)12})$
(1', 1)	28	$(n_{221}, \theta_{(2)21})$	6	$(n_{222}, \theta_{(2)22})$
				0
				$(n_{213}, \theta_{(2)13})$
				$(n_{223}, \theta_{(2)23})$

the γ gene rearrangement process is assumed to be independent from other rearrangement processes occurring at the same time, and therefore it will be studied alone. We will come back to this issue in the discussion.

Time Window Mechanism

The time window mechanism (TWM) is a purely probabilistic mechanism. During the time window the γ locus is accessible at both chromosomes and the recombination machinery is upregulated. Each specific allele or isotype rearrangement is attempted at most once. We consider independence between alleles of the same gene and between different γ isotypes. Notice that the observations of $V_\gamma 1+$ cells with no $V_\gamma 4$ rearrangements and of $V_\gamma 4+$ cells with no $V_\gamma 1$ rearrangements are evidence against ordered isotype rearrangements. Under these conditions, we only need to take into account the rearrangement status of the gene isotypes at the end of the time window and their respective probabilities. Taking into account that one precursor cell without any in-frame rearrangement will not mature to a $\gamma\delta$ cell, the classification probabilities of the observed data are viewed as conditional probabilities given their respective in-frame rearrangement event.

The TWM has the following parameters: p_1 and p_4 are the probabilities of occurring a $V_\gamma 1 J_\gamma 4$ or a $V_\gamma 4 J_\gamma 1$ rearrangement per chromosome during the time window, respectively; f_1 and f_4 are the probabilities of an in-frame rearrangement given a $V_\gamma 1 J_\gamma 4$ or a $V_\gamma 4 J_\gamma 1$ rearrangement, respectively.

Considering the above assumptions, the joint probabilities of (1,1,1) and (2,1,2) table entries are, respectively, expressed by

$$\theta_{111} = 2p_1 (1 - p_1) f_1 (1 - p_4)^2 \quad (2.2)$$

and

$$\theta_{212} = 2p_4 (1 - p_4) f_4 \times 2(1 - p_1) p_1 (1 - f_1). \quad (2.3)$$

Tab. 2.2: Predicted probabilities of the table entries under the time window mechanism, where $\theta_{2..}$ is given by (2.5).

$V_{\gamma}1+$ Cells		$V_{\gamma}4J_{\gamma}1$	
$V_{\gamma}1J_{\gamma}4$	(0, 0)	(1, 0)	(1, 1)
(1', 0)	$\frac{2(1-p_1)(1-p_4)^2}{2-p_1f_1}$	$\frac{4(1-p_1)(1-p_4)p_4}{2-p_1f_1}$	$\frac{2(1-p_1)p_4^2}{2-p_1f_1}$
(1', 1)	$\frac{p_1(2-f_1)(1-p_4)^2}{2-p_1f_1}$	$\frac{2p_1(2-f_1)(1-p_4)p_4}{2-p_1f_1}$	$\frac{p_1(2-f_1)p_4^2}{2-p_1f_1}$
$V_{\gamma}4+$ Cells		$V_{\gamma}1J_{\gamma}4$	
$V_{\gamma}4J_{\gamma}1$	(0, 0)	(1*, 0)	(1*, 1*)
(1', 0)	$\frac{2(1-p_4)(1-p_1)^2}{\theta_{2..}}$	$\frac{4(1-p_4)(1-p_1)p_1f_1}{\theta_{2..}}$	$\frac{2(1-p_4)p_1^2f_1^2}{\theta_{2..}}$
(1', 1)	$\frac{p_4(2-f_4)(1-p_1)^2}{\theta_{2..}}$	$\frac{2p_4(2-f_4)(1-p_1)p_1f_1}{\theta_{2..}}$	$\frac{p_4(2-f_4)p_1^2f_1^2}{\theta_{2..}}$

Using the same reasoning, the probability of having a $V_{\gamma}1+$ cell is simply

$$\begin{aligned}\theta_{1..} &= \sum_{j,k} \theta_{1jk} = 2p_1(1-p_1)f_1 + p_1^2(f_1^2 + 2(1-f_1)f_1) \\ &= p_1f_1(2-p_1f_1).\end{aligned}\quad (2.4)$$

Analogously, the probability of having a $V_{\gamma}4+$ cell is

$$\begin{aligned}\theta_{2..} &= \sum_{j,k} \theta_{2jk} \\ &= p_4f_4(2-p_4f_4) \left[(1-p_1)^2 + 2(1-p_1)p_1(1-f_1) + p_1^2(1-f_1)^2 \right].\end{aligned}\quad (2.5)$$

Consequently, the joint probabilities of (1,1,1) and (2,1,2) table entries are given by

$$\theta_{(1)11} = \frac{\theta_{111}}{\theta_{1..}} = \frac{2(1-p_1)(1-p_4)^2}{2-p_1f_1} \quad (2.6)$$

and

$$\theta_{(2)12} = \frac{2(1-p_4) \times 2(1-p_1)p_1(1-f_1)}{(2-p_4f_4) \left[(1-p_1)^2 + 2(1-p_1)p_1(1-f_1) + p_1^2(1-f_1)^2 \right]}. \quad (2.7)$$

The rest of the joint probabilities are constructed in same way as the previous ones (see table 2.2).

Pure Feedback Mechanism

It is known that in B cell precursors a productive heavy chain rearrangement inhibits further heavy chain gene rearrangements (Alt et al., 1981; Kitamura and Rajewsky, 1992). The same holds for TCR β gene rearrangement in $\alpha\beta$ T cell precursors (reviewed in von Boehmer et al. 1998). In both cases, there is some inherent mechanism that prevents further rearrangements when the first in-frame rearrangement occurs. This type of mechanism is generally known as feedback.

Tab. 2.3: States and their links of the PFM Markov chain. The absorbing states are denoted by — in their links and the rearrangement process status can be 1 (on) or 0 (off).

state					state				
allelic configuration		process			allelic configuration		process		
id	$V_{\gamma 1}J_{\gamma 4}$	$V_{\gamma 4}J_{\gamma 1}$	status	Links	id	$V_{\gamma 1}J_{\gamma 4}$	$V_{\gamma 4}J_{\gamma 1}$	status	Links
1	(G,G)	(G,G)	1	2,3,10,19	33	(R,R)	(F,G)	1	34,36,38
2	(G,G)	(R,G)	1	5,6,11,20	34	(R,R)	(F,G)	0	—
3	(G,G)	(F,G)	1	4,6,8,12,21	35	(R,R)	(R,R)	0	—
4	(G,G)	(F,G)	0	—	36	(R,R)	(R,F)	1	37
5	(G,G)	(R,R)	1	14,23	37	(R,R)	(R,F)	0	—
6	(G,G)	(R,F)	1	7,15,24	38	(R,R)	(F,F)	1	39
7	(G,G)	(R,F)	0	—	39	(R,R)	(F,F)	0	—
8	(G,G)	(F,F)	1	9,17,26	40	(R,F)	(G,G)	1	41,42,49
9	(G,G)	(F,F)	0	—	41	(R,F)	(R,G)	1	44,45,50
10	(R,G)	(G,G)	1	11,12,31,40	42	(R,F)	(F,G)	1	43,45,47
11	(R,G)	(R,G)	1	14,15,32,41	43	(R,F)	(F,G)	0	—
12	(R,G)	(F,G)	1	13,15,17,33,42	44	(R,F)	(R,R)	1	51
13	(R,G)	(F,G)	0	—	45	(R,F)	(R,F)	1	46
14	(R,G)	(R,R)	1	35,44	46	(R,F)	(R,F)	0	—
15	(R,G)	(R,F)	1	16,36,45	47	(R,F)	(F,F)	1	48
16	(R,G)	(R,F)	0	—	48	(R,F)	(F,F)	0	—
17	(R,G)	(F,F)	1	18,35,47	49	(R,F)	(G,G)	0	—
18	(R,G)	(F,F)	0	—	50	(R,F)	(R,G)	0	—
19	(F,G)	(G,G)	1	20,21,28,40,52	51	(R,F)	(R,R)	0	—
20	(F,G)	(R,G)	1	23,24,29,41,53	52	(F,F)	(G,G)	1	53,54,61
21	(F,G)	(F,G)	1	22,24,26,42	53	(F,F)	(R,G)	1	56,57,62
22	(F,G)	(F,G)	0	—	54	(F,F)	(F,G)	1	55,57,59
23	(F,G)	(R,R)	1	30,44,56	55	(F,F)	(F,G)	0	—
24	(F,G)	(R,F)	1	25,45,57	56	(F,F)	(R,R)	1	—
25	(F,G)	(R,F)	0	—	57	(F,F)	(R,F)	1	58
26	(F,G)	(F,F)	1	27,47,59	58	(F,F)	(R,F)	0	—
27	(F,G)	(F,F)	0	—	59	(F,F)	(F,F)	1	60
28	(F,G)	(G,G)	0	—	60	(F,F)	(F,F)	0	—
29	(F,G)	(R,G)	0	—	61	(F,F)	(G,G)	0	—
30	(F,G)	(R,R)	0	—	62	(F,F)	(R,G)	0	—
31	(R,R)	(G,G)	1	32,33	63	(F,F)	(R,R)	0	—
32	(R,R)	(R,G)	1	35,36					

Qualitatively, this mechanism, referred throughout as pure feedback mechanism (PFM), is an extension of the TWM in the sense that the first successful rearrangement makes the recombination machinery unavailable and/or makes the remaining gene segments inaccessible in both chromosomes. These processes are not instantaneous and therefore there is delay until rearrangements cease. Potentially, this delay allows for the production of cells with more than one in-frame rearrangement.

The stochastic modeling was based on the Markov chain formalism in continuous time (see, for example, Kulkarni 1995). Hence, the waiting time for $V_{\gamma 4}J_{\gamma 1}$ and $V_{\gamma 1}J_{\gamma 4}$ rearrangements is exponentially distributed with averages $1/\mu_4$ and $1/\mu_1$, respectively. The delay is also exponentially distributed with average $1/\mu_d$. The duration of the time window is unitary.

At the beginning of the time window the rearrangement process in the cell is on (status 1), and it stops after the first in-frame rearrangement (status 0). Each allele has 3 states: without rearrangement or germline (G), with an out-of-frame rearrangement (R) and with an in-frame rearrangement (F). Hence, the Markov chain has 63 states, corresponding to the configuration of the two alleles in each gene and the status of the rearrangement process in the cell. The cell starts in state ((G,G),(G,G),1). For example, it could do a $V_{\gamma 1}$ in-frame rearrangement switching to state ((F,G),(G,G),1). Then, during the delay period, a $V_{\gamma 4}$ out-of-frame rearrangement occurs, passing through to state ((F,G),(R,G),1). Finally, the rearrangement process stops,

reaching the state $((F,G),(R,G),0)$. In Table 2.3 are displayed all the states with their respective links. For sake of space, we do not provide the transition rate matrix, $\mathbf{Q} = \{q_{i,j}\}$ where $q_{i,j}$ represents the transition rate of moving from the i -th state to the j -th state. Instead, we provide some examples. The transition rates from state 1 and from state 3 are given by

$$q_{1,j} = \begin{cases} -2\mu_1 - 2\mu_4, & j = 1 \\ 2\mu_4(1 - f_4), & j = 2 \\ 2\mu_4f_4, & j = 3 \\ 2\mu_1(1 - f_1), & j = 10 \\ 2\mu_1f_1, & j = 19 \\ 0, & \text{otherwise} \end{cases} \quad \text{and } q_{3,j} = \begin{cases} -2\mu_1 - \mu_4, & j = 3 \\ \mu_d, & j = 4 \\ \mu_4(1 - f_4), & j = 6 \\ \mu_4f_4, & j = 8 \\ 2\mu_1(1 - f_1), & j = 12 \\ 2\mu_1f_1, & j = 21 \\ 0, & \text{otherwise} \end{cases} \quad (2.8)$$

where $q_{1,1} = -\sum_{j \neq 1} q_{1,j}$ and $q_{3,3} = -\sum_{j \neq 3} q_{3,j}$. The rates of the remaining state transitions are defined using the same reasoning. Note that the states in which the rearrangement process is off (status 0) are absorbing, and the corresponding rows of $\mathbf{Q}$ are set to zero.

According to the Markov chain theory, the state probability vector at the instant of time t_2 , $\mathbf{p}_F(t_2)$, is generically calculated by

$$\mathbf{p}_F(t_2) = \mathbf{p}_F(t_1) \exp \{(t_2 - t_1) \cdot \mathbf{Q}\}, \quad (2.9)$$

where $\mathbf{p}_F(t_1)$ is the state probability vector at a previous instant t_1 , and $\mathbf{P}_F(t_2 - t_1) = \exp \{(t_2 - t_1) \cdot \mathbf{Q}\}$ is the corresponding state transition probability matrix, whose entries are denoted $p_{i,j}(t_2 - t_1)$. The rearrangement Markov process starts in a cell without any rearrangement at time $t_1 = 0$, when $\mathbf{p}_F(0) = \{1, 0, 0, \dots, 0\}$, and ends at $t_2 = 1$, i.e. in the end of the unitary time window. Under this scenario, the probability of having at least one $V_\gamma 1$ rearrangement is given by

$$\begin{aligned} \theta_{1..} = & p_{1,22}(1) + p_{1,25}(1) + \sum_{j=27}^{30} p_{1,j}(1) + p_{1,43}(1) + p_{1,46}(1) + \sum_{j=48}^{51} p_{1,j}(1) \\ & + p_{1,55}(1) + p_{1,58}(1) + \sum_{j=60}^{63} p_{1,j}(1) \end{aligned} \quad (2.10)$$

and the probability of having a $V_\gamma 4+$ cell is

$$\begin{aligned} \theta_{2..} = & p_{1,4}(1) + p_{1,7}(1) + p_{1,9}(1) + p_{1,13}(1) + p_{1,16}(1) + p_{1,18}(1) \\ & + p_{1,34}(1) + p_{1,37}(1) + p_{1,39}(1). \end{aligned} \quad (2.11)$$

Referring to the same examples given in TWM, the joint probabilities of (1,1,1) and (2,1,2) table entries are, respectively, expressed by

$$\theta_{111} = p_{1,28}(1) \quad \text{and} \quad \theta_{212} = p_{1,13}(1). \quad (2.12)$$

The conditional probabilities $\theta_{(1)11}$ and $\theta_{(2)12}$ are given by the quotient between the corre-

sponding joint probabilities and the cell type probabilities, that is $\theta_{(1)11} = \theta_{111}/\theta_{1..}$ and $\theta_{(2)12} = \theta_{212}/\theta_{2..}$, respectively. The same reasoning is applied to determine the rest of the probabilities in the contingency table. As expected, there is no closed form expression for the classification probabilities of the contingency table under PFM.

Feedback With Asynchronous Chromosome Accessibility Mechanism

Access of the recombination machinery to loci in different chromosomes is modulated by the state of chromatin (Sleckman et al., 1998; Stanhope-Baker et al., 1996). Thus, one possible generalization of the PFM is to allow for asynchronous accessibility of the two chromosomes during the time window. It is very unlikely that both chromosomes become accessible at the same instant, so we assume that there is a first period when only one chromosome is susceptible to rearrangement and a second period in which both chromosomes can undergo recombination. During the first period only one allele in each isotype can rearrange - monoallelic period, while in the second period both alleles in both isotypes can rearrange - biallelic period. As before, at any time point, the first in-frame rearrangement stops the rearrangement process in both chromosomes, by a process that has a delay.

To model this mechanism, abbreviated as FACAM, there are two Markov chains: one for each period. The biallelic period is modeled as in PFM. In the case of the monoallelic Markov chain the states are the same as in PFM, but transitions to states with more than two rearrangements of the same isotypes (e.g., state 5 displayed in table 2.3) are not allowed by setting to zero the corresponding rows of the transition rate matrix $\mathbf{K} = \{K_{i,j}\}$. In the monoallelic period, equation (2.8) for the two examples converts to

$$K_{1,j} = \begin{cases} -\mu_1 - \mu_4, & j = 1 \\ \mu_4(1 - f_4), & j = 2 \\ \mu_4 f_4, & j = 3 \\ \mu_1(1 - f_1), & j = 10 \\ \mu_1 f_1, & j = 19 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad K_{3,j} = \begin{cases} -\mu_1 - \mu_4, & j = 3 \\ \mu_d, & j = 4 \\ \mu_4(1 - f_4), & j = 6 \\ \mu_4 f_4, & j = 8 \\ \mu_1(1 - f_1), & j = 12 \\ \mu_1 f_1, & j = 21 \\ 0, & \text{otherwise} \end{cases} \quad (2.13)$$

because we have only one allele accessible to recombination for each gene. The rest of the transition rates are similarly determined as in these examples. With the two Markov chains, we have one more parameter defining the relative duration of the monoallelic and biallelic periods, respectively t_m and $1 - t_m$. Note that, when $t_m = 0$, the FACAM model reduces to the PFM model, which is nested in the former. The state probability vector at the end of the monoallelic period is calculated by

$$\mathbf{p}_M(t_m) = \mathbf{p}_M(0) \exp \{t_m \mathbf{K}\}. \quad (2.14)$$

where $\mathbf{p}_M(0) = \{1, 0, \dots, 0\}$. Using $\mathbf{p}_M(t_m)$ as the initial state probability vector of the Markov chain associated with biallelic period, the state probability vector at the end of the unitary time window is

$$\mathbf{p}_B(1) = \mathbf{p}_M(t_m) \exp \{(1 - t_m) \mathbf{Q}\}, \quad (2.15)$$

where Q represents the transition rate matrix of the PFM Markov chains. The classification probabilities of the observed table under this mechanism are given by their corresponding state probabilities at the end of the time window, similarly described for the PFM.

Data Fitting and Parameter Estimation Procedures

The inference of the three rearrangement mechanisms were based on the maximum likelihood (ML) methodology with the help of several numerical procedures implemented in Mathematica 3.0 (Wolfram, 1996). The unconditional goodness-of-fit tests were the usual Pearson's and Wilks's likelihood ratio tests, whose statistics have under the null hypothesis an asymptotic chi-square distribution with the degrees of freedom given by the difference between the number of free parameters of the sampling distribution and the number of parameters of the models (see, for example, Agresti 2002). To compare two nested models (e.g. PFM and FACAM), the conditional goodness-of-fit tests were the same as above, but now the null distribution of the test statistics is asymptotically chi-square with the degrees of freedom given by the difference of dimensions between the parametric spaces of two models.

We fixed the in-frame probabilities f_1 and f_4 at 0.33 and 0.15, respectively. The rationale for the second probability is that an in-frame stop codon is present at the end of the $V_{\gamma 4}$ gene, which must be excised during recombination to produce a functional chain. The frequency of this excision has been estimated experimentally to be 50% (Capone et al., 1998; Kang et al., 1995).

The likelihood equations for the remaining TWM parameters are:

$$\begin{aligned} \frac{n_{12\cdot} + n_{2\cdot 2} + 2n_{2\cdot 3}}{p_1} - \frac{n_{11\cdot} + 2n_{2\cdot 1} + n_{2\cdot 2}}{1 - p_1} + \frac{n_{1\cdot\cdot} f_1}{2 - p_1 f_1} \\ + \frac{2n_{2\cdot\cdot} \left[(1 - p_1) - (1 - 2p_1)(1 - f_1) - p_1(1 - f_1)^2 \right]}{\left[(1 - p_1)^2 + 2(1 - p_1)p_1(1 - f_1) + p_1^2(1 - f_1)^2 \right]} = 0 \\ \frac{n_{22\cdot} + n_{1\cdot 2} + 2n_{1\cdot 3}}{p_4} - \frac{n_{21\cdot} + 2n_{1\cdot 1} + n_{1\cdot 2}}{1 - p_4} + \frac{n_{2\cdot\cdot} f_4}{2 - p_4 f_4} = 0. \end{aligned} \quad (2.16)$$

These are second order equations and therefore their solutions can be trivially derived. Since the expressions are complicated we opted not to present them here.

The ML estimation of the PFM and the FACAM parameters are cumbersome, since there are no closed form expressions for the classification probabilities of the observed table. So, we handled the fitting procedure with the steepest descent algorithm in its non-parametric way provided by Mathematica (Wolfram, 1996), which avoids a prior definition of gradient functions usually needed in other numerical methods (e.g., Newton-Raphson). We tested several initial conditions to minimize the problem of attaining local or global maximum. The best solution reported here appeared repeatedly, and thus we believe it corresponds to the "real" ML estimates. This is further substantiated by the fact that by setting to zero the rate of closing rearrangements we obtained with the PFM the same solution as with the probabilistic TWM model as expected (not shown).

Tab. 2.4: Results of goodness-of-fit test for the three rearrangement mechanisms.

Mechanism	Estimates				Pearson			Likelihood Ratio		
	μ_1	μ_4	μ_d	t_m	Observed	d.f.	p-value	Observed	d.f.	p-value
TWM	—	—	—	—	39.78	8	0.00	36.81	8	0.00
PFM	0.41	3.48	42.13	—	10.18	7	0.18	11.99	7	0.10
FACAM	0.58	7.00	93.86	0.38	2.30	6	0.89	2.87	6	0.83
PFM/FACAM	—	—	—	—	8.17	1	0.0025	9.12	1	0.0043

As mentioned above, PFM is a nested model of FACAM with $t_m = 0$. Therefore their goodness-of-fit was compared by conditional testing:

$$H_0 : t_m = 0 \text{ versus } H_1 : t_m \neq 0.$$

As described in Agresti (2002), the likelihood ratio and Pearson's test can be adapted to this conditional scenario. Now, the likelihood ratio and Pearson's statistic are respectively:

$$Q_v(PFM|FACAM) = -2 \sum_{i,j,k} n_{ijk} (\ln \tilde{\mu}_{ijk} - \ln \hat{\mu}_{ijk})$$

and

$$Q_P(PFM|FACAM) = \sum_{i,j,k} \frac{(\hat{\mu}_{ijk} - \tilde{\mu}_{ijk})^2}{\tilde{\mu}_{ijk}},$$

where $\hat{\mu}_{ijk}$ is the expected value of (i, j, k) table entry under the FACAM and $\tilde{\mu}_{ijk}$ is the expected value of (i, j, k) table entry under the PFM. Both statistics have under null hypothesis a chi-square distribution with 1 degree of freedom (one single restriction under H_0).

2.3 Results

2.3.1 Analysis of γ loci structure in mature thymocytes reveals a rearrangement mechanism involving both feedback and asynchronous chromosome accessibility

A simple time window for rearrangement in both chromosomes is the most parsimonious hypothesis for the γ genes rearrangement mechanism. The ML estimates of the TWM model parameters are $\hat{p}_1 = 0.12$ and $\hat{p}_4 = 0.62$ (Table 2.4). The observed frequencies of mature thymocytes with $V_{\gamma}4J_{\gamma}1$ and $V_{\gamma}1J_{\gamma}4$ rearrangements are strikingly different from the expected values (Table 2.5). The null p-values of the goodness-of-fit tests presented in Table 2.4 confirm this discrepancy and allow to confidently reject the TWM model. What stands out from Table 2.5 is that within mature $V_{\gamma}4+$ thymocytes considerable fewer double $V_{\gamma}4J_{\gamma}1$ rearrangements were observed than it would be expected according to the TWM. This result suggests that some additional process disfavours or prevents double $V_{\gamma}4J_{\gamma}1$ rearrangements.

The frequency of cells with double rearrangements is reduced by feedback mechanisms that prevent further rearrangements in the remaining loci once the first in-frame rearrangement is completed. Indeed, the PFM model improves the quality of the fitting to the experimental data as compared to the TWM model (the p-values are acceptable: 18% and 10% for Pear-

Tab. 2.5: Best fit of the three mechanisms using the maximum likelihood methodology.

Mechanism	$V_\gamma 1+$ Cells		$V_\gamma 4J_\gamma 1$		$V_\gamma 4+$ Cells		$V_\gamma 1J_\gamma 4$	
	$V_\gamma 1J_\gamma 4$	(0, 0)	(1, 0)	(1, 1)	$V_\gamma 4J_\gamma 1$	(0, 0)	(1, 0)	(1, 1)
Observed	(1*, 0)	17	50	54	(1*, 0)	49	4	0
TWM		18.25	59.08	47.82		29.40	5.21	0.23
PFM		24.51	39.44	58.53		42.11	3.65	0.15
FACAM		17.73	49.97	54.51		46.29	2.71	0.02
Observed	(1*, 1)	0	4	14	(1*, 1)	28	6	0
TWM		2.02	6.54	5.29		44.02	7.79	0.34
PFM		1.13	4.11	11.28		33.69	6.88	0.52
FACAM		0.11	3.23	13.45		29.20	8.32	0.46

son's and Wilks's likelihood ratio tests, respectively). The ML estimates $\hat{\mu}_1 = 0.41$, $\hat{\mu}_4 = 3.48$ and $\hat{\mu}_d = 42.13$, which suggest that the $V_\gamma 4J_\gamma 1$ gene rearranges much faster than $V_\gamma 1J_\gamma 4$. Notwithstanding, the expected values of n_{111} and n_{112} in the $V_\gamma 1+$ cells according to the model are not correct. In n_{111} , rearrangements of the $V_\gamma 4J_\gamma 1$ gene are less observed than expected, whereas in n_{112} more single rearrangements of the $V_\gamma 4J_\gamma 1$ gene are observed than expected. A mechanism, such as FACAM, in which only one of the chromosomes is accessible to recombination for a while allows for better predictions, because it can bias the output towards cells with single rearrangements.

The ML estimates of the FACAM parameters are $\hat{\mu}_1 = 0.58$, $\hat{\mu}_4 = 7.00$, $\hat{\mu}_d = 93.86$ and $\hat{t}_m = 0.38$. In comparison to the previous estimates with the PFM, we have now higher estimates of the rearrangement rates μ_1 and μ_4 . This was expected, since increasing the rearrangement rates will tend to increase the frequency of rearrangements during the monoallelic period. Consequently, the expected frequencies in $V_\gamma 1+$ table will tend to concentrate in n_{112} and n_{113} . According to the estimate of t_m , the monoallelic period lasts 38% of the time, whereas the rest of the time has biallelic accessibility. As we expected, the fit was improved in comparison to the two previous mechanisms (see Table 2.5). The expected values are in good agreement with the observed ones and the p-values are very high.

Since there is evidence favoring both feedback models, PFM and FACAM, we determined which is the best model using conditional testing (as described in methods). The results displayed in Table 2.4 give strong evidence against PFM model. Therefore, we conclude that the FACAM is the best model among the three models tested and for the data at hand, suggesting the importance of asynchronous accessibility of the chromosomes during the rearrangement of the TCR γ chain.

2.3.2 Further analysis of the rearrangement process according to FACAM

Since the FACAM model fits very well the experimental data we proceed to a more detailed biological analysis of its properties. This model makes additional predictions of the status of $V_\gamma 1$ and $V_\gamma 4$ loci in mature thymocytes that can be compared to other experimental observations (Table 2.6). According to this model, a cell has 44% chance of having at least one in-frame $V_\gamma 4$ or $V_\gamma 1$ gene rearrangement. The estimated ratio between $V_\gamma 4+$ and $V_\gamma 1+$ cells is around

Tab. 2.6: Other estimates derived from the ML estimates of the FACAM at the end of the time window.

Features	Estimates
Probability of an in-frame γ rearrangement	0.44
Ratio between $V_{\gamma}4+$ and $V_{\gamma}1+$ cells	1.39
Probability of double in-frame rearrangements in $V_{\gamma}1+$	
$V_{\gamma}1J_{\gamma}4$	5×10^{-3}
$V_{\gamma}4J_{\gamma}1$	3×10^{-6}
$V_{\gamma}1J_{\gamma}4$ and $V_{\gamma}4J_{\gamma}1$	8×10^{-3}
Probability of double in-frame rearrangements in $V_{\gamma}4+$	
$V_{\gamma}4J_{\gamma}1$	4×10^{-4}

1.39, and thus it is expected that there are more $V_{\gamma}4+$ thymocytes than $V_{\gamma}1+$ thymocytes. The thymocytes expressing two different receptors, i.e. allelically or isotypically included, should be very rare ($< 1\%$). Some of these quantitative predictions have been confirmed elsewhere (Boucontet et al., 2005).

Previous theoretical analysis of the rearrangement mechanism of antigen receptor loci were only concerned with the final outcome, and could not provide information about intermediate loci configurations in precursors (Claverie and Langman, 1984; Kitamura and Rajewsky, 1992; Wabl and Steinberg, 1992). The FACAM model follows in time all loci configurations so it can be used to infer composition of the $\gamma\delta$ T cell precursor pool. It can even provide estimates of the frequency of precursors that never completed differentiation and therefore could not be measured experimentally by examining mature thymocytes. Figure 2.2 displays the probabilities of some γ loci configuration as a function of the time elapsed since a $\gamma\delta$ T cell precursor initiated rearrangement. Except for the bottom curve in this figure, each curve corresponds to the summed probabilities of a set of loci-configurations that contains the subset of the curve right below it. Under these conditions, assuming that a constant number of precursors initiate rearrangement per unit of time, then the composition of the precursor pool corresponds to the areas delimited by the curves. There are a few predictions of the FACAM model that can be easily inferred from this figure. For example, the low frequency of allelic or isotypic inclusion above mentioned is reflected in the apparent overlap of the curves with only one productive rearrangement and the cells with more than one rearrangement (the second and third curve from bottom to top). Also, evident from this graph is that the frequency of cells that did not exhaust all rearrangement possibilities in the two γ loci, in particular in the $V_{\gamma}1$ gene, is significant. Thus it amounts to 23% (corresponding to difference between the 4th and 3rd curve from bottom to top at time=1.0). Finally, within the pool of precursors that have remodeled the chromatin to make γ loci accessible to the recombination machinery, about 55-62% cells will have both alleles accessible whereas the remaining will have only one allele accessible. The upper estimate for biallelic accessibility is simply the area to the right of the vertical separator, and assumes that cells that ceased rearrangement upon productive rearrangement are still precursors. The lower estimate corresponds to the area to the right of the vertical separator which is above the 3rd curve from the bottom, and assumes that cells which underwent at least one productive rearrangement before ceasing rearrangements are no longer precursors, i.e., they differentiated.

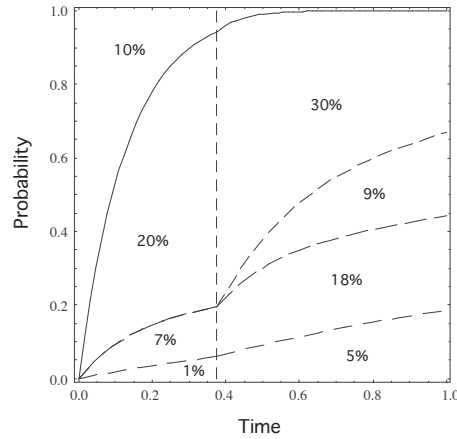


Fig. 2.2: Probabilities of some γ loci configurations as a function of the time elapsed since the $\gamma\delta$ precursor cell initiated rearrangement. From bottom to top the curves are: the probability of only one $V_{\gamma 1}$ in-frame rearrangement, the probability of only one in-frame rearrangement in either $V_{\gamma 1}$ or $V_{\gamma 4}$, the probability of more than one in-frame rearrangement (which overlaps with the previous curve), the previous probability added with the probability that all genes rearranged except one $V_{\gamma 1}$ that is in germline configuration, and at least one rearrangement (whether it is in-frame or not). The percentages indicate the areas in the graph, which are delimited by these curves and the vertical line separating the periods in which one or two chromosomes are accessible.

It is reasonable to assume that the parameter values of the rearrangement mechanism have been evolutionary selected to increase the fitness of the organism. Most likely the probability that a precursor gives rise to a mature lymphocyte containing at least one functional allele — the yield of the rearrangement process — will be under tight regulation. On the other hand, immunological common sense tells us that allelic or isotypic inclusion can be disadvantageous. In this context, we asked how the yield and allelic exclusion as predicted by the FACAM depend on model parameters. This analysis reveals a tradeoff between yield and allelic exclusion when both the $V_{\gamma 1}$ and $V_{\gamma 4}$ rearrangement rates are changing. As both rates increase, the rearrangement process yield increases until a plateau is reached at about 60% surviving cells (Figure 2.3A). But the increase in rearrangement rates increases the production of cells with allelic or isotypic inclusion, i.e. cells expressing more than one functional γ gene (Figure 2.3B). Interestingly, the estimated values are on the edge of the yield plateau, and at low frequencies of double expressing cells, as if this would be a compromise between the two outcomes. It should be noted that the same tradeoff is obtained with the PFM, but not with the TWM (not shown), but the percentage of cells with isotypic inclusion is higher with PFM than with the FACAM.

2.4 Discussion

This paper addressed the rearrangement process of the two major γ genes expressed in the TCR of $\gamma\delta$ T cells ($V_{\gamma 1}J_{\gamma 4}$ and $V_{\gamma 4}J_{\gamma 1}$). Three potential models were studied: a simple probabilistic time window mechanism, a pure feedback mechanism, and a feedback mechanism with asynchronous chromosome accessibility. According to experimental data, there is evi-

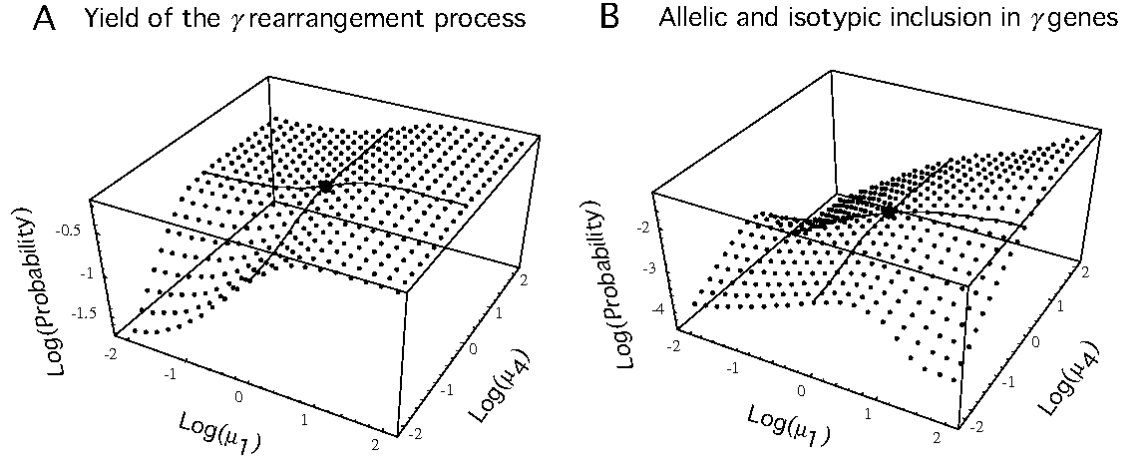


Fig. 2.3: Tradeoff between yield and allelic exclusion γ gene rearrangement. The logarithm of the probability that cells have at least one-in frame V_γ gene rearrangement (i.e. yield, A) and logarithm of the probability that cells have more than one-in frame V_γ gene rearrangement (i.e. allelic inclusion, B) are plotted as a function of the logarithm of the rates of μ_1 and μ_4 according to FACAM model. The lines and the larger dot indicate the maximum likelihood estimates of the parameters.

dence favoring the two feedback mechanisms and against the time window mechanism using unconditional goodness-of-fit testing. Moreover, the feedback mechanism with asynchronous chromosome accessibility shows an almost perfect fit to the data, which suggests that it is more realistic than the other candidates. This evidence was confirmed by conditional testing of the pure feedback model against its rival.

The above mechanisms were mathematically described by two different stochastic modeling approaches, namely a pure probabilistic formalism for the time window and a Markov chain formalism for the two feedback mechanisms. The latter formalism is a significant improvement concerning the former probabilistic model, or other probabilistic models available in the literature (Kitamura and Rajewsky, 1992; Wabl and Steinberg, 1992) since it describes explicitly the rates of concurrent processes and follows in time all locus-configurations. Probabilistic models are often simplified by making ad hoc assumptions allowing to ignore concurrent events. For example, Kitamura and Rajewsky (1992) and Wabl and Steinberg (1992) assumed that precursor cells would first try to rearrange one allele and only upon failure to make a productive rearrangement they would proceed to rearrange the second allele. Concurrent rearrangement processes in T and B cell precursors have been modeled before (Kalmanovich and Mehr, 2003; Louzoun et al., 2002; Mehr et al., 1999; Piper et al., 1999), using Monte-Carlo simulations of the fate of a finite number of cells. We adopted a continuous-time Markov chain approach, which allowed us to calculate the time-dependent probability distributions of the γ loci-configurations. Furthermore, using this approach we obtained maximum-likelihood parameter estimates, which could not be easily done using simulations.

Most previous models of rearrangement of Ig or TCR genes have assumed that lymphocyte precursors would try all rearrangement possibilities before they mature or die (Claverie and Langman, 1984; Kalmanovich and Mehr, 2003; Kitamura and Rajewsky, 1992; Mehr et al.,

1999; Piper et al., 1999; Wabl and Steinberg, 1992). Louzoun et al. (2002) introduced a "rigid time window" by assuming that B cell precursors would only make a limited number of rearrangement attempts. Analogously, in our continuous-time γ gene rearrangement models a precursor can only activate the rearrangement for a finite period of time. At the end of this time window, if a precursor did not assemble a functional gene it dies even if some germ-line genes remain. The FACAM model operating under the maximum-likelihood parameter values predicts that the number of precursors that made no rearrangement during the time window is negligible. However, about 23% of all precursors exit the rearrangement program with one $V_{\gamma}1$ gene still in germline configuration, and without a productive rearrangement. Therefore, at least in the case of the γ genes it would be unrealistic to assume that precursors would exhaust all rearrangement possibilities.

A possible criticism to the present analysis is the study of the γ gene rearrangement process in isolation from the other loci rearranging within the same time window. Despite an apparent independence from the δ locus (Itoharu et al., 1993), the γ locus can be affected by rearrangements in the β locus (Fenton et al., 1988; Krotkova et al., 1997; von Boehmer et al., 1999) and to a lesser extent in the α locus (Bruno et al., 1996; Fritsch et al., 1998; Terrence et al., 2000)). To address the role of competing α and β locus in the present analysis, we extended the TWM model for the γ genes to include the possibility of stopping γ rearrangement upon an in-frame rearrangement in competing loci. The predictions of this model diverged significantly from the data (p-values < 0.0001 for the Pearson's and likelihood ratio goodness-of-fit tests). We also extended the PFM Markov chains to include transitions out of the γ lineage upon rearrangement in the α and β locus. The fit was similar to that of PFM alone (p-values of 0.09 and 0.17 for the likelihood ratio and Pearson's statistic, respectively), and since the two models are nested, we can neglect the effect of the β locus. Finally, since the FACAM, as stated previously, is in close agreement with the data, any model that includes it cannot be considered the "best model" by conditional testing. Therefore, the rearrangements in the other loci do not influence the results and the conclusions presented in this paper. In other words, the γ gene rearrangement process seems to be independent of the α and β locus.

The mechanism underlying allelic exclusion in the genes encoding antigen receptors is an old immunological puzzle. A time window of availability of the RAG proteins during lymphocyte development is well documented and could potentially explain allelic exclusion (Cohn and Langman, 1990). Likewise, the feedback inhibition of rearrangement upon expression of the antigen receptor has been demonstrated using transgenic animals providing another explanation (Alt et al., 1981; Kitamura and Rajewsky, 1992). Our present analysis of γ gene rearrangements suggests that these two mechanisms isolated or together are unable to explain allelic exclusion, since they overestimate the number of double rearrangements. Thus, minimizing the occurrence of double rearrangements calls for a process of chromatin remodeling responsible for different accessibility of the γ gene loci to the recombination machinery, as suggested before (Stanhope-Baker et al., 1996; Yancopoulos and Alt, 1985).

Asynchronous replication might be involved in the asynchronous opening of the immunoglobulin κ locus (Mostoslavsky et al., 2001). We have shown that the TCR γ locus data can be explained if one of the chromosomes becomes accessible before the other. The most parsimonious

monious explanation would be that if the opening the locus is slow and stochastic then there will always be one locus opening before the other. However, B cell precursors undergoing recombination show mostly methylation of the κ locus in only one of the chromosomes (Goldmit et al., 2002). This has been interpreted as an indication that the opening of the loci is asynchronous, as we assumed here, but furthermore that the accessibility of the two chromosomes is mutually exclusive, i.e., at any time only one chromosome is accessible in the precursors. According to our model, both chromosomes could be simultaneously open, albeit one 'opened' first. Due to the relatively fast feedback mechanism and the rates of rearrangement, we estimated that about 60% of the $\gamma\delta$ T cell precursors would have both alleles accessible. This is yet another prediction that could be tested to further validate our model.

The parameter estimates of the model featuring asynchronous accessibility and feedback suggests an evolutionary tradeoff between the yield of the rearrangement process — the probability at which precursors reach maturity — and the frequency of cells expressing more than one functional receptor. We estimated that at least 23% of all precursor cells would die without even attempting to rearrange both $V\gamma 1$ alleles. This apparent waste could be potentially reduced by increasing the rate of $V\gamma 1$ rearrangements or prolonging the window for rearrangements. However, it might be impossible to reduce the waste by increasing the rate of $V\gamma 1$ rearrangements without compromising isotypic exclusion, which is facilitated by the differences in $V\gamma 4$ and $V\gamma 1$ productive rearrangement rates (the former being about 7 fold faster than the latter). Reducing waste by increasing the time window might also not be possible since this may interfere with the generation of $\alpha\beta$ T cells from the same precursor cells, or it may even allow RAG recombinases to mediate deleterious chromosomal translocations (Raghavan et al., 2004; Tycko and Sklar, 1990).

It is commonly assumed that allelic and isotypic exclusion are necessary to minimize the possibility that lymphocytes with two functional antigen receptors open the way for pathologic autoimmunity (Sarukhan et al., 1998; Zal et al., 1996). The rationale is that lymphocytes bearing autoreactive and pathogen-specific receptors could mount autoimmune responses if activated during an infection with that pathogen. Since the γ genes are also allelically and isotypically excluded one is tempted to reason that $\gamma\delta$ T cells could potentially cause autoimmune pathologies. The observation that α gene deficient animals show autoimmune dysregulation suggests that this might be the case (Wen et al., 1994), however, $\gamma\delta$ T cells have mostly been attributed a protective role in autoimmune disease models. Therefore, the selective pressure against double expressing $\gamma\delta$ T cells might be of a different nature. Several infection models suggest that $V\gamma 4$ and $V\gamma 1$ $\gamma\delta$ T cells may play different complementary functions (O'Brien et al., 2001). Allelic exclusion might be required to avoid interference between these functions, which can be independently regulated by controlling independent $V\gamma 4+$ $V\gamma 1-$ and $V\gamma 1+$ $V\gamma 4-$ subpopulations, but not a subpopulation of $V\gamma 4+$ $V\gamma 1+$ cells.

To conclude, it is worth noting that the present analysis benefited from the fact that in practice the universe of rearrangement possibilities in the murine γ locus amounts to one $V\gamma$ gene segment per isotype. A similar analysis would not be feasible in the cases of the loci encoding other TCR genes or Ig genes in which larger numbers of V gene segments can rearrange to each J segment. Furthermore, because the γ locus involves both isotypic and

allelic exclusion we could dissociate the contribution of chromatin accessibility from that of the feedback mechanism to minimizing the production of mature $\gamma\delta$ T cells expressing two γ chains.

Bibliography

- Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons, New York, 2nd edition.
- Aifantis, I., Buer, J., von Bohemer, H., and Azugui, O. (1997). Essential role of the pre-T cell receptor in allelic exclusion of the T cell receptor β locus. *Immunity*, 7:601–607.
- Alt, F. W., Rosenberg, N., Lewis, S., Thomas, E., and Baltimore, D. (1981). Organization and reorganization of immunoglobulin genes in a-MuLV-transformed cells: Rearrangement of heavy but not light chain genes. *Cell*, 27:381–390.
- Bergman, Y., Fisher, A., and Cedar, H. (2003). Epigenetic mechanisms that regulate antigen receptor gene expression. *Curr. Opin. Immunol.*, 15:176–181.
- Boucontet, L., Sepúlveda, N., Carneiro, J., and Pereira, P. (2005). Mechanisms controlling termination of V-J recombination at the TCR- γ locus: Implications for allelic and isotypic exclusion of TCR- γ chains. *J. Immunol.*, 174:3912–3919.
- Bruno, L., Fehling, H. J., and van Boehmer, H. (1996). The $\alpha\beta$ T cell receptor can replace the $\gamma\delta$ receptor in the development of $\gamma\delta$ lineage cells. *Immunity*, 5:343–.
- Burnet, F. M. (1957). A modification of Jerne's theory of antibody production using the concept of clonal selection. *Aust. J. Sci.*, 20:67–77.
- Capone, M., Hockett Jr., R. D., and Zlotnik, A. (1998). Kinetics of T cell receptor β , γ , and δ rearrangements during adult thymic development: T cell receptor rearrangements are present in CD44+CD25+ pro-T thymocytes. *Proc. Natl. Acad. Sci. USA*, 95:12522.
- Claverie, J.-M. and Langman, R. E. (1984). Models for the rearrangement of immunoglobulin genes: a computer view. *TiBS*, 9:293–296.
- Cohn, M. and Langman, R. E. (1990). The protection: the unit of humoral immunity selected by evolution. *Immunol. Rev.*, 115:11–147.
- Coleclough, C. (1983). Chance, necessity and antibody gene dynamics. *Nature*, 303:23–26.
- Feeney, A. J., Goebel, P., and Espinoza, C. R. (2004). Many levels of control of V gene rearrangement frequency. *Immunol. Rev.*, 200:44–56.
- Fenton, R. G., Marrack, P., Kappler, J. W., Kanagawa, O., and Seidman, J. G. (1988). Isotypic exclusion of $\gamma\delta$ T cell receptors in transgenic mice bearing a rearranged β -chain gene. *Science*, 241:1089–1092.

- Fritsch, M., Andersson, A., Petersson, K., and Ivars, F. (1998). A TCR α chain transgene induces maturation of CD4-CD8- $\alpha\beta$ T cells from $\gamma\delta$ T cell precursors. *Eur. J. Immunol.*, 28:828–837.
- Gilfillan, S., Dierich, A., LeMeur, M., Benoist, C., and Mathis, D. (1993). Mice lacking Tdt: mature animals with an immature lymphocyte repertoire. *Science*, 261:1175–1178.
- Goldmit, M., Schlissel, M., Cedar, H., and Bergman, Y. (2002). Differential accessibility at the κ chain locus plays a role in allelic exclusion. *EMBO J.*, 21:5255–5261.
- Gu, H., Kitamura, D., and Rajewsky, K. (1991). B cell development regulated by gene rearrangement: arrest of maturation by membrane-bound $D\mu$ protein and selection of DH element reading frames. *Cell*, 65:47–54.
- Itoharu, S., Mombaerts, P., Lafaille, J., Iacomini, J., Nelson, A., Clarke, A. R., Hooper, M. L., Farr, A., and Tonegawa, S. (1993). T cell receptor δ gene mutant mice: independent generation of $\alpha\beta$ T cells and programmed rearrangements of $\gamma\delta$ TCR genes. *Cell*, 72:337–348.
- Jerne, N. K. (1955). The natural-selection theory of antibody formation. *Proc. Natl. Acad. Sci. USA*, 41:849–857.
- Kalmanovich, G. and Mehr, R. (2003). Models for antigen receptor gene rearrangement. III. Heavy and light chain allelic exclusion. *J. Immunol.*, 170:182–193.
- Kang, J., Baker, J., and Raulet, D. H. (1995). Evidence that productive rearrangements of TCR γ genes influence the commitment of progenitor cells to differentiate into $\alpha\beta$ or $\gamma\delta$ T cells. *Eur. J. Immunol.*, 25:2706–2709.
- Khor, B. and Sleckman, B. P. (2002). Allelic exclusion at the TCR β locus. *Curr. Opin. Immunol.*, 14:230–234.
- Kitamura, D. and Rajewsky, K. (1992). Targeted disruption of μ chain membrane exon causes loss of heavy-chain allelic exclusion. *Nature*, 356:154–156.
- Komori, T., Okada, A., Stewart, V., and Alt, F. W. (1993). Lack of N regions in antigen receptor variable region genes of Tdt-deficient lymphocytes. *Science*, 261:1171–1175.
- Krotkova, A., von Boehmer, H., and Fehling, H. J. (1997). Allelic exclusion in $pt\alpha$ -deficient mice: no evidence for cell surface expression of two T cell receptor (TCR)- β chains, but less efficient inhibition of endogeneous $V\beta \rightarrow (D)J\beta$ rearrangements in the presence of a functional TCR-beta transgene. *J. Exp. Med.*, 186:767.
- Kulkarni, V. (1995). *Modeling and Analysis of Stochastic Systems*. Chapman & Hall.
- Livák, F., Tourigny, M., Schatz, D., and Petrie, H. (1999). Characterization of TCR gene rearrangements during adult murine T cell development. *J. Immunol.*, 162:2575–2580.
- Louzoun, Y., Friedman, T., Prak, E. L., Litwin, S., and Weigert, M. (2002). Analysis of B cell receptor production and rearrangement. Part I. light chain rearrangement. *Semin. Immunol.*, 14:169–190.

- Malissen, M., Trucy, J., Jouvin-Marche, E., Cazenave, P.-A., Scollay, R., and Malissen, B. (1992). Regulation of TCR α and β gene allelic exclusion during T-cell development. *Immunol. Today*, 13:315–322.
- Mehr, R., Shannon, M., and Litwin, S. (1999). Models for antigen receptor gene rearrangement. I. Biased receptor editing in B cells: Implications for allelic exclusion. *J. Immunol.*, 163:1793–1798.
- Mostoslavsky, R., Singh, N., Tenzen, T., Goldsmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B. E., Chess, A., Cedar, H., and Bergman, Y. (2001). Asynchronous replication and allelic exclusion in the immune system. *Nature*, 414:221–225.
- O'Brien, R. L., Lahn, M., Born, W. K., and Huber, S. A. (2001). T cell receptor and function cosegregate in $\gamma\delta$ T cell subsets. *Chem. Immunol.*, 79:1–28.
- Oettinger, M. A., Schatz, D. G., Gorka, C., and Baltimore, D. (1990). Rag-1 and rag-2, adjacent genes that synergistically activate V(D)J recombination. *Science*, 248:1517–1523.
- Pereira, P. and Boucontet, L. (2005). Rates of recombination and chain pair biases greatly influence the primary $\gamma\delta$ TCR repertoire in the thymus of adult mice. *J. Immunol.*
- Piper, H., Litwin, S., and Mehr, R. (1999). Models for antigen receptor gene rearrangement. II. Multiple rearrangement in the TCR: Allelic exclusion or inclusion. *J. Immunol.*, 163:1799–1808.
- Raghavan, S. C., Swanson, P. C., Wu, X., Hsieh, C. L., and Lieber, M. R. (2004). A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. *Nature*, 428:88–93.
- Rooney, S., Chaudhuri, J., and Alt, F. W. (2004). The role of the non-homologous end-joining pathway in lymphocyte development. *Immunol. Rev.*, 200:115–131.
- Sarukhan, A., Garcia, C., Lanoue, A., and von Boehmer, H. (1998). Allelic inclusion of t cell receptor alpha genes poses an autoimmune hazard due to low-level expression of autospecific receptors. *Immunity*, 8:563–570.
- Sleckman, B., Bassing, C., Bardon, C., Okada, A., Khor, B., Bories, J. C., Monroe, R., and Alt, F. W. (1998). Accessibility control of variable region gene assembly during T-cell development. *Immunol. Rev.*, 165:121–130.
- Stanhope-Baker, P., Hudson, K. M., Shaffer, A. L., Constantinescu, A., and Schlissel, M. S. (1996). Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. *Cell*, 85:887–897.
- Terrence, K., Pavlovich, C. P., Matechak, E. O., and Fowlkes, B. J. (2000). Premature expression of T cell receptor (TCR) $\alpha\beta$ suppresses TCR $\gamma\delta$ gene rearrangement but permits development of $\gamma\delta$ lineage. *J. Exp. Med.*, 192:537–548.

- Tonegawa, S. (1993). Somatic generation of antibody diversity. *Nature*, 302:575–581.
- Tycko, B. and Sklar, J. (1990). Chromosomal translocations in lymphoid neoplasia: a reappraisal of the recombinase model. *Cancer Cells*, 2:1–8.
- Uematsu, Y., Ryser, S., Dembic, Z., Borgulya, P., Krimpenfort, P., Berns, A., von Boehmer, H., and Steinmetz, M. (1988). In transgenic mice the introduced functional cell receptor gene prevents expression of endogenous genes. *Cell*, 52:831–841.
- von Boehmer, H., Aifantis, I., Azogui, O., Feinberg, J., Saint-Ruf, C., Zober, C., Garcia, C., and Buer, J. (1998). Crucial function of the pre-T-cell receptor (TCR) in TCR β selection, TCR β allelic exclusion and $\alpha\beta$ versus $\gamma\delta$ lineage commitment. *Immunol. Rev.*, 165:111–119.
- von Boehmer, H., Aifantis, I., Feinberg, J., Lechner, O., Saint-Ruf, C., Walter, U., Buer, J., and Azogui, O. (1999). Pleiotropic changes controlled by the pre-T-cell receptor. *Curr. Op. Immunol.*, 11:135–142.
- Wabl, M. and Steinberg, C. (1992). Allelic exclusion model questioned. *Nature*, 359:370–371.
- Wen, L., Roberts, S. J., Viney, J. L., Wong, F. S., Mallick, C., Findly, R. C., Peng, Q., Craft, J. E., Owen, M. J., and Hayday, A. C. (1994). Immunoglobulin synthesis and generalized autoimmunity in mice congenitally deficient in $\alpha\beta$ (+) T cells. *Nature*, 369:654–658.
- Wilson, A., Held, W., and MacDonald, H. R. (1994). Two waves of recombinase gene expression in developing thymocytes. *J. Exp. Med.*, 179:1355–1360.
- Wolfram, S. (1996). *The Mathematica Book*. Cambridge University Press, Cambridge, 3rd edition.
- Yancopoulos, G. D. and Alt, F. W. (1985). Developmentally controlled and tissue specific expression of unrearranged VH gene segments. *Cell*, 40:271–281.
- Zal, T., Weiss, S., Mellor, A., and Stockinger, B. (1996). Expression of a second receptor rescues self-specific T cells from thymic deletion and allows activation of autoreactive effector function. *Proc. Natl. Acad. Sci. USA*, 93:9102–9107.

3. MECHANISMS CONTROLLING TERMINATION OF V-J RECOMBINATION AT THE TCR γ LOCUS

Boucontet, L. Sepúlveda, N., Carneiro, N. and Pereira, P. (2005). Mechanisms controlling termination of V-J recombination at the TCR γ locus: Implications for allelic and isotypic exclusion of TCR γ chains. *J. Immunol.*, 174:3912-3919.

Abstract

Analyses of V γ -J γ rearrangements producing the most commonly expressed TCR γ chains in over 200 $\gamma\delta$ TCR⁺ thymocytes showed that assembly of TCR γ V-region genes display properties of allelic exclusion. Moreover, introduction of functionally rearranged TCR γ and δ transgenes results in a profound inhibition of endogeneous TCR γ rearrangements in progenitor cells. The extent of TCR γ rearrangements in these cells is best explained by a model in which initiation of TCR γ rearrangements at both alleles is asymmetric, occurs at different frequencies depending on the V or J segments involved and is terminated upon production of a $\gamma\delta$ TCR. Approximately 10% of the cells studied contained two functional TCR γ chains involving different V and J γ gene segments, thus defining a certain degree of isotypic inclusion. However, these cells are isotypically excluded at the level of cell surface expression possibly due to pairing restrictions between different TCR γ and δ chains.

3.1 Introduction

Exons encoding the variable regions of lymphocyte receptors for antigen are assembled during lymphocyte ontogeny from clusters of V (D) and J segments in a process known as V(D)J recombination (Tonegawa, 1983). T cell precursors can rearrange up to four different loci (α , β , γ and δ) and express either of two different TCR ($\alpha\beta$ and $\gamma\delta$), thus defining two T lymphocyte populations (the $\alpha\beta$ and the $\gamma\delta$ T cell populations). Most TCR δ , TCR γ and TCR β rearrangements take place in a population of thymocytes known as pre-T cells and that is characterized by low expression of CD44, high expression of CD25 and lack of expression of the CD4 or CD8 markers (Capone et al., 1998; Godfrey and Zlotnik, 1993; Livak et al., 1999), whereas most V α to J α rearrangements occur later in development in cells co-expressing the CD4 and CD8 molecules (Petrie et al., 1995; Wilson et al., 1996). Functionally rearranged TCR β chains associate with the non-rearranging surrogate α -chain (pre-T α) forming the pre-TCR (Fehling et al., 1995). Only cells expressing the pre-TCR efficiently traverse a developmental checkpoint usually referred to as TCR β selection (Dudley et al., 1994) and further differentiate into the $\alpha\beta$ T cell pathway (von Boehmer et al., 1999). Such differentiation is accompanied by downregulation of RAG-1/2 expression, silencing of TCR γ transcription (thus avoiding expression of

a $\gamma\delta$ TCR in $\alpha\beta$ lineage cells), acquisition of the CD4 and CD8 markers and upregulation of RAG-1/2 expression allowing the initiation of rearrangements at the *TCR α* locus (Wilson et al., 1994). In contrast, no surrogate chains for $\gamma\delta$ lineage cells have been identified and evidence suggest that simultaneous expression of functional *TCR γ* and *TCR δ* chains that can efficiently pair are required for differentiation along the $\gamma\delta$ T cell lineage (Kang et al., 1998; Passoni et al., 1997; Pereira and Boucontet, 2004).

Since lymphocytes are diploid cells, the recombination process could, theoretically, result in the production of two productively rearranged TCR alleles and therefore in the expression at the cell surface of more than one TCR specificity. Analyses of a relatively large number of $\alpha\beta$ T cells and clones have shown that, as a rule, mature T cells contain only one productive *TCR β* rearrangement, whereas the second allele is either nonproductively rearranged or contain their *V β* gene segments in germline configuration (Aifantis et al., 1997; Malissen et al., 1992; Seboun et al., 1992). In contrast, most mature $\alpha\beta$ T cells contain two *TCR α* rearrangements and 20-30% of them bear two productively rearranged *TCR α* chains (Aifantis et al., 1997; Malissen et al., 1992). Thus, assembly of *TCR β* chain V region is regulated in a way that allows only one of the two alleles to be expressed in each cell (a phenomenon usually referred to as allelic exclusion or genotypic allelic exclusion) whereas that of *TCR α* chain V-region genes is regulated differently permitting allelic inclusion. Post-translational mechanisms appear to limit T cells carrying two functional *TCR α* rearrangements to the expression of a single *TCR α* chain at the cell surface (Gascoigne and Alam, 1999) (a phenomenon usually referred to as phenotypic allelic exclusion). Together, these mechanisms ensure that most $\alpha\beta$ T cells express a unique TCR at the cell surface what may be essential for the generation of specific immune responses and for effective tolerance mechanism based on the elimination or inactivation of developing cells overtly reacting with self-antigens.

To explain allelic exclusion at the *TCR β* locus it has been postulated that the product of a successful *V β* to *DJ β* recombination prevents further rearrangements at the same locus via a feed back mechanism (Malissen et al., 1992), similar to what was proposed earlier for the Ig H chain locus in developing B cells (Alt et al., 1981; Coleclough, 1983). Together with the postulate that rearrangement starts initially in one of the two alleles, this hypothesis predicts correctly the extent of *TCR β* rearrangement found in mature T cells (Malissen et al., 1992). Also in support of a feedback mechanism are data showing that lymphocytes from *TCR β* transgenic (Tg) mice that express the transgene early in development are almost completely devoid of endogenous *V β* to *DJ β* rearrangements (Uematsu et al., 1988). More recent experiments analyzing the proportion of progenitor cells containing two completed *TCR β* gene rearrangements in progenitor cells from normal, pre-T α -deficient mice and SLP-76-deficient mice have indicated an essential role of signals mediated by the pre-TCR in *TCR β* allelic exclusion (Aifantis et al., 1997, 1999).

Analyses of *TCR δ* rearrangements in a panel of T cell hybridomas derived from splenic $\gamma\delta$ T cells showed a high percentage of cells carrying two productively rearranged *TCR δ* chains, indicating that assembly of *TCR δ* V-region genes, like that of *TCR α* V-region genes, is not regulated in the context of allelic exclusion (Sleckman et al., 1998). In support of this notion, Tg mice carrying a functionally rearranged *TCR δ* chain showed no evident inhibition of rear-

rangements at the endogenous TCR δ locus (Ferrero et al., 2001). Whether post-translational mechanisms limit the number of TCR δ chains that may be expressed at the cell surface in $\gamma\delta$ T cells is not known, although pairing preferences observed between different TCR γ and TCR δ chains likely contributes to the monoallelic expression of TCR δ chains at the cell surface (Bluestone et al., 1991; Pereira et al., 2000).

Whether assembly of TCR γ V-region genes is regulated in a context of allelic exclusion is a matter of debate. Analyses of human $\gamma\delta$ T cell lines have shown that a variable proportion (1-6%) of cells co-express two functional TCR γ chains at the cell surface (Davodeau et al., 1993) and these data has been taking to imply that assembly of TCR γ genes is not submitted to allelic exclusion mechanisms. In contrast, analyses of TCR γ rearrangements in mouse $\gamma\delta$ thymocyte clones showed that most $\gamma\delta$ thymocytes contained only one functionally rearranged allele for most of the TCR γ isotypes (Pereira and Boucontet, 2004), what suggest that assembly of TCR γ genes in the mouse may be regulated in the context of allelic exclusion.

To understand the mechanisms regulating assembly of TCR γ V-region genes one needs to take in consideration the genomic structure of the TCR γ locus, which differs greatly between humans and mice. Thus, whereas a human progenitor cell can produce a maximum of two functionally rearranged TCR γ chains, a mouse progenitor cell could, theoretically, produce six to eight functional TCR γ chains, depending on the mouse strain (Pereira and Boucontet, 2004; Pereira et al., 1996). This is due to the fact that the mouse TCR locus is organized in four clusters of V γ , J γ and C γ regions containing seven V γ gene segments, four J γ gene segments and four C γ regions (hereafter referred to as V1 to V7, J1 to J4 and C1 to C4 respectively, according to the nomenclature of Heilig and Tonegawa 1986). Each cluster contains a C γ region, linked to a single J γ element and one to four V γ gene segments, which rearrange preferentially to the J γ segment present in the same cluster. Thus, if expression of more than one TCR γ chain at the cell surface of $\gamma\delta$ T cells is to be prevented, TCR γ chains must be regulated not only to avoid expression of the two alleles, but also to avoid simultaneous expression of different TCR γ isotypes. Here we show that assembly of mouse TCR γ V-region genes is regulated in a manner compatible with allelic exclusion mediated by a $\gamma\delta$ TCR that can be expressed at the cell surface. Moreover, isotypic exclusion is regulated genotypically, at least in part, as evidenced by the fact that different V γ and J γ gene segments display distinct probabilities to participate in a recombination reaction and by the different frequencies at which a rearrangement involving particular V γ and J γ gene segments produce a functional chain. Further phenotypic isotypic exclusion results from the different capacity of TCR γ isotypes to pair with a more or less restricted diversity of TCR δ chains.

3.2 Materials and Methods

3.2.1 Mice

C57BL/6Jlco (B6), mice were obtained from Iffa Credo (L'Abresle, France). B6 mice Tg for a functionally rearranged V γ 1J γ 4C γ 4 (Tg- γ ; Gerber et al. 1999), B6 mice Tg for the same V γ chain together with a functionally rearranged V δ 6D δ 2J δ 1 chain (Tg- $\gamma\delta$; Malissen et al. 1997) and B6 TCR δ -deficient mice (Itohara et al., 1993) were maintained in our animal facilities. B6

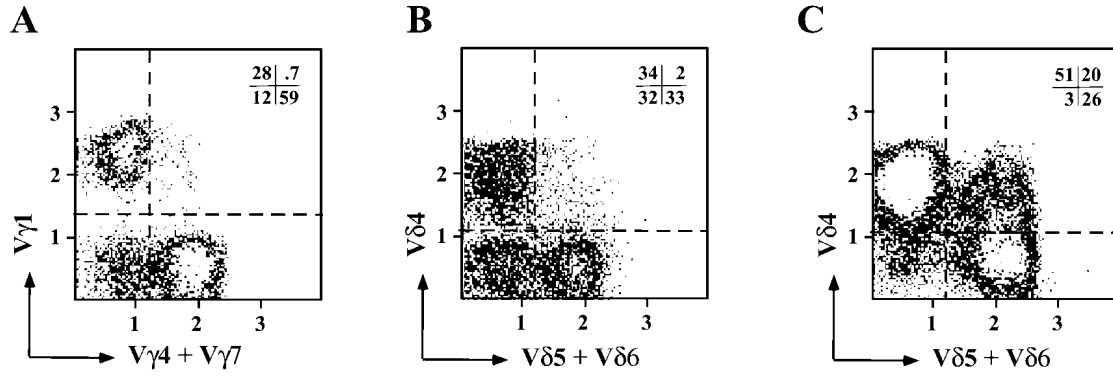


Fig. 3.1: Most $\gamma\delta$ thymocytes bear an unique $\gamma\delta$ TCR at the cell surface. CD4⁻CD8⁻ thymocytes were stained with anti- δ -PE, CD3-allophycocyanin, and either with 2.11-biotin and anti-V γ 4-FITC plus anti V γ 7- FITC (A) or anti-V γ 4-biotin and anti-V γ 5-FITC plus anti-V γ 6-FITC (B), followed by streptavidin-PerCP, and analyzed in a FACSCalibur. C, Putative dual-expressors in B (4% in this particular experiment) were sorted, expanded in vitro for 4 days, and reanalyzed as in B. Data are shown as dot-plots of the log10 of fluorescence intensity of the indicated mAbs in electronically gated CD3⁺ δ ⁺ cells. Numbers denote the fraction of cells in each quadrant.

mice deficient in the TCR β enhancer ($E\beta^{-/-}$; Leduc et al. 2000) were obtained from Dr. Pierre Ferrier (Centre d'Immunologie de Marseille-Luminy, Marseille, France) and also maintained in our animal facilities. All animals were used between six and eight weeks of age unless otherwise indicated.

3.2.2 Abs, Flow cytometry and cell sorting

Abs, FACS analyses and cell sorting were performed as described in (Pereira and Boucontet, 2004).

3.2.3 Cell cultures

Cloning of individually sorted $\gamma\delta$ thymocytes was performed as described in Pereira and Boucontet (2004). Sorted V1⁺ and V4⁺ cells were cloned in plates coated with anti-V1 mAb or anti- V4 mAb, respectively, thus providing a second control for the TCR γ chain expressed at the cell surface.

3.2.4 Single cell and single clone PCRs, cloning and sequencing

The single clone PCR to detect the rearrangement status of the TCR γ locus have been described in detail (Pereira and Boucontet, 2004). The protocol was identical except that only primers specific for the V1 and V4 V-regions were used and the germ line status was only analyzed for the V1 and the V4 regions. Reverse primers for the germline V4 region were: VG4GLext CTGAACAGCAGGTGGTTGCC and VG4GLint CCAAGCTAAGAAGGATGTGG. Single cell PCR was performed as described in Grigoriadou et al. (2003) using the VG1 CCGGCAAAAAGCAAAAAGTT and JG4 GCAAATATCTTGACCCATGA primers.

3.3 Results

3.3.1 Most $\gamma\delta$ T cells express one antigen receptor at the cell surface

To quantify the proportion of $\gamma\delta$ thymocytes expressing two different TCR γ or TCR δ chains at the cell surface we used available anti-V γ and anti-V δ mAbs. Cells staining simultaneously with an anti-V γ 1 and a mixture of anti-V γ 4 and anti-V γ 7 mAbs are rare, representing less than 1% of all $\gamma\delta$ thymocytes in B6 mice (Figure 3.1A). Because together, these Abs recognize >90% of the $\gamma\delta$ thymocytes in this strain, these data suggest that the vast majority of $\gamma\delta$ T cells express a unique TCR γ isotype at the cell surface.

In contrast, double staining of the same $\gamma\delta$ T cell population with Abs specific for the V δ 4 chain on one hand and a mixture of Abs recognizing V δ 5 and three of the four V δ 6 gene segment present in B6 mice on the other shows a small but reproducible population of double-stained cells that represent 2-4% of the total $\gamma\delta$ thymocytes (Figure 3.1B). Of those, $\sim$ 20% (i.e. 0.8% of the $\gamma\delta$ thymocytes) do really express two different TCR δ chains at the cell surface as evidenced by sorting and re-staining of the putative double-expresser cells (Figure 3.1C). Because about one third of the $\gamma\delta$ thymocytes bear TCR δ chains that are not recognized by these mAbs and because cells expressing two different TCR δ chains containing the same V-gene segment are not identified by these analysis, these data allows to estimate to less than 3% the maximum frequency of $\gamma\delta$ thymocytes expressing two different TCR δ chain at the cell surface.

3.3.2 Extent of allelic and isotypic inclusion of the functionally most relevant J1 and J4 gene segments

To quantify the extent of genotypic allelic and isotypic inclusion at the TCR γ locus and to gain a better understanding on the mechanisms controlling assembly of TCR γ V-region genes, we analyzed the rearrangement status of the V1 and V4 gene in progenies of a total of 234 $\gamma\delta$ thymocytes (141 V1⁺ and 93 V4⁺) using a previously described two-step PCR (Figure 3.2 and Pereira and Boucontet 2004). Cells bearing at the cell surface either V1 or V4 chains represent >80% of the $\gamma\delta$ thymocytes in B6 mice (Pereira et al., 2000). For simplicity we only analyzed V1 to J4 and V4 to J1 rearrangements, as well as the germline status of the V1 and V4 gene segments. This was based on our previous observation that in adult $\gamma\delta$ thymocytes >98% of the rearrangements involving the J4 gene segment also involved the V1 gene segment and >90% of the rearrangements involving the J1 gene segment also involved the V4 gene segment (Pereira and Boucontet, 2004). These analyses also showed that about half of the V1⁺ and V4⁺ thymocytes bear functional V2-J2 rearrangements and the mechanisms precluding detectable expression of V2 chains at the cell surface of these cells have been previously discussed (Pereira and Boucontet, 2004). The PCR amplification products of the rearrangements were sequenced to determine their functionality and a summary of the results is shown in Figure 3.2.

Approximately 1% of the cells contained two functionally rearranged J4 or J1 segments (1 of 141 V1⁺ clones (clone 66) and 1 of 93 V4⁺ clones (clone 60) contained the two J4 or the

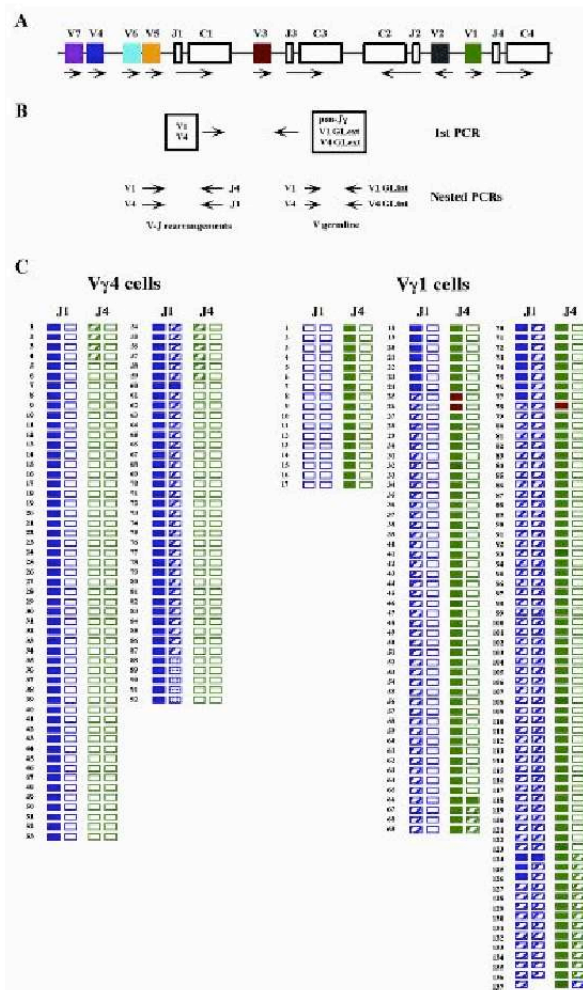


Fig. 3.2: Rearrangement status of the J1 and J4 regions in progenies of individual V4⁺ and V1⁺ thymocytes. **A**, Schematic representation of the genomic organization of the mouse TCR γ locus. The map is not drawn to scale. Arrows indicate transcriptional orientation. **B**, Primers used for the analyses of the rearrangement status of the TCR γ locus in progenies of individual $\gamma\delta$ T cells. Primers inside a box are meant to imply that they are used together in the same PCR, whereas isolated primers denote individual PCRs. **C**, Individually sorted V4⁺ and V1⁺ thymocytes were expanded and analyzed for the rearrangement status of the J1 and J4 regions as described in *Materials and Methods*. Each of the two boxes under every J gene segment represents one allele without any intentional order. The colors denote the V gene involved in the rearrangement. Lack of the box denotes deletion of a particular allele due to recombination of V and J segments present in different clusters. Filled boxes denote productive rearrangements. Hatched boxes denote unproductive rearrangements. Empty boxes denote lack of rearrangement at that allele. Light colored boxes in V4⁺ clones 88 and 92 denote that the rearrangement status of these alleles could not be determined unambiguously. These five clones were excluded from the statistical analyses.

two J1 segments productively rearranged, respectively), indicating that the vast majority of $\gamma\delta$ thymocytes contain only one functional rearrangement at the corresponding alleles. Contrasting with the paucity of allelically included cells, 18 clones (i.e., 8%; clones 18-24 and 70-77) were found to harbor two productive rearrangements involving two different J segments. Inter-

estingly, all these clones were found in the $V1^+$ population showing an apparent dominance of $V1$ chains over $V4$ chains for their expression at the cell surface as part of a $\gamma\delta$ TCR. As discussed previously (Pereira and Boucontet, 2004), such apparent dominance results, at least in part, from the fact that $V4$ chains pair with a more restricted pool of $TCR\delta$ chains than $V1$ chains.

3.3.3 Expression of a functionally rearranged $\gamma\delta$ TCR inhibits rearrangements at the endogenous $TCR\gamma$ locus

The low frequency of allelically included cells may suggest that the products of functionally rearranged $V1$ -J4 or $V4$ -J1 gene segments are involved in the termination of $TCR\gamma$ rearrangements in progenitor cells. They could do so by participating in a signaling complex either alone, together with a putative surrogate $TCR\delta$ chain or together with a $TCR\delta$ chain with which they can form a $\gamma\delta$ TCR. In either case, it will be expected that introduction of a functionally rearranged $TCR\gamma$ chain in progenitor cells will result in a certain degree of inhibition of $TCR\gamma$ rearrangements at the endogenous locus. Moreover, comparison of the extent of inhibition resulting from ectopic expression of either a functional $TCR\gamma$ chain alone or together with a functional $TCR\delta$ chain will be indicative of whether a surrogate receptor containing a $TCR\gamma$ chain or a complete $\gamma\delta$ TCR is involved in this process.

We have previously produced mice transgenic for a functionally rearranged $V1J4C4$ chain ($Tg-\gamma$) or for the same $TCR-\gamma$ chain together with a functionally rearranged $V\delta6D\delta J$ chain ($Tg-\gamma\delta$) (Gerber et al., 1999; Malissen et al., 1997). Because among all $TCR\gamma$ chains $V1$ chain are known to pair with the largest number of $TCR\delta$ chains (Bluestone et al., 1991; Pereira et al., 2000), these animals are most suitable to analyze this issue. Because a large majority of $V1^+$ thymocytes contain at least one $V4$ -J1 rearrangement (Figure 3.2) we analyzed and compare the occurrence of this type of rearrangement in $V1^+$ cells isolated from wild-type, $Tg-\gamma$ and $Tg-\gamma\delta$ mice. Introduction of a functionally rearranged $V1$ chain reduced the frequency of endogenous $V4$ -J1 rearrangements by about 3 fold (Figure 3.3). A further 3-fold reduction was observed when a functionally $TCR\delta$ chain was co-expressed with the same $TCR\gamma$ chain, resulting in a 84% inhibition of endogenous $V4$ -J1 rearrangements in $\gamma\delta$ T cells from $Tg-\gamma\delta$ mice (Figure 3.3). This extent of inhibition of endogenous $TCR\gamma$ rearrangements is of a similar magnitude than that observed in the endogenous $V\beta$ to $DJ\beta$ rearrangements as the result of early expression of a functionally rearranged $TCR\beta$ transgene (Ardouin et al., 1998). These data strongly suggest the participation of functionally rearranged $TCR\gamma$ and $TCR\delta$ chains in the regulation of $TCR\gamma$ V-region gene assembling. The inhibition observed by the expression of a Tg $TCR\gamma$ chain alone could reflect the increased probability of progenitor cells in $Tg-\gamma$ mice to express a $\gamma\delta$ TCR.

3.3.4 $\gamma\delta$ T cell progenitors do not attempt all possible $TCR\gamma$ rearrangements

The low frequency (12,8%) of $V1^+$ cells harboring rearrangements at the two J4 alleles (Figure 3.2) indicates that $\gamma\delta$ T cell progenitors do not attempt all possible $TCR\gamma$ rearrangements. This could result from the low probability at which the $V1$ and J4 gene segments rearrange

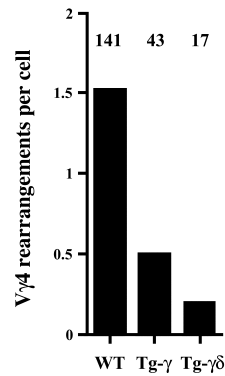


Fig. 3.3: Inhibition of endogenous V4-J1 rearrangements by $\gamma\delta$ TCR transgenes. Single sorted V1⁺ cells isolated from the indicated mouse strains were expanded in vitro and their DNA was analyzed for V4-J1 rearrangements and for the presence of unrearranged V4 genes as described in Materials and Methods. Data are shown as the number of V4-J1 rearrangements per cell. Numbers over the histograms denote the number of $\gamma\delta$ T cell clones analyzed. Clones from wild-type mice are the same as in Figure 3.2.

in progenitor cells in a given period of time (Pereira and Boucontet, 2004). Also, a receptor different from the $\gamma\delta$ TCR may signal back to stop further rearrangements at the TCR γ locus. An obvious candidate will be the pre-TCR, which may be expressed in $\gamma\delta$ T-cell precursors that produce a functional TCR β chain and that is known to be essential in the process of allelic exclusion at the TCR β locus (Aifantis et al., 1997). To investigate whether signals through the pre-TCR inhibit rearrangements at the TCR γ locus and, if so, to examine the relevance of such mechanism in $\alpha\beta$ and $\gamma\delta$ T-cell fate decisions, we quantify, by single cell PCR, the frequency of cells containing two V1-J4 rearrangements in sorted V1⁺ cells isolated from wild type mice and from mice deficient in the TCR β enhancer ($E\beta^{-/-}$). This mutation completely inhibits rearrangement at the TCR- β locus and, therefore, $\gamma\delta$ T cells in these mice develop in the absence of putative pre-TCR mediated signals. Moreover, comparison between these two strains of mice also eliminates competition between the TCR γ and the TCR β loci for components of the V(D)J recombinase as a plausible explanation for possible differences.

Nine of 71 (12,7%) V1⁺ cells from wild type mice contained two V1-J4 rearrangements (Table 3.1). This frequency is nearly identical to that found in the V1⁺ clones shown in Figure 3.2, indicating that the culture step did not bias the experimental sample. In contrast, 18 of 78 (23.1%) of the V1⁺ cells isolated from $E\beta^{-/-}$ mice harbored two V1-J4 rearrangements. The differences between the two V1⁺ populations are statistically significant ($p = 0.04$ according to the Fisher's exact test), demonstrating a quantitatively minor but evident role of the pre-TCR in stopping rearrangements at the TCR γ locus in T-cell progenitors.

3.3.5 *The extent of TCR γ rearrangements found in $\gamma\delta$ T cells is better explained by a model in which initiation of TCR γ rearrangements at both alleles is asymmetric and is terminated upon production of a functional $\gamma\delta$ TCR*

Whereas the feedback mechanism certainly contributes to the inhibition of rearrangements at the second allele, it alone cannot explain allelic exclusion unless the rearrangement step is

Tab. 3.1: Limited role of the pre-TCR in the termination of $V\gamma$ to $J\gamma$ rearrangements. ^a Differences between WT and $E\beta^{-/-}$ cells are statistically significant according to Fisher's exact test ($p = 0.04$).

Strain	Cell analyzed	Cells with two V1-J4 rearrangements ^a
WT	V1 ⁺ clones	17 of 141 (12.06%)
WT	V1 ⁺ single cells	9 of 71 (12.68%)
$E\beta^{-/-}$	V1 ⁺ single cells	18 of 78 (23.08%)

generally very inefficient (Coleclough et al., 1981). This is due to the fact that the product of a rearrangement must be tested for its ability to encode a functional chain. Attempting to calculate the rates of different $TCR\gamma$ rearrangements and to investigate whether other mechanisms regulate $TCR\gamma$ V-gene assembly, we constructed probabilistic models and compared the extent of V1 to J4 and V4 to J1 rearrangements found in V1⁺ and V4⁺ thymocytes with those predicted by the models (Sepúlveda et al., 2005). Models that only take into consideration the time given to a progenitor cell to rearrange and a feedback mechanism to stop further rearrangements, although compatible with the experimental data, only gives marginal statistical significance, strongly suggesting that additional mechanisms regulate TCR V-gene assembly. Interestingly, a very accurate match between the experimental data and the expected values was observed when a differential accessibility of the two $TCR\gamma$ alleles was imposed in the model, assuming that after a period during which only one allele is accessible both alleles become accessible (Sepúlveda et al., 2005). Thus, on statistical basis, the experimental data is compatible with the hypothesis that assembly of $TCR\gamma$ V-region genes is regulated by the products of functional $TCR\gamma$ rearrangements and the experimental data is better explained if the two alleles become accessible to the V(D)J recombinase asynchronously. Importantly, the models make a number of predictions that can be experimentally analyzed and used to test their robustness. Thus, the best fitting model predicts that a progenitor cell will rearrange the V4 gene 12 times more often than the V1 gene, what compares well with the value of 11 times found experimentally (Pereira and Boucontet, 2004). Moreover, the model also predicts a ratio between V4⁺ and V1⁺ cells of 1.40, which also compares reasonable well with the ratio of 1.67 obtained by staining thymocytes with $V\gamma$ -specific mAbs ex-vivo. Finally, the model predicts that $\sim 0.5\%$ of the V1⁺ cells will contain two functional V1-J4 rearrangements, compatible with the 0.7% found in the analyzed here (Figure 3.2). Interestingly, the model also predicts that $\sim 44\%$ of the progenitor cells will be rescued by functional V1 or V4 chains. Given that a few more progenitor cells could be rescued by functional V2 or V7 chains, the expected value of $\gamma\delta$ T cell precursors that will produce a functional $TCR\gamma$ chain may not differ significantly from the maximum of 56% of the $\alpha\beta$ T cell precursors expected to produce a functional $TCR\beta$ chain (Malissen et al., 1992).

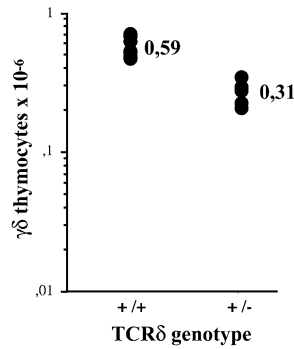


Fig. 3.4: The number of $\gamma\delta$ thymocytes correlates with the probability of producing a functional $\gamma\delta$ TCR in progenitor cells. Thymocytes from back-crossed (B6 $\times$ TCR $\delta^{-/-}$) $\times$ B6 mice were stained with anti- δ -PE and anti-CD3-allophycocyanin and analyzed in a FACSCalibur. Data are shown as the number of $\gamma\delta$ thymocytes present in individual mice analyzed at 4 week of age and separated on the basis of having one or two functional alleles of the C δ region. Numbers denote the mean values of $\gamma\delta$ thymocytes on each group.

3.3.6 The number of $\gamma\delta$ thymocytes correlates with the probability of producing a selectable $\gamma\delta$ TCR

A proper comparison of the models with the data requires that newly differentiated $\gamma\delta$ T cells do not expand in the thymus or, if they do, that this expansion is independent of the TCR γ isotype they express. Although it is generally believed that formation of a selectable $\gamma\delta$ TCR in $\gamma\delta$ -lineage progenitor cells induces maturation of the cells in a process that involve little or no proliferation (Penit et al., 1995), it was important to ascertain that this is indeed the case. If maturation of $\gamma\delta$ thymocytes occurs without expansion, it will be expected that the number of $\gamma\delta$ T cells present in the thymus is a direct function of the probability of producing a selectable $\gamma\delta$ TCR. Therefore, we quantified and compared $\gamma\delta$ thymocyte numbers in mice carrying only one functional allele of the C δ chain and in wild type littermates. As shown in Figure 3.4, TCR $\delta^{+/-}$ mice contained about half the number of $\gamma\delta$ thymocytes than their wild-type littermates, demonstrating that the number of $\gamma\delta$ thymocytes is a linear function of the probability of a progenitor cell to produce a functional $\gamma\delta$ TCR, and suggesting that no homeostatic mechanism regulates $\gamma\delta$ T cell numbers in the thymus.

3.4 Discussion

The data presented in this manuscript indicate that the vast majority of mature $\gamma\delta$ thymocytes exhibit properties of allelic exclusion at the TCR γ locus at the level of V-region gene assembly, at least in what concerns the most commonly used J1 and J4 regions. Such allelic exclusion is achieved, in part, through a feedback mechanism mediated by the products of the functionally rearranged TCR γ and TCR δ chains and requires a proper interaction between both chains to form a functional $\gamma\delta$ TCR. The existence of a signaling mechanism is strongly suggested by the inhibition of endogenous TCR γ rearrangements observed in Tg- $\gamma\delta$ mice. However, the possibility that a rapid development of T cells may not provide adequate time for RAG mediated recombination in the Tg model cannot be formally excluded at the moment. Independent

evidence consistent with the existence of a feedback mechanism came from the comparison of the extent of TCR γ rearrangements found in $\gamma\delta$ T cells with those predicted by statistical models (Sepúlveda et al., 2005).

For signaling through a $\gamma\delta$ TCR to be effective in ensuring allelic exclusion at the TCR γ locus it is required that, in general, TCR δ rearrangements precede TCR γ rearrangements in progenitor cells. Although complete rearrangements at both loci are found at the same developmental stage (Capone et al., 1998; Livak et al., 1999), indirect evidence suggest that this may indeed be the case. First, cells with rearrangements at the TCR δ locus with and the TCR γ locus in germline configuration were easily found in fetal thymocyte hybridomas (Chien et al., 1987). Moreover, analyses of thymocytes from SCID mice (Bosma and Carroll, 1991) showed that rearrangements at the TCR δ locus exceed by far those occurring at the TCR γ or TCR γ loci (Carroll and Bosma, 1991). SCID mice produce double-stranded DNA breaks mediated by the RAG proteins at the initiation of the recombination process at normal levels (Zhu and Roth, 1995) but their defect in the DNA-dependent protein kinase catalytic subunit protein, which is involved in the resolution of coding ends, precludes the formation of V(D)J coding joints at appreciable frequency. Therefore, the SCID mice data provide evidence for a different accessibility of the TCR δ and TCR γ loci to participate in a recombination reaction in early progenitor cells. It should be pointed out that we are not implying a strict temporal order in the rearrangements at both loci, but just an increased probability of a TCR δ chain to be expressed before a TCR γ chain in progenitor cells. Absence of a strict temporal order is suggested by the isolation of fetal hybridomas containing rearrangements at the TCR γ locus and the TCR δ locus in germline configuration (Chien et al., 1987) and by our data showing that ectopic expression of a functionally rearranged TCR γ chain only partially inhibits endogenous rearrangements at the TCR γ locus.

Although the feedback mechanism certainly contributes to the inhibition of rearrangements at the second allele, it alone cannot explain allelic exclusion unless the rearrangement step is generally very inefficient (Coleclough et al., 1981). This is due to the fact that the product of a rearrangement must be tested for its ability to encode a functional chain. The actual frequencies of B and T lymphocytes containing V(D)J rearrangements at the two IgH or TCR β alleles was shown to be very close to those expected if rearrangements at the two alleles were attempted sequentially and every progenitor cell had enough time to rearrange both alleles (Coleclough, 1983; Malissen et al., 1992). These results prompted the generalized notion that accessibility to the recombination machinery differs between two identical alleles. However, the same results would be predicted if the efficiency of recombination is generally low but the time given to a progenitor cell to recombine is long and this because these experiments analyze a unique type of recombination. The genomic organization of the TCR γ locus in the mouse, which permit multiple V to J rearrangements occurring in the same chromosome, allows to differentiate between these two possibilities. This is due to the fact that, although the initiation of recombination at different V and J segments is likely to be independent from each other (see below), the fact that the products of these recombination events terminates recombination in the whole locus results in their functional dependence. In other words, the extent of V1-J4 rearrangements will depend not only on the probability of producing a functional V1 chain, but

also on that of producing a functional V4 chain in the same progenitor cell and vice versa. In this scenario, a feedback mechanism together with a limited time given to the progenitor cell to rearrange their $TCR\gamma$ genes fail to fit the experimental data because they always overestimate the number of cells containing the two alleles rearranged. Fitting the extent of expected rearrangements to the experimental values at one J region invariably results in predicted values for rearrangements at the other J region very different from those observed experimentally, indicating that other mechanisms participate in the process of allelic exclusion at the $TCR\gamma$ locus. Imposing a different accessibility of the two alleles to the action of the V(D)J recombinase result in an almost perfect match between the observed and expected values suggesting that this could be one of the mechanisms by which allelic exclusion is achieved. In this line, recent experiments have shown the existence of differential epigenetic modifications at two $Ig\kappa$ alleles, established early during embryonic development, that lead to a preferential accessibility of one of the alleles to the V(D)J recombinase (Mostoslavsky et al., 1998). These epigenetic modifications are clonally transmitted and also correlate with the asynchronous replication of different TCR and Ig alleles in mature lymphocytes, with $Ig\kappa$ rearrangements in mature B cells found preferentially in the early replicating allele (Mostoslavsky et al., 1998, 2001). Interestingly, DNA demethylation, which is one of the epigenetic modifications increasing chromatin accessibility, was shown to occur with different kinetics in both $Ig\kappa$ alleles in a manner compatible with a different probability of each allele to be demethylated, more than with a specific mechanism imposing demethylation at a single allele (Goldmit et al., 2002). Independent evidence for the stochastic nature of the preferential accessibility of one of the alleles has also been obtained in studies analyzing the chromatin structure surrounding the transcriptional enhancers associated with the $Ig\kappa$ locus (Schlissel, 2002).

Even if the two chromosomes become accessible to the action of the V(D)J recombinase with different probabilities, that will only have a limited effect in the isotypic exclusion of mouse $TCR\gamma$ genes due to the genomic structure of the mouse $TCR\gamma$ locus (Figure 3.2). Possible mechanisms of isotypic exclusion of $Ig\Delta$ light chains have been extensively discussed and two general models (a sequential model and a probabilistic model) have been proposed (Louzoun et al., 2002). The most relevant difference between the two models is that whereas the sequential model proposes that λ rearrangements are regulated by κ rearrangements, the probabilistic model presumes complete independence of both loci. By analogy with the $Ig\Delta$ chain loci, the same two general models could be proposed to explain isotypic exclusion at the $TCR\gamma$ locus in the mouse. However, the fact that a fraction of $V1^+$ cells contain the two J1 alleles in germline configuration (Figure 3.2), argues against a strict temporal order in the rearrangements at different $J\gamma$ regions and strongly suggests that rearrangements at different $J\gamma$ regions initiate independently. In these conditions, concomitant rearrangement of two or more isotypes and, therefore, expression of two functionally rearranged $TCR\gamma$ chains becomes possible and their frequency will depend on the rearrangement rates at different J regions. Obviously, the chances that a progenitor cell will produce two functional chains decrease as the rates of rearrangement of the different $TCR\gamma$ chains diminish but with a concomitant decrease in the fraction of progenitor cells that will succeed in producing a functional $TCR\gamma$ chain. Interestingly, increasing the rate of recombination of one isotype relative to the other will have as a

consequence a relatively high rate of success keeping the probability of producing isotypically included cells at $\sim 1\%$ (Sepúlveda et al., 2005). In these conditions, however, the majority of the rescued cells will express the most frequently rearranged isotype (V4 in mouse $\gamma\delta$ thymocytes). An elevated production of $V1^+$ cells can be obtained by reducing the probability of obtaining functional V4 chains without altering the rate of rearrangement of the V4 gene segment. This is precisely the consequence of the presence of an in-frame stop codon at the end of the V4 gene, which decreases by about 2-fold the chance that a V4-J1 rearrangement will result in a functional V4 chain (Capone et al., 1998; Kang et al., 1995). Thus, the 11-fold difference in the rates at which V1 and V4 rearrange (Pereira and Boucontet, 2004), together with the presence of the stop codon at the end of the V4 gene segment are sufficient to ensure the development of relatively high numbers of $V4^+$ and $V1^+$ $\gamma\delta$ thymocytes from a common precursor, keeping the probability that a cell may contain two functionally rearranged isotypes below 1%. The selective advantage for the co-existence of both cell populations may relate to their different functions, as it has been suggested in several infectious models (reviewed in O'Brien et al. 2001).

Although assembly of $V\gamma$ genes displays properties of allelic exclusion, it is expected that a small fraction of $\gamma\delta$ cells will contain two functionally rearranged alleles. This is predicted by stochastic models of $TCR\gamma$ gene rearrangement and our data indicates that $\gamma\delta$ cells containing two functional V1 or V4 chains are in the order of 1%. More importantly, the models also predict a very small fraction of isotypically included cells, which is in apparent contradiction with the experimental data. Thus, $\sim 12\%$ of the $V1^+$ thymocytes also contain a functional V4 chain (Figure 3.2) and about half of the $\gamma\delta$ thymocytes contain a functional V2 chain in addition to the $V\gamma$ chain that is expressed at their cell surface (Pereira and Boucontet, 2004). These cells are isotypically included at the genetic level but excluded at the level of $TCR\gamma$ expression at the cell surface as demonstrated by their staining with $V\gamma$ -specific Abs (Figure 3.1 and Pereira and Boucontet 2004). At least part of this phenotypic exclusion is due to the restriction in $V\delta$ chain pairing displayed by V2 and possibly V4 chains (Pereira and Boucontet, 2004), reminiscent of the allelically included B cells that were shown to be allelically excluded at the level of pre-B cell receptor surface expression (ten Boekel et al., 1998). Altogether, these and previous data analyzing the extent of $TCR\gamma$ rearrangements in a smaller number of $\gamma\delta$ thymocytes (Pereira and Boucontet, 2004), are consistent with the hypothesis that cessation of rearrangements at the $TCR\gamma$ locus (and possibly at the $TCR\delta$ and $TCR\beta$ loci) and final development along the $\gamma\delta$ T cell pathway are mediated by signals through a $\gamma\delta$ TCR that can be expressed at the cell surface.

Previous analysis of human $\gamma\delta$ T cell clones with $V\gamma$ -specific Abs showed that a significant fraction of cells (1-7%) contained two functional $TCR\gamma$ chains, suggesting that assembly of human $V\gamma$ genes is not regulated in the context of allelic exclusion (Davodeau et al., 1993). This is in contrast with our data showing that vast majority of mouse $\gamma\delta$ thymocytes express a unique $TCR\gamma$ chain at the cell surface. However, as also shown here, up to 3% of the mouse $\gamma\delta$ thymocytes express two different $TCR\delta$ chains at the cell surface. If signals to terminate rearrangements at the $TCR\gamma$ locus, and possibly $TCR\delta$ locus as well, are mediated by a selectable $\gamma\delta$ TCR and not by a surrogate receptor, only the chain that rearranges later can be allelically excluded efficiently. Furthermore, a certain degree of isotypic exclusion of $TCR\gamma$

chains in the mouse is due to their restricted pairing with TCR δ chains. Thus, differences in the time at which TCR γ and TCR δ rearrangements take place in human progenitors and/or a less evident restriction for TCR δ chains displayed by human TCR γ chains could explain, at least in part, the different phenotype observed in mouse and human $\gamma\delta$ T cells with regard to allelic exclusion of TCR γ chains.

Because assembly of TCR δ chains does not exhibit properties of allelic exclusion it is expected that $\sim 20\%$ of the $\gamma\delta$ T cells containing complete V(D)J rearrangements at both alleles will contain two functional TCR δ chains (Sleckman et al., 1998). The actual number of $\gamma\delta$ T cells bearing two functional TCR δ chains is higher than that ($\sim 28\%$ in the combined data from Sleckman et al. 1998 and Pereira and Boucontet 2004). This is also a possible consequence of the restricted pairing of TCR γ and TCR δ chains and of the fact that $\gamma\delta$ T cell progenitors do not attempt all possible rearrangements at the TCR γ locus. In these circumstances, progenitor cells containing two functional TCR δ chains have a greater chance to be rescued by a TCR γ chain to enter the $\gamma\delta$ T cell pool than progenitor cells containing a single functional TCR δ chain. Because $\sim 37\%$ of the $\gamma\delta$ T cells contain only one complete TCR δ rearrangement (13, 22) it is expected that $\sim 18\%$ ($0.28 \times 0.63 \times 100$) of the $\gamma\delta$ T cells contain two functional TCR δ chains. Of those, a relatively large fraction will harbor a TCR δ chain unable to pair with the functional TCR γ chain and, therefore, will be allelically excluded at the cell surface expression. Consistent with this interpretation is the fact that, of 11 V4⁺ clones containing two functional TCR δ chains, 7 (63%) contained a V δ 6 chain (L. Boucontet and P. Pereira, unpublished observations), which is known to be rarely expressed with V4 (Pereira et al., 2000). Thus, in the absence of any other constraint it can be estimated that $\sim 6\%$ of the $\gamma\delta$ thymocytes may express two different TCR δ chains at the cell surface. This value is somewhat higher than the maximum of 3% shown by staining with V -specific mAbs (Figure 3.1) suggesting that additional mechanisms further restrict the expression of two functional TCR δ chains.

Finally, our results suggest a mechanism to explain, at least in part, how the absence of a functional pre-T cell receptor results in an increased number of $\gamma\delta$ thymocytes (Aifantis et al., 2002; Fehling et al., 1995). Thus, signaling through the pre-TCR may inhibit TCR γ rearrangements in cells that, otherwise, could still become $\gamma\delta$ T cells. This mechanism, however, appears to operate in a small number of progenitors, likely because V β -DJ β rearrangements mostly take place after the majority of TCR γ and TCR δ rearrangements have been completed (Capone et al., 1998; Livak et al., 1999).

Bibliography

- Aifantis, I., Borowski, C., Gounari, F., Lacorazza, H. D., Nikolich-Zugich, J., and von Boehmer, H. (2002). A critical role for the cytoplasmic tail of pT α in T lymphocyte development. *Nat. Immunol.*, 5:483–488.
- Aifantis, I., Buer, J., von Boehmer, H., and Azugui, O. (1997). Essential role of the pre-T cell receptor in allelic exclusion of the T cell receptor β locus. *Immunity*, 7:601–607.
- Aifantis, I., Pivniouk, V. I., Gartner, F., Freinberg, J., Swat, W., Alt, F. W., von Boehmer, H., and

- Geha, R. S. (1999). Allelic exclusion of the T cell receptor β locus requires the SH2 domain-containing leucocyte protein (SLP)-76 adaptor protein. *J. Exp. Med.*, 190:1093–1102.
- Alt, F. W., Rosenberg, N., Lewis, S., Thomas, E., and Baltimore, D. (1981). Organization and reorganization of immunoglobulin genes in A-MuLV-transformed cells: Rearrangement of heavy but not light chain genes. *Cell*, 27:381–390.
- Ardouin, L., Ismaili, J., Malissen, B., and Malissen, M. (1998). The CD3- $\alpha\beta\epsilon$ and CD3- $\zeta\eta$ modules are each essential for allelic exclusion at the T cell receptor β locus but are both dispensable for the initiation of V to (D)J recombination at the T cell receptor- β , γ - and - δ loci. *J. Exp. Med.*, 187:105–116.
- Bluestone, J. A., Cron, R. Q., Barrett, T. A., Houlden, B., Sperling, A. I., Dent, A., Hedrick, S., Rellahan, B., and Matis, L. A. (1991). Repertoire development and ligand specificity of murine TCR $\gamma\delta$ cells. *Immunol. Rev.*, 120:5–33.
- Bosma, M. J. and Carroll, A. M. (1991). The SCID mouse mutant: definition, characterization, and potential uses. *Annu. Rev. Immunol.*, 9:323–350.
- Capone, M., Hockett, Jr, R. D., and Zlotnik, A. (1998). Kinetics of T cell receptor and rearrangements during adult thymic development: T cell receptor rearrangements are present in CD44+CD25+ pro-T thymocytes. *Proc. Natl. Acad. Sci. USA*, 95:12522–12527.
- Carroll, A. M. and Bosma, M. J. (1991). T-lymphocyte development in scid mice is arrested shortly after the initiation of T-cell receptor delta gene recombination. *Genes Dev.*, 5:1357–1366.
- Chien, Y.-h., Iwashima, M., Wettstein, D. A., Kaplan, K. B., Elliot, J. F., Born, W., and Davis, M. M. (1987). T-cell receptor δ gene rearrangements in early thymocytes. *Nature*, 330:722–727.
- Coleclough, C. (1983). Chance, necessity and antibody gene dynamics. *Nature*, 303:23–26.
- Coleclough, C., Perry, R. P., Karjalainen, K., and Weigert, M. (1981). Aberrant rearrangements contribute significantly to the allelic exclusion of immunoglobulin gene expression. *Nature*, 290:372–378.
- Davodeau, F., Peyrat, M.-A., Houde, I., Hallet, M.-M., De Libero, G., Vié, H., and Bonneville, M. (1993). Surface expression of two distinct functional antigen receptors on human $\gamma\delta$ T cells. *Science*, 260:1800–1802.
- Dudley, E. C., Petrie, H. T., Shah, L. M., Owen, M. J., and Hayday, A. C. (1994). T cell receptor β chain gene rearrangement and selection during thymocyte development in adult mice. *Immunity*, 1:83–93.
- Fehling, H. J., Krotkova, A., Saint-Ruf, C., and von Boehmer, H. (1995). Crucial role of the pre-T cell receptor $\alpha\beta$ gene in the development of $\alpha\beta$ but not $\gamma\delta$ T cells. *Nature*, 375:795–798.

- Ferrero, I., Wilson, A., Beermann, F., Held, W., and MacDonald, H. R. (2001). T cell receptor specificity is critical for the development of epidermal $\gamma\delta$ T cells. *J. Exp. Med.*, 194:1473–1483.
- Gascoigne, N. R. J. and Alam, S. M. (1999). Allelic exclusion of the T cell receptor α -chain: developmental regulation of a post-translational event. *Sem. Immunol.*, 11:337–347.
- Gerber, D. J., Azuara, V., Levraud, J.-P., Huang, S. Y., Lembezat, M.-P., and Pereira, P. (1999). Interleukin-4 producing $\gamma\delta$ T cells that express a very restricted TCR repertoire are preferentially localized in the liver and in the spleen. *J. Immunol.*, 163:3076–3082.
- Godfrey, D. I. and Zlotnik, A. (1993). Control points in early T cell development. *Immunol. Today*, 14:547–553.
- Goldmit, M., Schlissel, M., Cedar, H., and Bergman, Y. (2002). Differential accessibility at the κ chain locus plays a role in allelic exclusion. *EMBO J.*, 21:5255–5261.
- Grigoriadou, K., Boucontet, L., and Pereira, P. (2003). Most IL-4-producing $\gamma\delta$ thymocytes of adult mice originate from fetal precursors. *J. Immunol.*, 171:2413–2420.
- Heilig, J. S. and Tonegawa, S. (1986). Diversity of murine γ genes and expression in fetal and adult T lymphocytes. *Nature*, 322:836–840.
- Itohara, S., Mombaerts, P., Lafaille, J., Lacomini, J., Nelson, A., Clarke, A. R., Hooper, M. L., Farr, A., and Tonegawa, S. (1993). T cell receptor δ gene mutant mice: independent generation of $\alpha\beta$ T cells and programmed rearrangements of $\gamma\delta$ TCR genes. *Cell*, 72:337–348.
- Kang, J., Baker, J., and Raulet, D. H. (1995). Evidence that productive rearrangements of TCR γ genes influence the commitment of progenitor cells to differentiate into $\alpha\beta$ or $\gamma\delta$ T cells. *Eur. J. Immunol.*, 25:2706–2709.
- Kang, J., Coles, M., Cado, D., and Raulet, D. H. (1998). The developmental fate of T cells is critically influenced by TCR $\gamma\delta$ expression. *Immunity*, 8:427–438.
- Leduc, I., Hempel, W. M., Mathieu, N., Verthuy, C., Bouvier, G., Watrin, F., and Ferrier, P. (2000). T cell development in TCR β enhancer-deleted mice: Implications for $\alpha\beta$ T cell lineage commitment and differentiation. *J. Immunol.*, 165:1364–1373.
- Livak, F., Tourigny, M., Schatz, D. G., and Petrie, H. T. (1999). Characterization of TCR gene rearrangements during adult murine T cell development. *J. Immunol.*, 162:2575–2580.
- Louzoun, Y., Friedman, T., Park, E. L., Litwin, S., and Weigert, M. (2002). Analysis of B cell receptor production and rearrangement. Part I. Light chain rearrangement. *Sem. Immunol.*, 14:169–190.
- Malissen, M., Pereira, P., Gerber, D. J., Malissen, B., and DiSanto, J. P. (1997). The common cytokine receptor chain controls survival of $\gamma\delta$ T cells. *J. Exp. Med.*, 186:1277–1285.

- Malissen, M., Trucy, J., Jouvin-Marche, E., Cazenave, P.-A., Scollay, R., and Malissen, B. (1992). Regulation of TCR α and β gene allelic exclusion during T-cell development. *Immunol. Today*, 13:315–322.
- Mostoslavsky, R., Singh, N., Kirillov, A., Pelanda, R., Cedar, H., Chess, A., and Bergman, Y. (1998). κ chain monoallelic demethylation and the establishment of allelic exclusion. *Genes Dev.*, 12:1801–1811.
- Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B. E., Chess, A., Cedar, H., and Bergman, Y. (2001). Asynchronous replication and allelic exclusion in the immune system. *Nature*, 414:221–225.
- O'Brien, R. L., Lahn, M., Born, W. K., and Huber, S. A. (2001). T cell receptor and functions cosegregate in $\gamma\delta$ T cell subsets. *Chem. Immunol.*, 79:1–28.
- Passoni, L., Hoffman, E. S., Kim, S., Crompton, T., Pao, W., Dong, M.-Q., Owen, M. J., and Hayday, A. C. (1997). Intrathymic selection events in $\gamma\delta$ cell development. *Immunity*, 7:83–95.
- Penit, C., Lucas, B., and Vasseur, F. (1995). Cell expansion and growth arrest phases during the transition from precursor (CD4-CD8-) to immature (CD4+CD8+) thymocytes in normal and genetically modified mice. *J. Immunol.*, 154:5103–5113.
- Pereira, P. and Boucontet, L. (2004). Rates of recombination and chain pair biases greatly influence the primary $\gamma\delta$ TCR repertoire in the thymus of adult mice. *J. Immunol.*, 173:3261–3270.
- Pereira, P., Gerber, D., Regnault, A., Huang, S. Y., Hermitte, V., Coutinho, A., and Tonegawa, S. (1996). Rearrangement and expression of V γ 1, V γ 2 and V γ 3 TCR γ genes in C57BL/6 mice. *Int. Immunol.*, 8:83–90.
- Pereira, P., Hermitte, V., Lembezat, M.-P., Boucontet, L., Azuara, V., and Grigoriadou, K. (2000). Developmentally regulated and lineage-specific rearrangement of T cell receptor V α/δ gene segments. *Eur. J. Immunol.*, 30:1988–1997.
- Petrie, H. T., Livak, F., Burtrum, D., and Mazel, S. (1995). T cell receptor gene recombination patterns and mechanisms: cell death, rescue, and T cell production. *J. Exp. Med.*, 182:121–127.
- Schlissel, M. (2002). Allelic exclusion of immunoglobulin gene rearrangement and expression: why and how? *Sem. Immunol.*, 14:207–212.
- Seboun, E., Joshi, N., and Hauser, S. L. (1992). Haplotypic origin of β -chain genes expressed by human T-cell clones. *Immunogenetics*, 36:363–368.
- Sepúlveda, N., Boucontet, L., Pereira, P., and Carneiro, J. (2005). Stochastic modeling of t cell receptor γ gene rearrangement. *J. Theor. Biol.*, 234:153–165.

- Sleckman, B. P., Khor, B., Monroe, R., and Alt, F. W. (1998). Assembly of productive T cell receptor δ variable region genes exhibits allelic inclusion. *J. Exp. Med.*, 188:1465–1471.
- ten Boekel, E., Melchers, F., and Rolink, A. G. (1998). Precursor B cells showing H chain allelic inclusion display allelic exclusion at the level of pre-B cell receptor surface expression. *Immunity*, 8:199–207.
- Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature*, 302:575–581.
- Uematsu, Y., Ryser, S., Dembic, Z., Borgulya, P., Krimpenfort, P., Berns, A., von Boehmer, H., and Steinmetz, M. (1988). In transgenic mice the introduced functional T cell receptor β gene prevents expression of endogenous β genes. *Cell*, 52:831–841.
- von Boehmer, H., Aifantis, I., Feinberg, J., Lechner, O., Saint-Ruf, C., Walter, U., Buer, J., and Azogui, O. (1999). Pleiotropic changes controlled by the pre-T-cell receptor. *Curr. Opin. Immunol.*, 11:135–142.
- Wilson, A., de Villartay, J.-P., and MacDonald, H. R. (1996). T cell receptor delta gene rearrangement and T early (TEA) expression in immature lineage thymocytes: implications for lineage commitment. *Immunity*, 4:37–45.
- Wilson, A., Held, W., and MacDonald, H. R. (1994). Two waves of recombinase gene expression in developing thymocytes. *J. Exp. Med.*, 179:1355–1360.
- Zhu, C. and Roth, D. B. (1995). Characterization of coding ends in thymocytes of scid mice: implications for the mechanism of V(D)J recombination. *Immunity*, 2:101–112.

Part III

PERIPHERAL DYNAMICS OF REGULATORY AND EFFECTOR T CELL REPERTOIRES

4. REPERTOIRE DYNAMICS OF PERIPHERAL REGULATORY AND EFFECTOR T CELLS WITHOUT INTERCLONAL COMPETITION

Carneiro, J., Léon, K., Caramalho, I., van den Dool, C., Gardner, R., Oliveira, V., Bergman, M.-L., Sepúlveda, N., Paixão, T., Faro, J., and Demengeot, J. (2007). When three is not a crowd: A crossregulation model of the dynamics and repertoire selection of regulatory CD4 T cells. *Immunol. Rev.* 216:48-68.

Abstract

Regulatory CD4 T (T_R) cells, enriched in the CD25 pool of healthy individuals, mediate natural tolerance and prevent autoimmune diseases. Despite their fundamental and potential clinical significance, T_R cells have not yet been incorporated in a coherent theory of the immune system. This article reviews experimental evidence and theoretical arguments supporting a model of T_R cell dynamics, uncovering some of its most relevant biological implications. According to this model, the persistence and expansion of T_R cell populations depend strictly on specific interactions they make with APCs and conventional T_E cells. This three-partner crossregulation imposes that T_R cells feed on the specific autoimmune activities they suppress, with implications ranging from their interactions with other cells, to their repertoire selection in the periphery and in the thymus, and to the relationship between these cells and the innate immune system. These implications stem from the basic prediction that the peripheral dynamics sorts the CD4 repertoire into two subsets: a less diverse set of small clones of autoreactive T_E and T_R cells that regulate each other's growth, and a more diverse set of barely autoreactive T_E clones, whose expansion is limited only by APC availability. It is argued that such partitioning of the repertoire sets the ground for self-nonself discrimination.

4.1 Introduction

Regulatory CD4 T cells, that express FoxP3 (Fontenot et al., 2003; Hori et al., 2003; Khattri et al., 2003) and are enriched in the CD25 pool of healthy individuals (Sakaguchi et al., 1995), have been gaining increasing relevance in immunology (Sakaguchi, 2004). Many lines of evidence indicate that these cells play a key role in the development of natural tolerance and in the prevention of autoimmune pathologies, by controlling the activation and proliferation of other autoreactive lymphocytes (Sakaguchi et al., 1995, 2001). The functional significance of these cells has broadened as they were shown to modulate the immune response against pathogens preventing the associated immunopathology (Belkaid et al., 2002; Cahill et al., 1997; Hori et al., 2002; Suvas et al., 2004), and to prevent rejection of transplants (Graça et al., 2002; Kingsley et al., 2002; Taylor et al., 2001). The hypothesis that suppressor or regulatory T cells orchestrate the immune system is not new (see for example the editorial in Sercarz et al. 1974) and

yet a theory incorporating these cells and their function is still lacking. The paucity of theoretically framed hypothesis adds up to other difficulties in dealing with regulatory T cells, namely the inability to isolate regulatory T cells, and the difficulty to design and interpret quantitative experiments to assess their function.

This article reviews a mathematical modeling approach to regulatory T cells, based on the hypothesis, put forward by Leon et al. (2000), according to which the persistence and expansion of regulatory T cell populations depend on the interactions these cells make both with APCs, and the T cells they suppress. This three-partner crossregulation model, which "forces" regulatory T cells to "feed" on the specific (auto)immune activities they suppress, has far reaching implications for the immunobiology of these cells, ranging from the interactions they make with other cells, to their population dynamics and repertoire selection in the periphery, to the constraints of thymic selection, or the relationship between these cells and the innate immune system.

4.2 Natural tolerance, clonal selection, and regulatory T cells

Natural tolerance refers to the status of absence of harmful immune responses against body components observed in healthy individuals. The developmental processes leading to natural tolerance are so robust, so "natural" that for many years they were cast aside from immunology (Grabar, 1950). The biological significance of natural tolerance becomes most patent upon its failure during autoimmune diseases. The risk of autoimmunity cannot be disassociated from the capacity of the immune system to cope with diverse and fast evolving pathogens (Langman and M., 1987). The latter is achieved by setting up a vast and diverse repertoire of antigen receptors expressed by lymphocytes, which as a whole is capable of recognizing any possible antigen. Indeed, most lymphocytes have a unique antigen receptor generated by VDJ recombination in lymphocyte precursors. The randomness in the generation of antigen receptors makes it inescapable that lymphocytes with receptors recognizing body antigens are also made. These autoreactive lymphocytes can potentially cause autoimmune diseases if their activation and clonal expansion in the periphery is not prevented. The question is how is this avoided in naturally tolerant individuals?

According to Burnet's original clonal selection theory (Burnet, 1957) expansion of autoreactive lymphocyte clones and autoimmunity would be avoided by deleting autoreactive lymphocytes from the repertoire once and for all during embryonic development. This possibility was dismissed by the fact that the generation of lymphocytes is a life long process in mammals. Following an early suggestion by Lederberg (1959) deletion of potentially self-destructive lymphocytes was reformulated as an aspect of lymphopoiesis, and accordingly immature lymphocytes expressing an autoreactive receptor are deleted either in the thymus or in the bone marrow (Acha-Orbea and MacDonald, 1995; Kisielow et al., 1988). Langman and M. (1987) solved the problem of tolerance to peripheral antigens not expressed in these lymphopoietic organs, through the two-signal model, which predicts efficient deletion of autoreactive lymphocytes during circulation.

Central or peripheral deletion of immature cells play certainly an important role, but alone cannot explain natural tolerance. The major shortcoming of deletion models is the well--

documented presence of significant numbers of mature autoreactive lymphocytes in the periphery of normal healthy animals. Many different experiments have demonstrated that these autoreactive cells could undergo clonal expansion and cause disease, if they were not under the control of specific regulatory T cells. It is worth recalling here some of these experiments as they reveal important properties of the processes leading to the development and maintenance of natural tolerance.

Animals thymectomized during a short time window after birth, but not thereafter, develop in adulthood a pathologic autoimmune syndrome characterized by infiltration of multiple organs (Nishizuka and Sakakura, 1969; Sakaguchi et al., 1982). Although these observations could suggest a special wave of regulatory T cell production in the thymus during the perinatal period (Modigliani et al., 1996) there is now evidence that, in mice, regulatory T cells are produced in the thymus throughout life (Fontenot et al., 2005b; Itoh et al., 1999; Jordan et al., 2001; Seddon and Mason, 1999a). Most likely, neonatal thymectomy produces an imbalance of regulatory T cells in the seed of cells colonizing the periphery (Dujardin et al., 2004), and this imbalance is critically amplified by the peripheral cell population dynamics (Carneiro et al., 1995). Indeed, reconstituting neonatally thymectomized animals with CD25+CD4+ regulatory T cells from healthy adult animals prevents the autoimmune syndrome triggered by neonatal thymectomy (Suri-Payer et al., 1998).

Adoptive transfers of the total peripheral CD4 T cell pool into syngeneic thymectomized animals reconstitute (partial) immunocompetence and also natural tolerance in the recipients. This indicates that if the thymus is fundamental for the development of natural tolerance it may play a lesser role in its maintenance during adulthood. Natural tolerance will not be reconstituted in the recipients upon transfer of CD4 T cell subsets that do not have a poised proportion of regulatory T cells. Thus, large clonal expansions and autoimmune pathology are observed in empty animals transferred with few numbers of CD4 T cells (our unpublished observations) or with CD25⁻ (4), CD45Rb^{high} (Morrissey et al., 1993; Powrie et al., 1993; Powrie and Mason, 1990), or Foxp3⁻ (Wan and Flavell, 2005) CD4 T cell subsets that are poor in regulatory T cells. This autoimmune lymphoproliferative pathology is not observed in animals reconstituted with the complementary CD4 subsets, which are enriched in regulatory T cells, or with poised mixtures of regulatory enriched and impoverished subsets. In these recipients, the clonal expansion of T cells is controlled, thus reaching a steady state (Almeida et al., 2002; Annacker et al., 2000, 2001). Regulatory T cell enriched pools from donors lacking specific tissues fail to prevent autoimmune responses against those tissues in the recipients (Seddon and Mason, 1999a), which indicates that regulatory T cells are antigen (or at least tissue) specific, and that their persistence as a population requires continuous stimulation by peripheral antigens. Collectively, these observations indicate that the robustness of natural tolerance in the adults is associated to the density-dependent interactions regulatory T cell populations make with other T cell and APC populations (Leon et al., 2000). It also shows that very strong perturbations to the T cell equilibrium proportions may be amplified through clonal expansion leading to autoimmunity.

This disequilibrium between regulatory T cells and their targets leading to autoimmunity may be elicited, not only by direct perturbations to the T cell proportions, but also indirectly by perturbing other leukocytes. Autoimmune responses to body antigens can be triggered

in adult animals by immunization with adjuvants (reviewed in Panoutsakopoulou and Cantor 2001) that entail massive local inflammation, and therefore strong perturbations to the innate immune system. Intriguingly, reverting a situation of overt autoimmunity in adulthood seems more difficult to achieve than to imbalance a situation of natural tolerance.

The conditions necessary to achieve tolerance to histo-incompatible transplants grafted during embryonic life, further emphasize the role of thymic selection and peripheral lymphocyte dynamics in the establishment of natural tolerance to body tissues. Most histo-incompatible tissues successfully grafted during embryonic life will be rejected once immunocompetence develops, the only exceptions being the hematopoietic tissue (Billingham et al., 1953) and thymic epithelium (Modigliani et al., 1995; Ohki et al., 1987). These two tissues are able to induce tolerance to themselves, and also to other tissues from the same donor (Billingham et al., 1953; Modigliani et al., 1995; Ohki et al., 1987). Both hematopoietic cells and thymic epithelium are involved in positive and negative selection and MHC restriction of CD4 T cells in the thymus, which suggests that thymic selection preconfigures the repertoire to facilitate the development of natural tolerance to the ensemble of peripheral MHC-peptides; while failure in preconfiguring the repertoire in this way may lead to tissue rejection.

Based on the Crossregulation Model described in the next section, we will sketch in the following sections a picture of the immune system, trying to portray and give a coherent tentative explanation to all the above properties of natural tolerance. Before doing so, it must be stressed that it will be assumed throughout this review that to establish natural tolerance it is sufficient to prevent clonal expansion of autoreactive CD4 T lymphocytes. This simple view is appropriate because it turns the regulation of autoimmunity into a problem of lymphocyte population dynamics, which can be clarified by studying the conditions for the equilibrium between potentially pathogenic and regulatory T cells. However, it is also important to acknowledge from the outset that some important aspects of natural tolerance might be lost by assuming this somewhat traditional "clonal selection" view, since autoimmune pathology can be prevented by regulatory T cells even in the presence of an already large "clone" of autoreactive T cells (Lafaille et al., 1994).

4.3 *The Crossregulation Model*

Leon et al. (2000) argued that the large body of data on adoptive transfers of tolerance implies a bistable dynamic system, and that this is a natural expectation if the persistence and expansion of regulatory T cell populations depend on the target T cells they suppress. Further studies led to the consolidation of a hypothesis for the dynamics of regulatory T cells and the role of these cells in natural tolerance and in self-nonself discrimination (Carneiro et al., 2005; Leon, 2002; Leon et al., 2004, 2003, 2001). This hypothesis is embodied in the Crossregulation Model outlined in this section, and succinctly proposes that specific autoreactive regulatory T cells "feed" on the very same autoimmune dynamics that they suppress.

4.3.1 *General Biological Principles*

Every model obeys some general principles that guide its design. The Crossregulation Model is based on two general principles that are essential for the integrity and viability of multicellular

organisms. First, the persistence of any cell lineage or tissue requires that its cells make recurrent interactions with other cells within the organism; cells that fail to make intercellular interactions will die by apoptosis. Second, the homeostatic turnover of cells in a lineage or tissue involves density dependent feedback mechanisms controlling cell cycle. These feedback mechanisms are mediated by indirect interactions among cells (such as competition for limited survival and growth factors of molecular or cellular nature) or by direct interactions, such as contact inhibition. In the immune system these general principles of multicellular organization must be reconciled with the capacity of leucocytes to undergo activation, proliferation, and differentiation in response to pathogens (Janeway Jr, 1992; Stewart, 1992).

4.3.2 Postulates of the Model

The Crossregulation Model describes the peripheral lymphocyte population dynamics taking into account three mutually interacting cell types: (a) antigen presenting cells (APCs) displaying membrane MHC-peptide complexes; (b) effector T (T_E) cells that can potentially induce autoimmunity or build immune responses to foreign pathogens depending on their specificity; and (c) regulatory T (T_R) cells which suppress proliferation of effector T cells with similar specificities, preventing their clonal expansion.

The model, in its conceptual or mathematical formulations, requires a set of postulates that summarize the life cycle of these cells and interactions they make with each other. These postulates are the following:

1. APCs in the body can be collectively classified as different populations of equivalent APCs. Thus, each APC in a particular population presents the same set of peptides, being regarded as equivalent as far as recognition by and conjugation with T cells is concerned.
2. Each APC population is in a stationary state being continuously renewed from precursors.
3. T_E and T_R cells are also classified as different populations according to their clonal specificity. For the purpose of this paper, it is more relevant to aggregate these cells into populations of equivalent clones with respect to their interactions with the APC populations.
4. T_E and T_R cells are exported as such by the thymus where they differentiate from precursors after productive rearrangement of their TCR genes, at some time during positive and negative selection. The quantitative contribution of peripheral differentiation is neglected here.
5. Quiescent or resting T_E and T_R cells are slowly lost by apoptosis in the periphery.
6. Activation of T_E and T_R cells to perform functions and to progress through the cell cycle requires interactions with APCs presenting cognate antigens (cognate APCs, for short), and depends on interactions these T cells make with each other.

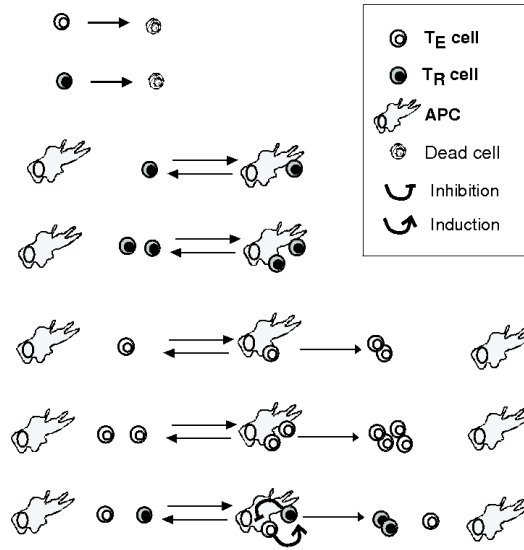


Fig. 4.1: The Crossregulation Model. The reactional diagram indicates the events and interactions underlying the dynamics of APCs, T_E cells and T_R cells as assumed in the model. In this simple scenario the APC can only form conjugates with a maximum of two T cells.

7. T_E and T_R cells interact indirectly by competition for access to cognate APCs and more directly by molecular processes that require the co-localization of both cells in physical domains in the vicinity of these cognate APCs. We will call these domains APC-dependent interaction foci (or foci, for short). The simplest form of these foci is the multicellular conjugates studied in Leon et al. (15). Interactions in the same foci guarantees some degree of specificity in these interactions: only T_E and T_R cells recognizing peptides on the same APC will partake the same foci and therefore interact with each other.
8. Proliferation of specific T_E cell populations is promoted by productive interactions with cognate APC populations, and may be suppressed by regulatory T cells if the APCs present also their cognate peptides.
9. T_R cell proliferation depends on interactions with both APCs and T_E cells co-localized in the same foci.

4.3.3 Minimal Mathematical Formulation of the Model

In this review we illustrate the insights gained with the Crossregulation Model using one of its simplest mathematical forms, namely the scenario where the APCs can only form multicellular conjugates with a maximum of two T cells, and crossregulatory interactions between T_E and T_R cells are restricted to these multicellular conjugates (fig. 4.1). Although this is clearly an unrealistic setting, it captures most of the qualitative properties of more realistic scenarios, as discussed in the next section. The reaction diagram in fig. 4.1 can be translated into the following set of differential equations describing the time evolution of the densities of T_E and T_R cell populations with the same specificity, respectively E and R, that are driven by a single

cognate APC population, and that is at a fixed density.

$$\frac{dE}{dt} = p_E E_A - d_E E \quad (4.1)$$

$$\frac{dR}{dt} = p_R R_A - d_R R \quad (4.2)$$

where E_A and R_A are the densities of activated T_E and T_R cells in multicellular conjugates. The parameters p_E and p_R are the proliferation rates of activated T_E and T_R cells, respectively. The parameters d_E and d_R are the death rates of T_E and T_R cells. The densities of activated T_E and T_R cells are calculated in a stepwise manner (Carneiro et al., 2005; Leon et al., 2000). First, the equilibrium density of T cells in conjugates, denoted C , is obtained as a function of the total density of T cells, $T = E + R$, and the total density of APC conjugation sites, A . This is computed considering that the reactions of conjugate formation and dissociation, which happen at a much faster time scale than the changes in T cell population sizes, are at equilibrium with constant K . The expression for the conjugates (equation (4.3)) implies that they initially increase linearly with both APC and T cell density, and saturate when either of the two populations becomes limited.

$$C = \frac{1 + K(A + T) - \sqrt{[1 + K(A + T)]^2 - 4ATK^2}}{2K} \quad (4.3)$$

The densities of T_E and T_R conjugates, respectively E_C and R_C , are obtained through the corresponding portions of the conjugate densities.

$$E_C = \frac{E}{T}C \text{ and } R_C = \frac{R}{T}C \quad (4.4)$$

Considering the fractions of conjugation sites per APC that are occupied by T_E and T_R cells at equilibrium, respectively ϵ and ρ :

$$\epsilon = \frac{E_C}{A} \text{ and } \rho = \frac{R_C}{A}, \quad (4.5)$$

the density of activated cells is finally obtained according to the following expressions:

$$E_A = E_C \left(1 - \frac{2\rho}{2 - \epsilon} \right) \quad (4.6)$$

$$R_A = R_C \left(\frac{2\epsilon}{2 - \rho} \right) \quad (4.7)$$

that take into account the stoichiometries of the conjugates indicated in the diagram of fig. 4.1. The factors in brackets in equations (4.6) and (4.7) are, respectively, the probability that a conjugated T_E cell has no neighboring T_R cell (i.e. being alone in the conjugate or with another T_E cell), and the probability that a conjugated T_R cell has a neighboring T_E cell. These expressions are based on a multinomial approximation that is valid given that the total number of sites (summed over the APCs) is much larger than the number of sites per APC. More rigorous and general expressions, based on the hypergeometric distributions, were derived in the original references (Leon et al., 2003, 2000).

4.3.4 Standard Behavior

Leon et al. (2000) have shown that the particular dynamical behavior set into play in the Cross-regulation Model is determined by two key composite parameters, representing the effective "growth indexes" of T_R and T_E cell populations. These two parameters are directly proportional to the basic parameters controlling population growth, namely conjugation constants, density of APCs, maximum proliferation rate per activated T cell, and inversely proportional to the death rate of the corresponding population. The values of T_R and T_E growth indexes define four parameter regimes according to the resulting dynamic behavior. A systematic analysis of the possible parameter regimes can be found in the original publication (Leon et al., 2000). For the present purposes, it suffices to focus on the parameter regime where a system composed of two populations of T_E and T_R cells, driven by a sufficiently large population of APCs, shows a bistable behavior. In particular, the system can evolve either into an equilibrium state in which only T_E but not T_R cells are present, or into a state in which both cell types coexist in a stable balance, depending on the composition of the seeding mixture of populations (fig. 4.2). The system develops into the T_R and T_E coexistence state, provided that the seeding population has sufficient T_R cells (fig. 4.2a). Otherwise, if T_R cells are initially "outnumbered", T_E cells will competitively exclude them from the system (fig. 4.2b). One of the key aspects of the model is that the interactions between T_E and T_R cells depend on the density of the APC population, and thus T_R cells can also be "outnumbered" by the APCs themselves. A sufficiently large number of APCs will dilute away the direct T cell interactions, giving T_E cells a chance to competitively exclude T_R cells (fig. 4.2c).

In this model, (auto)immunity and tolerance to a given antigen are interpreted, respectively, as the competitive exclusion of T_R cells by the expansion of the T_E cell population, which becomes limited only by APC availability, and the poised coexistence of both T_E and T_R cell populations. As argued before (Carneiro et al., 2005; Leon et al., 2000), healthy immune systems necessarily operate in the bi-stable regime. This is the only condition compatible with the observations that significant numbers of both T_R and T_E cells can be recovered from healthy animals, and that the incidence of autoimmune pathology or tolerance in adoptively reconstituted recipients can be modulated by changing the proportions and absolute numbers of T_R and T_E cell-enriched populations.

4.4 Cellular and molecular mechanisms of crossregulation between regulatory and effector T cells

Since the original hint at the crossregulation hypothesis (Leon et al., 2000), a considerable amount of information has accumulated concerning the life cycle of CD25+ T_R cells, and the interactions they make with other cells. In this section we will briefly review these findings. The bottom-line is that there is ample evidence for the existence of APC-dependent crossregulation of autoreactive T_R and T_E cell populations, as captured in the model, though the molecular and cellular details are hitherto unclear.

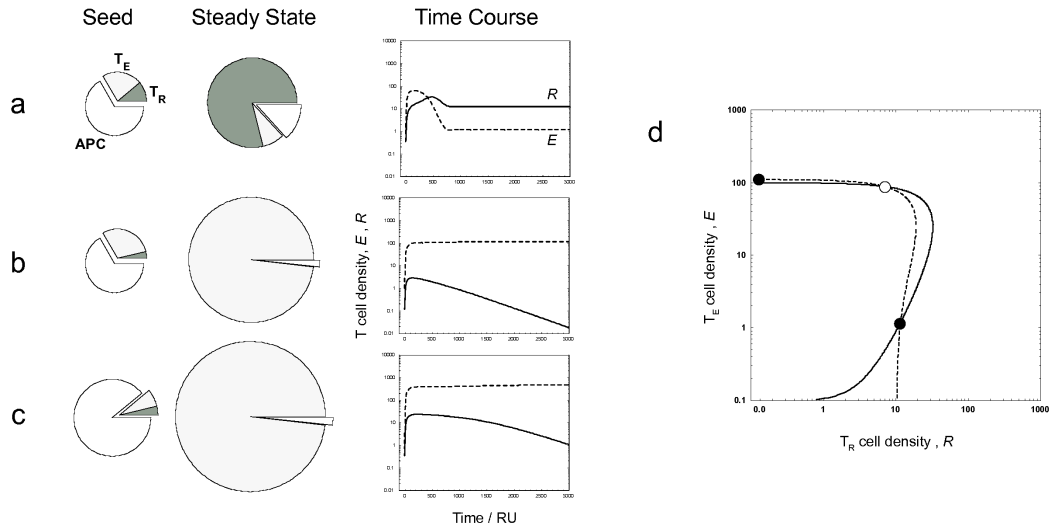


Fig. 4.2: Population dynamics according to the Crossregulation Model. a, b, and c- Numerical solutions of the model for different seeding values of variables E and R, and parameter A, respectively 2:2/3:1/3 (a), 2:8/9:1/9 (b) and 16:2/3:1/3 (c). The pies represent the proportions of the indicated cell types in the initial seeding and at the final steady states (the area of the pie is a linear function of the $\log(A+E+R)$). Reference parameter values: $K = 1$, $d_E = d_R = 0.01$, $p_E = 1.1$, and $p_R = 1.0$. d- Phase space of the model under the same parameters with $A = 2$ (the same as in a and b). The lines are the nullclines uniting points where the individual variables do not change ($dR/dt = 0$ and $dE/dt = 0$ are plain and dashed). The black dots indicate the two stable steady states, and the white dot indicates the unstable saddle point.

4.4.1 Crossregulatory interactions between regulatory and effector T cells in vitro

Irrespective of the underlying mechanisms, APC-dependent crossregulation of T_E and T_R cell proliferation can be readily observed in vitro, using an experimental design that allows one to follow the proliferation of both cell types independently (fig. 4.3). T_R-enriched CD25+CD4+ T cells (for short CD25+ T_R cells) isolated from healthy animals are unable to proliferate when stimulated in vitro with APCs and anti-CD3 (Takahashi et al., 1998; Thornton and Shevach, 1998). And yet, these regulatory cells do proliferate when T_E-enriched CD25-CD24+ T cells (for short CD25- T_E cells) are added to the culture, as can be assessed by CFSE-delabeling (fig. 4.3b). Moreover, the proliferation of T_R and T_E cell populations in these co-cultures is strongly correlated, indicating that the two cell types are using the same growth factors. Since the collective proliferation increases with the T_E:T_R cell proportion at which the co-cultures are seeded, it follows that T_E cells are producing growth factors shared by both cell types, and the production (or availability) of these growth factors is inhibited by T_R cells in a dose dependent manner.

IL-2 is one of the growth factors mediating the crossregulation between T_E and T_R cells observed in vitro (Scheffold et al., 2005; Takahashi et al., 1998; Thornton and Shevach, 1998). Most of the proliferation of freshly isolated T cells is dependent on autocrine/paracrine IL-2 produced by the T cells themselves. CD25+ T_R cells, which do not proliferate when stimulated in similar conditions, do not transcribe the IL-2 gene. Notwithstanding, CD25+ T_R cells, expressing a high affinity receptor for IL-2, do proliferate when supplied with exogenous IL-2

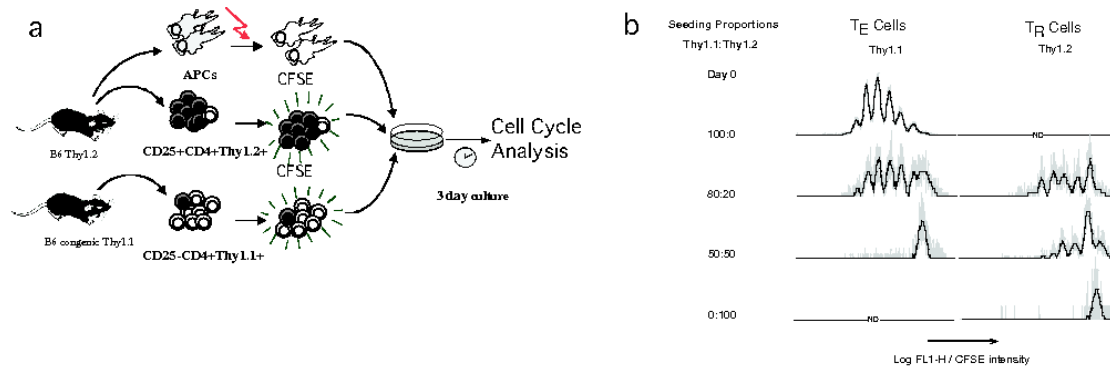


Fig. 4.3: Crossregulation among regulatory T cells and effector T cells in the presence of APCs in vitro. a- Experimental design. b- Proportions and CFSE-profiles of T_E and T_R cells. CD4+CD25+ T cells and CD4+CD25- T cells, purified by highspeed cell sorter respectively from Thy1.2 and Thy1.1 C57Bl/6 congenic mice, were used as T_R and T_E cell enriched populations. Both cell populations were labeled with CFSE and set into cultures at the indicated proportions fixing the total input of 10^5 T cells with 2×10^5 T-cell depleted splenocytes as APCs, and anti-CD3 (5 microgram/ml). Proportions and CFSE-fluorescence profiles of the Thy1.1 and Thy1.2 populations were measured after 3 days of culture by flow cytometry. Data from (Leon, 2002).

(Takahashi et al., 1998; Thornton and Shevach, 1998). Anti-IL-2 blocking antibodies inhibit the proliferation in co-cultures in a dose-dependent manner (de la Rosa et al., 2004). Finally, CD25+ T cells suppress IL-2 transcription by CD25- T cells (Thornton et al., 2004a) and/or the availability of this shared cytokine in co-cultures (de la Rosa et al., 2004), completing the crossregulation interaction scheme.

The molecular and cellular details of the suppression mechanism are controversial. Studies using transwell culture systems and specific blocking antibodies have been used to minimize the role of paracrine suppressive cytokines, concluding in favor of a pathway dependent on direct cell-to-cell interactions between activated CD25+ T cells and target CD25- T cells (Thornton and Shevach, 1998). This conclusion, however, is not fully warranted because suppressive cytokines can act in a juxtacrine manner that may not be effective at the relatively long distances of transwell cultures, or that may not be inhibited with antibody concentrations optimized to block paracrine pathways. Convincing evidence show that a significant part of the suppression observed in vitro is due to the consumption of IL-2 by the CD25+ T_R cells (de la Rosa et al., 2004). The fact that expression of the IL-2 gene is "autocatalytic" implies that the overall concentration of IL-2 available in cell cultures might be critically sensitive to IL-2 consumption by CD25+ T_R cells, as discussed by Scheffold et al. (2005) and suggested by mathematical models (Burroughs et al., 2006; de Boer and Hogeweg, 1987). However, suppression of IL-2 messenger RNA is still observed in co-cultures of CD25+ T_R cells and CD25- T_E cells supplemented with exogenous IL-2, despite the bypass of the proliferation blockade (Thornton et al., 2004a). This inhibition of IL-2 transcription implicates mechanisms of suppression other than IL-2 consumption.

In summary, T_E and T_R cells crossregulate each other's proliferation in vitro, through the crossregulation of IL-2 production and availability via yet unresolved interactions. Other growth

factors may play minor roles in crossregulation of proliferation. CD25+ T_R cells stimulated with anti-CD3, with or without APCs, proliferate in response to a variety of c-family cytokines, namely IL-4 (Thornton et al., 2004b) and IL-15 (our unpublished observations). Some of these cytokines, namely IL-4, can be directly produced by cells within the CD25- T_E subset (Gollob and Coffman, 1994). CD25- T_E cells may also induce APCs to upregulate growth promoting factors, as it has been documented for other stimulatory interactions among T cells (Ridge et al., 1998).

4.4.2 Crossregulatory interactions between regulatory and effector T cells in vivo

Several lines of evidence indicate that the mutual interactions between T_E and T_R cell populations, which underlie the Crossregulation Model, are also relevant in vivo. The immunopathology observed in immunodeficient animals reconstituted with CD25- T_E cells isolated from healthy donors is concomitant with uncontrolled proliferation (Annacker et al., 2001; Sakaguchi et al., 1995). In animals receiving CD25+ T_R cells alone, or animals receiving both CD25- T_E and CD25+ T_R cells, the T cell proliferation is less pronounced and controlled; consequently, the total T cell population reaches an equilibrium state with lower cell numbers (Almeida et al., 2002; Annacker et al., 2001). Furthermore, in recipients reconstituted with CD25- T_E and CD25+ T_R cells, the size of the T_R population recovered at steady state increases with that of the activated T_E population (Almeida et al., 2002; Annacker et al., 2001), indicating that T_R cell proliferation depends on T_E cells. Although the clear interpretation of these experiments is complicated by the fact the cell populations are clonally heterogeneous, the results are non-trivially compatible with the Crossregulation Model (Carneiro et al., 2005; Leon et al., 2000). Hence, the controlled expansion observed in recipients of CD25+ cells is only compatible with the model, if CD25+ T_E cells are co-purified and carried along with CD25+ T_R cells into the recipients, where they become the source of shared growth factors. In fact, recent observations on the relative frequency CD25+Foxp3- T cells, presumably bona fide activated T_E cells, within the pool of CD4+CD25+ T cells of healthy mice (Fontenot et al., 2005b) gives support to this additional assumption.

Corroborating the relevance of the observations in vitro, IL-2 has also been implicated in the crossregulation of T_E and T_R cell population dynamics in vivo. Mice genetically deficient in IL-2 signaling, namely IL-2^{-/-} (Sadlack et al., 1993), CD25^{-/-} (Willerford et al., 1995), IL-2 receptor γ -chain (Suzuki et al., 1995), STAT-5^{-/-} (Antov et al., 2003; Teglund et al., 1998), and JAK-3^{-/-} (Antov et al., 2003) mutants, display uncontrolled lymphoproliferation and autoimmune pathology. Regulatory T cell density is significantly decreased in these animals (Antov et al., 2003; Papiernik et al., 1998). Adoptive transfers of CD25+ T cells isolated from wild type donors restores regulation of T cell dynamics and prevents disease in mutant animals (Almeida et al., 2002; Antov et al., 2003; Malek et al., 2002). Collectively, these observations indicate that the persistence and expansion of T_R cell populations in vivo are strongly dependent on IL-2 signaling. As non-regulatory T cells are the major source of this cytokine in vivo, IL-2 must be one of the major T_E -dependent growth factors shared by T_E and T_R cells. Although it would be tempting to assume here a simple scenario where IL-2 is THE only growth factor

underlying all the postulates in the Crossregulation Model, this hypothesis would fall short. The expansion of non-regulatory T cells observed in animals with compromised IL-2 signaling is necessarily mediated by other, most likely autocrine, growth factors. Likewise, wild type T_E cells resort to other autocrine growth factors whose expression is likely coordinated with that of IL-2. In contrast, IL-2 seems to be the main growth factor that T_E cells provide to T_R cells. This was captured in the model by making the proliferation rate of activated T_E cells greater than that of activated T_R cells ($p_E > p_R$). Coherence with the postulates of the model requires the collective expression of all autocrine growth factors to be suppressed; otherwise, if T_R cells would only suppress the subset of growth factors they depend on, they would always be outcompeted by effector T cells (Leon et al., 2000).

As to the actual mechanism of suppression in vivo the information is still rather incomplete. It is unlikely that all autocrine growth factors produced by T_E cells would be suppressed by IL-2 consumption alone, unless their expression is dependent on IL-2. Accordingly, cytokines such as IL-10 (Annacker et al., 2001; Asseman et al., 1999) and TGF- (Powrie et al., 1996; Seddon and Mason, 1999b), which are potent modulators of gene expression in target cells, have been implicated as likely mediators of T_R -dependent suppression of T_E cell proliferation and function in vivo. Direct cell-to-cell interactions among T cells, as proposed for in vitro systems, may also play a role as suggested by the implication of GITR (McHugh et al., 2002; Shimizu et al., 2002) and CTLA-4 (Read et al., 2000; Takahashi et al., 2000) in suppression. However, direct cell-to-cell interactions among T cells have been recently questioned based on imaging of APC, T_R and T_E cell movement and encounters in vivo (Tang et al., 2006), which indicate that there are less frequent stable conjugates between T cells than between T cells and APCs. This suggests that APCs mediate signals between T cells serving as a temporal bridge (Tang et al., 2006). Alternatively, T_R cells may leave a transient trail of suppressive molecules in the foci surrounding the APC that suppress T_E cells when they later pass by.

4.4.3 APC-dependence of interactions among antigen specific T cells

As mentioned above, the Crossregulation Model postulates that growth, survival, and suppression signals exchanged by T_E and T_R cells require the physical co-localization of these cells in interaction foci, whose capacity depends on the APC density. On the one hand, this confers antigen specificity to the interactions, and on the other hand, brings forth the possibility that changes in the number of APCs affect the balance between T_E and T_R cells.

Recent advances in two-photon confocal microscopy support this assumption. Both antigen-specific T_E and T_R cells swarm around cognate APCs, with which they can form stable conjugates (Tang et al., 2006). This local increase in the density of antigen specific T cells is expected to favor and make possible interactions among T cells, which would otherwise not meet frequently enough, due to the low densities in the rest of the body. This need for increase in local density is required for direct cell-to-cell interactions as well as to those interactions that are mediated by soluble factors acting in paracrine or juxtacrine models.

Localization of T cell interactions, whether resulting in proliferation or suppression, around cognate APCs is also imposed by the transient, labile expression of their molecular media-

tors. Thus, most above-mentioned paracrine regulatory molecules, which are candidates to play the role of T_E cell derived growth factors shared with T_R cells, or to play the role of a suppressive signal, are transiently expressed by T cells upon productive interactions with cognate APCs. This is well documented for growth promoting cytokines, such as IL-2 and other c-family members, as well as for suppressive cytokines like IL-10 and TGF-. The expression of the corresponding receptors by T cells is also TCR-dependent and labile. The best example being the labile expression of CD25 itself (Almeida et al., 2002; Annacker et al., 2001; Zelenay et al., 2005) which is definitively involved in allowing T_R cells to benefit from IL-2 produced by T_E cells, and may also be involved in suppressing proliferation via IL-2 consumption, as discussed above. Even when interactions between T cells are mediated through the modification of APCs, irrespective of growth promoting or inhibiting interactions, the effect can be very localized provided that the APCs remain in place, or if their modification is transient.

Further support to the notion that antigen-specific interactions between T_R and T_E cells operating in vivo (Furtado et al., 2002; Lafaille et al., 1994; Tanchot et al., 2004) are restricted to APC-dependent foci is provided by several reports, showing that two T cell populations will only interfere with each other if they have the same specificity or if their cognate antigens are presented by the same APCs, or co-localized in the same tissue (Davies et al., 1996; Kedl et al., 2000; Moses et al., 2003; Troy and Shen, 2003; Wise et al., 1998).

Notwithstanding these facts and considerations, in vitro studies (Thornton and Shevach, 2000) originated a common assertion that regulatory T cells are antigen non-specific. Definitively, regulatory CD4 T cells do not recognize the antigen on their target T cells since the latter lack MHC class II molecules. And yet, suppression will be effectively antigen-specific in vivo if the rate-limiting step is the antigen-dependent co-localization of the suppressor and target cells. Antigen-nonspecific bystander suppression is most likely observed in vitro when the experimentalist forces the massive co-localization of regulatory and target cells in the bottom of the culture plate wells, reproducing artificially what in vivo could only be achieved by antigen- and TCR-dependent mechanisms.

4.4.4 Several molecular mechanisms may lead to the same cell population dynamics

In the previous subsections we evoked different hypotheses about the molecular details of the crossregulatory interactions between APCs, regulatory T cells and their targets. The different hypotheses for the mechanism of suppression are illustrated in fig. 4.4, where the temporal order and stoichiometry of the key intermediate steps are portrayed. Despite their differences, all the reactions have the same input and output. It is therefore expected that these different cell interaction mechanisms may produce the same collective behavior and the same cell population dynamics, provided that the intermediate steps are sufficiently fast.

In order to better understand how each cell interaction mechanism impinges on the collective behavior of the population, we explored individual cell-oriented simulation systems that take into account geometric and spatial constraints, not easily captured in differential equations. In these simulations, individual cells are regarded as hybrid microagents (Cho et al., 2005; Milutinovic, 2004), treated as soft spheres with internal state and state transition rules.

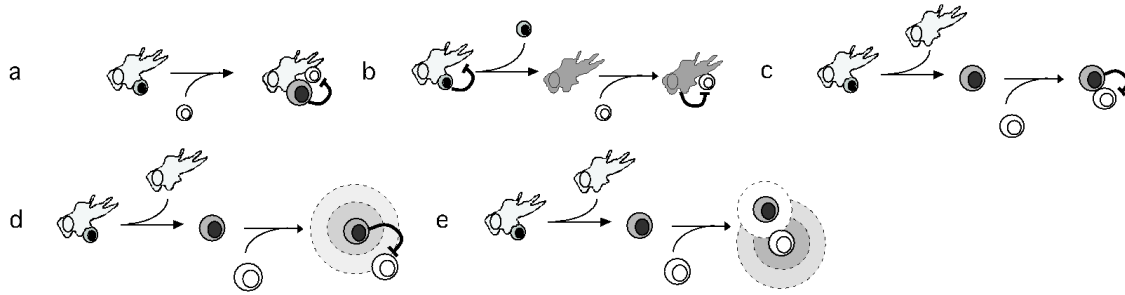


Fig. 4.4: Candidate molecular mechanisms for APC-dependent suppression of T_E cells by T_R cells. The reactional diagrams indicate the key intermediate steps involved in different hypotheses for suppression: a- the T_R cell delivers an inhibitory signal to T_E cell simultaneously conjugated with the APC; b- the T_R cell uses the APC as a temporal bridge to deliver a signal to the T_R cell; c- the T_R cell, upon activation by the APC, delivers a contact dependent signal to the T_E cell; d- the activated T_R cell delivers paracrine signal to the T_E cell; e- the activated T_R cell consumes autocrine/paracrine factors required by the activated T_E cell.

The spheres undergo random walks in Real 3D-space, and can adhere to each other, building non-stoichiometric aggregates. These aggregates move and rotate randomly, and the magnitude of translation and rotation decreases with the size of the aggregate, as suggested by experimental observations. To make these simulations realistic, the motion and adhesion parameters have been parameterized so that the simulations can reproduce the motion and aggregation patterns observed when APCs, T_E and T_R cells are cultured in vitro, and imaged by confocal microscopy (fig. 4.5a). In these simulations, the interactions each cell makes determine changes in its state transition rules. The internal state, in turn, determines if the cell dies or divides, and how it interacts with other cells. By defining the rules by which the internal state changes, one can simulate the different hypotheses. As it turns out, simulations based on each of the hypotheses listed in fig. 4.4, can qualitatively reproduce the experimental CFSE-profiles depicted in fig. 4.3b, under plausible parameter regimes. For example, the results illustrated in fig. 4.5b were obtained using simulations of the hypothesis portrayed in fig. 4.4c. According to this hypothesis the APC-dependent activation and production of growth factors by T_E cells is inhibited by T_R cells via direct cell-to-cell contact, and this inhibitory capacity is induced transiently in T_R cells following their contact with the APC.

A systematic analysis of the capacity of the different mechanisms to explain the quantitative details of in vitro suppression assays will be reported elsewhere (Gardner, et al. in preparation). For the present purpose it is only necessary to say that the Crossregulation Model describes, without loss of generality, the cell population dynamics that is expected from all these different molecular mechanisms. In other words, at the slow time scale at which the average T cell population densities change by death and proliferation one can bracket all the fast intermediate steps indicated in fig. 4.4 to obtain the "aggregated" suppression reaction of fig. 4.1.

Supported by these results, in the next section we use modeling and simulation of the Crossregulation Model, described via eqn.1-7, to understand the selection of the repertoire of T_E and T_R cells in vivo.

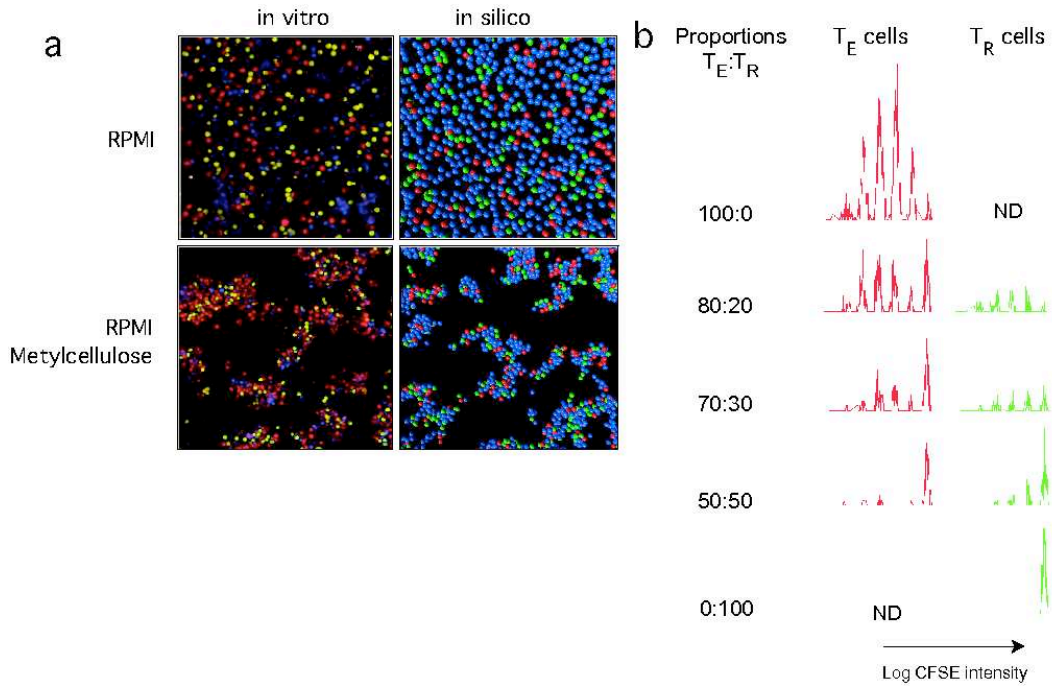


Fig. 4.5: Individual cell-oriented simulations of in vitro cultures of regulatory T cells, effector T cells, and APCs. a- Simulations in silico can reproduce patterns of aggregation of the three cell types as observed in real in vitro cultures under different media. Purified CD25+CD4+ T_R -enriched cells (labeled with CFSE-green), purified CD25-CD4+ T_E -enriched cells (labeled with CMTMR-red), and irradiated erythrocyte and T-cell depleted APC-enriched leucocytes from peritoneal cavity (labelled with DDAO-blue) were set into cultures in glass-bottom petri-dishes with RPMI-media, with or without methylcellulose, and with anti-CD3. Aggregates were imaged with a two-photon laser-scanning microscopy setup. b- Individual cell-oriented simulations of CFSE-delabeling using aggregation parameters corresponding to RPMI media and state transition rules for the cell life cycle corresponding to hypothesis of fig. 4.4c. Results adapted from (Gardner et al. in preparation).

4.5 Peripheral selection of the preimmune repertoire dependent on crossregulation

What maintains the size and diversity of regulatory T cell populations in a steady state immune system? Are all regulatory T cells specific for self-antigens? Or instead, are there regulatory T cells against any possible antigen, including pathogens? If so, how can protective immune responses be mounted? These are the kind of questions that are being debated in the literature, and for which immunological common sense and current theories provide no clear a priori answer. It turns out that the Crossregulation Model makes some qualitative predictions about the repertoire of regulatory T cells, thus generating testable answers to these questions.

4.5.1 Dependence on APC density

The Crossregulation Model posits that the persistence and expansion of T cell populations require recurrent interactions with APCs that lead to activation and, if growth factors are locally available, to cell proliferation. This proliferation is necessary to compensate the slow loss of quiescent cells. The model assumes that APCs are limited in number, and therefore T

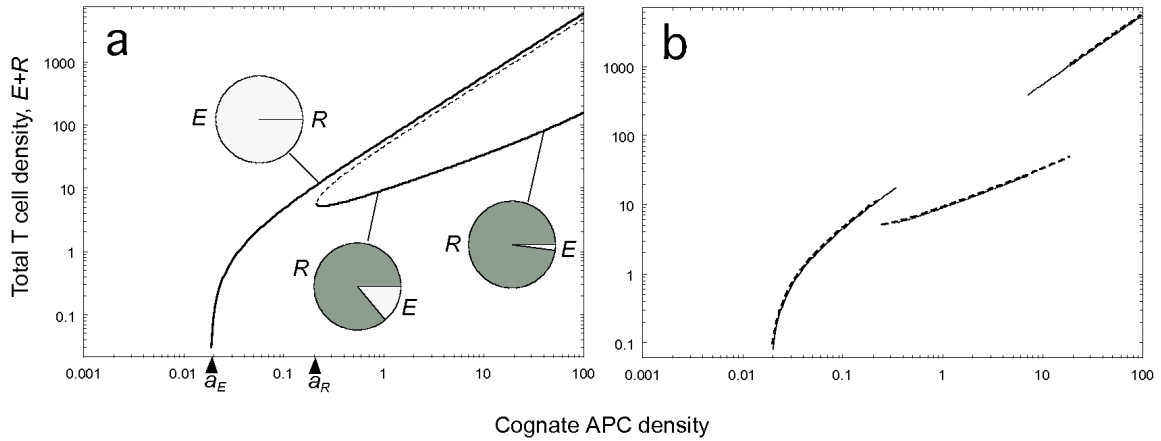


Fig. 4.6: Equilibrium densities of specific T_R and T_E cell populations as a function of the density of their cognate APCs. a- Bifurcation diagram of the model representing all possible equilibria. The lines indicate the total equilibrium density of T cells (sum of the variables $E + R$) as a function of the cognate APC density (parameter A). Solid lines indicate stable equilibria and the dashed line indicates unstable equilibria. The pies indicate the relative proportions of T_E and T_R cells at equilibrium (respectively $E/(E + R)$ and $R/(E + R)$) for the indicated values of A . b- The lines indicate the equilibria that are actually reached by solving the system with fixed initial conditions (solid line: $E = 2/3$ and $R = 1/3$; dashed line: $E = 1/3$ and $R = 2/3$), as a function of the density of cognate APCs. Remaining parameter values as in fig. 4.2.

cells will compete for this resource reaching a steady state (De Boer and Perelson, 1994, 1997; Leon et al., 2000). While both T_E and T_R cells need productive APC conjugations for activation, T_E cells produce autocrine growth factors and T_R cells do not. Therefore, T_E cells once activated by APCs can proliferate driven by their own autocrine growth factors, while T_R cells, in addition to a productive APC conjugation, will need paracrine or juxtacrine growth factors produced by T_E cells in their vicinity. These simple principles imply that there is a critical APC density, say a_E , that is necessary and sufficient to sustain a T_E cell population (fig. 4.6-top). The persistence of a T_R cell population requires a higher density of APCs, say a_R ; only APC densities above this value can sustain a sufficiently large density of a growth factor producing T_E cell population in the neighborhood of the T_R cells (fig. 4.6-top). Based on these simple cell population principles it is expected, as shown below, that the peripheral T cell repertoire can be naturally partitioned in fractions with and without regulatory T cells, according to the degree of autoreactivity.

4.5.2 Stationary state in clonally diverse T_E and T_R cell populations

Understanding the T cell population and repertoire dynamics seems a daunting task. T_E and T_R cells belong to many different clones. Each clone of T cells will have its own set of cognate APCs, which most likely will partially overlap with those other clones. Furthermore, APCs are also heterogeneous, consisting of many populations of equivalent cells. APCs in each population will present a particular set of peptides, and each peptide will be at its own specific concentration. This peptide set may overlap partially with the peptide set of other APC classes. How can we grasp the properties of this complicated network of interacting cell populations?

Many of these complications may be fairly reduced if one considers only the cell population dynamics at the steady state in a preimmune individual. At least this is what would be expected according to the principles of population dynamics underlying the Crossregulation Model. Leon et al. (2003) demonstrated that each population of equivalent APCs can either sustain a single clone (or set of equivalent clones) of T_E cells, or, alternatively, sustain a single clone (or set of equivalent clones) of T_E cells in equilibrium with a single clone (or set of equivalent clones) of T_R cells. This is so because, according to the model, all the T cells that recognize some peptides on the same APC will compete for conjugation sites, and the most efficient clones of each T_E or T_R type, will exclude their respective competitors.

Competitive exclusion implies that, to some approximation, one can regard the steady state composition of the peripheral repertoire as a set of "operationally" independent pairs of T_E and T_R clonal populations, in equilibrium with their own population of (equivalent) cognate APCs. How good is this approximation? To address this issue, we performed simulations where many T_E and T_R cells clones and APC populations classes, were initially networked in non-trivial ways, and allowed to reach equilibrium. The results of these simulations will be reported elsewhere (Sepulveda et al. In preparation). It turns out that in these simulations, after competitive exclusion has purged the repertoire from less efficient clones, the final connectivity is effectively one T_E cell clone to its cognate APCs, or one pair of T_E and T_R clones with their cognate APCs. A simplified setting where the whole repertoire is made of many independent pairs of T_E and T_R clones driven by a distinct cognate APC population, recapitulates the basic behavior of the more complex simulations, and is, therefore, used in the following sections to illustrate the implications of the Crossregulation Model.

4.5.3 Partitioning of the repertoire in subsets with and without regulatory T cells

The Crossregulation Model predicts that the peripheral repertoire of CD4 T cells in healthy preimmune individuals can be naturally partitioned in three subsets of lymphocytes according to the density of the corresponding cognate APCs (fig. 4.7).

The first subset is composed of all the lymphocytes that do not meet their rare cognate APCs frequently enough during circulation through the body. Essentially these would be recent thymic immigrants waiting to die, and their total number should be determined by the ratio between thymic production and peripheral death rates. Within this lymphocyte set, the average clonal size is 1 cell. The proportions of T_R and T_E cells should be the ones produced by the thymus if the two cell types have identical death rates, or otherwise balanced towards the cell type decaying at a slower pace. T_E cells produce autocrine growth factors, and can divide upon sporadic encounters with cognate APCs, and therefore, as a population, they will decay more slowly than T_R cells, even if the "intrinsic" life-span of the two cell types are the same. This set of peripheral lymphocytes would be lost in thymectomized adult animals.

The second subset of specificities corresponds to T cells that are driven to proliferate by populations of cognate APCs whose density in the body can sustain T_E cells but not T_R cells (i.e. APC density larger than a_E but lower than a_R) The clonal sizes within this T cell subset would be proportional to the limiting APC population, being small but larger than a single

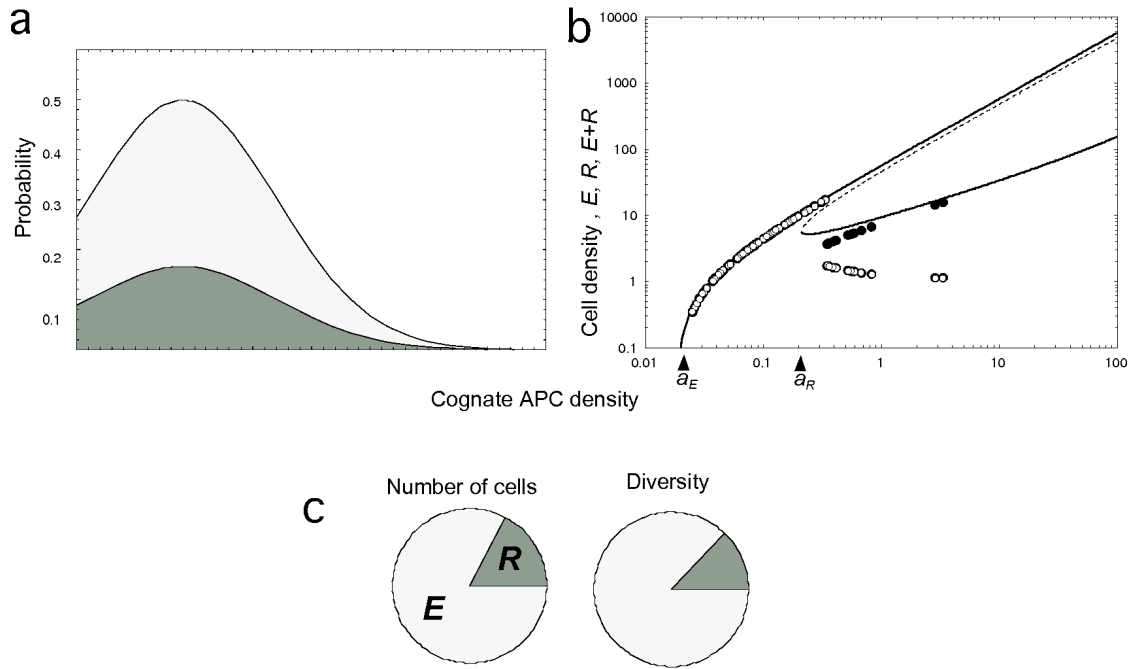


Fig. 4.7: Partition of the repertoire in specificity subsets with and without T_R cells as a function of the density of cognate APCs. a- Frequency distribution of specificities over the cognate APC density. $\text{Log}(A)$ is a Gaussian with mean -1.1 and standard deviation 0.6. The lighter and darker regions correspond respectively to T_E and T_R cells. b- Equilibrium composition of T_E and T_R populations in the periphery. White and black dots represent the equilibrium values of T_E and T_R cell density (variables E and R) in 100 independent pairs of populations driven by specific cognate APCs whose densities A were randomly sampled from the distribution depicted in (a). c- Equilibrium proportions of T_E and T_R cells in terms of total number of cells (sums of E and R over all populations) and diversity (counting only populations where E or R are greater than 0). Remaining parameter values as in fig. 4.2.

cell. Because lymphocytes compete for APCs, only one clone (or a class of equivalent clones) would remain in the system; the remaining clones would be competitively excluded or reduced to the baseline level determined by thymic export and death rate. Within this population, there would be mostly T_E cells. T_R cells with the same specificity would not be sustained because they would not encounter APCs and T_E cells, on which they depend, often enough to compensate their death rate. The T_E cells within this subset would persist upon adult thymectomy but the residual T_R cells would be lost.

The third subset of lymphocyte specificities corresponds to T cell clones that recognize a sufficiently dense cognate APC population that can sustain both T_E and T_R cells (i.e. cognate APCs at densities higher than a_R). The clonal sizes of either T_R or T_E cells in this subset would be relatively small because T_R cells would prevent clonal expansion. Diversity would also be limited, as in the previous subset, due to competitive exclusion. The proportions of T_E and T_R cells would be biased towards the latter, and this bias would increase with the number of stimulatory APCs that sustain them (fig. 4.2). Both T_E and T_R cells within this set should persist upon adult thymectomy.

Tolerance to self requires that T_R cells control all autoreactive T cells that would otherwise cause autoimmunity. It is reasonable to assume that T_E cells that are in equilibrium with

cognate APCs at low densities (lower than a_R) are not sufficiently expanded to cause autoimmunity, and only T_E cell specificities driven by cognate APCs at high density (larger than a_R) can trigger autoimmunity. Under these conditions, a globally tolerant steady state can only be reached if all the peripheral T clonal populations are seeded with enough regulatory T cells. In the simulations illustrated in fig. 4.7b this was guaranteed by seeding each population with 1/3 of T_R cells, irrespective of their specificity.

4.5.4 The Basic Ground for Self-nonself Discrimination

The partitioning of the preimmune repertoire described above is plausible given that it follows naturally from the experimentally supported assumptions on crossregulation, and that it implies that the peripheral repertoire of (Foxp3+) T_R cells should only partially overlap with that of (Foxp3-) T_E cell, and should be, on average, more biased towards body antigens, as recently demonstrated (reviewed by Kim and Rudensky 2006). But does this have any biologically meaningful consequences?

Consider that the peripheral repertoire of healthy adult animals would in fact be partitioned as portrayed in fig. 4.7. The frequency distribution of specificities over their cognate APCs produces (by construction) a reasonable output: none of the T cell populations driven by a high density of cognate APCs are locked in the equilibrium containing exclusively effector cells, the total T_R cells are a minor fraction of the total T cells (fig. 4.7c piechart on the left), and the diversity of T_R cells is only a fraction of the total diversity of the resident peripheral pool (fig. 4.7c piechart on the right).

Suppose now that a new antigen would be introduced in the body under conditions that generate a new ensemble of APCs presenting its peptides. Since this new antigen has not participated in the selection of the preimmune repertoire, it is fair to assume that all T cell populations in the repertoire, in the fractions with and without T_R cells, are likely candidates to recognize its peptides. Under these conditions, the pool of T cells responding to the new antigen corresponds to a random sample from the set of available clonal specificities. Most likely, T_R cells will be a minority in this pool, and therefore responding cells will proliferate and evolve towards the equilibrium where T_E cells eventually outcompete T_R cells.

This means that the repertoire partitioning, which follows naturally from the Crossregulation Model, gains a whole new meaning in terms of the capacity of the immune system to be tolerant to body antigens, and to respond to the remaining universe of antigens. In other words, the natural partition of the repertoire sets the ground for self-nonself discrimination.

4.6 Constraints on Thymic Selection and Maturation of Regulatory T cells

The thymus plays a key role in self-tolerance. This section addresses the possibility that thymic selection and differentiation of regulatory T cells from immature thymocytes are controlled in a way that they synergize with peripheral selection to produce an appropriate partitioning of the preimmune repertoire. The upshot is that there is some redundancy in thymic and peripheral selection, and this redundancy increases the robustness of repertoire partitioning and natural tolerance.

How can thymic selection influence the density of APCs that can stimulate a given clone of T cells in the periphery? The density of cognate APCs available to a T cell clone in the periphery increases with the crossreactivity of the TCR and with the ubiquity of the cognate peptides. On the one hand, the more crossreactive, degenerate, or promiscuous a TCR is the larger the set of its cognate peptides, and therefore the larger the set of APCs that can present cognate peptides. On the other hand, if the TCR of a clone recognizes ubiquitous peptides, then this clone can be, at least potentially, stimulated by many APCs. In the thymus, positive selection keeps in the repertoire only those immature thymocytes whose TCRs are sufficiently crossreactive to recognize some peptides expressed locally, and perhaps a minimum density of APCs presenting these peptides (Faro et al., 2004). In turn, negative selection deletes from the repertoire those immature thymocytes, bearing TCRs that either recognize ubiquitous antigens or that are excessively multireactive. The overall process of thymic selection is expected to produce and export to the periphery an emergent repertoire, in which clonal specificities are distributed over the density of peripheral cognate APCs as a bell-shaped curve (illustrated in fig. 4.7a and fig. 4.8). The median and the variance of this distribution must therefore depend on the thresholds and stringency of positive and negative selection processes.

The other relevant question is: how is the frequency of regulatory T cells at which the periphery is seeded controlled in the thymus? This is basically a question about the mechanism of generating T_R cells. Commitment of immature thymocytes to the regulatory phenotype may happen by many different non-mutually exclusive mechanisms, ranging from stochastic cell-fate decisions followed by selection, to directed differentiation or "instruction". Both selection and instruction may be mediated by a combination of TCR-dependent (Bensinger et al., 2001; Jordan et al., 2001; Modigliani et al., 1996) and independent mechanisms, including IL-2 receptor dependent signals (Malek et al., 2002). The fraction of regulatory T cells that seeds the periphery can be therefore modulated both in TCR-specificity dependent and independent ways.

4.6.1 *Constraints on positive and negative selection, and on thymic differentiation of regulatory T cells*

Thymic selection can potentially shape the frequency distribution of T cell clonal specificities over cognate APC densities in many ways that are compatible with a globally tolerant steady state (fig. 4.8). This is possible due to the robustness of the peripheral selection process that sorts the repertoire in portions with and without T_R cells, and ensures that the highest autoreactivities are restricted to the regulated portion.

In a scenario where thymic selection is less stringent and the avidity thresholds for positive and negative selection of the TCR are further apart, the distribution of the emergent repertoire over the density of peripheral cognate APCs is shifted to the right and broader than in the previous scenario (fig. 4.7). Less stringent selection means that thymopoiesis will have a better yield in terms of the number and diversity of mature cells per precursor. However, a less stringent selection also implies a higher frequency at which the thymus exports clonal specificities driven by more dense cognate APCs in the periphery - i.e. highly autoreactive or crossreac-

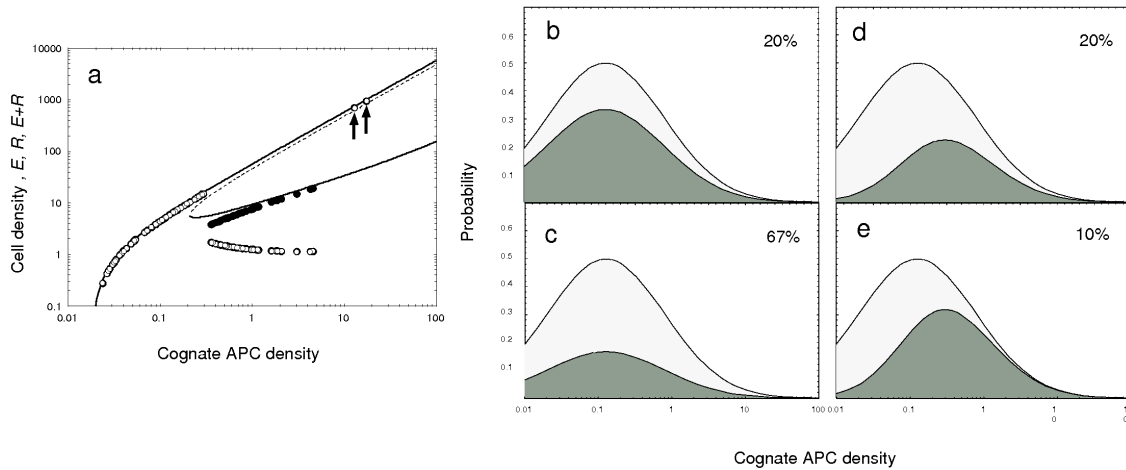


Fig. 4.8: Dependence of the incidence of autoimmunity on the shape of the distribution of clonal specificities over the density of cognate APCs and frequency of regulatory T cells. a- Illustration of "autoimmunity": at least one T_E clonal specificity, that could be controlled by regulatory T_R cells, reaches the APC limited state (indicated by the arrows). b,c,d,e- Shape of the emergent repertoire seeding the periphery in terms of the frequency of specificities distributed over the cognate APC density ($\text{Log}(A)$ is a Gaussian with mean -0.9 and standard deviation 0.8). The lighter and darker regions correspond respectively to T_E and T_R cells. Proportions were defined as fixed specificity-independent fraction of T_R cells set at 0.3(3) (c) or 0.6(6) (b), and variable specificity-dependent fraction of T_R defined as the following saturating function of A : $m \cdot 10^A / (0.1 + 10^A)$ where m is 0.6(6) (d) or 0.95 (e). The percentages indicated the incidence of autoimmunity, when the repertoire is seeded with 100 clones drawn randomly from these distributions. Remaining parameter values as in fig. 4.2.

tive clonal specificities. This in turn increases the frequency at which autoreactive populations are locked in the regulatory-cell free equilibrium, interpreted as autoimmunity (fig. 4.8a). The risk of autoimmunity might nevertheless be kept under control in conditions that relax selection stringency (fig. 4.8d compared to fig. 4.7) through a compensatory increase in the rate at which the thymus generates regulatory T cells (fig. 4.8c,d and e). Potentially, this compensatory increase in regulatory cell generation rate may be mediated by tuning the efficiency of TCR-specific or nonspecific mechanisms (fig. 4.2b, d, and e).

From the point of view of ensuring tolerance to self, a scenario in which the thymus exports regulatory T cells with a fixed (specificity-independent) probability (fig. 4.8b) might be equivalent to a scenario in which the thymus exports preferentially autoreactive or multireactive regulatory T cells (fig. 4.8d). And yet, for the same number of precursors, one expects that there will be less clonal specificities of T_E cells able to respond to invading pathogens, in the first scenario (fig. 4.8b) than in the second scenario (fig. 4.8d). This is so because the crossregulation dynamics in the periphery purges from the repertoire those regulatory T clones that are not sufficiently autoreactive, and therefore do not find dense enough cognate APCs populations. The TCR-dependent maturation of regulatory T cells in the thymus is therefore more flexible, and deals better with a tradeoff between generation of diversity and reliability of natural tolerance. Whether or not there is an advantage in generating regulatory T cells via a TCR-dependent mechanism depends therefore on how much diversity is necessary to protect

against pathogens, and also on how many clonal specificities would be lost by committing to regulatory function immature lymphocytes whose progeny will not persist in the periphery.

A mechanism in which thymic selection biases the repertoire of regulatory T cells towards those TCRs that are highly autoreactive or crossreactive, or recognize ubiquitous body antigens (as depicted in fig. 4.8) was proposed before (Modigliani et al., 1996), and more recently shown to operate in mice (Bensinger et al., 2001; Jordan et al., 2001). This indicates that at least in mice there might be a trade-off between diversity and tolerance induction. An intriguing possibility is that animals with larger bodies, which have higher absolute number of precursors and therefore can produce more TCR variants per unit of time, might get away without resorting to a TCR-dependent mechanism of generating regulatory T cells in the thymus. Therefore, it is not unlikely that there might be differences across species in terms of the constraints on thymic selection. For example humans might have lost, or perhaps never have evolved, mechanisms of regulatory T cell maturation in the thymus that depend on TCR signaling, as mice seem to have done.

It is worth stressing that, according to the model, specific T_R cells will only colonize the periphery if T_E cells of the same specificity are also there, which implies that the thymus should produce specificity-matched pairs of T_E and T_R cells, as it was shown in TCR-transgenic animals (Apostolou et al., 2002; Jordan et al., 2001).

Finally, this section would not be properly concluded without emphasizing that the thymus can only produce an appropriate distribution of specificities over the density of cognate APCs in the periphery if there is some correlation between what is presented in the thymus and what is presented in the periphery. This is achieved partially, as mentioned in the introduction and discussed before (Leon et al., 2003), by biasing the emergent TCR specificities towards the recognition of MHC peptide sets present both in the thymus and in the periphery. This correlation is further promoted by the AIRE gene product that makes constitutive tissue-specific antigens available in the thymus (Derbinski et al., 2001; Kyewski and Derbinski, 2004). This promiscuous expression of peripheral antigens mediates deletion of immature thymocytes (Liston et al., 2003), and may elicit the maturation to the regulatory phenotype (Coutinho et al., 2005). AIRE mutant mice (Anderson et al., 2002), which display generalized autoimmune pathology, would correspond in our model to a scenario where the repertoire distribution is broader and shifted to higher APC densities than the wild type (the scenarios illustrated in fig. 4.8a,c and fig. 4.7, respectively). The finding that AIRE protein mediates the expression of peripheral antigens in local clusters of epithelial cells implies that deletion or instruction to the regulatory phenotype is inherently stochastic: among all the immature thymocytes bearing a specific TCR only those that occasionally migrate close to these clusters of epithelial cells can be deleted or instructed to become regulatory.

4.7 *Coordination with the Innate Immune System*

Invertebrates have highly evolved mechanisms of defense against pathogens. Some of these mechanisms are still built-in in the innate immune system of vertebrates, most likely readapted to operate in coordination with the adaptive lymphocyte system. As discussed above, the

partitioning of the T cell repertoire, as expected from the Crossregulation Model, provides a robust mechanism underlying some kind of self-nonself discrimination, i.e. natural tolerance to self and adaptive immune responses to exogenous antigens unrelated to self. The model suggests also how innate mechanisms can synergize with crossregulation of T lymphocytes to fight infection, to reinforce tolerance to self, and to deal with conflicting situations such as infections with pathogens expressing antigens that mimic self.

The direct recognition of molecular structures of pathogens by innate receptors on APCs and their precursors plays a well-accepted role in promoting immune responses (Coutinho and Moller, 1974; Coutinho and Poltorack, 2003; Janeway Jr, 1989). From the onset of an infection, these receptors trigger the proliferation and differentiation of APC precursors causing an avalanche of new APCs that present antigens from the pathogen in the draining lymph nodes. These APCs are "new" not only because they contain new antigens from the pathogen, but also because they likely present profiles of self-peptides that are different from self-peptide profiles displayed by the preimmune APCs. These changes in self-peptide profiles may be instrumental in facilitating immune responses by rendering pre-existing dominant T_R clones "blind" to these new APCs (Leon et al., 2003). The local upsurge of these new APCs drives the expansion of a pool of T cell clones sampled from the equilibrium preimmune repertoire. In most of the cases, regulatory T cells will be outnumbered in this pool, and although they might interfere with the rate of clonal expansion of the responding T_E cells, they will not prevent the response. However, in those rare situations in which the microorganism mimics self, T_R cells may predominate in the pool of antigen-specific T cells. In these situations of mimicry, as discussed before (Leon et al., 2004), the outcome will depend on speed and magnitude of the rise in APC density, triggered by the innate mechanisms, relative to the proportion of T_R cells preset in the preimmune repertoire. Whatever the proportion of protective T_R cells is, they may be diluted in APC-dependent interaction foci, if the influx of APCs is sufficiently fast and large. In contrast, if the influx is slow and gradual enough, a fraction of T_R cells in the responding pool may slowly adapt its size and control the expansion of T_E cells. The innate APC responses and T cell crossregulation may be co-adapted in vertebrates, so that among the microorganisms mimicking the host, those that make fast and overwhelming infections will override self-tolerance and eventually trigger autoimmunity, while those that grow slowly within the body might be assimilated.

The causal relationship between infections and autoimmunity is a long-standing puzzle. A direct correlation between the incidence of autoimmunity and infections is to be expected from antigen mimicry, according to the classical models and also, as just seen, in the present models of regulatory T cells, in the case of acute overwhelming infections. A few examples of autoimmune disorders, which are documented as being caused by infection with particular pathogens, support this view (Cahill et al., 1997; Fujinami and Oldstone, 1985; Hori et al., 2002; Panoutsakopoulou and Cantor, 2001). And yet, there are several experimental animal models where infection appears to prevent the onset of autoimmunity (Cahill et al., 1997; Fujinami and Oldstone, 1985; Hori et al., 2002; Panoutsakopoulou and Cantor, 2001). Moreover, some epidemiological studies suggest an inverse correlation between the incidence of autoimmunity and the general prevalence of infections in human populations (Bach, 2001). The

Crossregulation Model provides a simple rational for these puzzling observations as discussed in Leon et al. (2004). The continuous, gradually increasing exposure to diverse subclinic infections, rapidly controlled by innate and adaptive immune responses, is expected to lead to a concomitant slow gradual increase in the density of APC presenting self-peptides to autoreactive T cells. This slow increase in APC density renders tolerance more robust, since it implies a gradual increase in the density and in the fraction of T_R cells (see for example fig. 4.6 and fig. 4.7).

The discovery that CD25+CD4+ T_R cells express some innate receptors for molecular structures of pathogens, such as toll-like receptors (TLR), and proliferate in response to the corresponding ligands (Caramalho et al., 2003; Crellin et al., 2005; Suttmüller et al., 2006) might have an important biological significance in this context of the coadaptation of innate and adaptive immune systems. This direct effect of microorganisms on regulatory T cells can synergize with the indirect effect that subclinic infections have on APC densities to render tolerance more robust. In particular it might have a fundamental role in keeping a proper balance of T_E and T_R within the relative minority of autoreactive T cells that happen to be driven by the growing population of APCs elicited during an acute infection by a pathogen poorly related to self. In the absence of such potential direct signals to T_R cells by the pathogen, as we have discussed above, the tendency within the pool of responding T cells would be towards a predominance of T_E cells, and eventually to the competitive exclusion of T_R cells. This predominance of T_E within the subset of autoreactive T cells swiping with the response, could switch the equilibrium associated with APCs presenting purely endogenous peptides towards autoimmunity (Leon et al., 2003). This sort of "epitope spreading" of immunity from pathogens to self-antigens, can be controlled through the direct stimulatory effect that pathogen structures like LPS have on T_R cells (Caramalho et al., 2003), via the maintenance of regulatory cells within the autoreactive cells swiping on the immune response. A similar protective effect could also be mediated by the upregulation of endogenous TLR-ligands locally in the infection foci.

4.8 *A Critique of the Crossregulation Model and Other Scenarios*

The Crossregulation Model describes regulatory T cells, from the cell interactions that control their life cycle to the selection of their repertoire in the thymus and in the periphery, and their contribution to the definition of self. The model submits specific answers to some of the currently open questions about regulatory T cell immunobiology. Thus, under the light of the model, one expects that the most important source of regulatory T cells should be the thymus, and not peripheral differentiation, given the functional implications of this assumption. The regulatory T cell repertoire emerging from the thymus should be complete and overlapping with that of conventional T cells, though the model is undecided on the fraction of regulatory T cells per specificity. In contrast with the emergent repertoire, the actual repertoire of regulatory T cells residing in the periphery should not cover all the specificities available within the pool of conventional T cells. Instead, it should be restricted to those specificities that see high densities of cognate APCs, and therefore could be associated to autoimmunity. Regula-

tory cells should therefore be all autoreactive, and only occasionally will some of these cells also recognize antigens from microorganisms. This means that adaptive immune responses to microorganisms are facilitated by the predominance of non-regulatory T cells within the responding pool. The model also indicates that protective immune responses can be mounted to microorganisms that mimic host antigens provided that infection induces locally a massive upsurge in APCs.

Some of these expectations derived from the Crossregulation Model contrast with some of the hypotheses available in the literature and deserve a critical discussion here.

4.8.1 Does size matter? Control of Lymphocyte Activation and Effector Function versus the Control of Clonal Expansion

As mentioned before, we assumed that T_R cells prevent autoimmunity by controlling the clonal expansion of autoreactive CD4 T lymphocytes. In a transgenic animal model for regulatory T cell prevention of spontaneous autoimmune encephalomyelitis (Lafaille et al., 1994) the peripheral repertoire mimics a situation where a single clone of autoreactive anti-myelin basic protein T cells represents more than 90% of all the peripheral CD4 pool, and still, a minor population of specific regulatory T cells is able to prevent autoimmune disease. This indicates that controlling clonal expansion of autoreactive T cells may be sufficient, though not necessary, to control autoimmunity. In these animals, other mechanisms must operate to ensure tolerance despite the high frequency of responding cells. These additional mechanisms are not captured in the Crossregulation Model. The crossregulation mechanism and the implied partitioning of the repertoire, must therefore be understood as a first filter before other mechanisms of peripheral tolerance induction, or class regulation, are allowed to operate.

4.8.2 Central versus peripheral Generation of Regulatory T cells

We neglected the contribution of peripheral differentiation of regulatory T cells, despite some indications that regulatory T cell differentiation may take place in the periphery (Apostolou and von Boehmer, 2004; Knoechel et al., 2005; Kretschmer et al., 2005; Mahnke et al., 2003). Often, one cannot rule out the possibility that thymic-derived T_R cells assist the de novo generation of regulatory T cells from naïve cells in the periphery. With this in mind, we have modeled this scenario by adding T_R -dependent peripheral differentiation of effector T cells to the regulatory phenotype, and the results are equivalent to the ones described here (Leon et al. 2003, 2000; unpublished results). However, in some reports de novo generation of T_R cells cannot be explained this way unless one assumes that TCR transgenics in a Rag deficient background may have residual amounts of regulatory T cells. Although this matter deserves further definitive clarification, the fact that thymic differentiation of T_R cells is so well-established, and together with our theoretical results, lead us to the conviction that the mechanisms of peripheral differentiation play a secondary role.

4.8.3 *Specificity of the growth factors regulatory T cell populations*

Burroughs et al. (2006) proposed a model of regulatory T cell dynamics that has some of the features of the Crossregulation Model presented here, but differs significantly in that they propose that regulatory T cells would be sustained by a hitherto unidentified, regulatory T cell specific growth factor. The availability of this growth factor in different tissues would allow each tissue to sustain a population of regulatory T cells at a particular density. This tissue specific density of regulatory T cells would set a threshold on the density of IL-2 producing cells that would have to be reached before an immune response would kick in. In our model, these effects might also exist though the regulatory T cell population is controlled indirectly via tissue specific effector T cells.

Regulatory T cells express specific transcription factors, such as Foxp3, which could lead to the upregulation of unique cytokine receptors, not expressed by naïve or effector T cells. However, this scenario has not been detected by microarray analysis of CD4 subsets (Fontenot et al., 2005a; Gavin et al., 2002; Lechner et al., 2001; McHugh et al., 2002). In addition to this empirical argument, one can raise an argument of principle against regulatory T cell specific growth factors. Pathogens are notorious for exploring and utilizing any "fault line" in the immune system of the host. A pathogen that would incorporate in its genome a regulatory-specific growth factor of the host, or would somehow promote its hyperexpression by the infected tissues, could compromise the immune response. This fault-line is not available to pathogens in the model we proposed here, because T_R cell growth depends on the same autocrine/paracrine factors (IL-2 or others) that drive the immune responses by T_E cells. A pathogen that would mess with these putative growth factors could find itself in trouble. On the one hand, decreasing the availability of growth factor might enhance the immune response on the long run following a transient decrease of regulatory T cells. On the other hand, increasing the availability of the growth factor might, at least transiently, enhance directly the expansion of effector T cells. Therefore, the Crossregulation Model, as formulated here, might embody a host protection against this kind of microorganism manipulation of the immune response.

4.9 *Concluding Remarks*

Despite the growing relevance of regulatory T cells and the hope that they might be applied in clinical immunology to improve the management of autoimmune diseases, the truth is that hitherto these cells have not been incorporated into a coherent view of the immune system. Many fundamental questions about regulatory T cells find no tentative answers within the frame of traditional immunological thinking. And yet, submitting tentative answers to open problems, which can be turned into precise testable hypotheses, is what makes theories useful. The absence of theoretically framed hypotheses turns regulatory T cell immunobiology into a cutting-edge, but particularly difficult field.

In this review we explored the hypothesis that specific autoreactive regulatory T cell populations "feed" on the very same autoimmune responses that they suppress, and show that this mechanism provides an integrated explanation for several features of the immune system. The

most important consequence of this Crossregulation Model is that the repertoire of the resident peripheral CD4 population is expected to partition naturally into two pools: a less diverse pool containing small clones of autoreactive T_E and T_R cells that regulate each other's growth, and a more diverse pool, containing many clones of barely autoreactive T_E cells, whose expansion is limited only by APC availability. The clones in the latter pool are therefore free to mount immune responses to microorganisms antigenically unrelated to the host.

This review was driven by the conviction that modeling is only productive when it manages to bridge theory and experimentation. We hope that the enouncing of several specific hypotheses during the present exploration of the crossregulation model, will serve as a basis for further modeling and experiments.

Why is it then, that three is not a crowd? It should be clear by now that all the described immune behaviors predicted by the model require the engagement of all three cell types - APC, regulatory T cells, and effectors T cells - and depend critically on their relative densities. Therefore, none of these cell types can be factored out without losing the capacity to explain some relevant aspects of immunobiology. Indeed, many controversies in the literature, ranging from the disparity between in vitro and in vivo observations, to the role of TLR-ligands in promoting or inhibiting regulatory function, may derive from the oversimplification of trying to eliminate any of these cell types, and their respective proportions, from the scene.

Bibliography

- Acha-Orbea, H. and MacDonald, H. R. (1995). Superantigens of mouse mammary tumor virus. *Annu. Rev. Immunol.*, 13:459–486.
- Almeida, A. R., Legrand, N., Papiernik, M., and Freitas, A. A. (2002). Homeostasis of peripheral CD4+ T cells: IL-2R alpha and IL-2 shape a population of regulatory cells that controls CD4+ T cell numbers. *J. Immunol.*, 169:4850–4860.
- Anderson, M. S., Venzani, E. S., Klein, L., Chen, Z., Berzins, S. P., Turley, S. J., von Boehmer, H., Bronson, R., Dierich, A., Benoist, C., and Mathis, D. (2002). Projection of an immunological self shadow within the thymus by the aire protein. *Science*, 298:1395–1401.
- Annacker, O., Burlen-Defranoux, O., Pimenta-Araújo, R., Cumano, A., and Bandeira, A. (2000). Regulatory CD4+ T cells control the size of the peripheral activated/memory CD4+ T cell compartment. *J. Immunol.*, 164:3573–3580.
- Annacker, O., Pimenta-Araújo, R., Burlen-Defranoux, O., Barbosa, T. C., Cumano, A., and Bandeira, A. (2001). CD25+ CD4+ T cells regulate the expansion of peripheral CD4 T cells through the production of IL-10. *J. Immunol.*, 166:3008–3018.
- Antov, A., Yang, L., Vig, M., Baltimore, D., and Van Parijs, L. (2003). Essential role for STAT5 signaling in CD25+CD4+ regulatory T cell homeostasis and the maintenance of self-tolerance. *J. Immunol.*, 171:3435–3441.

- Apostolou, I., Sarukhan, A., Klein, L., and von Boehmer, H. (2002). Origin of regulatory T cells with known specificity for antigen. *Nat. Immunol.*, 3:756–763.
- Apostolou, I. and von Boehmer, H. (2004). In vivo instruction of suppressor commitment in naive T cells. *J. Exp. Med.*, 199:1401–1408.
- Asseman, C., Mauze, S., Leach, M. W., Coffman, R. L., and Powrie, F. (1999). An essential role for interleukin 10 in the function of regulatory T cells that inhibit intestinal inflammation. *J. Exp. Med.*, 190:995–1004.
- Bach, J. F. (2001). Protective role of infections and vaccinations on autoimmune diseases. *J. Autoimmun.*, 16:347–353.
- Belkaid, Y., Piccirillo, C. A., Mendez, S., Shevach, E. M., and Sacks, D. L. (2002). CD4+CD25+ regulatory T cells control *Leishmania* major persistence and immunity. *Nature*, 420:502–507.
- Bensinger, S. J., Bandeira, A., Jordan, M. S., Caton, A. J., and Laufer, T. M. (2001). Major histocompatibility complex class II-positive cortical epithelium mediates the selection of CD4(+)25(+) immunoregulatory T cells. *J. Exp. Med.*, 194:427–438.
- Billingham, R. E., Brent, L., and Medawar, P. B. (1953). Activity acquired tolerance of foreign cells. *Nature*, 172:603–606.
- Burnet, F. (1957). A modification of Jerne's theory of antibody production using the concept of clonal selection. *Aust. J. Sci.*, 20:67–69.
- Burroughs, N. J., Miguel Paz Mendes de Oliveira, B., and Adrego Pinto, A. (2006). Regulatory T cell adjustment of quorum growth thresholds and the control of local immune responses. *J. Theor. Biol.*, 241:134–141.
- Cahill, R. J., Foltz, C. J., Fox, J. G., Dangler, C. A., Powrie, F., and Schauer, D. B. (1997). Inflammatory bowel disease: an immunity-mediated condition triggered by bacterial infection with *Helicobacter hepaticus*. *Infect. Immun.*, 65:3126–3131.
- Caramalho, I., Lopes-Carvalho, T., Ostler, D., Zelenay, S., Haury, M., and Demengeot, J. (2003). Regulatory T cells selectively express toll-like receptors and are activated by lipopolysaccharide. *J. Exp. Med.*, 197:403–411.
- Carneiro, J., Paixão, T., Milutinovic, D., Sousa, J., Leon, K., Gardner, R., and Faro, J. (2005). Immunological Self-Tolerance: Lessons from Mathematical Modeling. *J. Comp. Appl. Math.*, 184:77–100.
- Carneiro, J., Stewart, J., Coutinho, A., and Coutinho, G. (1995). The ontogeny of class-regulation of CD4+ T lymphocyte populations. *Int. Immunol.*, 7:1265–1277.
- Cho, K. H., Johansson, K. H., and Wolkenhauer, O. (2005). A hybrid systems framework for cellular processes. *Biosystems*, 80:273–282.

- Coutinho, A., Caramalho, I., Seixas, E., and Demengeot, J. (2005). Thymic commitment of regulatory T cells is a pathway of TCR-dependent selection that isolates repertoires undergoing positive or negative selection. *Curr. Top. Microbiol. Immunol.*, 293:43–71.
- Coutinho, A. and Moller, G. (1974). Editorial: Immune activation of B cells: evidence for 'one nonspecific triggering signal' not delivered by the Ig receptors. *Scand. J. Immunol.*, 3:133–146.
- Coutinho, A. and Poltorack, A. (2003). Innate immunity: from lymphocyte mitogens to Toll-like receptors and back. *Curr. Opin. Immunol.*, 15:599–602.
- Crellin, N. K., Garcia, R. V., Hadisfar, O., Allan, S. E., Steiner, T. S., and Levings, M. K. (2005). Human CD4⁺ T cells express TLR5 and its ligand flagellin enhances the suppressive capacity and expression of FOXP3 in CD4⁺CD25⁺ T regulatory cells. *J. Immunol.*, 175:8051–8059.
- Davies, J. D., Leong, L. Y., Mellor, A., Cobbold, S. P., and Waldmann, H. (1996). T cell suppression in transplantation tolerance through linked recognition. *J. Immunol.*, 156:3602–3607.
- de Boer, R. J. and Hogeweg, P. (1987). Immunological discrimination between self and non-self by precursor depletion and memory accumulation. *J. Theor. Biol.*, 124:343–369.
- De Boer, R. J. and Perelson, A. S. (1994). T cell repertoires and competitive exclusion. *J. Theor. Biol.*, 169:375–390.
- De Boer, R. J. and Perelson, A. S. (1997). Competitive control of the self-renewing T cell repertoire. *Int. Immunol.*, 9:779–790.
- de la Rosa, M., Rutz, S., Dorninger, H., and Scheffold, A. (2004). Interleukin-2 is essential for CD4⁺CD25⁺ regulatory T cell function. *Eur. J. Immunol.*, 34:2480–2488.
- Derbinski, J., Schulte, A., Kyewski, B., and Klein, L. (2001). Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nat. Immunol.*, 2:1032–1039.
- Dujardin, H. C., Burlen-Defranoux, O., Boucontet, L., Vieira, P., Cumano, A., and Bandeira, A. (2004). Regulatory potential and control of Foxp3 expression in newborn CD4⁺ T cells. *Proc. Natl. Acad. Sci. USA*, 101:14473–14478.
- Faro, J., Velasco, S., Gonzalez-Fernandez, A., and Bandeira, A. (2004). The impact of thymic antigen diversity on the size of the selected T cell repertoire. *J. Immunol.*, 172:2247–2255.
- Fontenot, J. D., Gavin, M. A., and Rudensky, A. Y. (2003). Foxp3 programs the development and function of CD4⁺CD25⁺ regulatory T cells. *Nat. Immunol.*, 4:330–336.
- Fontenot, J. D., Rasmussen, J. P., Gavin, M. A., and Rudensky, A. Y. (2005a). A function for interleukin 2 in Foxp3-expressing regulatory T cells. *Nat. Immunol.*, 6:1142–1151.
- Fontenot, J. D., Rasmussen, J. P., Williams, L. M., Dooley, J. L., Farr, A. G., and Rudensky, A. Y. (2005b). Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity*, 22(3):329–341.

- Fujinami, R. S. and Oldstone, M. B. (1985). Amino acid homology between the encephalitogenic site of myelin basic protein and virus: mechanism for autoimmunity. *Science*, 230:1043–1045.
- Furtado, G. C., Curotto de Lafaille, M. A., Kutchukhidze, N., and Lafaille, J. J. (2002). Interleukin 2 signaling is required for cd4(+) regulatory t cell function. *J. Exp. Med.*, 196:851–857.
- Gavin, M. A., Clarke, S. R., Negrou, E., Gallegos, A., and Rudensky, A. (2002). Homeostasis and anergy of CD4(+)CD25(+) suppressor T cells in vivo. *Nat. Immunol.*, 3:33–41.
- Gollob, K. J. and Coffman, R. L. (1994). A minority subpopulation of CD4+ T cells directs the development of naive CD4+ T cells into IL-4-secreting cells. *J. Immunol.*, 152:5180–5188.
- Grabar, P. (1950). Immunochemistry. *Annu. Rev. Immunol.*, 19:453.
- Graça, L., Thompson, S., Lin, C. Y., Adams, E., Cobbold, S. P., and Waldmann, H. (2002). Both CD4(+)CD25(+) and CD4(+)CD25(-) regulatory cells mediate dominant transplantation tolerance. *J. Immunol.*, 168:5558–5565.
- Hori, S., Carvalho, T. L., and Demengeot, J. (2002). CD25+CD4+ regulatory T cells suppress CD4+ T cell-mediated pulmonary hyperinflammation driven by *Pneumocystis carinii* in immunodeficient mice. *Eur. J. Immunol.*, 32:1282–1291.
- Hori, S., Nomura, T., and Sakaguchi, S. (2003). Control of regulatory T cell development by the transcription factor Foxp3. *Science*, 299:1057–1061.
- Itoh, M., Takahashi, T., Sakaguchi, N., Kuniyasu, Y., Shimizu, J., Otsuka, F., and Sakaguchi, S. (1999). Thymus and autoimmunity: production of CD25+CD4+ naturally anergic and suppressive T cells as a key function of the thymus in maintaining immunologic self-tolerance. *J. Immunol.*, 162:5317–5126.
- Janeway Jr, C. A. (1989). Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb. Symp. Quant. Biol.*, 54:1–13.
- Janeway Jr, C. A. (1992). The immune system evolved to discriminate infectious nonself from noninfectious self. *Immunol. Today*, 13:11–16.
- Jordan, M. S., Boesteanu, A., Reed, A. J., Petrone, A. L., Holenbeck, A. E., Lerman, M. A., Naji, A., and Caton, A. J. (2001). Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. *Nat. Immunol.*, 2:301–306.
- Kedl, R. M., Rees, W. A., Hildeman, D. A., Schaefer, B., Mitchell, T., Kappler, J., and Marrack, P. (2000). T cells compete for access to antigen-bearing antigen-presenting cells. *J. Exp. Med.*, 192:1105–1113.
- Khattari, R., Cox, T., Yasayko, S. A., and Ramsdell, F. (2003). An essential role for Scurfin in CD4+CD25+ T regulatory cells. *Nat. Immunol.*, 4:337–342.

- Kim, J. M. and Rudensky, A. (2006). The role of the transcription factor Foxp3 in the development of regulatory T cells. *Immunol. Rev.*, 212:86–98.
- Kingsley, C. I., Karim, M., Bushell, A. R., and Wood, K. J. (2002). CD25+CD4+ regulatory T cells prevent graft rejection: CTLA-4- and IL-10-dependent immunoregulation of alloresponses. *J. Immunol.*, 168:1080–1086.
- Kisielow, P., Bluthmann, H., Staerz, U. D., Steinmetz, M., and von Boehmer, H. (1988). Tolerance in T-cell-receptor transgenic mice involves deletion of nonmature CD4+8+ thymocytes. *Nature*, 333:742–746.
- Knoechel, B., Lohr, J., Kahn, E., Bluestone, J. A., and Abbas, A. K. (2005). Sequential development of interleukin 2-dependent effector and regulatory T cells in response to endogenous systemic antigen. *J. Exp. Med.*, 202:1375–1386.
- Kretschmer, K., Apostolou, I., Hawiger, D., Khazaie, K., Nussenzweig, M. C., and von Boehmer, H. (2005). Inducing and expanding regulatory T cell populations by foreign antigen. *Nat. Immunol.*, 6:1219–1227.
- Kyewski, B. and Derbinski, J. (2004). Self-representation in the thymus: an extended view. *Nat. Rev. Immunol.*, 4:688–698.
- Lafaille, J. J., Nagashima, K., Katsuki, M., and Tonegawa, S. (1994). High incidence of spontaneous autoimmune encephalomyelitis in immunodeficient anti-myelin basic protein T cell receptor transgenic mice. *Cell*, 78:399–408.
- Langman, R. E. and M., C. (1987). The E-T (elephant-tadpole) paradox necessitates the concept of a unit of B-cell function: the protecton. *Mol. Immunol.*, 24(7):675–697.
- Lechner, O., Lauber, J., Franzke, A., Sarukhan, A., von Boehmer, H., and Buer, J. (2001). Fingerprints of anergic T cells. *Curr. Biol.*, 11:587–595.
- Lederberg, J. (1959). Genes and antibodies. *Science*, 129:1649–1653.
- Leon, K. (2002). *A quantitative approach to dominant tolerance*. PhD thesis, University of Porto, Porto.
- Leon, K., Faro, J., Lage, A., and Carneiro, J. (2004). Inverse correlation between the incidences of autoimmune disease and infection predicted by a model of T cell mediated tolerance. *J. Autoimmun.*, 22:31–42.
- Leon, K., Lage, A., and Carneiro, J. (2003). Tolerance and immunity in a mathematical model of T-cell mediated suppression. *J. Theor. Biol.*, 225:107–126.
- Leon, K., Perez, R., Lage, A., and Carneiro, J. (2000). Modelling T-cell-mediated suppression dependent on interactions in multicellular conjugates. *J. Theor. Biol.*, 207:231–254.
- Leon, K., Perez, R., Lage, A., and Carneiro, J. (2001). Three-cell interactions in T cell-mediated suppression? A mathematical analysis of its quantitative implications. *J. Immunol.*, 166:5356–5365.

- Liston, A., Lesage, S., Wilson, J., Peltonen, L., and Goodnow, C. C. (2003). Aire regulates negative selection of organ-specific T cells. *Nat. Immunol.*, 4:350–354.
- Mahnke, K., Qian, Y., Knop, J., and Enk, A. H. (2003). Induction of CD4+/CD25+ regulatory T cells by targeting of antigens to immature dendritic cells. *Blood*, 101:4862–4869.
- Malek, T. R., Yu, A., Vincek, V., Scibelli, P., and Kong, L. (2002). CD4 regulatory T cells prevent lethal autoimmunity in IL-2Rbeta-deficient mice. Implications for the nonredundant function of IL-2. *Immunity*, 17:167–178.
- McHugh, R. S., Whitters, M. J., Piccirillo, C. A., Young, D. A., Shevach, E. M., Collins, M., and Byrne, M. C. (2002). CD4(+)CD25(+) immunoregulatory T cells: gene expression analysis reveals a functional role for the glucocorticoid-induced TNF receptor. *Immunity*, 16:311–323.
- Milutinovic, D. (2004). *A stochastic model of microagent population*. PhD thesis, Instituto Superior Técnico, Lisboa.
- Modigliani, Y., Bandeira, A., and Coutinho, A. (1996). A model for developmentally acquired thymus-dependent tolerance to central and peripheral antigens. *Immunol. Rev.*, 149:155–200.
- Modigliani, Y., Thomas-Vaslin, V., Bandeira, A., Coltey, M., Le Douarin, N. M., Coutinho, A., and Salaun, J. (1995). Lymphocytes selected in allogeneic thymic epithelium mediate dominant tolerance toward tissue grafts of the thymic epithelium haplotype. *Proc. Natl. Acad. Sci. USA*, 92:7555–7559.
- Morrissey, P. J., Charrier, K., Braddy, S., Liggitt, D., and Watson, J. D. (1993). CD4+ T cells that express high levels of CD45RB induce wasting disease when transferred into congenic severe combined immunodeficient mice. Disease development is prevented by cotransfer of purified CD4+ T cells. *J. Exp. Med.*, 178:237–244.
- Moses, C. T., Thorstenson, K. M., Jameson, S. C., and Khoruts, A. (2003). Competition for self ligands restrains homeostatic proliferation of naive CD4 T cells. *Proc. Natl. Acad. Sci. USA*, 100:1185–1190.
- Nishizuka, Y. and Sakakura, T. (1969). Thymus and reproduction: sex-linked dysgenesis of the gonad after neonatal thymectomy in mice. *Science*, 166:753–755.
- Ohki, H., Martin, C., Corbel, C., Coltey, M., and Le Douarin, N. M. (1987). Tolerance induced by thymic epithelial grafts in birds. *Science*, 237:1032–1035.
- Panoutsakopoulou, V. and Cantor, H. (2001). On the relationship between viral infection and autoimmunity. *J. Autoimmun.*, 16:341–345.
- Papiernik, M., de Moraes, M. L., Pontoux, C., Vasseur, F., and Penit, C. (1998). Regulatory CD4 T cells: expression of IL-2R alpha chain, resistance to clonal deletion and IL-2 dependency. *Int. Immunol.*, 10:371–378.

- Powrie, F., Carlino, J., Leach, M. W., Mauze, S., and Coffman, R. L. (1996). A critical role for transforming growth factor-beta but not interleukin 4 in the suppression of T helper type 1-mediated colitis by CD45RB(low) CD4+ T cells. *J. Exp. Med.*, 183:2669–2674.
- Powrie, F., Leach, M. W., Mauze, S., Caddle, L. B., and Coffman, R. L. (1993). Phenotypically distinct subsets of CD4+ T cells induce or protect from chronic intestinal inflammation in C. B-17 scid mice. *Int. Immunol.*, 5:1461–1471.
- Powrie, F. and Mason, D. (1990). OX-22high CD4+ T cells induce wasting disease with multiple organ pathology: prevention by the OX-22low subset. *J. Exp. Med.*, 172:1701–1706.
- Read, S., Malmstrom, V., and Powrie, F. (2000). Cytotoxic T lymphocyte-associated antigen 4 plays an essential role in the function of CD25(+)CD4(+) regulatory cells that control intestinal inflammation. *J. Exp. Med.*, 192:295–302.
- Ridge, J. P., Di Rosa, F., and Matzinger, P. (1998). A conditioned dendritic cell can be a temporal bridge between a CD4+ T-helper and a T-killer cell. *Nature*, 393:474–478.
- Sadlack, B., Merz, H., Schorle, H., Schimpl, A., Feller, A. C., and Horak, I. (1993). Ulcerative colitis-like disease in mice with a disrupted interleukin-2 gene. *Cell*, 75:253–261.
- Sakaguchi, S. (2004). Naturally arising CD4+ regulatory T cells for immunologic self-tolerance and negative control of immune responses. *Annu. Rev. Immunol.*, 22:531–562.
- Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M., and Toda, M. (1995). Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol.*, 155:1151–1164.
- Sakaguchi, S., Sakaguchi, N., Shimizu, J., Yamazaki, S., Sakihama, T., Itoh, M., Kuniyasu, Y., Nomura, T., Toda, M., and Takahashi, T. (2001). Immunologic tolerance maintained by CD25+ CD4+ regulatory T cells: their common role in controlling autoimmunity, tumor immunity, and transplantation tolerance. *Immunol. Rev.*, 182:18–32.
- Sakaguchi, S., Takahashi, T., and Nishizuka, Y. (1982). Study on cellular events in postthymectomy autoimmune oophoritis in mice. I. Requirement of Lyt-1 effector cells for oocytes damage after adoptive transfer. *J. Exp. Med.*, 156:1565–1576.
- Scheffold, A., Huhn, J., and Hofer, T. (2005). Regulation of CD4+CD25+ regulatory T cell activity: it takes (IL-)two to tango. *Eur. J. Immunol.*, 35:1336–1341.
- Seddon, B. and Mason, D. (1999a). Peripheral autoantigen induces regulatory T cells that prevent autoimmunity. *J. Exp. Med.*, 189:877–882.
- Seddon, B. and Mason, D. (1999b). Regulatory T cells in the control of autoimmunity: the essential role of transforming growth factor beta and interleukin 4 in the prevention of autoimmune thyroiditis in rats by peripheral CD4(+)CD45RC- cells and CD4(+)CD8(-) thymocytes. *J. Exp. Med.*, 189:279–288.

- Sercarz, E., Williamson, A. R., and Fox, C. F. (1974). *The Immune System. Genes, receptors, signals*. Academic Press, New York.
- Shimizu, J., Yamazaki, S., Takahashi, T., Ishida, Y., and Sakaguchi, S. (2002). Stimulation of CD25(+)CD4(+) regulatory T cells through GITR breaks immunological self-tolerance. *Nat. Immunol.*, 3:135–142.
- Stewart, J. (1992). Immunoglobulins did not arise in evolution to fight infection (with discussion). *Immunol. Today*, 13:396–400.
- Suri-Payer, E., Amar, A. Z., Thornton, A. M., and Shevach, E. M. (1998). CD4+CD25+ T cells inhibit both the induction and effector function of autoreactive T cells and represent a unique lineage of immunoregulatory cells. *J. Immunol.*, 160:1212–1218.
- Sutmoller, R. P., den Brok, M. H., Kramer, M., Bennink, E. J., Toonen, L. W., Kullberg, B. J., Joosten, L. A., Akira, S., Netea, M. G., and Adema, G. J. (2006). Toll-like receptor 2 controls expansion and function of regulatory T cells. *J. Clin. Invest.*, 116:485–494.
- Suvas, S., Azkur, A. K., Kim, B. S., Kumaraguru, U., and Rouse, B. T. (2004). CD4+CD25+ regulatory T cells control the severity of viral immunoinflammatory lesions. *J. Immunol.*, 172:4123–4132.
- Suzuki, H., Kundig, T. M., Furlonger, C., Wakeham, A., Timms, E., Matsuyama, T., Schmits, R., Simard, J. J., Ohashi, P. S., Griesser, H., Taniguchi, T., Paige, C. J., and Mak, T. W. (1995). Deregulated T cell activation and autoimmunity in mice lacking interleukin-2 receptor beta. *Science*, 268:1472–1477.
- Takahashi, T., Kuniyasu, Y., Toda, M., Sakaguchi, N., Itoh, M., Iwata, M., Shimizu, J., and Sakaguchi, S. (1998). Immunologic self-tolerance maintained by CD25+CD4+ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state. *Int. Immunol.*, 10:1969–1980.
- Takahashi, T., Tagami, T., Yamazaki, S., Uede, T., Shimizu, J., Sakaguchi, N., Mak, T. W., and Sakaguchi, S. (2000). Immunologic self-tolerance maintained by CD25(+)CD4(+) regulatory T cells constitutively expressing cytotoxic T lymphocyte-associated antigen 4. *J. Exp. Med.*, 192:303–310.
- Tanchot, C., Vasseur, F., Pontoux, C., Garcia, C., and Sarukhan, A. (2004). Immune regulation by self-reactive T cells is antigen specific. *J. Immunol.*, 172:4285–4291.
- Tang, Q., Adams, J. Y., Tooley, A. J., Bi, M., Fife, B. T., Serra, P., Santamaria, P., Locksley, R. M., Krummel, M. F., and Bluestone, J. A. (2006). Visualizing regulatory T cell control of autoimmune responses in nonobese diabetic mice. *Nat. Immunol.*, 7:83–92.
- Taylor, P. A., Noelle, R. J., and Blazar, B. R. (2001). CD4(+)CD25(+) immune regulatory cells are required for induction of tolerance to alloantigen via costimulatory blockade. *J. Exp. Med.*, 193:1311–1318.

- Teglund, S., McKay, C., Schuetz, E., van Deursen, J. M., Stravopodis, D., Wang, D., Brown, M., Bodner, S., Grosveld, G., and Ihle, J. N. (1998). Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses. *Cell*, 93:841–850.
- Thornton, A. M., Donovan, E. E., Piccirillo, C. A., and Shevach, E. M. (2004a). Cutting edge: IL-2 is critically required for the in vitro activation of CD4+CD25+ T cell suppressor function. *J. Immunol.*, 172:6519–6523.
- Thornton, A. M., Piccirillo, C. A., and Shevach, E. M. (2004b). Activation requirements for the induction of CD4+CD25+ T cell suppressor function. *Eur. J. Immunol.*, 34:366–376.
- Thornton, A. M. and Shevach, E. M. (1998). CD4+CD25+ immunoregulatory T cells suppress polyclonal T cell activation in vitro by inhibiting interleukin 2 production. *J. Exp. Med.*, 188:287–296.
- Thornton, A. M. and Shevach, E. M. (2000). Suppressor effector function of CD4+CD25+ immunoregulatory T cells is antigen nonspecific. *J. Immunol.*, 164:183–190.
- Troy, A. E. and Shen, H. (2003). Cutting edge: homeostatic proliferation of peripheral T lymphocytes is regulated by clonal competition. *J. Immunol.*, 170:672–676.
- Wan, Y. Y. and Flavell, R. A. (2005). Identifying Foxp3-expressing suppressor T cells with a bicistronic reporter. *Proc. Natl. Acad. Sci. USA*, 102:5126–5131.
- Willerford, D. M., Chen, J., Ferry, J. A., Davidson, L., Ma, A., and Alt, F. W. (1995). Interleukin-2 receptor alpha chain regulates the size and content of the peripheral lymphoid compartment. *Immunity*, 3:521–530.
- Wise, M. P., Bemelman, F., Cobbold, S. P., and Waldmann, H. (1998). Linked suppression of skin graft rejection can operate through indirect recognition. *J. Immunol.*, 161:5813–5816.
- Zelenay, S., Lopes-Carvalho, T., Caramalho, I., Moraes-Fontes, M. F., Rebelo, M., and Demengeot, J. (2005). Foxp3+ CD25- CD4 T cells constitute a reservoir of committed regulatory cells that regain CD25 expression upon homeostatic expansion. *Proc. Natl. Acad. Sci. USA*, 102:4091–4096.

5. REPERTOIRE DYNAMICS OF PERIPHERAL REGULATORY AND EFFECTOR T CELLS WITH INTERCLONAL COMPETITION

Abstract

A healthy immune system involves a fine balance between effector T (T_E) cells that mount immune responses, and regulatory T (T_R) cells that suppress them. When this balance is perturbed, immunopathologies arise. Understanding this balance requires to know how the repertoires of T_E and T_R cells are shaped and how they relate to each other. One can argue that a too large intersection between the repertoires could lead to deleterious inhibition of specific immune responses against harmful microorganisms, while a too small overlap may open the way to autoimmune responses. Here we address this issue by a Crossregulation model that describes the peripheral dynamics of a large number of clones recognizing a non-exclusive set of antigen-presenting cells (APCs). Assuming a constant thymic export, three distinct shapes for the T_R and T_E cell repertoires emerged from the simulations, even for the same parameter set. These shapes are fairly distinguished by the composition of the corresponding resident population. The first shape is related to an overdominance of T_E cells in the repertoire with T_R cells being solely maintained by the thymus export. Thus, this shape cannot cope with a healthy immune system. A second shape shows a similar partition to the one that suggested in a previous work (Immunol. Rev., 216:58-68, 2007) that assures peripheral tolerance by a small subset of highly cross-reactive T_R -cell clones, and fight infections with other subset of T_E -cell clones devoided of T_R cells. This shape seems the most adequate to represent a healthy immune system. The last shape shows an overdominance of T_R cells that are maintained by T_E recent thymic emigrants. In spite of assuring peripheral tolerance, this shape does not appear efficient to fight infections that might occur throughout life. For each repertoire shape, the model advances clear predictions for the relationship between T_R and T_E -cell diversities and overlap. For the most adequate repertoire shape, the model predicts a higher diversity of T_E cells than of T_R cells and a negative correlation between clonal size distributions for clones belonging to the intersection repertoire. These predictions should be then confronted with available experimental data, which would be done afterwards in the Chapter 7.

5.1 Introduction

A complete characterization of the T-cell repertoire is fundamental to understand how the immune system avoids deleterious autoimmune responses throughout life but, at the same time, reacts efficiently to harmful pathogens present in nature. This task has gained a tremendous interest with the discovery of contradictory immunological roles of the so-called regulatory $CD4^+CD25^+$ T (T_R) cells expressing the transcription factor Foxp3 (Fontenot et al., 2003; Hori et al., 2003). On the one hand, T_R cells are beneficial to host owing to the fact that they control autoimmunity, graft-vs-host rejection, cancer, and inflammation during an immune response

against pathogens (Sakaguchi et al., 2001). On the other hand, they can also be detrimental to the host by suppressing protective immune responses against potentially harmful pathogens (Belkaid, 2007; Mills, 2004). To tackle this problem, several studies attempted to dissect the main properties of T-cell receptor (TCR) repertoire of T_R and nonregulatory (T_E) T cells using different mouse models (Hsieh et al., 2004, 2006; Pacholczyk et al., 2006, 2007; Wong et al., 2007). All studies showed that TCR samples of T_R cells were more diverse than those of T_E cells, both in the thymus and periphery, suggesting a higher TCR diversity of the respective T_R -cell repertoires. Moreover, samples of both cell types show some similarity in terms of shared TCRs and respective sample abundances. Yet, similarity varies from one study to another and, therefore, it is difficult to ascertain the exact intersection between both repertoires.

Besides the knowledge of structural properties of the repertoire, it is important to determine which antigens are recognized by T_R cells. Two lines of evidence support the notion that T_R cells react to self-antigens. First, T_R cells could prevent autoimmune pathologies when transferred to animals lacking them (Sakaguchi et al., 1995, 2007). Second, T_R thymocytes appear to be focused on high self-reactivities (Modigliani et al., 1996), but more resistance to negative selection process (van Santen et al., 2004). Notwithstanding, T_R cells can also react to foreign antigens. In fact, a study showed that only heterologous and not autologous antigen-presenting cells (APCs) could activate T-cell hybridomas loaded with T_R -driven TCRs (Pacholczyk et al., 2007). Therefore, it is still unclear what are effectively the main targets of T_R cells, either self or foreign antigens.

Crossreactivity is yet another important property of T_R -cell repertoire. In spite of its great importance in the global understanding on the efficiency of the immune system, it is still poorly characterized due to the lack of accurate experimental techniques that can simultaneously deal with a large diversity of T cells and antigens. Notwithstanding, Mason (1998) argues that T cells should be highly crossreactive, because their number is dramatically low in comparison with the number of antigens. The same argument can be applied to T_R cells. Yet, the degree of crossreactivity of T_R cells and how it relates with that of T_E cells remains to be established.

Since good experimental techniques are still lacking to properly assess structural and functional properties of T-cell repertoires, mathematical modelling proves to be a powerful tool to perform this task. This was recently illustrated with the Crossregulation model in the study of different aspects of T_R -cell physiology (León et al., 2004, 2007, 2003, 2000, 2001). In its essence, the model describes a basic interaction network between T_R and T_E cells when conjugated with the same APC. As a starting point of the study of T_R - and T_E -cell repertoires, we previously formulated the model dividing both T cells and APCs into distinct populations (Carneiro et al., 2007). Yet, a strong assumption was made in the model: each T-cell clone has its own APC niche and, thus, there is interclonal competition shaping the repertoire. Nonetheless, the model predicts that T-cell repertoire can be partitioned in three subsets. The first subset is composed of few clones with short life-span that interact with few APCs and, thus, cannot be maintained in the periphery. The second subset refers to a diverse set of barely autoreactive T_E cell clones, whose expansion is limited only by APC availability. The third subset is related to a less diverse set of small self-reactive T cell clones, exhibiting both autoreactive T_R and T_E cells that regulate each other's growth. Since the majority of T_R cells should belong

to the third subset, T_R cells should be less diverse than their T_E counterparts, in contrast to what was observed in the above-mentioned studies. Moreover, because the density of APCs recognized by each T-cell clone can be correlated with the degree of crossreactivity, the model predicts that T_R cells would be focused on high crossreactivities. Nevertheless, it is still unclear whether these predictions would hold when T-cell clones effectively compete for APCs.

This work aims then to extend previous model in order to contemplate competition among T-cell clones showing different specificities. To this end, we setup a model where T-cell clones recognize a non-exclusive set of APCs with different cell densities in the body. We studied relevant structural and functional properties of T-cell repertoire.

5.2 The Crossregulation Model

The Crossregulation model describes the peripheral lymphocyte population dynamics taking into account three mutually interacting cell types: APCs displaying membrane MHC-peptide complexes; T_E cells that can potentially induce autoimmunity or mount immune responses against foreign pathogens depending on their specificity; and T_R cells which might suppress proliferation of target T_E cells sharing similar APC specificity, preventing their clonal expansion.

The model satisfies a set of postulates concerning the life cycle of the above-mentioned three cell types and the interactions they make with each other. The postulates are the following (Carneiro et al. 2007 and previous chapter):

1. There are $\mathcal{A}$ distinct APC populations in the body, each one presenting a given but non-exclusive set of peptides, as also done in a stochastic model describing dynamics of the naive T-cell repertoire dynamics (Stirk et al., 2008). Every APC of a particular population presents the same set of peptides, being equivalent as far as recognition by and conjugation with T cells is concerned.
2. Each APC population is in a stationary state being continuously renewed from precursors. Thus, the density of each APC population will be considered a fixed parameter in the model.
3. T_E and T_R cells are also classified as different populations according to their APC specificity, defined by the set of different APC population they recognize. To study T-cell repertoire dynamics, it is useful to aggregate these cells into populations of equivalent clones with respect to their interactions with the APC populations.
4. Thymus export clones to periphery at a constant rate s_c .
5. Free T_E and T_R cells die with constant rate δ in the periphery.
6. Activation of T_E and T_R cells requires interactions with APCs presenting cognate antigens, and depends on interactions these T cells make with each other.
7. T_E and T_R cells interact indirectly by competition for access to cognate APCs and directly by molecular processes that require the co-localization of both cells in physical domains in the vicinity of these cognate APCs.

8. Proliferation of specific T_E cell populations is promoted by productive interactions with cognate APC populations, and might be suppressed by T_R cells if the APCs present also their cognate peptides.
9. T_R -cell proliferation depends on interactions with both APCs and T_E cells co-localized in the same physical domain.

A more detailed discussion about these postulates can be found elsewhere (Chapter 4 and Carneiro et al. 2007).

5.2.1 Mathematical formulation

Let $E_{i,t-t_i^*}$ and $R_{i,t-t_i^*}$ be the densities of T_E and T_R cells of clone i at time t given that this clone had entered the periphery at time t_i^* , respectively. Let also $E_{i,t-t_i^*}^*$ and $R_{i,t-t_i^*}^*$ be their activated cell densities. The basic dynamics of each clone is described by the following ordinary differential equations

$$\begin{cases} \frac{dE_{i,t-t_i^*}}{dt} = \pi_e E_{i,t-t_i^*}^* - \delta E_{i,t-t_i^*} \\ \frac{dR_{i,t-t_i^*}}{dt} = \pi_r R_{i,t-t_i^*}^* - \delta R_{i,t-t_i^*} \end{cases}, i = 1, \dots, \mathcal{D}_{T,t} \quad (5.1)$$

where $\mathcal{D}_{T,t}$ is the number of different T-cell clones in the periphery at time t , δ is the T-cell death rate, π_e and π_r are the proliferation rates of activated T_E and T_R cells, respectively. The above equations describe the situation where T_E and T_R cells die at rate δ , and proliferate with rate π_e and π_r , respectively. Proliferation is promoted upon T cell activation, which, in its turn, requires productive conjugation with APCs. For simplicity in the simulations, we generate all T-cell clones with similar initial conditions, $E_{i,t_i^*} = E$ and $R_{i,t_i^*} = R$, $\forall i$. Moreover, when the overall density $T_{i,t-t_i^*} = E_{i,t-t_i^*} + R_{i,t-t_i^*}$ of a clone falls below a certain threshold, we assume that that clone is extinct and the corresponding equations are removed from the system, as done previously (Carneiro et al., 1996; León et al., 2003). Our condition for clonal removal is $T_{i,t-t_i^*} = E_{i,t-t_i^*} + R_{i,t-t_i^*}$ is less than a certain value (e.g., 10^{-3}).

The densities of activated T_E and T_R cells of each clone are calculated in a stepwise manner (Carneiro et al., 2005; León et al., 2000). Let i -th T-cell clone recognize the j -th APC population. Let $C_{ij,t}$ be the total density of conjugates formed with T cells from clone i and APCs from population j . The dynamics of these conjugates is generically described by the following system of equations

$$\frac{dC_{ij,t}}{dt} = \gamma_c (A_j - \sum_i C_{ij}) (T_{i,t-t_i^*} - \sum_j C_{ij}) - \gamma_d C_{ij}, i = 1, \dots, \mathcal{D}_{T,t}, j = 1, \dots, \mathcal{A} \quad (5.2)$$

where γ_c and γ_d are the conjugation and deconjugation rates between APCs and T cells, respectively. In the above equation, new conjugates are formed by the conjugation of free APCs of population j with free T cells of clone i at rate λ_c , while existing conjugates deconjugate at rate λ_d . Assuming that the conjugation and deconjugation of T cells from the APCs is an

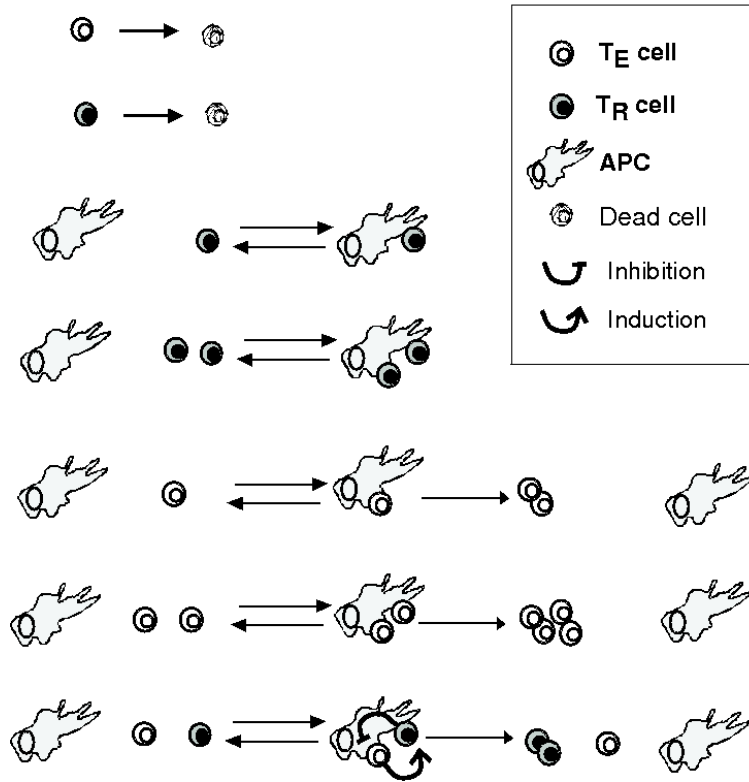


Fig. 5.1: The Crossregulation Model. The reactional diagram indicates the events and interactions underlying the dynamics of APCs, T_E cells and T_R cells as assumed in the model. In this simple scenario the APC can only form conjugates with a maximum of two T cells, which can be either from the same T-cell clone or from a different one, if it recognizes the same APC population.

extremely fast process with respect to the overall T cell clone dynamics, we solve in each time step the steady state of the above system of equations by the Euler method, which is slow but with guaranteed convergence.

Given the conjugate density $C_{ij,t}$, we next calculate the density of conjugated T_E and T_R cells as being proportional to the relative frequency of T_E and T_R in the clone, i.e.,

$$e_{ij,t} = C_{ij,t} \times \frac{E_{i,t-t_i^*}}{T_{i,t-t_i^*}} \text{ and } r_{ij,t} = C_{ij,t} \times \frac{R_{i,t}}{T_{i,t-t_i^*}}. \quad (5.3)$$

To calculate the amount of conjugated T cells being activated, we assume that an APC has s independent conjugation sites, each one with possibility of being occupied by a T cell (León et al., 2001). With this assumption, we can determine, respectively, the fractions of sites per APC occupied by T_E and T_R cells

$$\tilde{e}_{ij,t} = \frac{e_{ij,t}}{A_j s} \text{ and } \tilde{r}_{ij,t} = \frac{r_{ij,t}}{A_j s}. \quad (5.4)$$

Having this fraction calculated, the next step is to compute the fraction of activated T_E and T_R cells in the conjugates. As in previous works, it is assumed that a T_E cell can only be activated by an APC when there are no T_R cells in the same conjugate, otherwise a single T_R

cell has the ability to suppress all activation signals acting on conjugated T_E cells (León et al., 2001). In this situation, the fraction of activated T_E cells is equivalent to the fraction of T_E cells conjugated alone or with other T_E cells in the same APC. After some algebra, this fraction can be expressed by

$$e_{ij,t}^* = \frac{(1 - \tilde{r}_{ij,t})^s - (1 - \tilde{e}_{ij,t} - \tilde{r}_{ij,t})^s}{1 - (1 - \tilde{e}_{ij,t})^s}. \quad (5.5)$$

In its turn, a T_R cell is only activated when there is at least one T_E cells conjugated in the same APC. Thus, the fraction of activation T_R cells in the conjugated is given by the complementary fraction of having only T_R cells in the conjugates, i.e.,

$$r_{ij,t}^* = 1 - \frac{(1 - \tilde{e}_{ij,t})^s - (1 - \tilde{e}_{ij,t} - \tilde{r}_{ij,t})^s}{1 - (1 - \tilde{r}_{ij,t})^s}. \quad (5.6)$$

For sake of simplicity, and as done previously we only consider in this work APCs with just two sites ($s = 2$; Chapter 4 and Carneiro et al. 2007), and thus the above equations simplify to into

$$e_{ij,t}^* = 1 - \frac{2\tilde{r}_{ij,t}}{2 - \tilde{e}_{ij,t}} \text{ and } r_{ij,t}^* = \frac{2\tilde{e}_{ij,t}}{2 - \tilde{r}_{ij,t}}. \quad (5.7)$$

Finally, we calculate the fractions of activated T_E and T_R cells of i -th T-cell clone summed over all APC populations,

$$e_{i,t}^* = \sum_{j=1}^A \alpha_{ij} e_{ij,t} \times e_{ij,t}^* \text{ and } r_{i,t}^* = \sum_{j=1}^A \alpha_{ij} r_{ij,t} \times r_{ij,t}^*, \quad (5.8)$$

which are then transformed into densities by

$$E_{i,t}^* = E_{i,t} \times e_{i,t}^* \text{ and } R_{i,t}^* = R_{i,t} \times r_{i,t}^*. \quad (5.9)$$

5.2.2 Relevant properties of the basic Crossregulation model

Before we describe how the crossreactivity distribution is generated in the model, it is better to recall some of the basic dynamical properties of the model when T cells and APCs are considered as homogeneous populations, respectively. This would be particularly useful to understand how parameters regarding the APC population were specified throughout simulation. For this work, it is crucial to know the dependence of T_R - and T_E -cell steady states on the APC density; further details on the properties of this model can be found in Chapter 4. In this case, APC density can be divided in three parameter regimes related to qualitatively distinct behaviours of the system (Figure 5.2A). The first parameter regime refers to the trivial situation where the APC density is too low to sustain either T_E or T_R cells in the system (that is, T_R - and T_E -cell densities are in steady state equal to 0). When the APC density reaches a critical point (a_E in Figure 5.2A), the system will go to a parameter regime where only T_E cells can be maintained in the organism. T_R cells are excluded by competition. When a second critical point for APC density is reached (a_R in Figure 5.2A), the system enters in a bistable

Tab. 5.1: Summary of parameters/variables related to basic T-cell dynamics, characterization of APC population, and interaction between T cells and APCs. For some parameters, we provide values used throughout simulations; justification for their choice can be found elsewhere (Chapter 4 and Carneiro et al. 2007). The remaining parameters were determined in order to satisfy certain conditions of crossreactivity distribution, which will be explained throughout the text.

Param./Var.	Values	Description
Basic T-cell dynamics		
$\mathcal{D}_t$	—	Number of T-cell clones at time t (T-cell diversity)
t_i^*	—	Time of entrance of the i -th T-cell clone in the periphery
E_{i,t_i^*}, R_{i,t_i^*}	—	T_R - and T_E -cell densities from T-cell clone i when entering the periphery
$E_{i,t-t_i^*}, R_{i,t-t_i^*}$	—	T_E and T_R cell densities from T-cell clone i at time t given that that clone entered in the periphery at time t_i^*
$E_{i,t}^*, R_{i,t}^*$	—	Activated T_E - and T_R -cell densities from i -th clone at time t
π_e	1.00	T_E -cell proliferation rate upon stimulation by APCs
π_r	1.10	T_R -cell proliferation rate upon stimulation by APCs
δ	0.02	T-cell death rate
s_c	—	Number of T-cell clones entering the periphery per unit of time
Characterization of APC population		
$\mathcal{A}$	—	Number of different APCs in the periphery (APC diversity)
A_j	—	Cell density of j -th APC population
s	2	Number of conjugation sites per APC
λ_a	—	Mean density of each APC population (from an Exponential distribution)
Interaction between T cells and APCs		
$C_{ij,t}$	—	Density of APC-T-cell conjugates formed by i -th T-cell clone and j -th APC population
γ_c	1.00	Rate by which T cells conjugate with APCs
γ_d	1.00	Rate by which T cells deconjugate from APCs
k_i	—	Number of APC populations recognized by i -th T-cell clone (crossreactivity of i -th T-cell clone)
θ	0.90	Initial probability of connecting an APC population with a new clone
λ_c	—	Exponential decay in the probability of increasing crossreactivity of a clone given that it recognizes a certain number of APC populations
$A_{T,i}$	—	Total APC density recognized by the i -th T-cell clone

regime where, depending on the initial condition for T_R and T_E cells, the steady states can be either the coexistence of both cell types cells or the competitive exclusion of T_R cells by their T_E counterparts.

For a matter of simplicity, we perform all simulations by generating every T-cell clone under the same initial conditions for T_E and T_R cells (that is, $E_{i,t_i^*} = E$ and $R_{i,t_i^*} = R$, $\forall i$). To understand these simulations, it is crucial to know the behaviour of the system in the bistable parameter regime, as illustrated in Figure 5.2B. When specifying a particular initial condition for T_R and T_E cells in a clone, the parameter region where T_R cells cannot be sustained in the system is extended, from a_E to $\tilde{a}_R$ in Figure Figure 5.2B. When the APC density reaches $\tilde{a}_R$, T_R and T_E cells coexist in steady state. Finally, when the APC density is too high, T_E cells outcompete their T_R counterparts (not shown in Figure 5.2B; Carneiro et al. 2007).

5.2.3 Modeling the APC population

In the model, the APC pool is divided into different populations, each one presenting a distinct but yet overlapping set of peptides, as similarly done by Stirk et al. (2008). In this scenario, the abundance of each APC population is somehow dependent on the availability of the respective

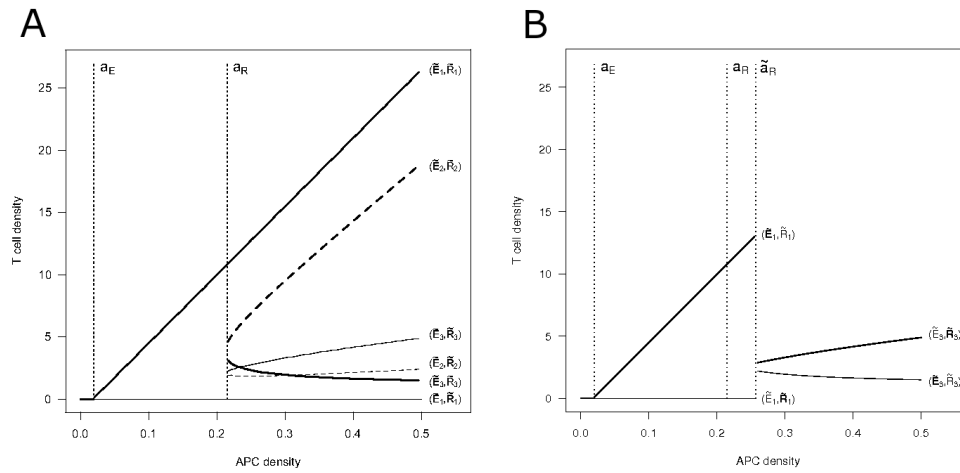


Fig. 5.2: Steady state analysis of the Crossregulation model as a function of APC density per T-cell clone. **A.** Bifurcation diagram as a function of APC density per T-cell clone, where stable and unstable steady states for T_E and T_R cells (thick and thin lines, respectively) are represented (solid and dashed lines, respectively). The system changes qualitatively its dynamic behavior in the critical points a_E and a_R , defining three distinct parameter regimes. First, for APC density less than a_E , the stable steady state is $(\tilde{E}_0, \tilde{R}_0) = (0, 0)$ (not shown). Second, for APC density between a_E and a_R , the stable steady state is given by $(\tilde{E}_1, \tilde{R}_1)$ with $\tilde{R}_1 = 0$. Third, for APC density higher than a_R , the system enters in a bistable regime that, depending on the initial conditions for T_R and T_E cells, the system either goes to $(\tilde{E}_1, \tilde{R}_1)$ or $(\tilde{E}_3, \tilde{R}_3)$. In this regime, there is another steady state, $(\tilde{E}_2, \tilde{R}_2)$, which is unstable. **B.** Stable steady states of T_E and T_R cells for a given initial condition, $E(0) = R(0) = 0.50$. Parameters used to build the plots are given in Table 5.1.

set of peptides in the periphery. Moreover, by postulate 2, the abundance of each APC population in the periphery is assumed to be in equilibrium. Therefore, every APC "species" and its respective density in the body is generated in the beginning of the simulations. For a matter of simplicity, we assume that there are many APC subsets at low densities reflecting the presentation of rare antigens in the body while there are just a few APC subsets at high densities that present ubiquitous antigens. With this rationale in mind, we consider that the density of each APC subset in the periphery follows an Exponential distribution with mean value λ_a , i.e.,

$$A_j | \lambda_a \rightsquigarrow \text{Exp}(\lambda_a), j = 1, \dots, \mathcal{A}. \quad (5.10)$$

5.2.4 Connecting T cells with APCs

When a new T-cell clone is exported to the periphery, we need to define the set of APC populations recognized by it (Figure 5.3A). Since a newly-generated T cell must pass through thymic negative and positive selection processes, it is reasonable to assume that it would recognize a limited set of APC populations when reaching periphery. In fact, thymic positive selection seems to assure that any T cell that comes out from the thymus would recognize some APC population in the periphery. In its turn, negative selection purges from the repertoire T-cell clones that recognize many APC populations. With this in mind, we setup an appropriate Markov chain model with discrete steps to generate a connectivity distribution extremely nar-

row around the mean. The states of this Markov chain are the number of APC populations recognized by a clone (number of connections, for short), ranging from 0 to $\mathcal{A}$. The initial state of a clone is zero connections. In each step of the chain, we test whether an APC population will be or not recognized by the newly-generated T-cell clone. Therefore, there are many chain steps as APC populations. Moreover, the number of APC recognized by a given T-cell clone can only be maintained or increase throughout the steps of the chain. We assume that the transition matrix $(\mathcal{A} + 1) \times (\mathcal{A} + 1)$ is given by

$$p_{i,j} = \begin{cases} 1 - \theta e^{-\lambda_c \times j}, & i = j \\ \theta e^{-\lambda_c \times j}, & i = j + 1 \\ 0, & \text{otherwise} \end{cases}, i, j = 0, \dots, \mathcal{A} \quad (5.11)$$

where θ is the probability of making a connection when a clone has zero connections, λ_c is the exponential rate by which the probability of making a new connection decreases (Figure 5.3B). By standard Markov chain theory, the connectivity distribution is then given by the vector that indicates the process started with zero connections multiplied by transition matrix after $\mathcal{A}$ steps. Figure 5.3C gives an example of the connectivity distribution, which is skewed to the left by construction. In practical terms, we connect the different APC populations to the new clone as follows: (i) sample successively without replacement each APC population; (ii) connect randomly each one to the new clone with probability of increasing connectivity included in equation (5.11).

In all simulations, we setup the parameter θ at 0.90, because in this way we assure that every newly-generated clone recognize with a high probability at least one APC species, in agreement with thymic positive selection, even when the mean of the connectivity distribution is very low. Furthermore, the parameter λ_c can be determined in order to obtain a connectivity distribution with a certain mean value. Since connectivity distribution cannot be calculated analytically, except for the case of few APC populations, we use the software *Mathematica* and its numerical equation-solving methods to tackle this issue.

As seen above, the persistence of a T-cell clone in the periphery is critically dependent on the total APC density recognized by that T-cell clone, denoted as $A_{T,i}$. Therefore, it is of interest to determine the probability distribution of $A_{T,i}$ in order to control how many T-cell clones would be in the different parameter regimes of the system, as illustrated in Figure 5.3D. To this end, we first note that the density of each APC population comes from an exponential distribution. Thus, the total APC density given the number of connections x is a sum of independent and identical exponential distributions, which leads to a Gamma distribution with shape parameter x and scale parameter λ_a , that is,

$$A_{T,i}|X = x \rightsquigarrow Ga(x, \lambda_a), x = 1, \dots, \mathcal{A}. \quad (5.12)$$

We then apply the well-known total probability theorem to obtain the probability density distribution of $A_{T,i}$ given that a clone recognizes at least one APC population

$$f_{A_{T,i}}(a) = \sum_{x=1}^{\mathcal{A}} f_{A_{T,i}}(a|x) \frac{P[X_i = x|\theta, \lambda_c]}{1 - P[X_i = 0|\theta, \lambda_c]}, \quad (5.13)$$

where $f_{A_{T,i}}(a|x)$ is the probability density function of the Gamma distribution given in equation (5.12), and $P[X_i = x|\theta, \lambda_c]$ is the probability of having a clone that recognizes x distinct APC populations. The above probability distribution is then a mixture of appropriate Gamma distributions, which also resembles a Gamma distribution (Figure 5.3D). In the simulations, we control the parameter λ_a to obtain a density probability of distribution $A_{T,i}$ that imply, for example, a 10% probability of generating a clone in the bistable regime (Figure 5.3D). To this end, we used again the software *Mathematica* to determine λ_a that satisfy the above condition.

The model has many parameters summarized in Table 5.1. Some of the parameters were chosen according to previous work (Carneiro et al., 2007). The remaining ones are specified in order to satisfy different conditions, such the proportion of generated clones falling in the bistable regime (Figure 5.3D).

5.3 Results

5.3.1 Peripheral T-cell repertoire dynamics in the absence of thymic influx

We first study the most simple situation where the whole T-cell repertoire is generated in the beginning of the simulations. This situation is conceptually equivalent to an adoptive transfer experiment, where a pool of T cells is transferred into an animal lacking T cells, such as nude or $RAG^{-/-}$ mouse (Sakaguchi et al., 1995; Zelenay et al., 2005). This subsection identifies what properties of the T-cell repertoire arise in this scenario.

Less crossreactive T-cell clones are outcompeted from the repertoire

It has been suggested that T-cell competition is one of the major forces shaping the T-cell repertoire (Freitas and Rocha, 2000). In the past, some mathematical models have been proposed to study this issue but not accounting for the presence of T_R cells in the repertoire (de Boer and Perelson, 1994, 1997). As stated above, T cells should form multicellular conjugates with APCs in order to survive and proliferate (Figure 5.1). Thus, competition among T cells from the same clone but also from different clones partaking the same APC is included in the model by the availability of APCs to form these multicellular conjugates. In general, it is expected that T-cell clones that recognize higher APC densities will outcompete the remaining ones because they have more APC that can support them in the periphery. Yet, this notion might not be completely true owing to the fact that T-cell clones share different set of APC populations, forming a complex network of interactions between these two cell types (Figure 5.3). In this scenario, the persistence of a given T-cell clone in the periphery is not only dependent on APC density, but also on the different APC subsets shared with the remaining clones. It is then reasonable to conceive a situation where a T-cell clone recognizing a large density of APCs can be outcompeted if a large APC density is shared by many different T-cell clones.

It is clear from the simulations that mean APC density recognized per T-cell clone increases throughout time (Figure 5.4A). This result suggests that clones that conjugate with more APCs tend to persist throughout simulations. Similar result can be extracted from the temporal evolution of mean crossreactivity of T-cell clones, which increases throughout simulation until reach-

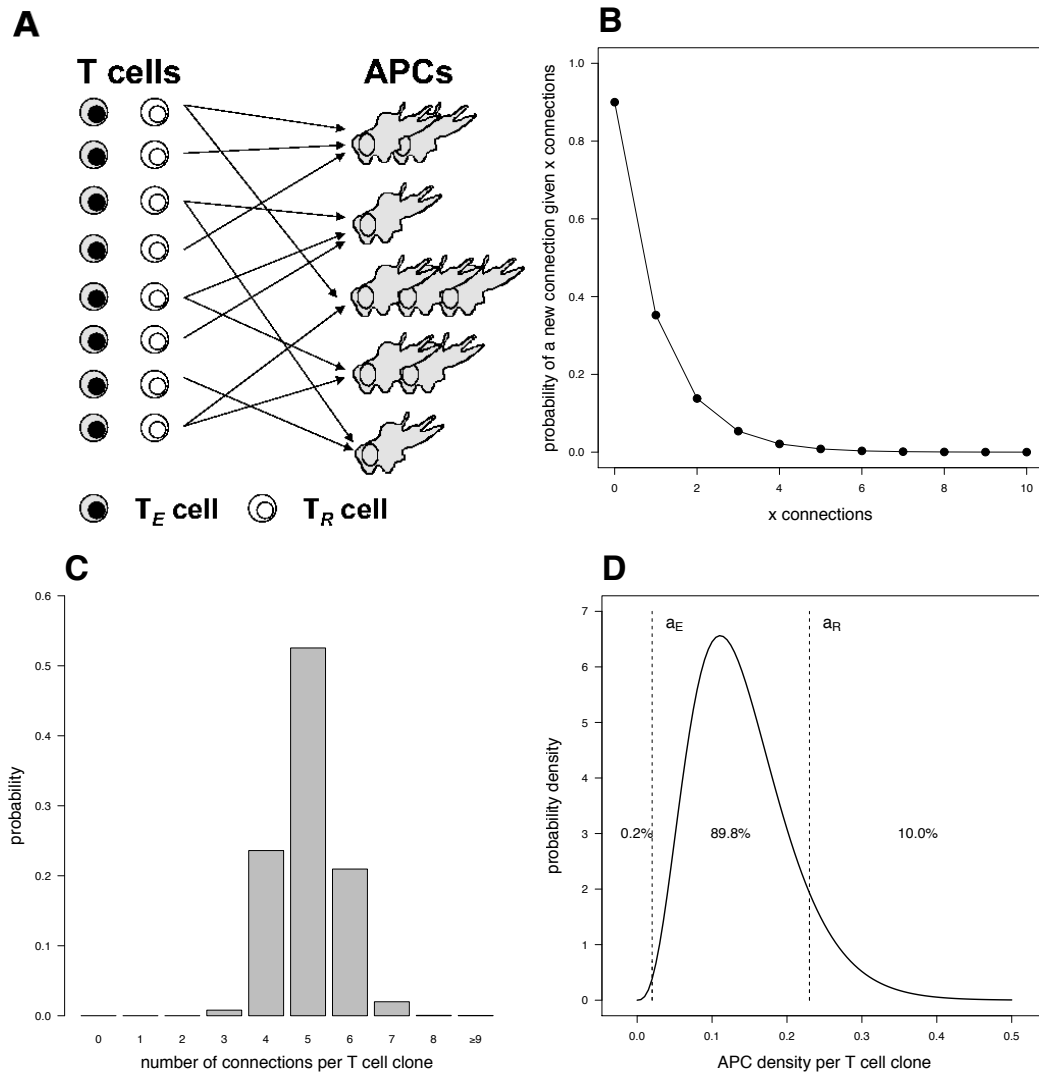


Fig. 5.3: Network between T-cell clones and APCs. **A.** T cell clones recognize a different subset of APCs; **B.** Probability of having a new connection between an APC subpopulation and a T-cell clone given that a T-cell clone has already x connection with APCs (that is, it recognizes already x APC populations); **C.** Probability distribution of the number of connections per T cell clone; **D.** Probability density distribution of APC density recognized per T cell clone. Parameters were setup in order to generate 0.2% T cell clones in an extinction regime ($A_t < a_E$), 89.8% T cell clones in immunity regime ($a_E \leq A_t \leq a_R$) and 10% T cell clones in the bistable regime ($A_t > a_R$). Parameters of plots B, C and D: $\mathcal{A} = 100$, $\lambda_a = 2.8 \times 10^{-3}$, $\theta = 0.90$, and $\lambda_c = 1.07$.

ing a plateau (Figure 5.4B). Because less crossreactive T-cell clones are purged from the repertoire, T-cell diversity decreases in time (Figure 5.4C), more pronounced in the beginning of the simulation, where the difference between T-cell clones with low and high level of cross-reactivity is larger, and less pronounced in the end of the simulation, where this difference is smaller. In terms of T-cell density, the respective steady-state is reached very early in the simulation (Figure 5.4D), maintaining it even when T-cell diversity is decreasing (Figure 5.4C).

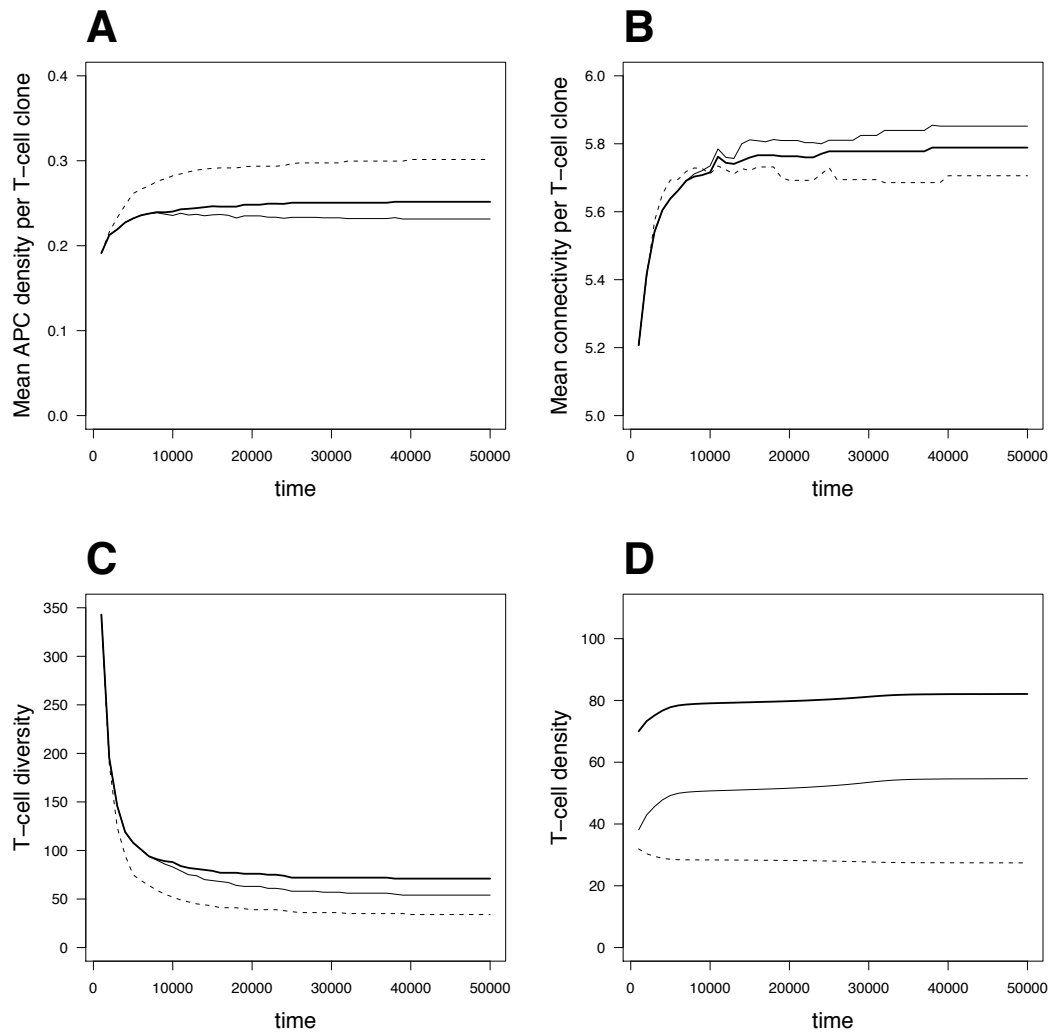


Fig. 5.4: Dynamical behavior of some variable of interest in a representative simulation without thymic export, where 500 distinct clones were generated in the beginning of the simulations, with similar initial conditions, competing for $\mathcal{A} = 100$ APC populations (thick line - T cells; solid line - T_E cells; dashed line - T_R cells; Initial condition for each clone: $t_i^* = 0$, $E_{i,0} = R_{i,0} = 0.50, \forall i = 1, \dots, 500$). Parameters λ_a and λ_c were specified in order to generate T-cell clones with mean crossreactivity of 5 APC species and a 10% probability of recognizing APC densities in the bistable regime, as illustrated in Figure 5.3 ($\lambda_a = 2.8 \times 10^{-3}$ and $\lambda_c = 1.07$). The remaining parameters are given in Table 5.1.

Partition of T-cell repertoire

We previously studied the T-cell repertoire assuming that T-cells recognize independent mutually exclusive APC subsets (Chapter 4 and Carneiro et al. 2007). In this case, if the clones are generated with similar initial conditions, the structure of the repertoire is straightforwardly predicted from the bifurcation diagram shown in Figure 5.2B. This shaping involves a partition of the repertoire in three distinct T-cell clone subsets. The first subset contains T-cell clones that eventually go extinct in the periphery, because they recognize low levels of APCs (less than a_E). The members of the second subset are T-cell clones that recognize APC densities

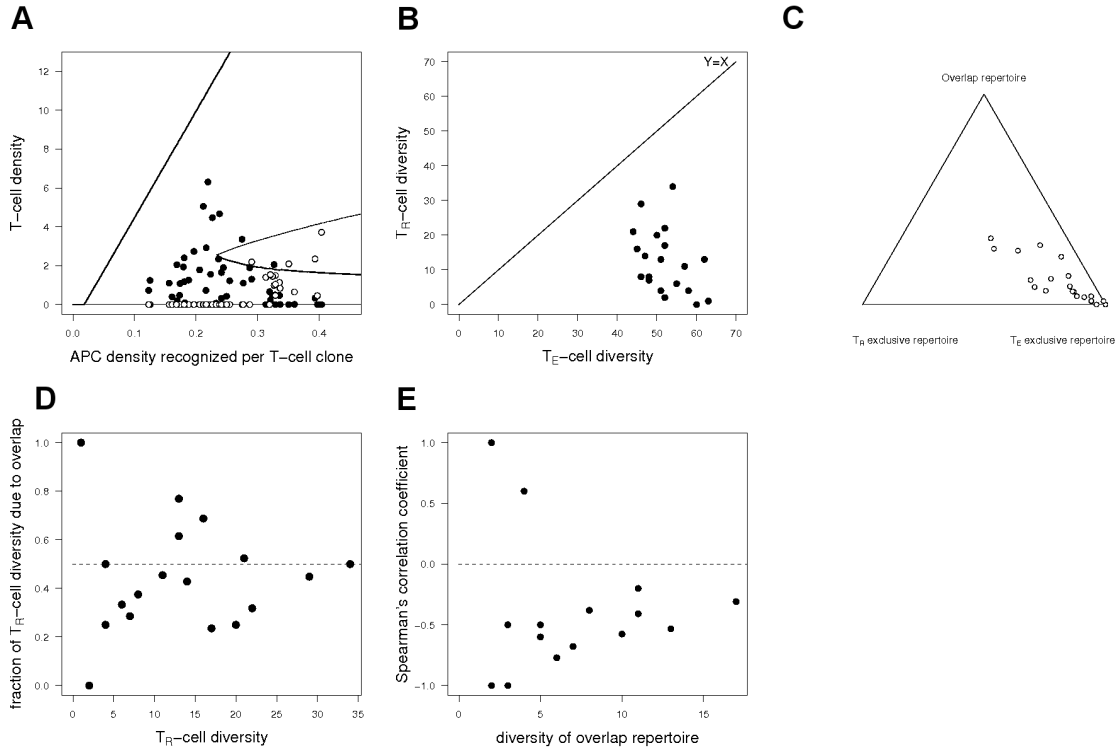


Fig. 5.5: The partition of the repertoire and some structural properties in simulations without thymic export. **A.** A representative snapshot of a T-cell repertoire at equilibrium ($t = 50000$), where solid and empty circles represent T_E and T_R clones, respectively. **B.** Relationship between T_R - and T_E -cell diversity, where the diagonal line reflects the situation where T_R - and T_E -cell diversities are equal. **C.** Ternary diagram that represent the contribution of each repertoire diversity for the total diversity. **D.** Relationship between T-cell diversities of T_R -cell exclusive and intersection repertoires as fraction of total T-cell diversity, where the diagonal line reflects the situation where T_R -cell exclusive and intersection repertoires are equally diverse. **E.** Spearman's correlation coefficient calculated between T_R and T_E -cell densities of clones in the intersection repertoire. Each dot in plots B, C, D and E represents a distinct simulation of the T-cell repertoire (20 simulations in total). The parameters of these simulations are given in Figure 5.4 and Table 5.1.

between a_E and $\tilde{a}_R$. At the equilibrium, these clones are composed of T_E cells only, because T_R cells cannot be maintained in this parameter regime. Therefore, this subset represents an exclusive T_E -cell repertoire. The third subset includes clones that recognize APCs in densities higher than $\tilde{a}_R$. Because of the specified initial conditions, T_R cells are sustained, coexisting with their T_E counterparts. Therefore, this third subset of clones is the origin of the intersection repertoire between T_E and T_R cells. In this repertoire, the average T_R -cell density is higher than that of T_E cells. Moreover, T_R and T_E -cell densities would be negatively correlated (Figure 5.2B). Since there is no parameter regime where T_R cells can be sustained alone, the previously-studied model provides no mean of forming an exclusive T_R -cell repertoire. This cannot happen because, on the one hand, T-cell clones are assumed to be independent of each other and, on the other hand, T_R -cell proliferation is strictly dependent on T_E cells with similar specificity.

At this point, we ask whether the above results are robust under interclonal competition for APCs. In particular, we ask whether a similar partition of the repertoire can be achieved or if T-cell competition is enough to form an exclusive T_R -cell repertoire. Figure 5.5 sheds some light to these questions. It is clear that, at the equilibrium, a partition of the repertoire is also obtained when T-cell clones with different specificities compete for APCs (Figure 5.5A). Since there is no thymic export in these simulations, clones lesser than a_E vanish. Moreover, since highly crossreactive T-cell clones can outcompete less crossreactive ones (Figure 5.4A and B), the effective limit for T-cell survival is higher than a_E , as shown by the respective APC density associated to the clone that recognizes less APCs. The exact value for this threshold cannot be computed because it is dependent on the connectivity structure between APCs and T cells generated (randomly) in each simulation. Similar result is obtained for the threshold that allows T_R cells to be maintained in a clone. In this case, clones recognizing APCs at densities higher than $\tilde{a}_R$ will help to maintain T_R cells in less crossreactive clones. As a consequence, the threshold for T_R -cell persistence decreases, being close to a_R , the lower APC density limit that defines the bistable regime in Figure 5.2A. It is worth noting that, in this scenario, T-cell density of each clone with a given APC density is, in equilibrium, decreased when comparing with the situation where different clones do not compete for the same APCs.

Since a similar partition is obtained in these less stringent simulations, some structural properties of T_R - and T_E -cell repertoires suggested in Carneiro et al. (2007) are maintained. Particularly, T_R cells should be less diverse than their T_R -cell counterparts (Figure 5.5B) due to the same rationale underlying the analysis of Figure 5.2B. In agreement with this observation, the diversity of T_E -cell exclusive repertoire has the largest contribution to total diversity (Figure 5.5C). Interestingly, T_R -cell repertoire can now be divided in exclusive and intersection subsets (Figure 5.5C). Therefore, in contrast to the previous model (Carneiro et al., 2007), competition between T-cell clones with different specificities led to the formation of an exclusive T_R -cell repertoire. The contribution of this repertoire to the overall T_R -cell diversity varies with the simulations, being sometimes lower, equal or higher than the respective T_R diversity of the intersection repertoire (Figure 5.5D). Finally, as predicted by Figure 5.2B, T_R and T_E -cell densities of clones belonging to the intersection repertoire tend to be negatively correlated according to Spearman's coefficient (Figure 5.5E).

5.3.2 Disentangling the T-cell repertoire in the case of a constant thymic influx

We now present the results of simulations with continuous thymic export of s_c clones per unit of time. These simulations are more complex than previous ones, and as we will see the resident and recent thymic emigrant populations in the T-cell repertoire can be devised. In this scenario, it is then expected that the repertoire of resident clones shows some of the above-mentioned structural properties exhibited in simulations without thymic export. Moreover, the repertoire of recent thymic emigrant clones will have a shape tightly close to its thymic repertoire, in our case, the initial conditions that we impose for the generation of each clone.

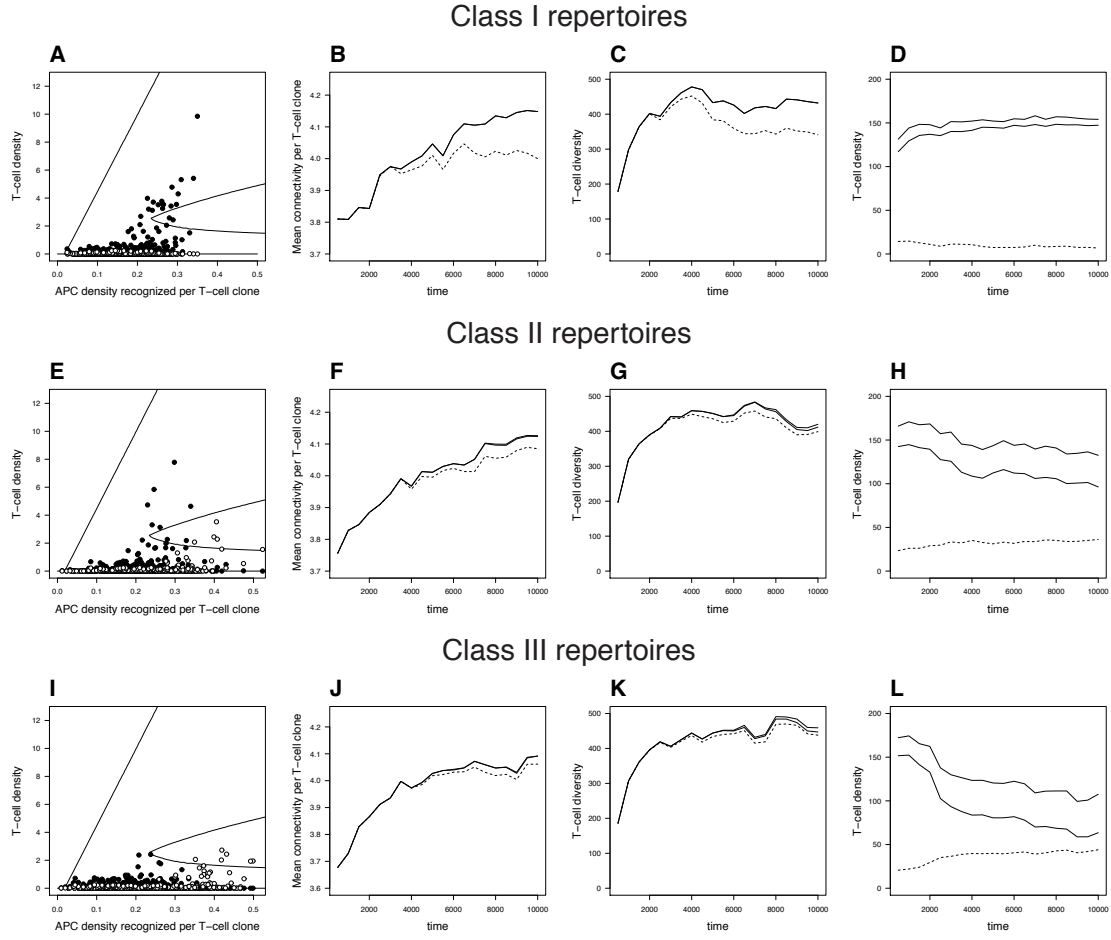


Fig. 5.6: Three distinct classes of repertoire arise in the simulations with thymic export $s_c = 0.3$ (thick line - T cells; solid line - T_E cells; dashed line - T_R cells; Initial condition for each clone: $E_{i,t_i^*} = 0.75$, $R_{i,t_i^*} = 0.25$, $\forall i$). Plots in first row refer to a phenotype where T_E -cell clones dominate the repertoire. In this scenario, T_R cells are mostly maintained by the continuous thymic export. Plots in second row reflect a repertoire showing a similar partition to the one shown in Figure 5.5A. Plots in third row refer to a phenotype where T_R cells dominate the repertoire. In this case, T_E cells are mostly maintained by the constant thymic export. Parameters λ_a and λ_c were setup in order to generate T-cell clones with a average connectivity of 3.75 APC species and a 10% probability of recognizing APC densities in the bistable regime ($\lambda_a = 3.6 \times 10^{-3}$ and $\lambda_c = 0.72$). The remaining parameters are given in Table 5.1.

Three distinct repertoire patterns arise from the simulations with continuous thymic export

Simulations with thymic export give rise to different T-cell repertoire patterns, even for the same parameter set (Figure 5.5). These patterns can be classified in three distinct classes, depending on the type of partition presented by the repertoires (Figures 5.6A, E and I).

The first class refers to repertoires dominated by T_E cells, even in the bistable parameter region (Figures 5.6A). In this scenario, T_R cells undergo extinction when reaching the periphery. Nonetheless, T_R -cell mean connectivity, density, and diversity increase in the initial phase of the simulations until reaching a plateau (Figures 5.6B, C and D). This is explained by the fact that, even that all T_R cells eventually become extinct, T_R -cell clones that recognize low

levels of APC will vanish faster than the ones recognizing higher levels of APCs, giving an apparent increase in these variables. Therefore, in this class of repertoire, T_R -cell repertoire is crucially maintained by the thymic output. Its shape is similar to that of the recent thymic emigrant population that in its turn, is given by the initial conditions by which every clone is generated in the simulations. Because T_R -cell clones cannot be sustained much longer in the system, this class of repertoire might be interpreted as a massive, or systematic, autoimmune disease, as observed in Foxp3-deficient mice (Brunkow et al., 2001).

Simulations in the second class show a partition similar to that shown in Figure 5.5A. Therefore, the repertoire is partitioned in three subsets, even though recent thymic emigrants are present in all of them (Figure 5.6E). The first subset comprises T-cell clones that recognize low levels of APCs and, thus, undergo extinction when entering periphery. This subset is maintained by thymic export. The second subset refers to T-cell clones that recognize intermediate levels of APCs and form the T_E -cell exclusive repertoire. The third subset encompasses T-cell clones that recognize high levels of APCs. In most cases, these clones are regulated by T_R cells. Occasionally, this third subset might include some clones devoided of T_R cells, which might be interpreted as an organ-specific autoimmune disease. As for previous class, mean connectivity increases in time (Figure 5.6F), but now due to a highly crossreactive T_R and T_E -cell resident population. Therefore, mean connectivity of both cell types increases not only due to elimination of less crossreactive clones from the repertoire, but also to the formation of a T_R and T_E resident pool with increasingly high crossreactivity. As in Figure 5.4C, T_E -cell diversity is higher than that of T_R cells (Figures 5.6G). Interestingly, T-cell density reaches its peak very early in the simulation but after that, with the maintenance of a resident T_R -cell pool, decreases slowly in time (Figures 5.6H), eventually reaching a steady state determined by an overall regulation of all resident clones by T_R cells. It is worth noting that this decrease in T-cell density is accompanied by a slow increase on T_R -cell proportion. This is due to an increase of the resident T_R -cell pool throughout the simulations.

The third class is when the repertoires are mostly dominated by T_R cells (Figures 5.6I and L). Therefore, these repertoires can be interpreted as a situation where the immune responses are globally regulated by T_R cells. It is then expected that individuals with this type of repertoire would not show any symptoms of autoimmunity but more susceptible to infections. Structurally, the repertoire is mainly composed of a recent thymic emigrant population and a resident one, mostly of T_R cells (Figure 5.6I). Few T_E -cell clones are maintained in the intermediate APC density region. However, T_R cells need to receive a growth factor from T_E cells and, thus, they persist in the periphery due to the source of T_E cells coming from the thymus. Since T-cell clones are export to the periphery with both cell types, T_R -cell diversity tends to be close to that of T_E cells (Figure 5.6K). In some extreme cases, T_R -cell diversity can be even higher (simulations not shown). Finally, as in the previous class, the evolution of total T-cell density has a biphasic behavior: it reaches a peak early in the simulations, followed by an abrupt drop. This drop in T-cell density correlates with the overtaking of the repertoire by T_R -cell clones (Figure 5.6L).

At this point, it is worth noting that the first and third classes of repertoires imply deleterious consequences for the host. On the one extreme, the first class is characterized by

the overdominance of T_E cells and, because of that, immune responses against self antigens would not be under control, generating eventually systemic autoimmune pathologies. On the other extreme, the third class shows a predominance of T_R cells, which implies that immune responses would be regulated by these cells. Therefore, the intermediate second class seems the most adequate to describe a T-cell repertoire of a healthy immune system, as suggested in Carneiro et al. (2007) and in Chapter 4. Yet, there is a pay to price in this class: some individuals might show organ-specific autoimmune pathologies due to the putative presence of T_E -cell clones without regulation in the parameter regime that allow T_R -cell maintenance in the system.

The three repertoire classes are distinguished by the shape of their resident populations

We next characterize the structural diversity of above three classes of repertoires. As mentioned above, the repertoires are composed of a resident population and a recent thymic emigrant one. The resident population is represented by a small number of clones, as shown in Figures 5.6A, E and I, where only a few clones are above the initial values for the generation of T_R and T_E cells in each clone. Nevertheless, it is difficult to define exactly which clones form effectively the resident population. For a matter of simplicity, the 25 oldest clones in the simulations are considered representative of the resident population. In contrast, recent thymic emigrant clones represent the majority of the repertoire, as also seen in Figures 5.6A, E and I by the large number of clones concentrated around the initial conditions. It is worth noting that most of these recent thymic emigrant clones will ultimately vanish.

Because of this overdominance of recent thymic emigrant clones, all three repertoire classes show similar shape when the whole repertoire is analyzed (Figures 5.7A, B and C). That is, T_E -cell diversity tends to be higher than that of T_R cells but with no clear distinction between different repertoire structures (Figure 5.7A). The same happens when analyzing the composition of the repertoire (Figure 5.7B). As expected, the main fraction of the repertoires is due to the overlap between T_R and T_E recent thymic emigrant clones. This implies a positive correlation between T_R and T_E -cell clones belonging to the intersection repertoire (Figure 5.7C).

As expected, significant structural differences in the resident population are observed in the three repertoire classes. In fact, the first class has a resident population with a large difference between T_R and T_E -cell diversities (Figure 5.7D). This is a consequence of having a resident repertoire mainly composed of a T_E -cell exclusive repertoire. The second class of repertoire has an intermediate difference between T_R and T_E -cell diversities. Using ternary diagrams, which allows to represent the composition of T-cell diversity in triangle, simulations of this repertoire class are positioned almost in the center region of the triangle, which means that exclusive T_R -cell and T_E -cell diversities and the respective shared diversity contribute almost equally to the overall T-cell diversity (Figure 5.7E). Since the third class of repertoires is characterized by an overdominance of T_R cells, the respective resident repertoire tends to exhibit similar or higher diversity of those cells (Figure 5.7D). The intersection repertoire contributes then greatly to the overall T-cell diversity of the resident repertoire, as opposed to T_R -cell exclusive repertoire (Figure 5.7E). Notwithstanding, every repertoire class agrees

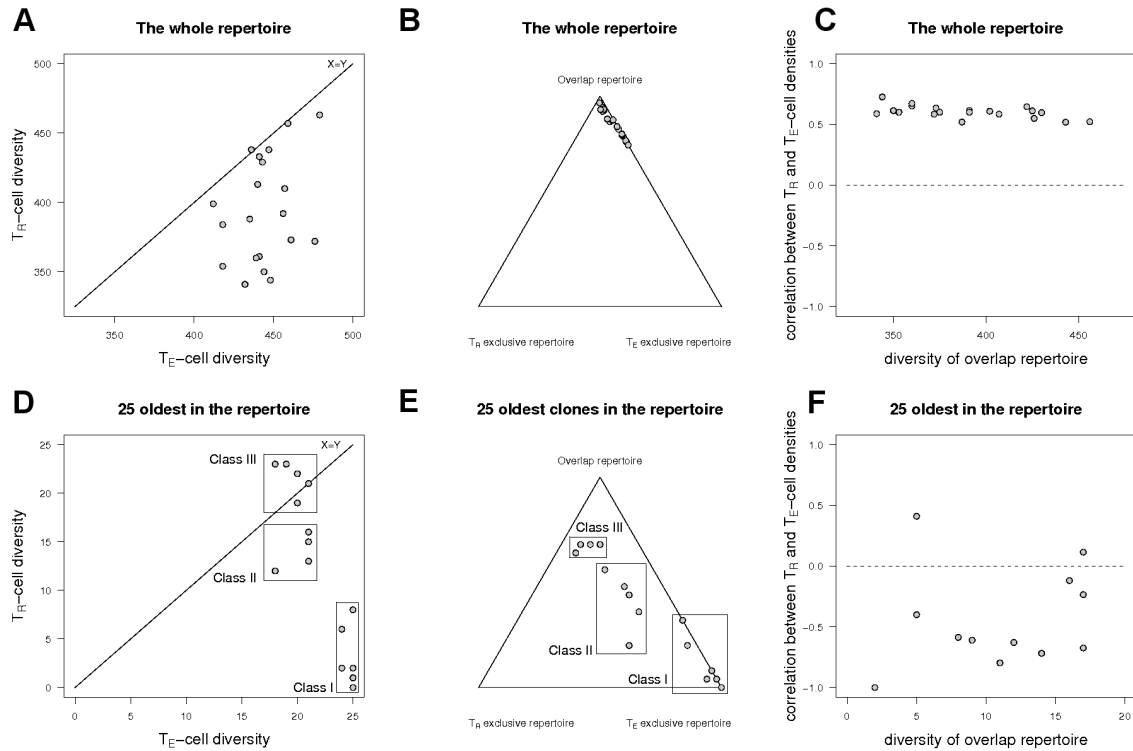


Fig. 5.7: Three repertoire classes are distinguished by structural properties of the resident populations. Plots A, B, and C are for the whole repertoire, while plots D, E, and F are for the resident population, defined by the 25 oldest clones in the repertoire. Each dot in the plots represents a distinct simulation (20 simulations in total). See Figure 5.6 for further details.

with a higher probability of having a negative correlation between T_R and T_E -cell densities of clones in the intersection repertoire. A positive correlation can also be obtained when some of the clones in the resident population undergo extinction due to the appearance of more-fitted T-cell clones in the simulations, which eventually replace them in the respective population.

5.3.3 Effect of some parameters in shaping the repertoires

We then studied the effect of the average connectivity by which the clones are initially generated, the APC diversity in the body, and the number of clones exported by the thymus in shaping the peripheral T-cell repertoire (Figure 5.8). We pay special attention to the probability of obtaining each of the above partition classes, the diversity of each T-cell repertoire components, and the diversity composition of the resident population. To calculate the proportion of each partition class for a given parameter set, we visually classified the repertoires according to the diversity composition exhibited by their 25 oldest clones, as illustrated in Figure 5.7E. As previously, we evaluated all simulations at time $t = 10000$.

Given a particular value for the APC diversity and thymic export, increasing the average connectivity by which the clones enter the periphery has an effect of producing more repertoires of class III than those of the remaining classes (Figure 5.8A). Because class III repertoires are dominated by T_R cells, overall T-cell diversity decreases with the parameter at hand

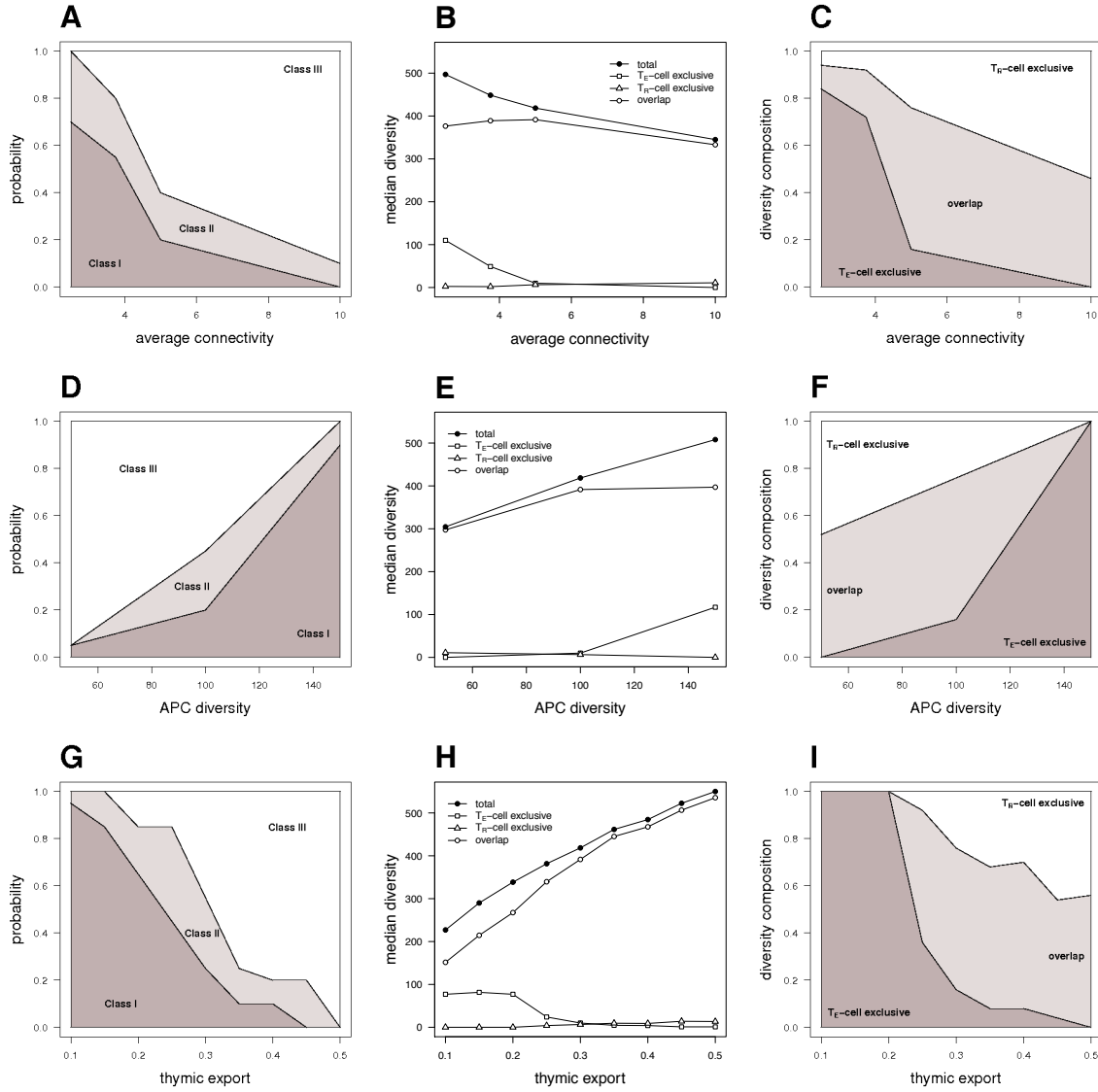


Fig. 5.8: Effect of some parameters on the structure of the repertoire. First, second and third rows refer to the effect of mean connectivity (by which each clone are initially generated), APC diversity, and thymic export, respectively. First, second, and third column refer to the probability of having each partition class, the diversity of each T-cell repertoire, and diversity composition for the 25 oldest clones in the repertoires. Simulations were evaluated at time $t = 10000$. Parameters: $s_c = 0.30$ and $\mathcal{A} = 100$ (plots **A**, **B**, and **C**); $s_c = 0.30$, $\lambda_a = 2.08 \times 10^{-3}$ and $\lambda_c = 1.07$ (plots **D**, **E**, and **F**; thymic average connectivity of 5 APC populations); $\mathcal{A} = 100$, $\lambda_a = 2.08 \times 10^{-3}$ and $\lambda_c = 1.07$ (plots **G**, **H**, and **I**; thymic average connectivity of 5 APC populations). The parameter λ_a was calculated in order to generate 10% of clones in the bistable regime. For each parameter set, we performed 20 simulations.

(Figure 5.8B). This is accomplished by a higher contribution of the resident T_R -cell diversity to the overall diversity of the resident population (Figure 5.8C). In fact, a higher average connectivity generated for the clones tend to increase the competition between different clones in the beginning of the simulations. This competition would then decrease T_E -cell density of highly crossreactive clones, allowing T_R cells coming later from the thymus to be sustained in the system.

For a given connectivity parameter set and thymic export, a higher APC diversity increases the overall T-cell diversity (Figure 5.8E). This is just a consequence of a higher combination of different APCs that can be recognized by the T-cell clones. In theory, this parameter when increased should decrease transiently interclonal competition, unless compensated by an increase of the thymic export. T_E cells would then benefit of this diminished interclonal competition in the beginning of the simulation to proliferate, which would "block" T_R -cell sustainability in bistable regime. Therefore, individuals with a higher number of APC species but with similar thymic output tend to exhibit repertoires of class I (Figure 5.8D). For this reason, diversity of T_E -cell exclusive repertoire tend also to increase with this parameter (Figure 5.8E) as well as its contribution to the overall diversity of the resident population (Figure 5.8F).

As with the number of APC species, overall T-cell diversity is also positively correlated with the thymic export (Figure 5.8H). This happens because there is a larger flux of diversity coming from the recent thymic emigrant population. Because clones enter the periphery with both T_R and T_E cells, diversity of the intersection repertoire also increases with thymic export (Figure 5.8H). In theory, a higher thymic export has two major effects. On the one hand, it increases the chance of generating highly crossreactive clones. On the other hand, competition between different clones for APCs is increased in the beginning of the simulations, which decreases T_E -cell density of highly crossreactive clones, allowing T_R cells coming later from the thymus to be sustained in the system. Because of these effects, high values for thymic export are prone to produce repertoires of class III (Figure 5.8G), exhibiting a high contribution of T_R -cell diversity in the resident population (Figure 5.8I).

5.4 Discussion

We here report the results of simulations of peripheral T-cell repertoires with and without thymic export where T-cell clones and APCs show an intricate network of interactions. Previously, we studied a situation where the peripheral T-cell repertoire was composed of independent clones, each one recognizing an exclusive set of APCs (Carneiro et al., 2007). In this early work, we showed that T_R cells are always maintained in a clone with their T_E -cell counterparts. A T_R -cell exclusive repertoire seemed then impossible to be generated under this simple model. To overcome this apparent limitation of the model, each T-cell clone in the repertoire was interpreted as an independent functional class that exhibiting similar TCR specificity but might contain distinct T_R -cell exclusive receptors. In this line of thought, a T_R -cell exclusive repertoire was implicitly included in the model.

In this work, we show that shared specificity between different clones is enough to originate a T_R -cell exclusive repertoire. Shared specificity is then a sufficient but not a necessary condition for the generation of a T_R -cell exclusive repertoire, since thymic T_R -cell ontogeny might do it in first place. In fact, it was observed a small overlap between thymic T_R and T_E TCR samples in recent studies on animal models with limited TCR diversity (Hsieh et al., 2006; Pacholczyk et al., 2006). At a first glance, this observation suggests that thymus is actually able to seed the periphery with a T_R -cell exclusive repertoire. Yet, a small overlap in the samples does not necessarily imply similar interpretation for the whole repertoire of an individual, as

discussed in depth later in Chapter 7. A small overlap between pairs of samples can be solely due to either small or differential sample sizes. In fact, Venturi et al. (2008) has alluded to this problem when studying the evolution of TCR repertoire of an individual during an influenza infection, and provided a method to standardize the calculations. Similar small overlap was obtained between peripheral T_R and T_E TCR samples, suggesting the existence of a peripheral T_R -cell exclusive repertoire (Hsieh et al., 2004; Pacholczyk et al., 2006). Yet, the above comments also hold in this case. We then conclude that current statistical analysis of available data is not sufficiently accurate to determine whether or not exists an exclusive T_R -cell repertoire.

Notwithstanding, our simulations with thymus exporting clones with both cell types suggest that the peripheral reshaping of the repertoire would originate a small T_R -cell exclusive repertoire, irrespective of the repertoire structure prototype studied here (Figure 5.7E). Our result can be further extended to the more general situation where thymus exports T_R -cell exclusive specificities. In fact, we demonstrate that peripheral reshaping of the repertoire would purge those specificities that are less crossreactive, maintaining only the highly crossreactive ones. At this point, it is worth noting that current models of thymic T_R -cell ontogeny agree that thymic T_R cells are focused on high crossreactivities towards self-antigens (Modigliani et al., 1996). In this line of thought, thymic T_R -cell generation seems optimize to putting out to the periphery T_R -cell clones that, in theory, would have a higher chance to survive to peripheral reshaping.

Three distinct repertoire structures have emerged in our simulations. On the one extreme, there is a repertoire structure with an overdominance of T_E cells (class I repertoires), which can be interpreted as the devastating systemic autoimmune pathology observed in Foxp3-knockout mice. On the other extreme, there is a repertoire structure in which T_R cells are highly represented (class III repertoires). In spite of assuring tolerance to self antigens, this type of repertoire might not be totally inefficient to fight infections due to the overdominance of T_R cells in the repertoire. In this scenario, this type of repertoire would benefit mild and moderate infections that often appear throughout life. Those infections would then become chronic due the inhibition of proper immune responses by (specific) T_R cells. Notwithstanding, strong infections might break the global regulation state of the repertoire, as demonstrated in Leon et al (León et al., 2004). The last repertoire structure lies in the middle of the above two extremes, showing a similar shape to that discussed in Carneiro et al. (2007) and obtained in simulations without thymic export (class II repertoire). Its shape is intimately to a partition of the repertoire in three subsets of clones: i. a small resident subset avoiding autoimmune responses by the presence of T_R cells; ii. a relatively large subset lacking T_R cells that allows T_E cells to respond to infrequent foreign antigens; iii. a third subset solely maintained by the thymus that would otherwise go extinct.

As argued throughout the text, class II repertoires seem the most adequate to describe a healthy immune system. Assuming that parameters are more or less realistic, this kind of repertoire has a small chance of being generated when there is a continuous thymic export, as shown in Figure 5.8. In fact, this repertoire structure seems more a transient behaviour than a true steady state. Since T_R cells are present in the resident population of this repertoire structure, a continuous thymic export would help the system to reach asymptotically the

above-mentioned global regulation repertoire structure. This is evident by the steadily increase of the fraction of T_R cells throughout time in the simulations classified as class II repertoire structures. As a consequence, we expect that aged mice would be in general more susceptible to infections due to a higher fraction of T_R cells in the peripheral repertoire. Interestingly, T_R cells seem to be at a higher fraction in older individuals than in young ones both in humans and mice (Dejaco et al. 2006; Nishioka et al. 2006; R. Paiva and colleagues, unpublished results). For example, Nishioka et al. (2006) reported that aged C57BL/6 mice have a fraction of T_R cells in the $CD4^+$ pool around 30%, which is statistically higher than the 14% observed in young mice. Moreover, these T_R cells in older mice seem to maintain their suppressive function. In agreement with our above interpretation, it was suggested that the increased fraction of T_R cells in the peripheral repertoire might be an explanation for the increased incidence of infectious diseases in the elderly (Dejaco et al., 2006; Nishioka et al., 2006).

At this point, it is important to note that thymic output might be controlled in order to maintain the repertoire as much time as possible in its healthy structure. Since it was more easy to obtain such repertoire structures in simulations without thymic export (data not shown), we speculate that thymus export should be divided in two distinct dynamical behaviours. First, periphery should be filled by a strong thymic export in the "early" days of life in order to generate T_R -cell clones that can be sustained in the system. Then, thymic export should decrease to avoid the filling of the periphery with further T_R -cell clones that, "helped" by early-generated T_R -cell clones, ultimately push the system to a global regulation repertoire structure. In fact, a decrease in thymic output throughout time is observed both in humans and mice by a mechanism called thymic involution. Yet, thymus involution is generally viewed as an undesirable process, as remarked by Dowling and Hodgkin (2009), because it is usually associated with a decline in immune function in the elderly. In the vein of the above discussion, we suggest that thymic involution might be crucial to maintain the repertoire in a healthy structure as much time as possible. Dowling and Hodgkin (2009) also argue that thymic involution is an important process shaping the T-cell repertoire in young individuals. In their view, T-cell clones exhibiting TCR close positive and negative thresholds have a shorter lifespan. In this scenario, thymus involutes to allow to peripherally select T-cell clones exhibiting TCR with affinities lying in the middle of positive and negative thresholds and, thus, have longer lifespan and are more useful for immunity.

Interestingly, a piecewise thymus export has been previously proposed but using a different argument (Modigliani et al., 1996). In this work, T_R cells against ubiquitous peptides are produced in the thymus during perinatal period until they cover every available tissue-specific antigen. Then, it is assumed that thymus stops producing T_R cells, exporting only naive T_E cells, which undergo a process of education upon encountering antigen that convert them to the T_R pool in the periphery. In this line of thought, this process of peripheral education assumes a key role in the maintained of the T_R pool in adulthood. Although peripheral education is now known to occur in the periphery (Apostolou and von Boehmer, 2004; de Lafaille et al., 2004; Thorstenson and Khoruts, 2001; Zelenay et al., 2005), the exact contribution in shaping the peripheral repertoire is still unclear, because current estimates for the fraction of peripherally-converted T_R cells are disparate across different experimental settings. Moreover, T_R cells are known to be continuously produced in the thymus (reviewed in Annacker et al. 2001).

It is well known that autoimmune diseases are affected by multiple genetic and environmental factors, which lead to a complex inheritance pattern (Gregersen and Behrens, 2006). Complexity in these diseases arise from the fact that most of them show incomplete penetrance, that is, only a proportion of individuals exhibiting risk genotypes will manifest disease. Traditionally, one invokes environmental factors to explain this observation, such as the occurrence of certain infections. Interestingly, twins maintained in the same environment do not show a 100% concordance rate for disease status. Moreover, different autoimmune-disease animal models show also incomplete penetrance, even under controlled lab environmental conditions. This is particularly evident in the NOD mouse strain that develops spontaneously type I diabetes with incomplete penetrance both in males and females (Anderson and Bluestone, 2005). To explain this phenomenon, we and others have previously argued that a fraction of incomplete penetrance could be attributed to an internal stochastic component governing the expression of the phenotype, such as a stochastic allelic expression (Alper and Awdeh, 2000; Alper et al., 2006; Sepúlveda et al., 2007, 2009). In the case of autoimmune diseases, we invoked the stochasticity in the T-cell repertoire generation as a putative explanation for this internal component of penetrance (Sepúlveda et al., 2009), following the observations in monozygotic twins when studying different autoimmune diseases (Haegert et al., 2003; Hohler et al., 1999). In this line of discussion, the simulations here performed showed that the three above-mentioned repertoire structures can arise in simulations with the same parameter set (Figure 5.8). If one interprets these simulations as different individuals with the same genetics, they indicate that the stochasticity in the T-cell repertoire formation is indeed an important factor in the understanding of incomplete penetrance in the context of complex autoimmune diseases. Recently, a study reported a lesser T_R -cell TCR diversity in NOD mice than in non-diabetic C57BL6 mice, suggesting important differences between the T-cell repertoires of these mouse models (Ferreira et al., 2009). However, the comparison of TCR diversity between unaffected and affected NOD mice was not performed, which would confirm or reject the hypothesized stochastic role of the T-cell repertoire formation in incomplete penetrance. In this scenario, we expect to observe differences in the T-cell repertoire between unaffected and affected NOD mice.

At this point, it is worth noting that the basic Crossregulation model is mathematically equivalent to a model that accounts for peripheral education (León et al., 2001). Therefore, the results here presented can help to refine the above model. The first stage of a high thymic export is then not only important to fill the periphery with T_R -cell clones that assure peripheral tolerance, but also to warrant they can be sustained in the system. The clones that would be maintained in this phase are expected to be highly crossreactive, possibly covering all autoantigen space. The second stage of a low thymic export, which might include generation of T_R cells, is crucially to avoid the repertoire be overtaken by T_R cells. It also assures that thymus does not provoke a big impact on the normal shape of the peripheral repertoire, maintaining it in a "healthy" status.

The long-term outcome of our simulations was somehow dependent on the early history of the repertoire formation. In this regard, if T_E -cell clones take over the repertoire in first place, it would then be difficult to maintain later-generated T_R -cell clones in the repertoire. This would

lead to a class I repertoire with its deleterious consequences to the host. In contrast, if early-generated T_R -cell clones could be sustained in the system, they would help later-generated ones to be kept in the repertoire, giving rise to a class III repertoire. This dependence on the early history of repertoire formation seems in close agreement with early studies on neonatally thymectomized animals (Asano et al., 1996), where thymus ablation between 2 and 4 days of life would only lead to an autoimmune condition in adult. This short time window for thymic ablation suggests that very early events of repertoire formation are crucial to determine the respective long-term outcome. In this line of thought, the experimental observation that thymic export decreases throughout life by a process called thymic involution would then be beneficial to the host, because it would help the repertoire to fall and be maintained in a class II structure.

Our simulations advance simple predictions for some structural properties of the peripheral repertoire. These predictions are distinct for each repertoire class, namely, in what concerns to the resident T-cell population. For instances, class I repertoires should be exhibited a higher diversity of resident T_E cells than that of T_R cells, and the overlap between the respective resident repertoires should be very small. Class II repertoires also predict a higher diversity of resident T_E cells but the difference between T_E and T_R -cell diversity should be smaller than in previous case. In their turn, class III repertoires show comparable diversity between T_R and T_E cells. Besides these predictions, our simulations suggest that the clonal size distributions of T_R and T_E -cell specificities belonging to the intersection repertoire tend to be negatively correlated.

Since the above repertoire classes are fairly distinguished by their resident population, a way to discriminate them experimentally is to study the evolution of the T-cell repertoire throughout time. In theory, one can detect the resident population by the TCR specificities that appear consistently in different time points. Notwithstanding, current peripheral T_R -cell repertoire studies just take a snapshot of the repertoire at a given time point (Hsieh et al., 2004; Lathrop et al., 2008; Pacholczyk et al., 2006, 2007; Wong et al., 2007), which might obscure the underlying complexities (Magurran and Henderson, 2003). These studies show T_R TCR samples more diverse than those from their T_E counterparts, which prompted the authors to take similar conclusion for the whole repertoire of an individual (Hsieh et al., 2004; Pacholczyk et al., 2006, 2007; Wong et al., 2007). Yet, we demonstrate that this conclusion might not hold (Sepúlveda et al. 2008 and in Chapter 7 of this thesis). In fact, when re-analyzing data from Hsieh et al. (2004, 2006), we could not reject the hypothesis of equal diversity between peripheral T_R and T_E cells at the level of the whole peripheral repertoire. Since there is no information available on the resident T-cell population, the equal diversity of peripheral T_R and T_E cells might be just due to a situation where the thymus exports preferentially clones with both cell types, as in simulations performed here. If this holds true, nothing can be said about the underlying repertoire structure, as illustrated in Figures 5.7A and B. In fact, calculations given in Chapter 7 will demonstrate that available data of Hsieh and colleagues (Hsieh et al., 2004, 2006) could not rule out a complete intersection between thymic T_R and T_E -cell repertoires and, thus, little information about the underlying structure can be extracted by a snapshot of the repertoire. In the light of these results, we conclude that current T_R -cell studies need to be redesigned in order to gain more insight of T_R -cell repertoire structure.

We then conclude that current model is not only useful to provide interpretations for the available data but also to suggest alterations in current experimental designs. In this regard, we recommend that future TCR repertoire studies should include samples taken in different time points. Using proper statistical analysis, one can confront the predictions of the model with the data.

Bibliography

- Alper, C. A. and Awdeh, Z. (2000). Incomplete penetrance of MHC susceptibility genes. *Tissue Antigens*, 56:199–206.
- Alper, C. A., Husain, Z., Larsen, C. E., Dubey, D. P., Stein, R., Day, C., Baker, A., Beyan, H., Hawa, M., Ola, T. O., and Leslie, R. D. (2006). Incomplete penetrance of susceptibility genes for MHC-determined immunoglobulin deficiencies in monozygotic twins discordant for type diabetes. *J. Autoimmun.*, 27:89–95.
- Anderson, M. S. and Bluestone, J. A. (2005). The NOD mouse: a model of immune dysregulation. *Annu. Rev. Immunol.*, 23:447–485.
- Annacker, O., Pimenta-Araujo, R., Burlen-Defranoux, O., and Bandeira, A. (2001). On the ontogeny and physiology of regulatory T cells. *Immunol. Rev.*, 182:5–17.
- Apostolou, I. and von Boehmer, H. (2004). In vivo instruction of suppressor commitment in naive T cells. *J. Exp. Med.*, 199:1401–1408.
- Asano, M., Toda, M., Sakaguchi, N., and Sakaguchi, S. (1996). Autoimmune disease as a consequence of developmental abnormality of a T cell subpopulation. *J. Exp. Med.*, 184:387–396.
- Belkaid, Y. (2007). Regulatory T cells and infection: a dangerous necessity. *Nat Rev Immunol*, 7:875–888.
- Brunkow, M. E., Jeffery, E. W., Hjerrild, K. A., Paeper, B., Clark, L. B., Yasayko, S. A., Wilkinson, J. E., Galas, D., Ziegler, S. F., and Ramsdell, F. (2001). Disruption of a new forkhead/winged-helix protein, scurf, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nat. Genet.*, 27:68–73.
- Carneiro, J., Coutinho, A., Faro, J., and Stewart, J. (1996). A model of the immune network with B-T cell cooperation I - Prototypical structures and dynamics. *J. Theor. Biol.*, 182:513–529.
- Carneiro, J., Léon, K., Caramalho, I., van den Dool, C., Gardner, R., Oliveira, V., Bergman, M.-L., Sepúlveda, N., Paixão, T., Faro, J., and Demengeot, J. (2007). When three is not a crowd: a Crossregulation model of the dynamics and repertoire selection of regulatory CD4+ T cells. *Immunol. Rev.*, 216:48–68.
- Carneiro, J., Paixão, T., Milutinovic, D., Sousa, J., Leon, K., Gardner, R., and Faro, J. (2005). Immunological Self-Tolerance: Lessons from Mathematical Modeling. *J. Comp. Appl. Math.*, 184:77–100.

- de Boer, R. J. and Perelson, A. S. (1994). T cell repertoires and competitive exclusion. *J. Theor. Biol.*, 169:375–390.
- de Boer, R. J. and Perelson, A. S. (1997). Competitive control of the self-renewing T cell repertoire. *Int. Immunol.*, 9:779–790.
- de Lafaille, M. A. C., Lino, A. C., Kutchukhidze, N., and Lafaille, J. J. (2004). CD25⁺ T cells generate CD25⁺Foxp3⁺ regulatory T cells by peripheral expansion. *J. Immunol.*, 173:7259–7268.
- Dejaco, C., Duftner, C., and Schirmer, M. (2006). Are regulatory T cells linked with aging? *Exp. Gerontol.*, 41:339–345.
- Dowling, M. R. and Hodgkin, P. D. (2009). Why does the thymus involute? A selection-based hypothesis. *Trends Immunol.*, 30:295–300.
- Ferreira, C., Singha, Y., Furmanskia, A. L., Wong, F. S., Gardena, O. A., and Dyson, J. (2009). Non-obese diabetic mice select a low-diversity repertoire of natural regulatory T cells. *Proc. Natl. Acad. Sci. USA*, 106:8320–8325.
- Fontenot, J. D., Gavin, M. A., and Rudensky, A. Y. (2003). Foxp3 programs the development and function of CD4⁺CD25⁺ regulatory T cells. *Nat. Immunol.*, 4:330–336.
- Freitas, A. A. and Rocha, B. (2000). Population biology of lymphocytes: the flight for survival. *Annu. Rev. Immunol.*, 18:83–111.
- Gregersen, P. K. and Behrens, T. W. (2006). Genetics of autoimmune diseases — disorders of immune homeostasis. *Nat. Rev. Genet.*, 7:917–928.
- Haegert, D. G., Galutira, D., Murray, T. J., O'Connor, P., and Gadag, V. (2003). Identical twins discordant for multiple sclerosis have a shift in their t-cell receptor repertoire. *Clin. Exp. Immunol.*, 134:532–537.
- Hohler, T., Hug, R., Schneider, P. M., Krummenauer, F., Gripenberg-Lerche, C., Granfors, K., and Merker-Hermann, E. (1999). Ankylosing spondylitis in monozygotic twins: studies on immunological parameters. *Ann. Rheum. Dis.*, 58:435–440.
- Hori, S., Nomura, T., and Sakaguchi, S. (2003). Control of regulatory T cell development by the transcription factor Foxp3. *Science*, 299:1057–1061.
- Hsieh, C.-S., Liang, Y., Tyznik, A. J., Self, S. G., Liggitt, D., and Rudensky, A. Y. (2004). Recognition of the peripheral self by naturally arising CD25⁺ CD4⁺ T cell receptors. *Immunity*, 21:267–277.
- Hsieh, C.-S., Zheng, Y., Liang, Y., Fontenot, J. D., and Rudensky, A. Y. (2006). An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nat. Immunol.*, 7:401–410.

- Lathrop, S. K., Santacruz, N. A., Pham, D., Luo, J., and Hsieh, C.-S. (2008). Antigen-specific peripheral shaping of the natural regulatory T cell population. *J. Exp. Med.*, 205:3105–3117.
- León, K., Faro, J., Lage, A., and Carneiro, J. (2004). Inverse correlation between the incidences of autoimmune disease and infection predicted by a model of T cell mediated tolerance. *J. Autoimmun.*, 22:31–42.
- León, K., Garcia, K., Carneiro, J., and Lage, A. (2007). How regulatory CD25+CD4+ T cells impinge on tumor immunobiology: the differential response of tumors to therapies. *J. Immunol.*, 179:5659–5668.
- León, K., Lage, A., and Carneiro, J. (2003). Tolerance and immunity in a mathematical model of T-cell mediated suppression. *J. Theor. Biol.*, 225:107–126.
- León, K., Pérez, R., Lage, A., and Carneiro, J. (2000). Modelling T-cell-mediated suppression dependent on interactions in multicellular conjugates. *J. Theor. Biol.*, 207:231–254.
- León, K., Pérez, R., Lage, A., and Carneiro, J. (2001). Three-cell interactions in T cell-mediated suppression? A mathematical analysis of its quantitative implications. *J. Immunol.*, 166:5356–5365.
- Magurran, A. E. and Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, 422:714–716.
- Mason, D. (1998). A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol. Today*, 19:395–404.
- Mills, K. H. G. (2004). Regulatory T cells: friend or foe in immunity to infection? *Nat. Rev. Immunol.*, 4:841–855.
- Modigliani, Y., Bandeira, A., and Coutinho, A. (1996). A model for developmentally acquired thymus-dependent tolerance to central and peripheral antigens. *Immunol. Rev.*, 149:155–200.
- Nishioka, T., Shimizu, J., Yamazaki, S., and Sakaguchi, S. (2006). CD4⁺CD25⁺Foxp3⁺ T cells and CD4⁺CD25[−]Foxp3⁺ T cells in aged mice. *J. Immunol.*, 176:6586–6593.
- Pacholczyk, R., Ignatowicz, H., Kraj, P., and Ignatowicz, L. (2006). Origin and T cell receptor diversity of Foxp3⁺CD4⁺CD25⁺ T cells. *Immunity*, 25:249–259.
- Pacholczyk, R., Kern, J., Singh, N., Iwashima, M., Kraj, P., and Ignatowicz, L. (2007). Nonself-antigens are the cognate specificities of Foxp3⁺ regulatory T cells. *Immunity*, 27:493–504.
- Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M., and Toda, M. (1995). Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol.*, 155:1151–1164.

- Sakaguchi, S., Sakaguchi, N., Shimizu, J., Yamazaki, S., Sakihama, T., Itoh, M., Kuniyasu, Y., Nomura, T., Toda, M., and Takahashi, T. (2001). Immunologic tolerance maintained by CD25⁺ CD4⁺ regulatory T cells: their common role in controlling autoimmunity, tumor immunity, and transplantation tolerance. *Immunol. Rev.*, 182:18–32.
- Sakaguchi, S., Wing, K., and Miyara, M. (2007). Regulatory T cells - a brief history and perspective. *Eur. J. Immunol.*, 37 Suppl 1:S116–S123.
- Sepúlveda, N., Paulino, C. D., and Carneiro, J. (2008). Diversity of regulatory and effector T lymphocytes (in portuguese). In Hill, M. M., Ferreira, M. A., Dias, J. G., Salgueiro, M. F., Carvalho, H., Vicente, P., and Braumann, C. A., editors, *Estatística: da Teoria à Prática*, pages 513–523, SPE editions, Lisboa.
- Sepúlveda, N., Paulino, C. D., Carneiro, J., and Penha-Gonçalves, C. (2007). Allelic penetrance approach as a tool to model two-locus interaction in complex binary traits. *Heredity*, 99:173–184.
- Sepúlveda, N., Paulino, C. D., and Penha-Gonçalves, C. (2009). Bayesian analysis of allelic penetrance models for complex binary traits. *Comp. Stat. Data Anal.*, 53:1271–1283.
- Stirk, E. R., Molina-París, C., and van den Berg, H. A. (2008). Stochastic niche structure and diversity maintenance in the T cell repertoire. *J. Theor. Biol.*, 255:237–249.
- Thorstenson, K. M. and Khoruts, A. (2001). Generation of anergic and potentially immunoregulatory CD25⁺CD4⁺ T cells in vivo after induction of peripheral tolerance with intravenous or oral antigen. *J. Immunol.*, 167:188–195.
- van Santen, H.-M., Benoist, C., and Mathis, D. (2004). Number of T reg cells that differentiate does not increase upon encounter of agonist ligand on thymic epithelial cells. *J. Exp. Med.*, 200:1221–1230.
- Venturi, V., Kedzierska, K., Tanaka, M. M., Turner, S. J., Doherty, P. C., and Davenport, M. P. (2008). Method for assessing the similarity between subsets of the T cell receptor repertoire. *J. Immunol. Methods*, 329:67–80.
- Wong, J., Obst, R., Correia-Neves, M., Losyev, G., Mathis, D., and Benoist, C. (2007). Adaptation of TCR repertoires to self-peptides in regulatory and nonregulatory CD4⁺ T cells. *J. Immunol.*, 178:7032–7041.
- Zelenay, S., Lopes-Carvalho, T., Caramalho, I., Moraes-Fontes, M. F., Rebelo, M., and Demengeot, J. (2005). Foxp3⁺ CD25⁺ CD4⁺ T cells constitute a reservoir of committed regulatory cells that regain CD25 expression upon homeostatic expansion. *Proc. Natl. Acad. Sci. U.S.A.*, 102:4091–4096.

Part IV

ASSESSING THE SHAPE OF REGULATORY AND EFFECTOR T CELL REPERTOIRES WITH EXPERIMENTAL DATA

6. POISSON ABUNDANCE MODELS TO ESTIMATE T CELL RECEPTOR DIVERSITY AND CLONAL SIZE DISTRIBUTIONS

Sepúlveda, N., Paulino, C. D. and Carneiro, J. (2009). Estimation of T-cell repertoire diversity and clonal size distribution by Poisson abundance models. Submitted to *J. Immunol. Methods*.

Note: This chapter gave rise to the package PAM for R language, which is available for free download at <http://qobweb.igc.gulbenkian.pt/>.

Abstract

The answer to many fundamental questions in immunology requires an accurate estimation of the T cell receptor (TCR) diversity and the respective clonal size distribution. Hitherto, estimation of TCR diversity has been tackled either by a "standard" method that assumes an homogeneous clonal size distribution, or by non-parametric methods, such as the well-known abundance-coverage estimator. However, both methods show caveats. On one hand, many TCR data exhibit heavy-tailed distributions, which cannot cope with the assumption of an equal frequency of every TCR sequence in the population. Thus, this "standard" method produces inaccurate estimates. On the other hand, non-parametric estimators are robust in a wide range of situations, but per se provide no information about the clonal size distribution. This paper deploys Poisson abundance models from ecology to estimate TCR diversity. These models assume that each TCR variant is sampled according to a Poisson distribution with a specific sampling rate, itself distributed, such as Exponential, Gamma, Lognormal and an appropriate mixture of Exponential distributions. From these models, one can also elicit the respective clonal size distribution. A procedure is suggested to evaluate robustness of diversity estimates with respect to the most abundant sampled TCR sequences. The package *PAM* for the R software is made available to help in the fitting and estimation of the models. For illustrative purposes, previously published data on mice with limited TCR diversity are analyzed.

6.1 Introduction

T cells recognize and respond to antigens by T cell receptors (TCRs) on the cell membrane. TCRs are heterodimers with two chains: α and β in $\alpha\beta$ T cells and γ and δ in $\gamma\delta$ T cells. These receptors are generated by the V(D)J recombination process during thymic T-cell development. In this developmental step, T-cell precursors randomly recombine different V, D and J gene segments that encode each TCR chain. Additional diversity is obtained by an imprecise joining of those gene segments. In theory, there are 10^{18} different possibilities of generating an $\alpha\beta$ TCR in humans (Janeway et al., 2005) and 10^{15} in mice (Davis and Bjorkman, 1988). However,

at a given time, $\alpha\beta$ TCR diversity has been estimated to be $> 2 \times 10^7$ in humans (Arstila et al., 1999; Naylor et al., 2005) and around 2×10^6 in mice (Casrouge et al., 2000).

Many fundamental questions in Immunology are intimately related to T cell repertoire diversity, such as: what is the general shape of T cell repertoire (Correia-Neves et al., 2001)?, what is the diversity of a T cell response to a given viral infection (Naumov et al., 2003; Pewe et al., 2004)?, how diverse is the memory T cell pool (Kedzierska et al., 2006)?, or how much diverse can be the repertoire of regulatory $CD4^+CD25^+Foxp3^+$ T cells (Hsieh et al., 2006; Pacholczyk et al., 2006; Wong et al., 2007)?

To answer the above questions, one generally collects a sample of T cells from an individual and sequences their TCRs by some experimental technique. In this scenario, one defines a clonotype as a distinct TCR sequence, either at the nucleotide or amino-acid level, while the clonal size is the number of cells in the body bearing the same TCR sequence. The so-called clonal size distribution is the frequency of clonotypes with a certain clonal size, and embodies all relevant information about the shape of the repertoire.

There are several statistical approaches to analyze TCR sequence data. Simpson's diversity and the Morisita-Horn similarity indexes have been used (Venturi et al., 2008, 2007). However, these two measures are just summary statistics of the samples and, thus, provide limited information about the diversity and clonal size distributions of the cell populations from which the samples were obtained. Another approach is to estimate diversity by a maximum likelihood method that assumes all clonotypes are equally represented in the population (Barth et al., 1985; Behlke et al., 1985). This method is usually seen as a standard tool to estimate TCR diversity (Casrouge et al., 2000; Hsieh et al., 2006). However, TCR sequence data often show long right tails, which are inconsistent with the above assumption, as noted by Naumov et al. (2003) and Pewe et al. (2004). With this in mind, some authors estimated diversity with the non-parametric abundance- and incidence-coverage estimators, taken from ecology, that take into account some heterogeneity in the clonal size distributions (Hsieh et al., 2006; Pacholczyk et al., 2006). Although these estimators may provide robust estimates in a wide range of situations, they give no clue about the underlying clonal size distribution. To overcome above limitations, this paper presents the Poisson abundance models, also from ecology, that can estimate not only the TCR diversity, but also the underlying clonal size distributions. In this way, one might have a better view of T cell repertoire. For illustrative purposes, the models are fitted to data obtained from mice with limited TCR diversity (Correia-Neves et al., 2001).

6.2 Poisson abundance models

TCR sequence data refer to the frequency of clonotypes that appear a certain number of times in the sample. In statistical terms, the respective sampling distribution is usually described by the following multinomial law (Sanathanan, 1972):

$$P[\{m_i\} | D, \eta] = \frac{D!}{(D-M)! \prod_{i=1}^n m_i!} [p_\eta(0)]^{D-M} \prod_{i=1}^n [p_\eta(i)]^{m_i}, \quad (6.1)$$

where D is the number of clonotypes in the population (TCR diversity), m_i is the number of clonotypes with i members in the sample, $p_\eta(i)$ is the probability of a clonotype being sampled

i times that, by its turn, is described by a model with parameters η , $M = \sum_{i=1}^n m_i$ is the number of distinct clonotypes obtained in the sample and $n = \sum_{i=1}^n i \times m_i$ is the sample size. The first goal of the analysis is to estimate the diversity D .

The true sampling scheme of a cell belonging to a certain clonotype is described by a hypergeometric distribution with parameters N (population size), m (the clonal size of the clonotype), and n (sample size). However, when the sample size is much smaller than the population size, which often occurs in practice, the Poisson distribution can be used to approximate the hypergeometric distribution (Dewdney, 1998). By modeling appropriately the parameter of Poisson distribution, one obtains the class of Poisson abundance models (PAMs). The simplest case is to consider that all clonotypes are equally represented in the population, which leads to a homogeneous Poisson distribution

$$p_\lambda(i) = \frac{e^{-\lambda} \lambda^i}{i!}, \quad (6.2)$$

where λ is the sampling rate, defined as the expected value of the hypergeometric distribution (i.e., $\lambda = n \times m/N$). In this scenario, Barth et al. (1985) have derived the conditional distribution of $\{m_i\}$ given the sample size n

$$P[\{m_i\} | D, n] = \binom{D}{M} \frac{M!}{\prod_{i=1}^n m_i!} \frac{n!}{\prod_{i=1}^n (i!)^{m_i}} \frac{1}{D^n}, \quad (6.3)$$

which might be used alternatively to estimate D . Regarding this same model, Behlke et al. (1985) have determined the sampling distribution of M . By statistical sufficiency arguments, one can show that M embodies all sampling information to estimate D , and thus estimation of D can be performed on its sampling distribution, without any loss of information.

Clonotypes have different sampling rates if they differ in their clonal sizes, i.e. clonal sizes are distributed. Thus, different models can be obtained by considering a probability distribution for the sampling rate λ . By doing this, one is modeling indirectly the clonal size distribution, as we see later on. The Gamma-Poisson model assumes a Gamma distribution for λ (Fisher et al., 1943). In this case, one needs to integrate the respective Poisson probabilities on all possible values for λ , that is,

$$p_{\alpha,\beta}(i) = \int_0^\infty \frac{e^{-\lambda} \lambda^i}{i!} \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)} d\mu = \frac{\Gamma(i+\alpha)}{\Gamma(i+1)\Gamma(\alpha)} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^i, \quad (6.4)$$

where α and β are the shape and scale parameters of the Gamma distribution, respectively. Note that, when α is an integer, the above equation converts into the well-known negative-binomial distribution. Furthermore, if $\alpha = 1$, the sampling rate λ is exponentially distributed, and one obtains the Geometric model

$$p_\beta(i) = \left(\frac{1}{\beta+1}\right)^i \frac{\beta}{\beta+1}. \quad (6.5)$$

Finally, the log-series distribution, which has already been applied to TCR sequence data (Pewe et al., 2004), is the limit distribution when α tends to zero (Fisher et al., 1943). However, this model cannot describe the probability of not sampling a clonotype, $p_\eta(0)$, included in

equation (6.1). Therefore, under this framework, the log-series distribution should not be used to estimate diversity. To do that, one must follow an alternative approach, described in detail in Pewe et al. (2004).

The Lognormal-Poisson model considers a Lognormal distribution for the sampling rate λ (Bulmer, 1974). That is, the logarithm of λ follows a normal distribution with mean value μ and variance σ^2 . Under this model, the probability of having a clonotype with i members in the sample is described by the following integral

$$p_{\mu, \sigma^2}(i) = \int_0^\infty \frac{e^{-\lambda} \lambda^{i-1}}{i!} \frac{e^{-\frac{(\ln \lambda - \mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} d\lambda, \quad (6.6)$$

which has no analytical expression. To overcome this problem, Bulmer (1974) suggests a numerical algorithm to calculate the above probabilities. It is worthy noting that this model shows right tail heavier than of the above models, and this feature may be useful to fit many TCR sequence data.

Another model with heavy tails is the Yule distribution that has been used to characterize sexual network formation (Jones and Handcock, 2003) and internet growth (Barabási and Albert, 1999). For what concerns diversity estimation, the Yule distribution can be derived from the following hierarchical structure: the sampling rate λ has an exponential distribution with parameter given by $\mu = e^{-\omega}/(1 - e^{-\omega})$, where ω is another exponential distribution but with parameter ρ . After integration in all possible values for λ , one obtains the following distribution

$$p_\rho(i) = \rho \frac{\Gamma(i+1)\Gamma(\rho+1)}{\Gamma(i+\rho+2)}. \quad (6.7)$$

For sufficiently large i , the above probability is approximately proportional to $1/i^{\rho+2}$, typical of a power-law model. Given this feature, one expects that, under this model, samples are formed by many clonotypes with just one (singletons) or two members (doubletons) and few clonotypes with many individuals.

6.3 Estimation of Diversity

Diversity and parameters of the models can be estimated by the maximization of the likelihood function given by equation (6.1), with $p_\eta(i)$ defined properly for each model: equations (6.2), (6.4), (6.5), (6.6) and (6.7) for the homogeneous Poisson, the Gamma-Poisson, the Geometric, the Lognormal-Poisson and the Yule model, respectively. Since the parameter D is an integer and the likelihood function for these models is well behaved and has a single maximum, one can use the following profile likelihood algorithm: estimate first diversity by the number of clonotypes in the sample, $\hat{D} = M$, and then obtain the respective estimates of the parameters of the models; increase the diversity estimate in one unity and re-estimate the parameters of the models; repeat previous step until the (log-)likelihood function starts to decrease. Considering fixed $\hat{D}$, the estimate of λ in the homogeneous Poisson model is given by $n/\hat{D}$. In the case of the Geometric model, the maximum likelihood estimate of β is given $\hat{D}/(\hat{D} + N)$. To estimate α and β in the Gamma-Poisson, one need to use a numerical method, such as the

Newton-Raphson. With respect to the Yule model, an estimation procedure of ρ can be as follows: (i) assume ρ integer, (ii) find the integer ρ^* that maximize the log-likelihood given $\hat{D}$, (iii) define the interval $(\rho^* - 1, \rho^* + 1)$, (iv) apply the well-known bisection method to this interval. Finally, the estimation of the Lognormal-Poisson model is a bit cumbersome, because there is no analytical expression for the integral (6.6). To overcome this problem, Bulmer (1974) also suggests to apply the scoring-fisher method using an appropriate approximation to the right tail of the Lognormal-Poisson distribution.

After a proper estimation of the models, one should then evaluate their goodness-of-fit. One must say that the equation (6.1) includes the number of clonotypes not observed in the sample, $D - M$. Since this number is estimated rather than observed, it must be excluded when evaluating the quality of fit of the models. For this purpose, one should use alternatively the conditional distribution $\{m_i\}$ given the total number of clonotypes observed in the sample, M , i.e.,

$$P[\{m_i\} | M, \eta] = \frac{M!}{\prod_{i=1}^n m_i!} \prod_{i=1}^n \left[\frac{p_\eta(i)}{1 - p_\eta(0)} \right]^{m_i}. \quad (6.8)$$

Then, one should use a goodness-of-fit test, such as, the Pearson χ^2 test and select models with p-values higher than 5%. To perform the test, it is recommended to divide data in few categories in order to have a reasonable χ^2 approximation for the test statistic under the null hypothesis. A good choice seems to consider only five categories: four referring to 1, 2, 3 and 4 members in the sample and a fifth category for clonotypes with more than 4 members.

TCR sequence data show often right heavy-tail distributions (Correia-Neves et al., 2001; Hsieh et al., 2006; Naumov et al., 2003). In this situation, the whole data might not be well described by the Poisson abundance models. Moreover, diversity estimates might be sensitive to the presence of most abundant clonotypes in the sample. To tackle these problems, the data at hand should be divided in two parts, one referring to the rare clonotypes and other to the most abundant ones, as recommended by some ecological studies (Chao and Bunge, 2002; Chao et al., 1993). The above models are fitted to the rare clonotype data only. Under this analysis, the overall diversity is the estimated diversity of rare clonotypes added to the observed frequency of most abundant ones in the sample. A critical point of the analysis is then the determination of the optimal threshold, say τ , that divides data in two parts. To this end, one should vary τ by considering first the highest value for it (i.e., the number of members of the most abundant clonotype) and then decreasing it until diversity estimates are reasonably stable and a good fit of the models is achieved. In this analysis, one would like to have a model that fits most of the data, while producing at the same time stable diversity estimates.

6.4 Estimation of the clonal size distribution

From the Poisson abundance models, one can infer the underlying clonal size distribution. The Poisson sampling rate of a clonotype is described as $\lambda = n \times m/N$, where N is the number of cells in the population and m is a random variable describing the clonal size of a clonotype (Dewdney, 1998). The quotient between n and N is called the sampling ratio, which is a constant if N is known or if one has a good estimate for it. Therefore, the distribution of

the random variable m is calculated by the sampling rate distribution times the inverse of the sampling ratio ($m = \lambda \times N/n$). Theoretically, the clonal size distribution of an individual is a random sample of D clonotypes taken from the random variable m . For sake of simplicity, the random variable m will be referred to as the true clonal size distribution.

We now give the clonal size distribution predicted by the different Poisson abundance models. They are easily obtained by standard probability techniques to transform random variables. In the homogeneous Poisson model, the sampling rate is constant among clonotypes, and therefore the clonal size distribution is also constant and given by the inverse of the sampling ratio times the sampling rate. In the Gamma-Poisson model, a Gamma distribution with parameters α and β is considered for the sampling rate. This implies also a Gamma distribution for the clonal size distribution with the parameters α and $\beta n/N$. As a special case of the Gamma-Poisson, the Geometric model assumes an exponential distribution with parameter μ for the sampling rate. In this case, the clonal size distribution is also an exponential distribution but with parameter $\mu n/N$. The Lognormal-Poisson describes the sampling rate by a Lognormal distribution with parameters μ and σ^2 . Thus, the clonal size distribution is also a Lognormal distribution but with parameters $\mu + \ln N/n$ and σ^2 . Finally, the Yule model describes the sampling rate by an appropriate hierarchical structure of two exponential distributions. The clonal size distribution is also a exponential distribution with $\mu = n/N \times e^{-w}/(1 - e^{-w})$, where w is an exponential distribution with parameter ρ . In this case, the clonal size distribution has no analytical expression, but can be determined by monte carlo simulation: (i) draw a value w from an exponential distribution with parameter ρ , (ii) calculate $\mu = n/N \times e^{-w}/(1 - e^{-\mu})$, (iii) draw a value x from a exponential distribution with parameter μ , (iv) repeat this process r times (say $r = 10000$). In the end one can determine the cumulative distribution function of the clonal sizes by the respective empirical cumulative function of the simulated sample $x_1, \dots, x_r$.

As suggested above, one should estimate TCR diversity with and without the most abundant clonotypes. Since many TCR data show long right tails, it is likely some clonotypes might be excluded from the analysis in order to have stable diversity estimates and good fits of the models. The excluded clonotypes are then considered as outliers. In this case, the above estimation of the clonal size distribution needs to be adapted. For simplicity, consider that the most abundant clonotype was discarded from the data set; the same reasoning is easily extend to more than one excluded clonotype. This clonotype has m_a members in the sample. If the population size is N cells, then the clonal size of that clonotype is simply proportional to the respective sample abundance, that is, $N \times m_a/n$. The population size of the remaining clonotypes following a certain model is then $N \times (1 - m_a/n)$. This number is now inserted in the clonal size distribution predicted by the respective fitted model.

6.5 Package PAM

The R software, freely available at <http://www.cran.r-project.org/>, was used to analyze data with the Poisson abundance models. In this regard, a package called *PAM* (*Poisson Abundance Models*) is available to help the user in the analysis of these models (see supplementary material). The package provides simple command lines to estimate and test the

Tab. 6.1: Number of distinct CDR3 TCR α sequences represented i times in the samples, where M is the total number of distinct sequences (clonotypes) in the samples, n is the respective sample size, and D_S is the Simpson's diversity index. DP CD3low data are of mouse 17, while remaining data of mouse 57. After proper standardization of sample sizes, Simpson's diversity index was estimated by the median of the diversity distribution resulting from 10000 random samples of a subset without replacement.

i	Thymus			Lymph nodes	
	DP CD3low	SP CD4 ⁺	SP CD8 ⁺	LN CD4 ⁺	LN CD8 ⁺
1	79	33	16	34	17
2	17	6	3	8	8
3	6	2	3	2	1
4	5	2	5	1	2
5	1	0	3	0	1
6	1	0	1	0	0
7	1	0	1	0	0
8		0	1	1	0
10		1	0	1	0
11		0	1	0	0
16		1		0	0
20		0		1	0
21		0			1
28		1			0
52					1
M	110	46	34	48	31
n	169	113	98	98	122
D_S	0.99	0.91	0.96	0.94	0.79

goodness-of-fit of the models, and also to estimate the respective clonal size distribution. The package has been tested using previously published data sets, not only from immunology but also from ecology, where diversity estimates are available for some models. Details on the usage of the package can be found in the respective documentation within R environment after proper installation.

6.6 Example

Correia-Neves et al. (2001) attempted to obtain a broad but manageable view of the thymic and peripheral CD4⁺ and CD8⁺ T cell repertoires. With this purpose, a mouse line was engineered with a very restricted TCR diversity. In this mouse line, there is a TCR α minilocus transgene composed of a single V α region and two J α segments, and a TCR β transgene. The expression of the endogeneous TCR α was prevented by the insertion of an appropriate mutation at the endogeneous α locus, while the same was achieved for the endogeneous TCR β , due to the allelic exclusion of the β locus. In this experimental setting, one expects that T cell diversity is solely generated by the imprecise joining of the V α region with either J α segments. Further experimental details can be found in the original reference (Correia-Neves et al., 2001).

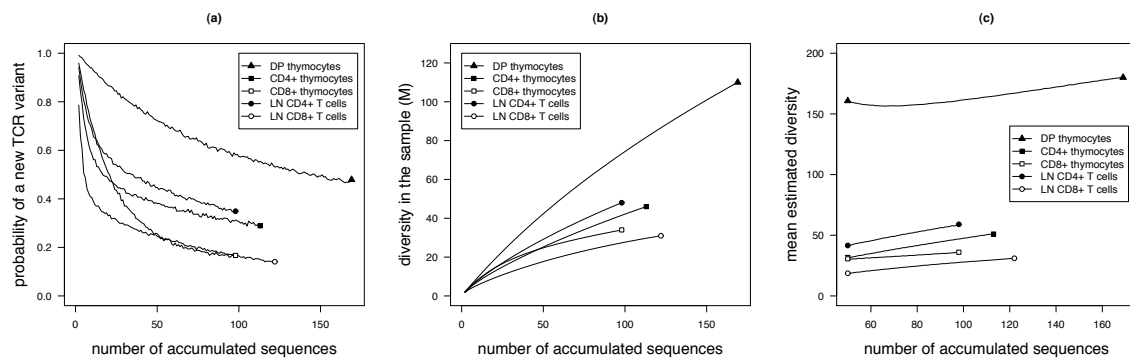


Fig. 6.1: Assessment of the information contained in the samples assuming that they were obtained sequentially: (a) probability of obtaining a new TCR variant, (b) mean sampled diversity, and (c) mean estimated diversity according to the homogeneous Poisson model as a function of the number of TCR sequences accumulated in the samples. The plots were obtained by simulating 10000 random sequential sampling schemes by which the TCR sequences are accumulated in the samples.

A large collection of thymic and peripheral CDR3 TCR α sequences is available from different mice (see supplementary material of Correia-Neves et al. 2001). Data under analysis will be of double positive CD4⁺CD8⁺ thymocytes expressing low levels of CD3 (DP CD3low) from mouse 17, and of single positive (SP) thymocytes, either CD4⁺ or CD8⁺, and lymph node (LN) T cells from mouse 57. Data from other mice were not included in the analysis due to their small sample sizes. Experimentally, TCR α 's of SP thymocytes and of LN T cells were sequenced by single-cell RT-PCR, while of DP CD3low by RT-PCR and cloning. TCR uniqueness was defined at the amino-acid level. No statistical analysis was previously done.

6.6.1 Tackling the information contained in the samples

Before estimating TCR diversity in each T-cell compartment, we perform a preliminary data analysis with the aim of knowing how much information on the repertoire is contained in samples. With this in mind, we suppose that sequences in the original samples were collected by a sequential sampling scheme (i.e., TCR sequences are added sequentially to the samples). In this scenario, we compute a curve that shows how the probability of obtaining a new TCR variant evolves throughout sampling. It is then expected that this probability decreases with the number of accumulated sequences in the samples. More importantly, if the samples are representative of an individual's repertoire, the above-mentioned probability should be near zero in the end of the sampling scheme. We also determine the so-called "species accumulation curve" that reflects how the number of distinct TCR sequences (sampled diversity) evolves throughout sampling (Casrouge et al., 2000; Pacholczyk et al., 2006; Wong et al., 2007). As above, one predicts that species accumulation curve reaches a plateau when the samples contains a large information of an individual's repertoire. It is worth noting that similar curves have been reported for the diversity predicted by homogeneous Poisson model (Casrouge et al., 2000) or for the some non-parametric estimators (Pacholczyk et al., 2006).

To compute these different curves, we use stochastic simulation via Monte Carlo method. We first obtain a random sequence by which TCR sequences are successively added to the

samples. For each sampling step, we calculate the proportion of simulations that the added TCR sequence increased the sampled diversity (Figure 6.1(a)), the mean of sampled diversity (Figure 6.1(b)), and the mean of diversity estimates according to the homogeneous Poisson model (Figure 6.1(c)).

The probability of obtaining a new TCR variant is a decreasing function of the number of accumulated sequences as expected (Figure 6.1(a)). However, it is far from zero at the current sample sizes. This demonstrates that the diversity in the samples at hand is not representative of that in the respective repertoires from which they were obtained. Similar evidences can be extracted from both sampled and estimated diversity curves (Figure 6.1(b) and (c)). In fact, these curves are increasing functions of the sample sizes, not reaching a plateau. It is worth noting the estimated diversity curves may not be very accurate since the homogeneous Poisson model does not hold on data, as we will see below. For all of this, the step of diversity estimation assumes in this scenario a critical role in the success of the whole analysis.

6.6.2 Estimating TCR diversity

Samples of different T cell compartments are mostly composed of many clonotypes with few members and few clonotypes with many members (Table 6.1). This is most evident in LN CD8⁺ T cells, where a single clonotype, with 52 members, represents 40% of the sample. A model with a power-law tail, such as the Poisson-Lognormal or Yule, might capture this feature. The DP CD3low thymocytes exhibits a not so long right tail distribution as the remaining populations, which might reflect a moderate selection in the early stages of thymocyte development.

Simpson's diversity index is first used to study diversity (Venturi et al., 2007). It is defined as the probability that any two TCR sequences chosen at random from the sample belong to different clonotypes, i.e.,

$$D_S = 1 - \frac{\sum_{i=1}^n m_i(m_i - 1)}{n(n - 1)}. \quad (6.9)$$

As a probability, it ranges from 0 to 1, where 0 means that all TCR sequences belong to the same clonotype (minimal diversity), while 1 indicates that every TCR sequence is a distinct clonotype (maximal diversity). When one wants to compare several populations with different sample sizes, the recommended approach is to standardize them to the smallest number of TCR sequences in any sample (Venturi et al., 2007). Simpson's diversity index is then applied directly to the smallest sample(s), while for the remaining samples it is estimated by the median of the diversity distribution resulting from a random sampling of a subset without replacement (Table 6.1). All thymic populations show values close to 1, which suggest almost maximal diversity for these T cell compartments. As expected, the lesser mature DP CD3low population shows the highest diversity among all thymic compartments. In lymph nodes, CD4⁺ T cells show diversity estimates close to the thymic SP populations, while CD8⁺ T cells are less diverse than the respective thymic SP population. This decrease in diversity might be attributed to a great expansion of a few clones in the lymph nodes upon encountering peripheral antigens.

Poisson abundance models were then fitted to data (Table 6.2). DP CD3low thymocytes are best described by the Gamma-Poisson, Lognormal-Poisson and Yule. However, the Gamma-Poisson seems an aberrant estimate for TCR diversity as compared to those obtained from

Tab. 6.2: Diversity estimates of DP CD3low, SP CD4⁺ and CD8⁺ thymocytes, and LN CD4⁺ and CD8⁺ T cells, and the corresponding p-value of Pearson's goodness-of-fit test for different PAMs when fitted to the whole data.

Population	Model	Diversity	p-value
Thymic DP CD3low	Homogeneous Poisson	180	$< 10^{-3}$
	Geometric	314	0.09
	Gamma-Poisson	4670	0.30
	Lognormal-Poisson	510	0.32
	Yule	625	0.52
Thymic SP CD4 ⁺	Homogeneous Poisson	51	$< 10^{-3}$
	Geometric	77	$< 10^{-3}$
	Gamma-Poisson	2388	0.01
	Lognormal-Poisson	16820	0.69
	Yule	176	0.30
Thymic SP CD8 ⁺	Homogeneous Poisson	36	$< 10^{-3}$
	Geometric	51	0.10
	Gamma-Poisson	69	0.07
	Lognormal-Poisson	49	0.05
	Yule	93	0.09
LN CD4 ⁺	Homogeneous Poisson	59	$< 10^{-3}$
	Geometric	93	0.02
	Gamma-Poisson	3074	0.09
	Lognormal-Poisson	1182	0.76
	Yule	206	0.54
LN CD8 ⁺	Homogeneous Poisson	31	< 0.001
	Geometric	41	$< 10^{-3}$
	Gamma-Poisson	1721	0.02
	Lognormal-Poisson	260	0.21
	Yule	91	0.44

the other models ($\hat{D} = 4670$). The Lognormal-Poisson and Yule led to more reasonable TCR diversity estimates, $\hat{D} = 510$ and 625, respectively.

SP CD4⁺ thymocytes are well described either by the Lognormal-Poisson or the Yule, predicting very different estimates for TCR diversity. One expects that SP thymocytes have a lower TCR diversity than more numerous DP thymocytes from which they derived. Therefore, the Lognormal-Poisson does not seem a good model, because it predicts diversity of SP CD4⁺ thymocytes around 17000, which is far beyond the estimates obtained for the DP population. The Yule distribution shows a good fit for the data and a diversity estimate lesser than of DP population. Thus, it is considered the best model to the SP CD4⁺ data.

In the case of SP CD8⁺ thymocytes, TCR diversity estimates seem similar in all models and lesser than of DP population. Yet no model shows a particular good fit (p-values around the significance level). Even though, the Geometric and Yule seem to be the most adequate models for this thymocyte population.

In both lymph node populations, the best fitted models seem to be again the Lognormal-Poisson and Yule. For CD4⁺ T cells, Lognormal-Poisson predicts a TCR diversity much higher

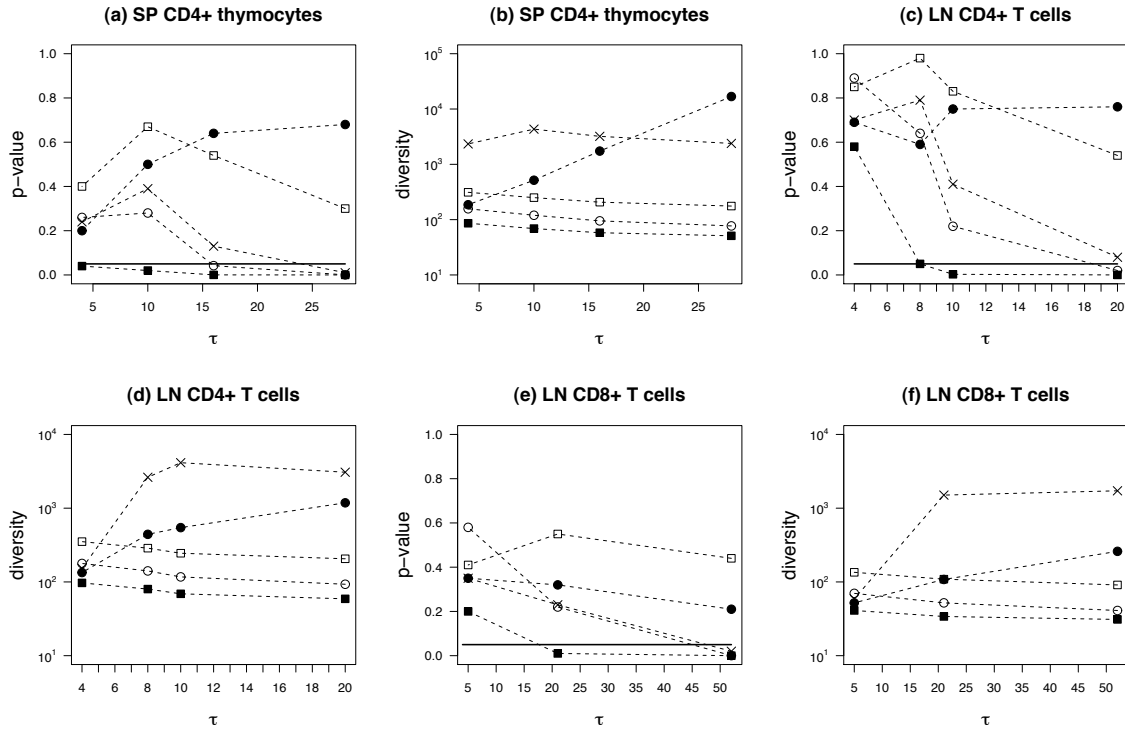


Fig. 6.2: Robustness of TCR diversity estimates according to the PAMs (■ — homogeneous Poisson, ○ — Geometric, × — Gamma-Poisson, ● — Lognormal-Poisson, □ — Yule). The τ 's axes of all plots refers to the threshold τ that divides the corresponding data in two parts when using the procedure described in Chao et al. (1993). Plots (a), (c) and (e) show the p-value of the Pearson's χ^2 goodness-of-fit test as a function of τ , where the solid line is the 5% significance level. In this regard, models fit the data well when p-values are above the reference value. Plots (b), (d) and (f) show how diversity estimates change with τ .

than of their thymocyte counterparts. This might be explained by an accumulation of diversity in the periphery. With respect to the Yule model, TCR diversity is estimated more closely to what was estimated in the thymus. For CD8⁺ T cells, the Lognormal-Poisson leads once again to TCR diversity higher than in thymus, while the Yule predicts similar TCR diversity as in SP thymocytes.

6.6.3 Robustness of TCR diversity estimates

As stated before, diversity estimation might be affected by the most abundant clonotypes in the sample, typically with more than 10 members (Chao et al., 1993). Thus, the next step of the analysis is to evaluate the robustness of diversity estimates, namely, of SP CD4⁺ thymocytes and of both LN CD4⁺ and CD8⁺ T cell populations. To this end, we exclude from the data, one by one, the most abundant clonotypes, and fit again the models and obtain new diversity estimates. In this analysis, one wants a model that produces reasonably stable diversity estimates as well as a good fit for the data.

In general, the homogeneous Poisson model cannot describe the data well, even when discarding the most abundant TCR sequences from the analysis (Figure 6.2(a), (c) and (e)).

The only exception is the good fit of the LN CD8⁺ T cells, but only for clonotypes with at maximum 5 members (Figure 6.2(e)). Therefore, as expected, a homogeneous clonal size distribution is not realistic. Other model with peculiar results is the Gamma-Poisson, producing the highest TCR diversity estimates in most T cell compartments. This is a consequence of the right heavy tails exhibited by the data, since diversity estimate tends to decrease dramatically, when leaving out the most abundant clonotypes (Figures 6.2(b), (d) and (f)).

For SP CD4⁺ thymocytes, the Lognormal-Poisson is the best fitted model with or without the most abundant clonotype (Figure 6.2(a)). However, the respective diversity estimates are higher than of DP thymocytes, and thus this model seems to produce unreliable TCR diversity estimates (Figure 6.2(b)). When the most abundant clonotype was not considered in the analysis, the Yule model improved its fit, being as adequate as the Lognormal-Poisson. A slightly improvement of the fit was also obtained when excluding additionally the second most abundant clonotype, but not significantly (Figure 6.2(a)). The diversity estimates are always smaller than that of DP population and reasonable stable (Figure 6.2(b)). With respect to the remaining models, no significant improvement of the fit was obtained when excluding from the analysis, one by one, the most abundant clonotype. For all of this, best model for the SP CD4⁺ thymocytes is the Yule, where the most abundant clonotype is viewed as an outlier.

For LN CD4⁺ T cells, most models can fit the data well if one discards the most abundant clonotype from the data set (Figure 6.2(c)). However, as in SP CD4⁺ thymocytes, the Yule and Lognormal-Poisson show, in most cases, the highest p-values. In the latter, the TCR diversity estimate decreases dramatically without the most abundant clonotype, but remains stable when discarding further clonotypes (Figure 6.2(d)). The fit of the Yule distribution was improved without the most abundant clonotype. Excluding further clonotypes increases the quality of fit but not significantly. In this model, diversity estimates are almost unaffected by the most extreme abundant clonotypes (Figure 6.2(d)). The Geometric model shows a particular good fit without the three most abundant clonotypes. Thus, the Yule and Lognormal-Poisson can describe the data well, but the most abundant clonotype should be considered as an outlier. The same can be said for the Geometric model, but with the three most abundant clonotypes as outliers.

Finally, LN CD8⁺ T cells are well described by the Yule and Lognormal-Poisson. Their fits do not dramatically improved when discarding, one by one, the most abundant clonotypes (Figure 6.2(e)), while the respective TCR diversity estimates are reasonably stable (Figure 6.2(f)). Therefore, both models should be considered for the whole data. As in the LN CD4⁺ T cells, the Geometric model also exhibits a good fit without the three most abundant clonotypes. This model will also be further analyzed.

6.6.4 Estimating clonal size distributions

All T cell compartments could be explained by the Yule distribution that predicts an appropriate mixture of exponential distributions for the clonal size distribution. Other good model was the Lognormal-Poisson that implies a Lognormal distribution for the clonal sizes. Finally, the Geometric model could describe either lymph node population and SP CD8⁺ thymocytes. This model entails an exponential distribution for the clonal sizes.

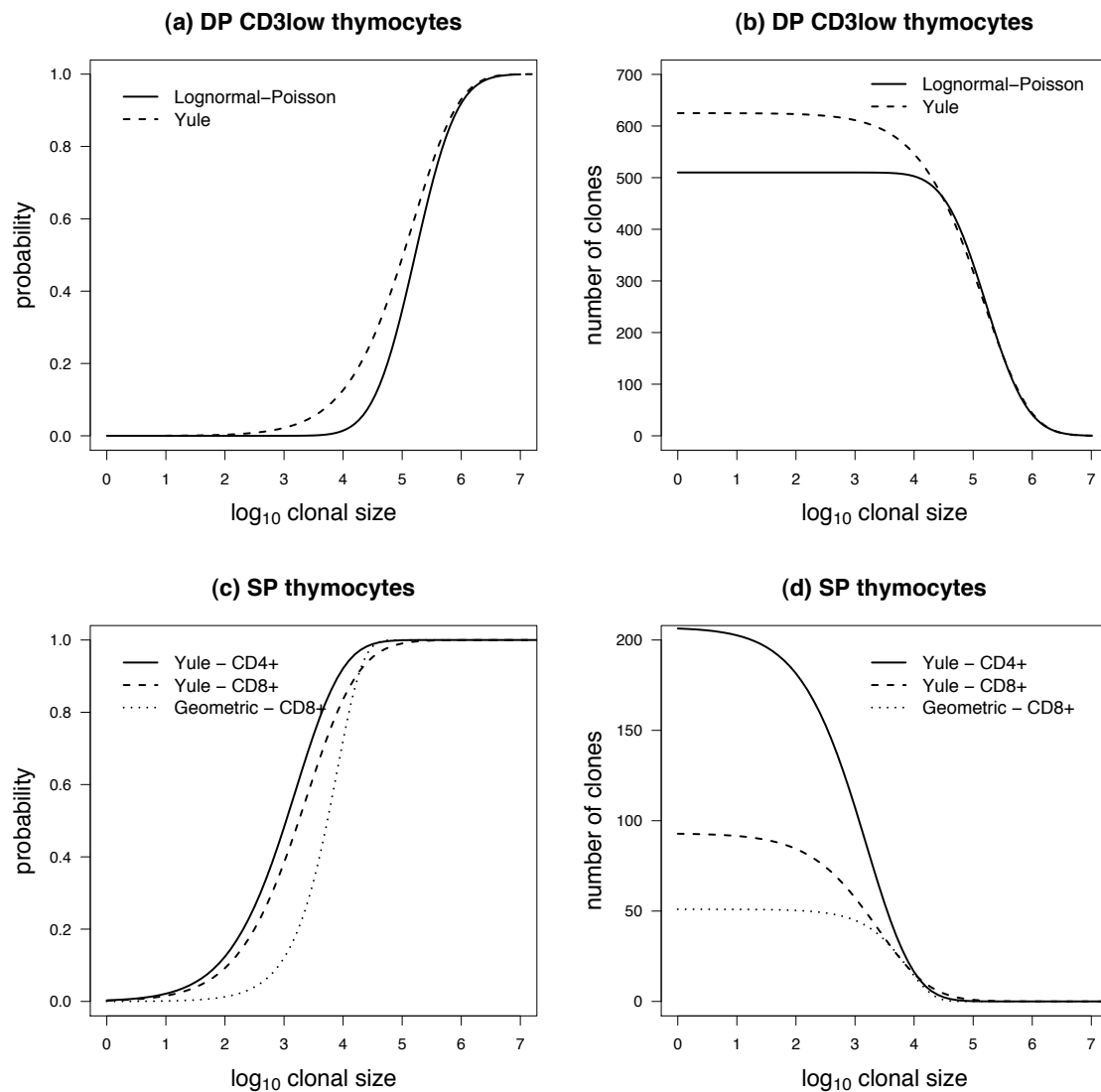


Fig. 6.3: Clonal size distributions of thymocyte populations (plots (a) and (c)). Number of clones with clonal size above a certain value (plots (b) and (d)). The clonal size distributions predicted by the Yule were obtained by Monte Carlo simulation with 10000 replicates.

To transform the Poisson sampling rate distribution into the respective clonal size distribution, one needs to know the total number of cells in the different T cell compartments. In the thymus, there are roughly 200 million thymocytes, where 95%, 0.05% and 0.02% of them are DP CD3low, SP CD4+ and SP CD8+, respectively. Lymph nodes have around 15 million cells. As shown in Correia-Neves et al. (2001), the frequency of CD4+ and CD8+ T cells changes with time. At week 6, when samples were collected, CD4+ and CD8+ T cells represent approximately 4% and 10% of total lymph node cells, respectively. Table 6.3 summarizes all relevant information.

In every T cell compartment, the clonal size distributions have the common feature of being highly skewed to the right (Figures 6.3 and 6.4 (a) and (c)). In this case, the mean and median provide very different descriptions of the clonal size distributions (Table 6.3). The best-fitted models entail different TCR diversity estimates. This difference is reflected in the clonal

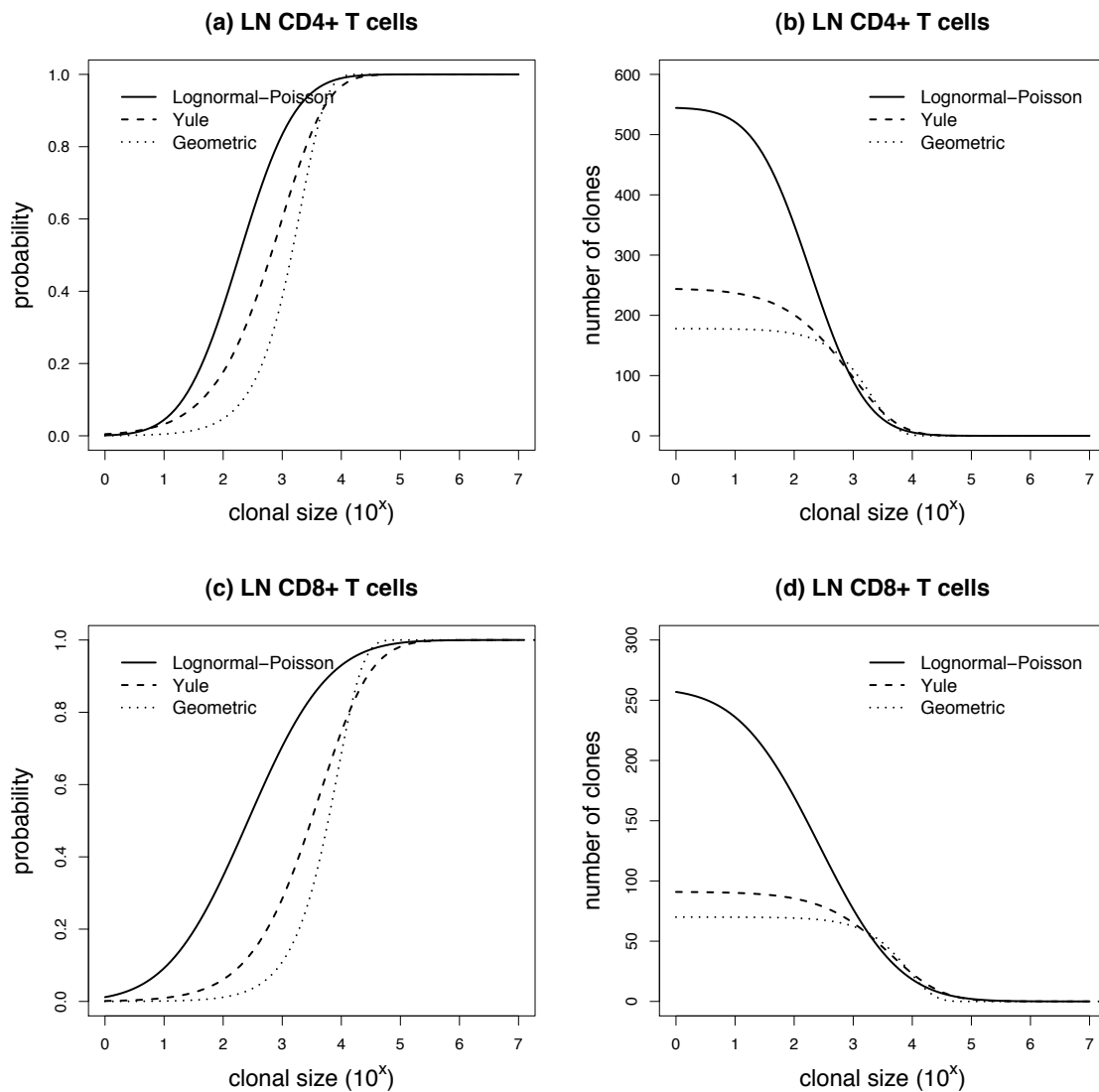


Fig. 6.4: Clonal size distributions of LN T cells (plots (a) and (c)). Number of clones with clonal sizes above a certain value (plots (b) and (d)). The clonal size distributions predicted by the Yule were obtained by monte carlo simulation with 10000 replicates.

size distribution by more clonotypes with "few" members in models with high TCR diversity estimates (Figure 6.3(b) and (d) for thymocyte population and Figure 6.4(b) and (d) for LN T cells). This is most evident for LN CD8⁺ T cells when comparing the Yule with the Lognormal-Poisson, where the latter model provides a higher TCR diversity estimates than the former (Figure 6.4(d)). In spite of this difference, similar number of extremely abundant clonotypes is predicted from all models.

6.6.5 Large sample sizes are required to discriminate the models

As seen before, each T-cell compartment could be fitted well by more than one PAM. For example, data from LN CD4⁺ T cells are reasonably described either by Geometric, Yule or Lognormal-Poisson (Table 6.3). Facing this result, it is of great interest to know what is the

Tab. 6.3: Characterization of clonal size distributions for selected PAMs: N is the total number of cells in the respective population and $\hat{D}$ is the diversity estimate. Number of outlier TCR sequences are also given as well as their contribution to the respective T cell compartment (in parenthesis). ^aBase 10 logarithm; ^bvalues refer to estimates without the outliers.

Population	N^a	Model	$\hat{D}$	Mean ^a	Median ^a	SD	Outliers
DP CD3low	8.28	Lognormal-Poisson	510	5.58	5.22	5.89	0 (0%)
		Yule	625	5.48	5.02	5.81	0 (0%)
SP CD4 ⁺	6.00	Yule	207	3.53 ^b	3.04 ^b	3.91 ^b	1 (25%)
SP CD8 ⁺	5.60	Geometric	51	3.89	3.73	3.89	0 (0%)
		Yule	93	3.73	3.07	4.37	0 (0%)
LN CD4 ⁺	5.78	Geometric	178	3.31 ^b	3.16 ^b	3.31	3 (39%)
		Lognormal-Poisson	545	2.92 ^b	2.27 ^b	3.55 ^b	1 (20%)
		Yule	245	3.28 ^b	2.80 ^b	3.64 ^b	1 (20%)
LN CD8 ⁺	6.18	Geometric	70	3.94 ^b	3.78 ^b	3.94 ^b	2 (60%)
		Lognormal-Poisson	260	3.74	2.42	5.06	0 (0%)
		Yule	91	4.10	3.49	4.73	0 (0%)

"true" model for the data. A simple way of tackling this question is to increase the sample size. In theory, it is expected that only the true model hold when compared to data. With this in mind, we perform a small simulation study to determine the sample size required to discriminate the above-mentioned models for the LN CD4⁺ T cells. To this end, we simulate data from each model using as true diversity the corresponding estimate of the original data set (Table 6.3). We then fit the remaining models to the simulated data, and test their goodness-of-fit according to Pearson's test. Since PAMs assume a Poisson sampling scheme, the sample size is not fixed when simulating a data set from the repertoire. To increase the sample size, we vary the parameters of each model in order to increase the respective sampling rates of each TCR variant in the repertoire. We then divided the simulated sample sizes in distinct categories, calculating in each one the proportion of simulations that reject each model at 5% significance level (Figure 6.5(a), (b) and (c)). As a control of the simulation, the model that generated the data is expected to be rejected in approximately 5% of the simulations. This was in fact obtained, as shown in Figures 6.5(a), (b) and (c). To complement this analysis, we also compute the median of the respective distribution of estimated diversities in each sample size category. In this regard, it is worth noting that the median of estimated diversities obtained from the true model agrees with the respective TCR diversity specified in that model (Figure 6.5(d), (e) and (f)).

In general, large sample sizes are required to reject all wrong models (Figure 6.5). This is clearly shown in Yule-generated data, where the Lognormal-Poisson can be at best rejected in 20% of the simulated samples of nearly 10000 TCR sequences (Figure 6.5(c)). This result might be explained by two facts. On the one hand, it is known that the Lognormal distribution can give good fit even when the totality of the left tail of the Lognormal distribution is absent from the data, the so-called "veiled line" phenomenon (Dewdney, 1998). On the other hand, the right tails of Yule and Lognormal-Poisson distributions are approximately power-laws. Putting both facts together the great acceptance proportion of the Lognormal-Poisson might

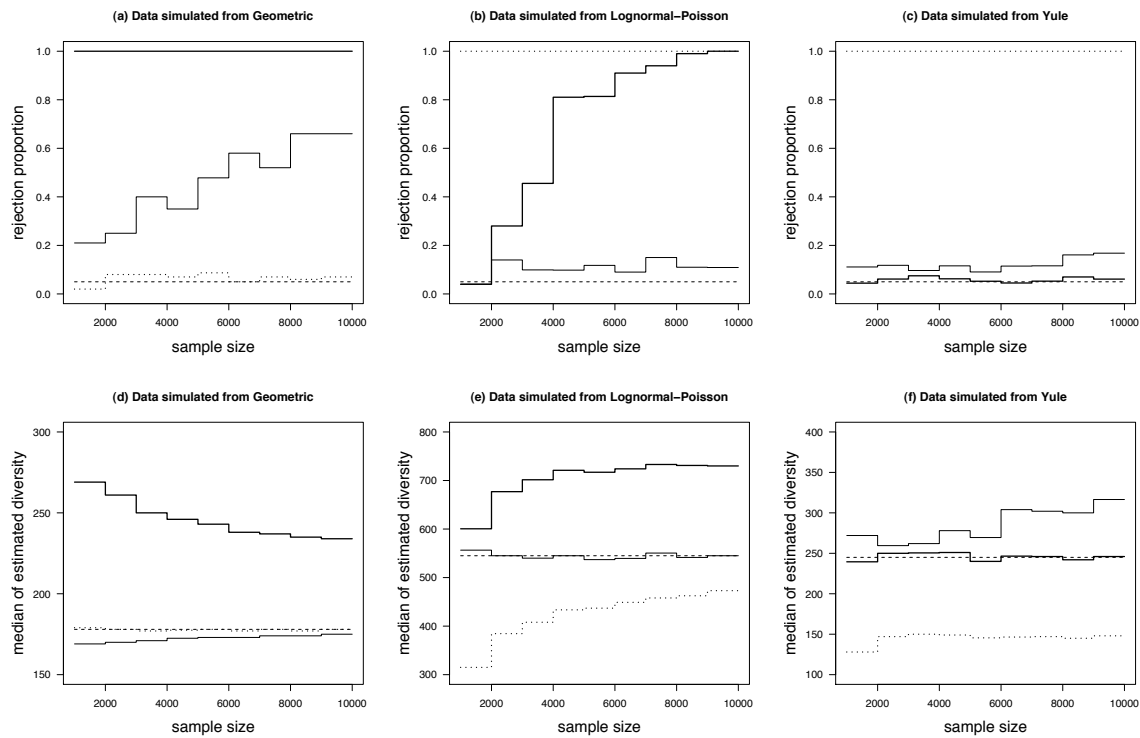


Fig. 6.5: Large sample sizes are required to discriminate the Geometric (dotted line), Yule (thick solid line) and Lognormal-Poisson (thin solid line) models from each other in the case of LN CD4⁺ T cell compartment. (a), (b) and (c) Rejection proportions of each model, where the dashed lines represent the 95% reference line. (d), (e) and (f) Median of estimated diversity predicted by each model, where the dashed lines represent the diversity estimates for the original data set predicted the respective model ($\hat{D} = 178, 545$ and 245 for the Geometric, Lognormal-Poisson and Yule, respectively). Each step in the plots was obtained according to 100 simulated samples with random sample sizes within each sample size category.

reflect a sort of veiled-line phenomenon, where one can only infer the power-law-like right tail of the Lognormal sampling rate distribution from the data. Notwithstanding, the TCR diversity predicted by Lognormal-Poisson is usually an overestimate of the true diversity, being worse with the increase of the sample size (Figure 6.5(f)). In contrast, it is enough 8000 TCR sequences to completely reject the Yule model for data generated from the Lognormal-Poisson (Figure 6.5(b)). This might be due to the fact that increasing the sample size leads to a higher information on the whole Lognormal sampling rate distribution, specially its left tail, which cannot be captured by the Yule distribution. Moreover, Yule-based TCR diversity estimates tend to be higher than underlying true diversity (Figure 6.5(e)). Notwithstanding, when the true model is either Yule or Lognormal-Poisson, samples of 1000 TCR sequences seem sufficient to reject a Geometric distribution (Figure 6.5(b) and (c)). This is related to the fact that Yule and Lognormal-Poisson have power-law right tails, much heavier than the exponential one of the Geometric model. Finally, the Lognormal-Poisson has a higher proportion of being wrongly accepted when data are generated from a Geometric distribution. More importantly, this model has the ability to estimate correctly the underlying TCR diversity (Figure 6.5(d)). In contrast, Geometric-generated data higher than 1000 TCR sequences cannot be captured by the Yule distribution.

Tab. 6.4: Diversity estimates $\hat{D}$ of LN CD4⁺ T cells in three mice and in their pooling data, where p is the corresponding p-value of Pearson's goodness-of-fit test for each PAM. Sample sizes for each mice and for pooled data are 98 (mouse 57), 69 (mouse 63), 98 (mouse 91) and 212 (pooled). The respective number of sampled different TCR variants are 48, 33, 28 and 95.

Model	Mouse 57		Mouse 63		Mouse 91		Pooled	
	$\hat{D}$	p	$\hat{D}$	p	$\hat{D}$	p	$\hat{D}$	p
Homogeneous Poisson	59	$< 10^{-2}$	39	$< 10^{-2}$	42	$< 10^{-2}$	111	$< 10^{-2}$
Geometric	93	0.02	62	0.01	73	0.03	171	$< 10^{-2}$
Gamma-Poisson	3074	0.08	2046	0.02	2615	0.06	4978	$< 10^{-2}$
Yule	206	0.54	153	0.23	158	0.18	385	0.01
Lognormal-Poisson	1182	0.76	7619	0.10	2998	0.29	32342	0.46

Facing the above results, we recommend that future studies on TCR repertoire should contemplate a large number of TCR sequences in the samples. Only in this way one could have a clear insight of the shape of the TCR repertoire. However, these large sample sizes should not be obtained by pooling data together from different animals as often done in TCR repertoire studies (Lathrop et al., 2008; Pacholczyk et al., 2006, 2007). Moreover, pooling data from different animals has the effect of increasing TCR diversity estimates by the inclusion of TCR variants belonging the respective exclusive repertoire of each individual. To demonstrate this effect, we pooled together LN CD4⁺ data from three distinct animals and fit different PAMs to them (Table 6.4). As expected, higher diversity estimates are predicted when analyzing pooled data for all models, even when they are rejected by the data. The most dramatic effect is observed in the Lognormal-Poisson, which can fit well every mouse data as well as pooled data. In this model, diversity estimates for each animal range from 1182 to 7619, as opposing to 32342 different TCR sequences predicted for the pooled data, which is much higher than the sum of the diversity of each individual mouse. Interestingly, Yule can fit well every individual mouse data but not the pooled data. This result shows that pooling data has also the effect of blurring the clonal size distribution present in an individual.

6.7 Discussion

Several Poisson abundance models were presented and used to estimate TCR diversity and clonal size distributions in mice with limited TCR diversity, based on a maximum likelihood framework. We are now in a better position to clarify the shape of T cell repertoire in these mice. After a productive imprecise joining of V and J of the α chain, DP CD3^{low} have no more than 600 distinct TCR sequences. With positive and negative selection, this number drops to less than 100 in both SP populations, even though SP CD4⁺ thymocytes seem more diverse than their CD8⁺ counterparts. In the lymph nodes, TCR diversity of both T cell populations might be higher than in the thymus, suggesting an accumulation of TCR diversity in the periphery. There are few clonotypes that will expand greatly when reaching periphery, maybe due to a better usage of resources. In this regard, CD8⁺ T cells are more prone to have clonotypes with large clonal sizes than CD4⁺ T cells.

The nature of the clonal size distribution can provide clues about the mechanisms governing the generation and maintenance of TCR diversity. The above analysis reveals that clonal size

distribution might be either an exponential, an appropriate mixture of exponentials, or a Lognormal, depending on the T cell compartment under analysis. To explain the origin of these distributions, one might think the immune system as a community divided in niches (Carneiro et al., 1995; Freitas and Rocha, 2000). A niche can be defined either as an anatomical site or as the availability of a certain resource. Different mechanisms for niche apportionment entail different clonal size distributions.

The Exponential clonal size distribution can be derived from the "broken stick" model (MacArthur, 1957). In the model, a community is compared to a stick with unit length. If there are S species in a community, the stick is broken simultaneously into S pieces (niches), each one occupied by a single species. The relative abundance of the species is proportional to the lengths of the segments they belong to. The simultaneous breaking of the stick implies that the niches are already predefined in the community. It is known that the architectures of thymic microenvironments and peripheral lymphocyte environment change with time (Freitas and Rocha, 2000; van Ewijk et al., 1999). Therefore, a static definition of thymic niche is rather difficult to envisage. For all of this, the "broken stick" model might not be a good candidate mechanism for the generation and maintenance of T cell diversity.

A natural way to extend the "broken stick" model is to consider a sequential niche apportionment. That is, new niches are generated in a sequence of stages in which the preexisting ones are further broken. If the number of stages is not too small, one expects the clonal size distribution to be Lognormal (Bulmer, 1974; Sugihara, 1980). A sequential niche apportionment seems reasonable for periphery, since the arrival of new clones in periphery coming from the thymus can be regarded as the stages of the model. For the thymus, this might be also true. It is known that thymocytes affect greatly thymic microenvironments. Thus, new thymic microenvironments might be created with the production of clones with new TCR sequences during the V(D)J recombination process. It is worth noting that other mechanisms generating a Lognormal distribution can be put forward (Diserud and Engen, 2000; Engen and Lande, 1996), but their relevancy to the immune system will be discussed elsewhere.

One can also extend the "broken-stick" model by breaking first the stick into a certain number of pieces. Those pieces define superniches that will be further divided according to their length in the stick. In this regard, larger superniches will tend to be more broken than small ones. With this two-stage mechanism, one might expect an appropriate mixture of exponential distributions for the clonal size distribution, as predicted by the Yule model. A superniche might be defined by the expression of certain homing receptors, such as CCR4 and CCR7 (Sallusto et al., 1999; Tang and Cyster, 1999), or by a certain motif in the TCR sequence (Wallace et al., 2000).

This discussion on the rationale and interpretation of the clonal size distributions put the emphasis on the necessity to better define and characterize immunological niches experimentally. The work of Lathrop et al. (2008) can be regarded as a first step in this direction by studying the repertoire in different anatomical locations. However, one can anticipate that characterizing static or sequential niche apportionment process would require not only the study of the T-cell repertoire at different anatomical locations but at different time points.

Data at hand can be fitted by more than one model. A straight forward way to decide

which model is more likely to T-cell biology is to increase sample size. By doing this, one should expect that only the true model would fit the data. Our study indicates that sample sizes should be large and in the order of 10^4 independent sequences. This sample size which was prohibitive based on the experimental techniques used in the original reference, are now within reach using the new high-throughput sequencing methods (Freeman et al., 2009; Holt and Jones, 2008; Shendure and Ji, 2008).

One must say that the TCR repertoire analysis does not end with the estimation of diversity and clonal size distributions. The next step is to quantify the intersection among different TCR repertoires. In this regard, one can calculate the well-known Morisita-Horn similarity index that takes into account not only the number of clonotypes shared by several samples, but also their clonal sizes (Venturi et al., 2008). Yet this measure should be viewed as a summary statistic that does not truly quantify the intersection at the population level. In theory, the Poisson abundance models might be used to study the intersection between TCR repertoires. To this end, models might include parameters that reflect how clonal sizes of the clonotypes belonging to the intersection set are related to each other. The Lognormal-Poisson seems to be the easiest model to do such, since the relationship between clonal sizes could modeled by parameters embodying the traditional Pearson's correlation coefficient. The Yule model can also be extended to the bivariate case, but entails for both variables a Yule distribution with the same parameter (Xekalaki, 1986), which might be unrealistic. For the remaining models, the relationship between clonal sizes may require an extra level of modeling. For all of this, additional work needs to be done in order to fully apply the Poisson abundance models to the analysis of TCR repertoires.

Bibliography

- Arstila, T. P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J., and Kourilsky, P. (1999). A direct estimate of the human $\alpha\beta$ T cell receptor diversity. *Science*, 286:958–961.
- Barabási, A. L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.
- Barth, R. K., Kim, B. S., Lan, N. C., Hunkapiller, T., Sobieck, N., Winoto, A., Gershenfeld, H., Okada, C., Hansburg, D., Weissman, I. L., and Hood, L. (1985). The murine T-cell receptor uses limited repertoire of expressed V_β gene segments. *Nature*, 316:517–523.
- Behlke, M. A., Spinella, D. G., Chou, H. S., Sha, W., Hartl, D. L., and Loh, D. Y. (1985). T-cell receptor β -chain expression: dependence on relatively few variable region genes. *Science*, 229:566–570.
- Bulmer, M. G. (1974). On fitting the Poisson Lognormal distribution to species-abundance data. *Biometrics*, 30:101–110.
- Carneiro, J., Stewart, J., Coutinho, A., and Coutinho, G. (1995). The ontogeny of class-regulation of CD4+ T lymphocyte populations. *Int. Immunol.*, 7:1265–1277.

- Casrouge, A., Beaudoin, E., Dalle, S., Pannetier, C., Kanellopoulos, J., and Kourilsky, P. (2000). Size of the $\alpha\beta$ TCR repertoire of naive mouse splenocytes. *J. Immunol.*, 164:5782–5787.
- Chao, A. and Bunge, J. (2002). Estimating the number of species in a stochastic abundance model. *Biometrics*, 58:531–539.
- Chao, A., Ma, M.-C., and Yang, M. C. K. (1993). Stopping rule and estimation for recapture debugging with unequal detection rates. *Biometrika*, 80:193–201.
- Correia-Neves, M., Waltzinger, C., Mathis, D., and Benoist, C. (2001). The shaping of the T cell repertoire. *Immunity*, 14:21–32.
- Davis, M. M. and Bjorkman, P. J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature*, 334:395–402.
- Dewdney, A. K. (1998). A general theory of the sampling process with applications to the veil line. *Theoretical and Population Biology*, 54:294–302.
- Diserud, O. H. and Engen, S. (2000). A general and dynamic species abundance model, embracing the lognormal and the gamma models. *The American Naturalist*, 155:497–511.
- Engen, S. and Lande, R. (1996). Population dynamic models generating the lognormal species abundance distribution. *Mathematical Biosciences*, 132:169–183.
- Fisher, R. A., Corbet, A. S., and Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, 12:42–58.
- Freeman, J. D., Warren, R. L., Webb, J. R., Nelson, B. H., and Holt, R. A. (2009). Profiling the T-cell receptor β -chain repertoire by massively parallel sequencing (ahead of print). *Genome Res.*
- Freitas, A. A. and Rocha, B. (2000). Population biology of lymphocytes: the flight for survival. *Annu. Rev. Immunol.*, 18:83–111.
- Holt, R. A. and Jones, S. J. (2008). The new paradigm of flow cell sequencing. *Genome Res.*, 18:839–846.
- Hsieh, C.-S., Zheng, Y., Liang, Y., Fontenot, J. D., and Rudensky, A. Y. (2006). An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nat. Immunol.*, 7:401–410.
- Janeway, C. A., Travers, P., Walport, M., and Shlomchik, M. J. (2005). *Immunobiology*. Garland, New York.
- Jones, J. and Handcock, M. S. (2003). An assessment of preferential attachment as a mechanism for human sexual network formation. *Proc. Roy. Soc. London B*, 270:1123–1128.

- Kedzierska, K., Venturi, V., Field, K., Davenport, M. P., Turner, S. J., and Doherty, P. C. (2006). Early establishment of diverse T cell receptor profiles for influenza-specific CD8⁺ CD62Lhi memory T cells. *Proc. Natl. Acad. Sci. USA*, 103:9184–9189.
- Lathrop, S. K., Santacruz, N. A., Pham, D., Luo, J., and Hsieh, C.-S. (2008). Antigen-specific peripheral shaping of the natural regulatory T cell population. *J. Exp. Med.*, 205:3105–3117.
- MacArthur, R. H. (1957). On the relative abundance of bird species. *Proc. Nat. Acad. Sci. USA*, 43:293–295.
- Naumov, Y. N., Naumova, E. N., Hogan, K. T., Selin, L. K., and Gorski, J. (2003). A fractal clonotype distribution in the CD8⁺ memory T cell repertoire could optimize potential for immune responses. *J. Immunol.*, 170:3994–4001.
- Naylor, K., Li, G., Vallejo, A. N., Lee, W. W., Koetz, K., Bryl, E., Witkowski, J., Fullbright, J., Weyand, C. M., and Goronzy, J. J. (2005). The influence of age on T cell generation and TCR diversity. *J. Immunol.*, 70:7446.
- Pacholczyk, R., Ignatowicz, H., Kraj, P., and Ignatowicz, L. (2006). Origin and T cell receptor diversity of Foxp3⁺CD4⁺CD25⁺ T cells. *Immunity*, 25:249–259.
- Pacholczyk, R., Kern, J., Singh, N., Iwashima, M., Kraj, P., and Ignatowicz, L. (2007). Nonself-antigens are the cognate specificities of Foxp3⁺ regulatory T cells. *Immunity*, 27:493–504.
- Pewe, L. L., Netland, J. M., Heard, S. B., and Perlman, S. (2004). Very diverse CD8 T cell clonotypic responses after virus infections. *J. Immunol.*, 172:3151–3156.
- Sallusto, F., Lenig, D., Forster, R., Lipp, M., and Lanzavecchia, A. (1999). Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature*, 401:708–712.
- Sanathanan, L. (1972). Estimating the size of a multinomial population. *Ann. Math. Stat.*, 43:142–153.
- Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing. *Nat. Biotechnol.*, 26:1135–1145.
- Sugihara, G. (1980). Minimal community structure: an explanation of species abundance patterns. *Amer. Naturalist*, 116:770–787.
- Tang, H. T. and Cyster, J. G. (1999). Chemokine up-regulation and activated T cell attraction by maturing dendritic cells. *Science*, 284:819–822.
- van Ewijk, W., Wang, B.-P., Hollander, G., Kawasoto, H., Spanopoulou, Itoi, M., Amagai, T., Jiang, Y.-F., Germeraad, W. T., Chen, W.-F., and Katsura, Y. (1999). Thymic microenvironments, 3-D versus 2-D? *Semin. Immunol.*, 11:57–64.
- Venturi, V., Kedzierska, K., Tanaka, M. M., Turner, S. J., Doherty, P. C., and Davenport, M. P. (2008). Method for assessing the similarity between subsets of the T cell receptor repertoire. *J. Immunol. Methods*, 329:67–80.

- Venturi, V., Kedzierska, K., Turner, S. J., Doherty, P. C., and Davenport, M. P. (2007). Methods for comparing the diversity of samples of the T cell receptor repertoire. *J. Immunol. Methods*, 321:182–195.
- Wallace, M. E., Bryden, M., Cose, S. C., Coles, R. M., Schumacher, T. N., Brooks, A., and Carbone, F. R. (2000). Junctional biases in the naive TCR repertoire control the CTL response to an immunodominant determinant of HSV-1. *Immunity*, 12:547–556.
- Wong, J., Obst, R., Correia-Neves, M., Losyev, G., Mathis, D., and Benoist, C. (2007). Adaptation of TCR repertoires to self-peptides in regulatory and nonregulatory CD4⁺ T cells. *J. Immunol.*, 178:7032–7041.
- Xekalaki, E. (1986). The bivariate yule distribution and some of its properties. *Statistics*, 17:311–317.

7. DIVERSITY AND OVERLAP BETWEEN REGULATORY AND EFFECTOR T CELL REPERTOIRES IN MICE

Partially based on the following papers:

Sepúlveda, N., Paulino, C. D. and Carneiro, J. (2008). Diversity of regulatory and effector T lymphocytes (in portuguese). In *Estatística: da teoria à prática* (M. M. Hill, M. A., Ferreira, J. G. Dias, M. F. Salgueiro, H. Carvalho, P. Vicente and C. Braumann, eds.), SPE Editions, Lisbon, p. 513-524.

Sepúlveda, N. and Carneiro, J. (2009). Biodiversity estimation: quantifying the immune repertoire in the body's cellular ecosystem. To appear in the Proceedings of 16th European Young Statistician Meeting (CD-Rom).

Abstract

The statistical analysis of the regulatory (T_R) and effector (T_E) TCR repertoires is a key point to understand how the immune system assures peripheral tolerance to self-antigens but, at the same time, is able to respond to harmful pathogens. Hitherto, this task has been done with simple statistical tools that either provide inaccurate TCR diversity estimates or do not quantify the diversity and overlap in terms of number of distinct clones present and shared among repertoires. Thus, current quantification of regulatory and effector T cell repertoires is still poor. To tackle this problem, we perform a meta-analysis of previously published data on T_R and T_E TCR repertoires using the Lognormal-Poisson model. Our results suggest that effector thymocytes exhibit higher TCR diversity than their regulatory counterparts, in agreement with the hypothesis that T_R cells are targeted to self-reactivity. More importantly, the correlation between clonal size distributions of thymic TCRs belonging to the respective intersection repertoire is positive, suggesting that T_R -cell differentiation is critically dependent on the abundance of each TCR sequence. We also demonstrate that thymic and peripheral T_R repertoires might intersect completely, and their respective TCR abundance distributions might be positively correlated. We argue that this complete intersection might be mainly due to mature T_R cells re-entering the thymus. Peripheral T_R and T_E cells exhibit similar TCR diversity, and their repertoires might share 5% to 100% TCR sequences. The respective clonal size distributions for TCRs belonging to the intersection repertoire might be either uncorrelated or positive correlated. To reach a more clearer conclusion for the intersection of these two repertoires, we recommend to determine the relative contribution of several populations, such as recent emigrants, peripheral converted T_R cells, and long-lived T cell clones, to the peripheral repertoires. We then conclude that this type of analysis improved the current knowledge on the diversity and shape of T_R and T_E cell repertoires in the thymus and periphery.

7.1 Introduction

It is well known that, even in healthy individuals, some autoreactive T cells might escape negative selection in the thymus, and enter the periphery. However, the proliferation and activation of these potentially disease-causing cells are in check by the so-called regulatory $CD4^+CD25^+$ T cells (T_R) expressing the forkhead transcription factor Foxp3 (Fontenot et al., 2003; Hori et al., 2003). T_R cells are mainly produced in the thymus, although they can also be induced in the periphery from other $CD4^+$ T cells, here collectively referred to as effector T cells (T_E), by homeostatic or antigen-driven cues [3–7]. In spite of their important role in the control of autoimmunity, accumulating data are now showing that T_R cells are able to dampen excessive inflammation and immune responses to many different microbial agents (Cabrera et al., 2008; Demengeot et al., 2006). Moreover, a recent study using T cell hybridomas shows that T_R cells are also prone to recognize and react to nonself-antigens (Pacholczyk et al., 2007). Facing these opposing evidences, it is difficult to envisage how the immune system can be efficient in the presence of those cells.

The solution to this conundrum seems to require a detailed description of T_R and T_E cell repertoires, usually done in terms of T cell receptor (TCR) these cells express. Because of the potentially large diversity of the TCR repertoires, extreme experimental systems are often used, in which TCR diversity is highly limited by some genetic-engineered trick [10–15]. Notwithstanding, available data are compelling to show that T_R and T_E cells are highly polyclonal [10–13], even under strict thymic selection (Pacholczyk and Kern, 2008; Wong et al., 2007). Moreover, TCR diversity seems higher in T_R cells than in their T_E counterparts, either in the periphery (Hsieh et al., 2004; Pacholczyk et al., 2007; Wong et al., 2007) or in the thymus (Hsieh et al., 2006; Pacholczyk et al., 2006). The repertoire of T_R cells partially overlaps with that of conventional T cells, but the extent of this intersection varies across different animal models (Pacholczyk et al., 2006; Wong et al., 2007) and anatomical regions (Pacholczyk and Kern, 2008). In healthy humans, it was also observed some overlap between T_R and T_E TCR samples in agreement with mouse data (Fazilleau et al., 2007).

In spite of the great effort in the above-mentioned TCR studies, the accompanying statistical analysis is not sufficiently rich to extract all relevant information from the data. On one hand, most of these reports use, and abuse of, the so-called MLE estimator (Barth et al., 1985; Behlke et al., 1985), which underestimates TCR diversity (Chapter 6). On the other hand, the statistical evaluation of the intersection repertoire is usually carried out by some similarity measure, such as the Jacard index or the Morisita-Horn index (Hsieh et al., 2004; Venturi et al., 2008). Yet, they provide limited information on the intersection repertoire, such as, how many distinct TCRs belong to that repertoire, and what is the correlation between clonal size distributions of TCRs shared by both cell types. Therefore, there is an urgent need to re-assess available data with alternative statistical methods in order to get a more accurate view of the underlying T_R and T_E cell repertoires.

Previously, we suggest the usage of the Poisson abundance models to estimate TCR diversity, and the underlying clonal size distributions. The Lognormal-Poisson model is of particular interest, because it can describe most TCR data, and therefore can be used to dissect the

main features of the intersection repertoire. Using this model, we perform a meta-analysis of previously published data on TCR repertoires of T_R and T_E cells (Hsieh et al., 2004, 2006), aiming to better quantify the diversity and the relationship between T_R and T_E -cell repertoires in the thymus and periphery.

7.2 Materials and Methods

7.2.1 Experimental data

Hsieh *et al.* (2004,2006) Hsieh et al. (2004, 2006) make use of the TCli TCR β transgenic mice to investigate the shape of peripheral and thymic T_R and T_E cell repertoires. In this experimental system, animals are on a TCR $\alpha^{+/-}$ background, which restricts T cells to express only one TCR α . Moreover, TCR β locus is composed of an endogeneous gene, and a fully rearranged transgene (V β 6). Since this locus operates under an allelic exclusion regime, only the transgene is expressed in each T cell. In both studies, it was studied the repertoire of regulatory and effector TRAV14 $^{+}$ (V α 2) CD4 $^{+}$ T cells by cloning and sequencing the endogeneously rearranged TCR α . TCR α sequences were defined at the amino acid level, and restricted to the highly variable CDR3 region, particularly important to determine the antigen specificity of the TCR. T_R and T_E cells were then identified as CD4 $^{+}$ CD25 $^{+}$ and CD4 $^{+}$ CD25 $^{-}$ T cells, respectively. Data under analysis are presented in Table 7.1.

7.2.2 Estimation of TCR diversity

TCR diversity in each sample is first analyzed by the Simpson's diversity index D_S (equation (6.9)). Since samples have different sizes, estimates for D_S should be standardized (Venturi et al., 2007). This can be obtained by the following bootstrap procedure (Venturi et al., 2007): (i) take the minimum sample size among samples under analysis, (ii) generate a sample, with that sample size, taken without replacement from an original sample, (iii) calculate D_S in the simulated sample, (iv) repeat previous steps several times (throughout 10000 runs per each original sample), (v) estimate D_S by the median of the respective simulated distribution.

We then estimate TCR diversity via Lognormal-Poisson model by maximizing the (log-)likelihood function. This hierarchical model belongs to the class of Poisson abundance models, which assumes that sampling of each TCR sequence follows a Poisson distribution with parameter λ , the so-called sampling rate (Chapter 6). Heterogeneity on the abundances of each TCR sequence is then introduced in the model by considering that the sampling rate is distributed as a Lognormal distribution. Other Poisson abundance models were also tested (homogeneous Poisson, Geometric, Gamma-Poisson, and Yule), but led invariantly to bad fits for the data. To evaluate the goodness-of-fit of the Lognormal-Poisson model, we apply Pearson's χ^2 test. Since χ^2 approximation might be poor in unbalanced samples (Agresti, 1992), such as those under analysis, we consider in most cases five abundance categories for the data: 1, 2, 3, 4, and > 4 repeats. Yet, in the analysis of thymic T_E cell samples, we use six abundance categories to improve the quality of the fit (i.e., 1, 2, 3, 4, 5, and > 5 repeats). We also study the robustness of TCR diversity estimates, as in previous chapter. The calculations were done in Mathematica and the R software. The level of significance was setup at 5%.

Tab. 7.1: Frequency of unique TCR sequences as a function of their repeats in peripheral T_R and T_E TCR samples from two experiments (Hsieh et al., 2004), and in peripheral and thymic T_R and T_E TCR samples from two distinct mice (Hsieh et al., 2006), where M is the total number of distinct TCR sequences in each sample, n is the respective sample size, and D_S is the Simpson's diversity index (in percentage).

Repeats	Mouse 1				Mouse 2				Peripheral T cells			
	Thymus		Periphery		Thymus		Periphery		Exp. 1		Exp. 2	
	T_R	T_E	T_R	T_E	T_R	T_E	T_R	T_E	T_R	T_E	T_R	T_E
1	123	105	102	80	113	127	107	79	72	63	83	65
2	23	16	15	14	23	7	13	15	17	8	15	9
3	7	1	2	7	6	4	1	6	1	5	8	6
4	0	4	5	4	1	1	1	3	2	2	0	2
5	1	1	1	3	0	0	2	1	2	3	1	2
6	0	1	0	2	0	0	1	1	—	1	0	0
7	0	0	1	1	0	0	0	0	—	—	1	1
8	2	0	0	—	0	1	0	0	—	—	—	1
9	0	0	0	—	0	0	1	0	—	—	—	0
10	0	0	1	—	0	1	—	1	—	—	—	0
11	0	0	—	—	2	—	—	—	—	—	—	0
13	0	1	—	—	0	—	—	—	—	—	—	0
14	0	—	—	—	0	—	—	—	—	—	—	1
17	0	—	—	—	1	—	—	—	—	—	—	—
18	1	—	—	—	—	—	—	—	—	—	—	—
M	175	129	127	111	146	141	126	108	94	82	108	87
n	229	180	180	179	220	175	165	160	127	123	149	148
D_S	99.0	99.1	99.2	99.1	98.8	99.4	99.3	99.1	99.4	98.9	99.4	0.98.2

7.2.3 Quantification of the intersection repertoire

The traditional approach to quantify the intersection between any two TCR repertoires is to compute measures that aim to capture the degree of similarity among samples. In this regard, the Jacard and the Morisita-Horn indexes are popular choices in TCR repertoire studies (Hsieh et al., 2004, 2006; Venturi et al., 2008). The Jacard index is calculated by the ratio between the number of TCR variants shared by two samples and the total number of distinct TCR variants observed in both samples. Note that this index does not take into account the frequency of TCR variants in the samples, which is done in the Morisita-Horn similarity index (Venturi et al., 2008). Mathematically, the Morisita-Horn index is defined by

$$C_{MH} = \frac{2 \sum_{i=1}^m f_i g_i}{\sum_{i=1}^m (f_i^2 + g_i^2)}, \quad (7.1)$$

where $f_i = n_{1i}/n_1$ and $g_i = n_{2i}/n_2$, n_{1i} and n_{2i} are the frequencies of i -th TCR variant in samples 1 and 2, n_1 and n_2 are the respective sample sizes, and m is the total number of distinct TCR variants observed in both samples. This index ranges from 0 and 1 (minimal and maximal similarity, respectively).

Here we adopt a parametric approach to study the intersection between any two TCR repertoires. With this approach, we expect to extract more information from the data, namely, the actual number of TCR variants shared by two repertoires and the correlation between

abundances of those shared TCR variants in each population. To this end, we need to divide each T cell repertoire in two subsets: one composed of TCR variants that are exclusive to each population (exclusive repertoire) and other composed of TCR variants that are common to both populations (the intersection repertoire). Since each sample is well described by the Lognormal-Poisson model, we extend it to two-sample case, as follows.

Let D_1 and D_2 (with $D_1 \leq D_2$) be the diversity of TCR variants in populations 1 and 2, respectively. Let π be the proportion of TCR diversity in population 1 due to the intersection repertoire. Thus, the diversity of the intersection repertoire is simply $O = D_1 \times \pi$. Diversity of each exclusive repertoire is then $E_1 = D_1 - O$ and $E_2 = D_2 - O$. Finally, $D_t = E_1 + E_2 + O$ is the total diversity of both populations.

The bivariate Lognormal-Poisson model for any two TCR overlapping repertoires assumes that the log sampling rates of TCR variants from each exclusive repertoire are random draws from two independent Normal distributions, i.e.,

$$\log \epsilon_{1i} \rightsquigarrow \mathcal{N}(\mu_1, \sigma_1^2), i = 1, \dots, E_1 \text{ and } \log \epsilon_{2j} \rightsquigarrow \mathcal{N}(\mu_2, \sigma_2^2), j = 1, \dots, E_2, \quad (7.2)$$

where ϵ_{1i} and ϵ_{2j} are the sampling rates of i -th and j -th TCR sequences in the respective exclusive repertoires of populations 1 and 2. To derive the model for the intersection repertoire, we assume that the marginal distributions of sampling rates are identical to the respective distribution in the exclusive repertoires. Since we are dealing with Normal distributions for the log sampling rates, the relationship between these rates in the intersection repertoire can be captured by a correlation coefficient ρ . Thus, the log sampling rates of TCR variants in the intersection repertoire follow a bivariate Normal distribution

$$\log \begin{bmatrix} \omega_{1k} \\ \omega_{2k} \end{bmatrix} \rightsquigarrow \mathcal{N} \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right), k = 1, \dots, O. \quad (7.3)$$

where ω_{1k} and ω_{2k} are the sampling rates of k -th TCR sequence belonging to the intersection repertoire in populations 1 and 2, respectively. Since the univariate Lognormal-Poisson model has no closed-form expression (Bulmer, 1974), the same happens for this two-sample version. Because of this, estimation of the model is cumbersome. Additional difficulty arises from the fact that most T cell repertoire studies do not provide the raw data (with frequency of each TCR variant in both samples), which are crucial to estimate properly the model. Therefore, a direct estimation of the model seems prohibited.

To tackle this problem, we estimate the above model in two stages: (i) obtain estimates for D_1 , D_2 , μ_1 , μ_2 , σ_1^2 and σ_2^2 by fitting the Lognormal-Poisson model to each individual TCR sample, and then (ii) perform massive stochastic simulation to determine values for π and ρ that are compatible with the observed Jacard or Morisita-Horn estimates, which are often available in this kind of studies. For sake of simplicity, the sampling rates belonging to different repertoires are placed in the following $D_t \times 2$ matrix

$$\Lambda = \begin{bmatrix} \epsilon_1 & \mathbf{0} \\ \omega_1 & \omega_2 \\ \mathbf{0} & \epsilon_2 \end{bmatrix}, \quad (7.4)$$

where $\epsilon'_1 = \{\epsilon_{11}, \dots, \epsilon_{1E_1}\}$, $\epsilon'_2 = \{\epsilon_{21}, \dots, \epsilon_{2E_2}\}$, $\omega'_1 = \{\omega_{11}, \dots, \omega_{1O}\}$ and $\omega'_2 = \{\omega_{21}, \dots, \omega_{2O}\}$. The simulation algorithm obeys then to the following steps: (i) obtain a sample for the sampling rates Λ given in equation (7.4); (ii) compute the probability vectors $\theta_1 = \Lambda_1 / \sum_{i=1}^{D_t} \lambda_{1i}$ and $\theta_2 = \Lambda_2 / \sum_{i=1}^{D_t} \lambda_{2i}$, where $\Lambda_1 = \{\lambda_{1i}\}$ and $\Lambda_2 = \{\lambda_{2i}\}$ refer to the first and second column of matrix Λ , respectively; (iii) generate two independent samples of TCR variant frequencies, one from a Multinomial distribution with sample size n_1 and probability vector θ_1 , and other also from a Multinomial distribution, but with sample size n_2 and probability vector θ_2 ; (iv) calculate the Jacard or the Morisita-Horn similarity index for the respective "paired" sample; (v) repeat the above steps K times to obtain a simulated distribution for the above indexes (throughout, $K = 1000$); (vi) test the goodness-of-fit of the model by the following empirical p-value

$$p = 2 \min \left\{ \hat{F}_X(x_{obs}), 1 - \hat{F}_X(x_{obs}) \right\}, \quad (7.5)$$

where $\hat{F}_X(x_{obs})$ is the simulated distribution function of either Jacard or the Morisita-Horn index, evaluated at the respective observed value x_{obs} . The algorithm is implemented in R language, and available from the authors upon request.

7.3 Results

7.3.1 Similar features among all TCR samples

All TCR samples show similar features as they are mainly composed of singletons and doubletons (TCR sequences with one and two repeats in the samples, respectively), and few highly abundant TCR variants (Tables 7.1). This pattern is captured by the Simpson's diversity index, leading to estimates close to maximal diversity. Same kind of pattern has been observed in previous chapter and in other studies (Naumov et al., 2003; Pewe et al., 2004; Wong et al., 2007). As discussed in previous chapter, the heavy tails observed in all TCR abundance distributions appear to contradict the standard assumption of an equal abundance of every TCR in the population. This was confirmed by the bad fit of the homogeneous Poisson model to data with p-values less than 10^{-5} for Pearson's goodness-of-fit test (results not shown). Interestingly, thymic T_R and T_E cells exhibit some TCRs in higher abundance than their peripheral subsets. In spite of different samples sizes among data sets, which may bias the analysis, this result indicates that the periphery has an effect of homogenizing the contribution of each TCR sequence for the respective clonal size distribution.

7.3.2 The number of distinct TCR sequences observed in the samples is far from representing TCR diversity at the population level

As in any statistical problem, the usage of adequate sample sizes is crucial to obtain accurate results. With this in mind, we investigate how far the number of distinct TCR sequences observed in the samples is from the respective TCR diversity at the population level. To this end, we simulate samples by adding sequentially to them a new TCR sequence, taken without replacement from the original samples, and testing whether this new sequence have already

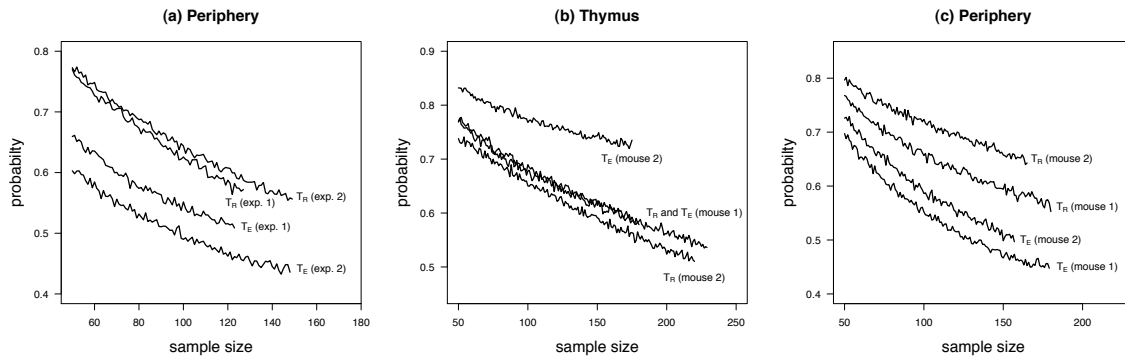


Fig. 7.1: Probability of obtaining a new TCR variant as a function of the respective sample size: (a) peripheral TCR samples from Hsieh et al. (2004), (b) and (c) thymic and peripheral TCR samples from Hsieh et al. (2006).

appeared in the generated sample. For a given sample size, the proportion of simulations in which the added TCR sequence increased the diversity in the generated sample is an estimate for the above-mentioned probability.

As expected, the probability of obtaining a new TCR variant in the sample is a decreasing function of the sample size (Figure 7.1). More importantly, this probability is in its best around 45% for peripheral T_E cells of mouse 1 (Figure 7.1(c)). This clearly demonstrates that sample sizes at hand are not sufficient to obtain all TCR sequences in the populations. Therefore, TCR diversity estimation assumes a key role in the whole analysis.

7.3.3 Thymic T_E cells show higher TCR diversity than their thymic T_R counterparts

To tackle the diversity of thymic populations, we first test whether the respective samples could have the same TCR abundance distribution. This hypothesis is in the borderline of statistical significance ($p = 0.04$, Pearson's independence test). To dissect this result, we then ask if each thymic subset in either mouse could exhibit similar TCR abundance distribution. This hypothesis seems true for thymic T_R cells ($p = 0.87$), but in the borderline of statistical significance for their T_E counterparts ($p = 0.06$). The latter result suggests a certain mouse-to-mouse heterogeneity in the data in the sense of having more singletons in mouse 1 than in mouse 2. Nonetheless, for a matter of simplicity, we assume that samples of the same thymic subset have similar TCR abundance distributions, but different ones in each cell type.

We then fit the Lognormal-Poisson model to the data. To this end, we consider that model with the same parameters for each thymic sample. The respective fits have good statistical quality, even when discarding the highly abundant TCR variants from the analysis (Figure 7.2(a)). Yet, for thymic T_E samples, the respective TCR diversity estimates are extremely large and sensitive to highly abundant TCR variants in the samples (Figure 7.2(b)). In fact, TCR diversity estimate drop almost one-fold when the highest TCR sequence with 13 members in the sample from mouse 1 was not included in the estimation of the model. Yet, the quality of the fit of the model decreases with the absence of the high abundant TCR sequences in the estimation stage. To complement the analysis of this case, we calculate the minimum

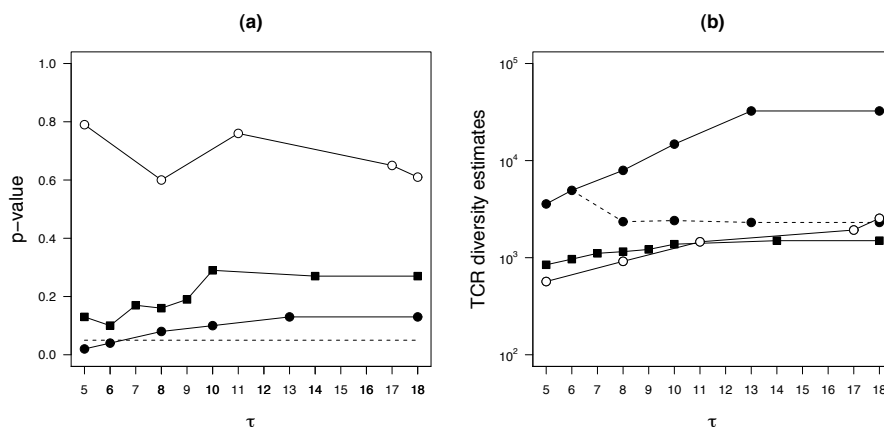


Fig. 7.2: Analysis of TCR diversity estimates based on the Lognormal-Poisson model ($\circ$ — thymic T_R cells, $\bullet$ — thymic T_E cells, and $\blacksquare$ — peripheral T cells), where τ is the maximum number of TCR repeats included in the estimation of the models. Plots (a) and (b) refer to the p-value of the Pearson's goodness-of-fit test and the underlying TCR diversity estimates, respectively. Dashed lines in (a) and (b) are the 5% significance level and lowest TCR diversity estimates for thymic T_E that provide a good fit of the model at 5% significance level, respectively.

Tab. 7.2: Outlier TCR sequences in thymic T_R samples and their number of repeats across samples (Hsieh et al., 2006).

Sequence	Mouse 1				Mouse 2			
	Thymus		Periphery		Thymus		Periphery	
	T_R	T_E	T_R	T_E	T_R	T_E	T_R	T_E
G25	18	1	4	0	17	0	5	0
G41	8	0	2	0	11	0	2	0
G69	8	3	7	1	11	2	9	2

TCR diversity that accept the model at 5% significance level (dashed line in Figure 7.2(b)). Interestingly, this lower limit for thymic T_R TCR diversity is robust to the high abundant TCR sequences, being almost constant along the different values of τ . More importantly, this limit is always higher than the corresponding TCR diversity estimate for thymic T_R cells. Thus, thymic T_E cells seem more diverse than their T_R counterparts.

Finally, thymic T_R subset is well described by the Lognormal-Poisson model, being the respective p-values higher than 0.60 (Figure 7.2(a)). Nonetheless, the underlying TCR diversity estimates decreases steadily from 2543 ($\tau = 18$) to 568 ($\tau = 5$), when excluding the high abundant TCR sequences from the fit of the model. Interestingly, the top 3 abundant TCR are common to both thymic T_R samples (Table 7.2), being responsible for the above-mentioned decrease in the TCR diversity estimates. Because of this, we discarded them from the estimation of the model. By doing this, the remaining thymic T_R TCR diversity is estimated in 568 unique TCR sequences in the population.

Tab. 7.3: Estimates of Morisita-Horn index between thymic and peripheral T cell compartments. Further estimates can be found in the original reference (Hsieh et al., 2006).

		Thymic T_E	Peripheral T_R	Peripheral T_E
Mouse 1	Thymic T_R	0.16	0.38	0.06
	Thymic T_E	—	0.08	0.50
Mouse 2	Thymic T_R	0.16	0.44	0.14
	Thymic T_E	—	0.09	0.46

7.3.4 Peripheral T_R and T_E cells exhibit similar TCR diversity

In the periphery, the observed number of unique TCR variants, and also the Simpson's diversity index, are higher in T_R samples than in their T_E counterparts (Table 7.1). Following this observation, Hsieh et al. (2004) suggest that peripheral T_R cells are more diverse than peripheral T_E ones. Yet, this observation was not previously confirmed by any supporting statistical analysis. To improve previous analysis, we first test if all peripheral samples could have the same TCR abundance distribution. This hypothesis could not rule out by the data ($p = 0.36$, Pearson's independence test). Because of this, we fit the Lognormal-Poisson model with the same parameters to all peripheral samples. The results show that the model is satisfactory to describe the pattern observed in all peripheral samples (Figure 7.2(a)). Yet, the quality of the respective fits is not as high as in thymic T_R samples. In fact, the respective p-values decreases as the high abundant TCR sequences are successively excluded from the estimation of the model. Notwithstanding, TCR diversity estimates show some robustness to abundant TCR sequences (Figure 7.2(b)). Since the best fit was obtained for all data, we conclude that peripheral T_R and T_E cells show similar TCR diversity, being estimated in 1497 distinct TCR sequences in these populations.

7.3.5 Thymic T_R -cell repertoire might share all TCR variants with that of thymic T_E cells

The next step of the analysis is to study the intersection between different T cell compartments. We first compare the intersection between thymic T_R and T_E cell repertoires. In this regard, Hsieh et al. (2006) compare these repertoires via Morisita-Horn similarity index (Table 7.3). The respective estimates from each mouse are in agreement with each other. The extent of the intersection seems small, but this might only be due to a higher TCR diversity in thymic T_E cells than in their T_R counterparts, as discussed above. The different samples sizes might also be another reason to decrease similarity between samples.

To dissect the intersection between these two repertoires, we fit the bivariate Lognormal-Poisson to data using massive Monte Carlo simulation. Previously, we consider the three most abundant TCR sequences in thymic T_R samples as potential outliers. Because of this, when we compared thymic T_R cell repertoire with the remaining ones, we set fixed the original abundances of these outlier TCR sequences in the generated TCR samples (Table 7.2). Furthermore, for thymic T_E repertoires, we use the lower limit of TCR diversity in the simulations, since the respective maximum likelihood diversity estimate seems unrealistic.

For mouse 1, our results suggest that 35% to 100% of TCR variants belonging to thymic T_R repertoire are also present in their thymic T_E counterpart (Figure 7.3(a)). Equivalently, 9%

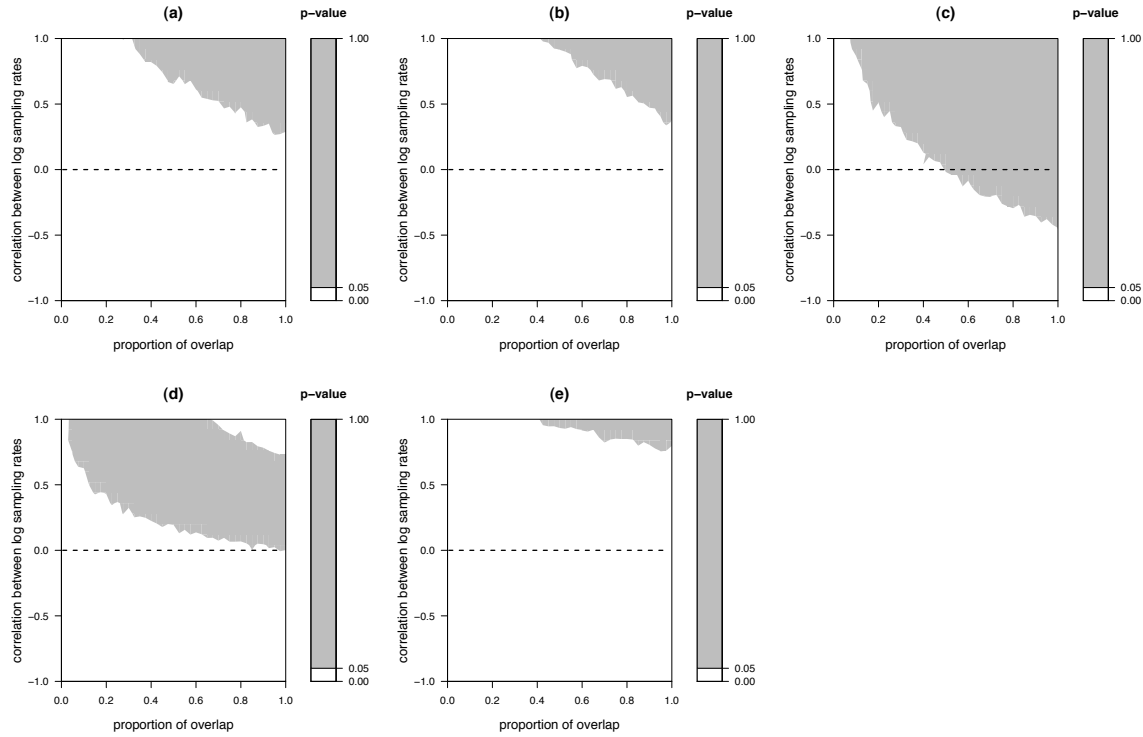


Fig. 7.3: Analysis of the intersection repertoire based on mouse 1 data: contour plots of p-values for fitting the Lognormal-Poisson model to the data as a function of the proportion of TCR sequences shared between pairs of repertoires, and correlation between log sampling rates. Data from mouse 2 follow similar patterns (results not shown). (a), (b) and (c) Comparison of thymic T_R repertoire with thymic T_E , peripheral T_R , and peripheral T_E cell repertoires, respectively. (d) and (e) Comparison of thymic T_E repertoire with peripheral T_R and T_E cell repertoires, respectively. Since diversity is higher in thymic T_E cells than in the periphery, the proportions of the shared TCR diversity were calculated in relation to each peripheral T cell subset. Parameters: thymic T_R cells (with three outliers, see table 7.2) — $\hat{D} = 568$, $\hat{\mu} = -1.35$ and $\hat{\sigma}^2 = 0.49$; thymic T_E cells — $\hat{D} = 2307$ (lower limit for diversity), $\hat{\mu} = 3.95$ and $\hat{\sigma}^2 = 2.72$; and peripheral T cells — $\hat{D} = 1497$, $\hat{\mu} = -3.92$ and $\hat{\sigma}^2 = 3.34$.

to 25% of thymic T_E TCR sequences are shared with the respective thymic T_R TCR repertoire. Interestingly, data show strong evidences for a positive correlation between log sampling rates for the TCR variants belonging to the intersection repertoires, which might reflect the same kind of relationship between the underlying log clonal size distributions. For mouse 2, similar results were obtained (results not shown). Yet, it is worth noting that the extent of TCR sharing seems higher in this animal: 65% to 100% of TCR diversity is shared between thymic T_R and T_E cell repertoires. Since we use the lower limit for thymic T_E TCR diversity, the above-mentioned estimates are clearly an underestimation of the true size of the intersection repertoire. Therefore, we conclude that thymic T_R cell repertoire overlaps in great extent its thymic T_E cell counterpart.

7.3.6 Thymic repertoire of the same T cell subset also overlaps greatly its peripheral counterpart

We then compare thymic T cell repertoires with their peripheral ones. Accordingly to Morisita-Horn index, the largest similarities are obtained when comparing thymic samples with pe-

ipheral ones for the same cell type (Table 7.3). This was somehow confirmed by simulation of bivariate Lognormal-Poisson model. In fact, for mouse 1, peripheral T_R TCR diversity accounts 50% to 100% for thymic T_R diversity (Figure 7.3(b)). Equivalently, thymic T_R TCR diversity represents 19% to 38% of that in the periphery. Moreover, thymic T_E diversity contributes 42.5% to 100% to that in the periphery (Figure 7.3(e)). In this case, it is expected that the size of the intersection repertoires between T_E repertoires might be greater than 42.5%, because we used in the simulations the lower limit for TCR diversity of thymic T_E cells. Interestingly, the log sampling rates of TCR sequences belonging to the above intersection repertoire are once again strongly and positively correlated.

The shared TCR diversity between thymic T_R and peripheral T_E cell repertoires ranges from 10% to 100% (Figure 7.3(c)). The correlation between log sampling rates is dominated by positive values, but it might also be negative for a large intersection between the repertoires. We could not rule out that all peripheral T_R TCR sequences might be common to its thymic T_E counterpart. The correlation between log sampling rates is always positive irrespective of the proportion of the overlap compatible with the data. Yet, when the intersection is extremely high, the model predicts an almost null correlation between log sampling rates.

7.3.7 T_R and T_E cells might share few or all TCR variants in the periphery

The study of the intersection between peripheral T_R and T_E -cell repertoires will be done in data from Hsieh et al. (2004), since Hsieh et al. (2006) did not provide any estimate for the similarity between samples of these repertoires. Previously, Hsieh et al. (2004) apply the Jacard index to dissect the similarity among samples. This measure was estimated around 0.025 for T_R and T_E data from either experiment, indicating that few TCR variants are shared among samples. Similarity increases when comparing samples of the same T cell subset, but from different experiments (0.15 for T_R samples, and 0.08 for T_E samples).

Massive stochastic simulation of the bivariate Lognormal-Poisson was then performed to determine the proportion of overlap and correlation between log sampling rates compatible with the above Jacard index estimates. We focus our analysis on the intersection between T_R and T_E cell repertoires. The results are inconclusive (Figure 7.4(a) and (b)). On one hand, the proportion of overlap might be as low as 5-10% with a correlation between log sampling rates close to 1. On the other hand, we cannot exclude the possibility of all TCR sequences being present in both repertoires, but this would require uncorrelated log sampling rates of TCR sequences belonging to the intersection repertoire.

It was shown above that some TCR variants are common to thymic and peripheral TCR repertoires. Thus, it is reasonable to expect that the peripheral T_R and T_E repertoires would share some TCR variants due to the above-mentioned overlap. We then calculate the limits for the number of TCR sequences shared across all T cell compartments. To this end, we multiply the limits for the intersection repertoire size between thymic T_R cells and the remaining T cell compartments, as shown in Figure 7.4(c). The calculation reveals that the overlap size due to thymic TCRs might account 0.6% to 38% for the peripheral TCR diversity.

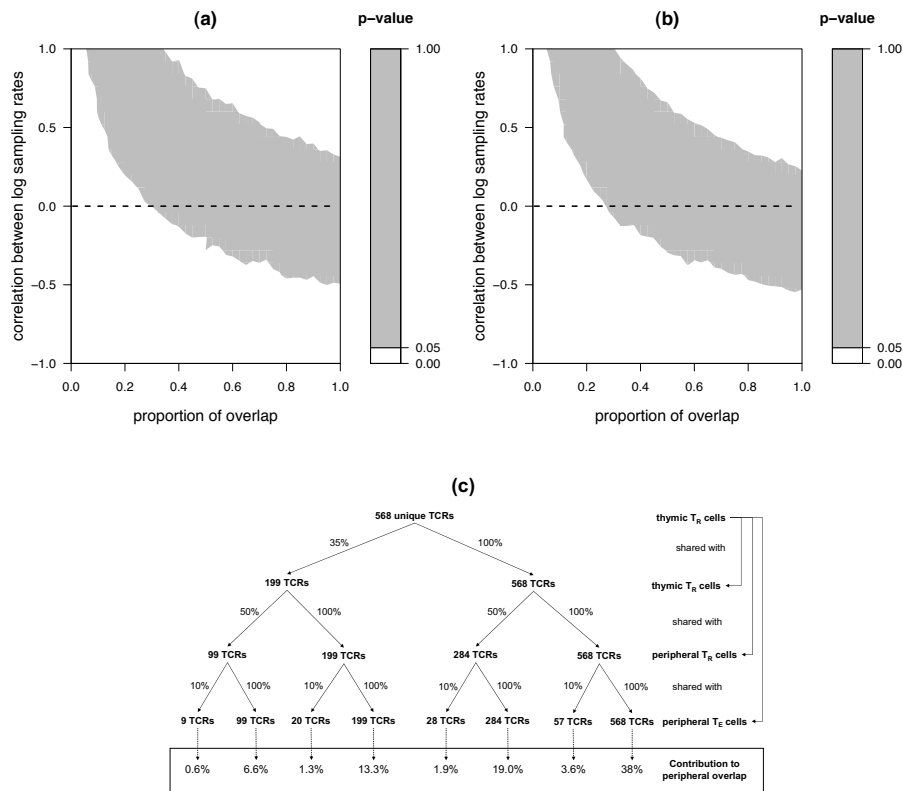


Fig. 7.4: Comparison between peripheral T_R and T_E TCR repertoires (data from Hsieh et al. 2004). (a) and (b) Contour plots for data from experiments 1 and 2, respectively. See legend of Figure 7.3 for parameter values and further information. (c) Limits for the contribution of thymic TCR sequences to peripheral intersection repertoire.

7.4 Discussion

In this chapter we perform a meta-analysis of previously published data on thymic and peripheral T_R and T_E cell repertoires. An interesting observation is that all analyzed samples are composed of many rare clonotypes and just a few highly abundant clonotypes. In the thymus, this pattern might be in part explained by a bias in the V(D)J recombination towards certain TCR sequences. In fact, in experimental settings where TCR diversity is somehow limited (Correia-Neves et al., 2001), we could rule out the homogeneous Poisson model for double positive thymocytes (see Table 6.2), way before positive and negative selections occur, suggesting that some TCR sequences are more probable of being produced than others. In the same line of evidence, some studies show that the imprecise joining of different gene segments tends to prefer certain amino acid motifs (Nadel and Feeney, 1995, 1997). With respect to the periphery, a recent study shows that some TCR sequences are only present in certain anatomical compartments, while others appear in different peripheral locations (Lathrop et al., 2008). Since peripheral samples here analyzed result from pooling together splenic and lymph node T cell data, it is reasonable to assume that rare TCR are distributed across distinct anatomical sights, while the highly abundant TCRs are common to different locations.

We use the Lognormal-Poisson model to study the diversity and overlap between thymic and peripheral T_R and T_E TCR repertoires, because every data set here analyzed could be well described by this model. Similar good fit of the Lognormal-Poisson model was obtained in the analysis of thymic and peripheral $CD4^+$ and $CD8^+$ T cells from other TCR-limited experimental settings (Chapter 6). Other models were tested, but fail to show good statistical quality when fitted to the data. When each TCR sequence represents unequivocally a single cell of the population under study, the theory developed in previous chapter predicts that the underlying clonal size distribution for the Lognormal-Poisson model is a Lognormal distribution. However, TCR data at hand were collected from a PCR-amplified TCR cDNA library, and thus one cannot distinguish whether two repeats of the same TCR sequence represent a single or two cells. A Lognormal distribution is then predicted for the TCR abundance distribution in the amplified cDNA library. To discuss the shape of the underlying clonal size distribution of an individual, we consider a simple exponential growth model for the PCR amplification stage, that is, $x_i = x_{i,0}e^{p_i k}$, where $x_{i,0}$ and x_i are the initial and final amplified quantity of i -th TCR sequence, respectively, p is the amplification rate, k is the number of PCR cycles. Because of the small difference in CDR3 length among TCR sequences, it seems reasonable to assume similar amplification efficiency across the population (that is, $p_i = p$). Under this assumption, the Lognormal clonal size distribution of the amplified populations reflects then that of an individual, but scaled by the constant e^{pk} . Similar probability distribution was also obtained in previous chapter with the analysis of single-cell PCR data, where the clonal size distribution can actually be extrapolated from the data. Thus, we speculate that the Lognormal clonal size distribution might be a general feature of T-cell repertoires. A more detailed discussion on the mechanisms generating a Lognormal clonal size distribution can be found in Chapter 6 and in the General discussion of this thesis.

T_R commitment in the thymus is currently under debate (reviewed in Pacholczyk and Kern 2008). One point that everyone agrees is that TCRs of T_R precursors are biased to high self-reactivity. Yet, the mechanisms by which this occurs is still controversial. On the one hand, some authors support a TCR-independent instructive model where T cell precursors divert to T_R lineage either by trans-acting factors affecting the double-negative $CD4^-CD8^-$ subset (Pennington et al., 2006) or upon encountering self-antigens or their agonist ligand (Apostolou et al., 2002; Ribot et al., 2006). On the other hand, other authors suggest a stochastic-selective model in which the bias towards high self-reactivity is explained by a strong resistance of T_R precursors to deletion during negative selection stage [31-33]. A two-step process has been recently proposed for thymic T_R development (Lio and Hsieh, 2008), in which pre-committed T_R precursors received a first TCR-dependent signal followed by an additional TCR-independent signal to sustain Foxp3 expression in those cells. Yet, the true nature of the first signal is still unclear (Lio and Hsieh, 2008), and thus it is difficult to predict the relationship between T_R and T_E TCR diversities.

According to the Lognormal-Poisson model, TCR diversity for thymic T_R cells is estimated around 570 TCR unique sequences (with 3 outlier TCR variants), a slightly higher estimate than that previously reported (Hsieh et al., 2006). The same model predicts a higher TCR diversity for thymic T_R cells, but the accuracy of this estimate is unclear. Nonetheless, it

suggests that thymic TCR repertoire is more diverse than its T_R counterpart. Moreover, we could not rule out the possibility that all thymic T_R TCRs are present in its thymic T_E TCR counterpart. This seems in agreement with the above-mentioned feature of T_R cells being particularly prone to recognize self-antigens, and therefore less diverse than thymic T_E cells. More importantly, a positive correlation between log clonal size distributions in the respective intersection repertoire is predicted across the different overlap sizes compatible with the data. This positive correlation is a statement on the T_R development program itself. In fact, a null correlation suggests that the abundance of each TCR sequences in the thymus does not play a role in the process of T_R differentiation. In this line of thought, the specificity of the TCRs would be a good way to divert T precursors to T_R lineage. When a positive or a negative correlation is obtained, it points to a T_R selection process somehow dependent on the abundance of each TCR sequence in the thymus. On the one hand, a positive correlation is compatible to the above-mentioned instructive-selection model that thymic encounter with cognate antigens facilitates T_R cell development (Apostolou et al., 2002; Ribot et al., 2006). In this regard, more abundant clones should have higher probabilities of encountering their cognate antigens than less abundant ones. On the other hand, a negative correlation is predicted by the stochastic-selective model, since T_R cells are thought to resist more to negative selection. Thus, more deletion of T_E precursors should occur during negative selection with the increase of self-reactivity of their TCRs. For all of this, we conclude that the thymic T_R development is based more on an instructive-selection model rather than a stochastic one.

The analysis of peripheral TCR repertoires suggests that T_R cells are as diverse as their T_E counterparts. Yet, this result should not be taken literally because we analyzed jointly data taken from two distinct studies (Hsieh et al., 2004, 2006), which used different strategies to obtain TCR samples. In this regard, each sample of Hsieh et al. (2004) is a TCR pool from different mice, as done also in other studies (Lathrop et al., 2008; Pacholczyk et al., 2006, 2007; Wong et al., 2007). By doing this, the subsequent TCR diversity estimates are for the overall TCR diversity of two mice in each peripheral compartment, $D_T = O + E_1 + E_2$, as defined in Subsection 7.2.3. In contrast, samples of Hsieh et al. (2006) are discriminated by mice. Thus, it is expected that peripheral TCR diversity in samples of Hsieh et al. (2004) should be in theory higher than those in Hsieh et al. (2006). Because of the small samples sizes used in both studies, we could not confirm this hypothesis.

We showed that thymic T_R TCR diversity accounts 19%-38% for the peripheral T_R one. The respective clonal size distributions in the intersection repertoire seems strongly and positively correlated. A fraction of the above intersection might be due to recent thymic emigrants (Figure 7.4). However, they represent a small fraction ($\approx 1\%$) of the peripheral T cell pool (Berzins et al., 1998; Scollay et al., 1980). Since thymus exports T cells to the periphery as an intermittent rather than a continuous process (Foss et al., 2001), it is reasonable to conceive that, when thymic "gates" are open, some T cells bearing the same TCR might go out to the periphery, while others would be retained inside the thymus. If one assumes that TCR does not affect the export of a T cell to periphery, it is expected that the clonal size distribution of recent thymic emigrants is positively correlated with that of the "retained" thymocyte counterparts. Other population that might also contribute to overlap is of mature T cells that

re-enter the thymus (Bosco et al., 2009; Hale and Fink, 2009). Estimates for the fraction of these recirculating T_R cells in $CD25^+CD4^+$ thymocyte compartment vary from 20% in one study (Zhan et al., 2007) to 50% in another study (McCaughy et al., 2007). In spite of the great discrepancy between estimates, these observations suggest that a large extent of the overlap between thymic and peripheral T_R TCR repertoires might be due to these recirculating T cells. T cell recirculation does not appear to be a TCR-driven process (Kirberg et al., 2008), and thus it is reasonable to assume that the probability of a given T cell re-entering the thymus seems intimately dependent on its relative frequency in the periphery, in agreement with a positive correlation between thymic and peripheral T_R clonal size distributions of the intersection repertoire. Finally, since TCR repertoire formation in the thymus of these mice seems biased towards some TCR sequences, it is also expected that a fraction of the overlap might be attributed to these recurrent TCR sequences.

Our results suggest that 10% to 100% TCR sequences might be shared between both peripheral T_R and T_E TCR repertoires, predicting a positive and a null correlation between clonal size distributions, respectively. As discussed above, a fraction of the overlap repertoire might be due to recent thymic emigrants that can range from 1% to 38% (Figure 7.4(c)). Since recent thymic emigrants represent a small fraction ($\approx 1\%$) of the peripheral T cell pool (Berzins et al., 1998; Scollay et al., 1980), it is unlikely a contribution of 38% of these cells for the intersection repertoire. Other fraction might be attributed to peripheral converted T_R cells. The extent of the peripheral conversion varies significantly across experimental systems (de Lafaille et al., 2004; Liang et al., 2005), but yet it only accounts for 4-7% of peripheral T_R subset in the mouse model here analyzed (Lathrop et al., 2008). Interestingly, peripheral conversion seems a TCR-specific mechanism (Lathrop et al., 2008). With this observation in mind, peripheral conversion might decrease the size of the intersection repertoire by converting some T_E cells bearing TCRs of the respective exclusive T_E repertoire. On the other hand, peripheral conversion might also decrease the correlation between clonal size distributions in the intersection repertoire by converting some T_E cells which express TCRs belonging to the intersection repertoire. Both explanations are equally possible, and need to be tackled experimentally. Another fraction of the peripheral intersection repertoire might come from TCRs of memory-like T cell clones with both cell types. In this case, the Crossregulation model predicts that T_R and T_E cell densities of these "long-lived" clones are negatively correlated (Figure 5.7F of Chapter 5). Putting all these observations together, it is reasonable to think that the overall overlap size and correlation between peripheral T_R and T_E TCR clonal size distributions reflect the relative contribution of TCRs belonging to recent thymic emigrants, peripheral converted T_R cells and "long-lived" T clones. However, available data are not enough to disentangle the contribution of each of the above populations to the intersection repertoire. Nonetheless, it seems unlikely a full overlap between peripheral T_R and T_E cell repertoires because, if this happens, most immune responses would be under control of T_R cells, even those against harmful microorganisms. A low overlap seems also unlikely because, as discussed above, several populations might contribute to the intersection repertoire. Therefore, the intersection between T_R and T_E TCR repertoires should be moderate. Further TCR studies with increased sample sizes are then needed to provide a more accurate estimate for the overlap size of these repertoires.

In this paper we did not focus on the mouse-to-mouse heterogeneity and its implications. Yet, it is interesting to observe that thymic T_R cell samples from different mice show the same top three abundant TCR sequences, labeled as G25, G41 and G69 in (Hsieh et al., 2006). Because of this, the respective Morisita-Horn index estimate is very high ($C_{MH} = 0.71$), but yet 88% of its value comes from those three TCR sequences. In spite of observing some mouse-to-mouse heterogeneity in the remaining TCR sequences, this yet points to a certain selection bias towards some TCR sequences in the formation of thymic T_R cell repertoires. A simple explanation for this bias is that these "public" TCRs (that are shared among mice) might be important to assure peripheral tolerance, being either highly cross-reactive towards self-antigens or targeted to ubiquitous self-antigens. Yet, this possibility is not clearly confirmed when comparing thymic frequencies of these TCR sequences with peripheral ones. In fact, in spite of all being present in the peripheral samples, only one of these highly abundant sequences (G69) increases its (relative) frequency in peripheral samples. The (relative) frequency of the remaining sequences drops in the periphery. Moreover, we could rule out the hypothesis of equal frequency of these three among thymic and peripheral samples from mouse 2 but not in mouse 1 ($p = 0.01$ and $p = 0.07$, respectively; Pearson's χ^2 test). To blur even further the analysis, G25 and G69 are on the top of most abundant sequences in the peripheral T_R sample from experiment 1 in Hsieh et al. (2004), but not in its respective counterpart from experiment 2. Notwithstanding, when T cells were retrovirally transduced with T_R TCR α genes and were adoptively transferred into RAG1-deficient hosts, only T cells transduced with G69 TCR sequence could expand in the periphery (Hsieh et al., 2004). This result suggests that the high frequency of some thymic T_R sequences is not directly correlated to their high affinity to self antigens. For all of these, it is still unclear what is the reason and the role of these public TCR sequences in the T_R repertoire.

In summary, this chapter reflects a first attempt to quantify thymic and peripheral T_R and T_E repertoires in terms of number of distinct TCRs in the populations, the number of TCRs shared among repertoires, and the relationship between the abundances of each TCR belonging to the intersection repertoire of the respective populations. We demonstrate that a more quantitative analysis provides more information on the diversity and shape of T_R and T_E cell repertoires. We conclude that this type of approach should be followed in future TCR studies to achieve a more clear view of how T_R and T_E cell repertoires are shaped towards an efficient immune system.

Bibliography

- Agresti, A. (1992). A Survey of Exact Inference for Contingency Tables. *Statistical Science*, 7:131–177.
- Apostolou, I., Sarukhan, A., Klein, L., and von Boehmer, H. (2002). Origin of regulatory T cells with known specificity for antigen. *Nat. Immunol.*, 3:756–763.
- Apostolou, I. and von Boehmer, H. (2004). In vivo instruction of suppressor commitment in naive T cells. *J. Exp. Med.*, 199:1401–1408.

- Barth, R. K., Kim, B. S., Lan, N. C., Hunkapiller, T., Sobieck, N., Winoto, A., Gershenfeld, H., Okada, C., Hansburg, D., Weissman, I. L., and Hood, L. (1985). The murine T-cell receptor uses limited repertoire of expressed V_{β} gene segments. *Nature*, 316:517–523.
- Behlke, M. A., Spinella, D. G., Chou, H. S., Sha, W., Hartl, D. L., and Loh, D. Y. (1985). T-cell receptor β -chain expression: dependence on relatively few variable region genes. *Science*, 229:566–570.
- Berzins, S. P., Boyd, R. L., and Miller, J. F. (1998). The role of the thymus and recent thymic migrants in the maintenance of the adult peripheral lymphocyte pool. *J. Exp. Med.*, 187:1839–1848.
- Bonasio, R., Scimone, M. L., Schaerli, P., Grabie, N., Lichtman, A. H., and von Andrian, U. H. (2006). Clonal deletion of thymocytes by circulating dendritic cells homing to the thymus. *Nat. Immunol.*, 7:1092–1100.
- Bosco, N., Kirberg, J., Ceredig, R., and Agens, F. (2009). Peripheral T cells in the thymus: have they just lost their way or do they do something? *Imm. Cell Biol.*, 87:50–57.
- Bulmer, M. G. (1974). On fitting the Poisson Lognormal distribution to species-abundance data. *Biometrics*, 30:101–110.
- Cabrera, G., Burzyn, D., Mundiano, J., Courreges, M. C., Camicia, G., Lorenzo, D., Costa, H., Ross, S. R., Nepomnaschy, I., and Piazzon, I. (2008). Early increases in superantigen-specific Foxp3+ regulatory T cells during mouse mammary tumor virus infection. *J. Virol.*, 82:7422–7431.
- Correia-Neves, M., Waltzinger, C., Mathis, D., and Benoist, C. (2001). The shaping of the T cell repertoire. *Immunity*, 14:21–32.
- de Lafaille, M. A. C., Lino, A. C., Kutchukhidze, N., and Lafaille, J. J. (2004). CD25- T cells generate CD25+Foxp3+ regulatory T cells by peripheral expansion. *J. Immunol.*, 173:7259–7268.
- Demengeot, J., Zelenay, S., Moraes-Fontes, M. A., Caramalho, I., and Coutinho, A. (2006). Regulatory T cells in microbial infection. *Springer Semin. Immun.*, 28:41–50.
- Fazilleau, N., Bachelez, H., Gougeon, M. L., and Viguier, M. (2007). Cutting edge: size and diversity of CD4⁺CD25^{high} Foxp3⁺ regulatory T cell repertoire in humans: evidence for similarities and partial overlapping with CD4⁺CD25 T cells. *J. Immunol.*, 179:3412–3416.
- Fontenot, J. D., Gavin, M. A., and Rudensky, A. Y. (2003). Foxp3 programs the development and function of CD4⁺CD25⁺ regulatory T cells. *Nat. Immunol.*, 4:330–336.
- Foss, D. L., Donskoy, E., and Goldschneider, I. (2001). The importance of hematogeneous precursors by the thymus is a gated phenomenon in normal adult mice. *J. Exp. Med.*, 193:365–374.

- Hale, J. S. and Fink, P. J. (Hale:2009). Back to the thymus: peripheral T cells come home. *Imm. Cell Biol.*, 87:58–64.
- Hori, S., Nomura, T., and Sakaguchi, S. (2003). Control of regulatory T cell development by the transcription factor Foxp3. *Science*, 299:1057–1061.
- Hsieh, C.-S., Liang, Y., Tyznik, A. J., Self, S. G., Liggitt, D., and Rudensky, A. Y. (2004). Recognition of the peripheral self by naturally arising CD25⁺ CD4⁺ T cell receptors. *Immunity*, 21:267–277.
- Hsieh, C.-S., Zheng, Y., Liang, Y., Fontenot, J. D., and Rudensky, A. Y. (2006). An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nat. Immunol.*, 7:401–410.
- Kirberg, J., Bosco, N., Deloulme, J.-C., Ceredig, R., and Agens, F. (2008). Peripheral T lymphocytes recirculating back into the thymus can mediate thymocyte positive selection. *J. Immunol.*, 181:1207–1214.
- Komatsu, N., Mariotti-Ferrandiz, M. E., Wang, Y., Malissen, B., Waldmann, H., and Hori, S. (2009). Heterogeneity of natural Foxp3⁺ T cells: A committed regulatory T-cell lineage and an uncommitted minor population retaining plasticity. *Proc. Natl. Acad. Sci. USA*, 106:1903–1908.
- Lathrop, S. K., Santacruz, N. A., Pham, D., Luo, J., and Hsieh, C.-S. (2008). Antigen-specific peripheral shaping of the natural regulatory T cell population. *J. Exp. Med.*, 205:3105–3117.
- Liang, S., Alard, P., Zhao, Y., Parnell, S., Clark, S. L., and Kosiewicz, M. M. (2005). Conversion of CD4⁺ CD25⁻ cells into CD4⁺ CD25⁺ regulatory T cells in vivo requires B7 costimulation, but not the thymus. *J. Exp. Med.*, 201:127–137.
- Lio, C.-W. J. and Hsieh, C.-S. (2008). A two-step process for thymic regulatory T cell development. *Immunity*, 28:100–111.
- Liston, A., Lesage, S., Wilson, J., Peltonen, L., and Goodnow, C. C. (2003). Aire regulates negative selection of organ-specific T cells. *Nat. Immunol.*, 4:350–354.
- McCaughy, T. M., Wilken, M. S., and Hogquist, K. A. (2007). Thymic emigration revisited. *J. Exp. Med.*, 204:2513–2520.
- Nadel, B. and Feeney, A. J. (1995). Influence of coding-end sequence on coding-end processing in V(D)J recombination. *J. Immunol.*, 155:4322–4329.
- Nadel, B. and Feeney, A. J. (1997). Nucleotide deletion and P addition in V(D)J recombination: a determinant role of the coding-end sequence. *Mol. Cell. Biol.*, 17:3768–3778.
- Naumov, Y. N., Naumova, E. N., Hogan, K. T., Selin, L. K., and Gorski, J. (2003). A fractal clonotype distribution in the CD8⁺ memory T cell repertoire could optimize potential for immune responses. *J. Immunol.*, 170:3994–4001.

- Pacholczyk, R., Ignatowicz, H., Kraj, P., and Ignatowicz, L. (2006). Origin and T cell receptor diversity of Foxp3⁺CD4⁺CD25⁺ T cells. *Immunity*, 25:249–259.
- Pacholczyk, R. and Kern, J. (2008). The T-cell receptor repertoire of regulatory T cells. *Immunology*, 125:450–458.
- Pacholczyk, R., Kern, J., Singh, N., Iwashima, M., Kraj, P., and Ignatowicz, L. (2007). Nonself-antigens are the cognate specificities of Foxp3⁺ regulatory T cells. *Immunity*, 27:493–504.
- Pennington, D. J., Silva-Santos, B., Silberzahn, T., Escrcio-Correia, M., Woodward, M. J., Roberts, S. J., Smith, A. L., Dyson, P. D., and Hayday, A. C. (2006). Early events in the thymus affect the balance of effector and regulatory T cells. *Nature*, 444:1073–1077.
- Pewe, L. L., Netland, J. M., Heard, S. B., and Perlman, S. (2004). Very diverse CD8 T cell clonotypic responses after virus infections. *J. Immunol.*, 172:3151–3156.
- Ribot, J., Romagnoli, P., and van Meerwijk, J. P. M. (2006). Agonist ligands expressed by thymic epithelium enhance positive selection of regulatory T lymphocytes from precursors with a normally diverse TCR repertoire. *J. Immunol.*, 177:1101–1107.
- Scollay, R., Butcher, E. C., and Weissman, I. L. (1980). Thymus cell migration. Quantitative aspects of cellular traffic from thymus to the periphery in mice. *Eur. J. Immunol.*, 10:210–218.
- Thorstenson, K. M. and Khoruts, A. (2001). Generation of anergic and potentially immunoregulatory CD25⁺CD4⁺ T cells in vivo after induction of peripheral tolerance with intravenous or oral antigen. *J. Immunol.*, 167:188–195.
- van Santen, H.-M., Benoist, C., and Mathis, D. (2004). Number of T reg cells that differentiate does not increase upon encounter of agonist ligand on thymic epithelial cells. *J. Exp. Med.*, 200:1221–1230.
- Venturi, V., Kedzierska, K., Tanaka, M. M., Turner, S. J., Doherty, P. C., and Davenport, M. P. (2008). Method for assessing the similarity between subsets of the T cell receptor repertoire. *J. Immunol. Methods*, 329:67–80.
- Venturi, V., Kedzierska, K., Turner, S. J., Doherty, P. C., and Davenport, M. P. (2007). Methods for comparing the diversity of samples of the T cell receptor repertoire. *J. Immunol. Methods*, 321:182–195.
- Wong, J., Obst, R., Correia-Neves, M., Losyev, G., Mathis, D., and Benoist, C. (2007). Adaptation of TCR repertoires to self-peptides in regulatory and nonregulatory CD4⁺ T cells. *J. Immunol.*, 178:7032–7041.
- Zelenay, S., Lopes-Carvalho, T., Caramalho, I., Moraes-Fontes, M. F., Rebelo, M., and Demengeot, J. (2005). Foxp3⁺ CD25⁺ CD4⁺ T cells constitute a reservoir of committed regulatory cells that regain CD25 expression upon homeostatic expansion. *Proc. Natl. Acad. Sci. U.S.A.*, 102:4091–4096.

- Zhan, Y., Bourges, D., Dromey, J. A., Harrison, L. C., and Lew, A. M. (2007). The origin of thymic CD4⁺CD25⁺ regulatory T cells and their co-stimulatory requirements are determined after elimination of recirculating peripheral CD4⁺ cells. *Int. Immunol.*, 19:455–463.

Part V

GENERAL DISCUSSION

8. GENERAL DISCUSSION

In this final chapter, we recapitulate and discuss the main results of the thesis. Some issues will be excluded because they appear to us small details of the overall picture. Interested reader in such detail should read the specific discussions in each chapter. Nonetheless, we take the chance to discuss, here with some detail, the reason for the good fitting of the Poisson-Lognormal distribution for the experimental data analyzed in Chapters 6 and 7. This topic deserves a special discussion, because similar distribution is often observed in ecological studies. Making parallels between immunological and ecological phenomena proved to be a useful exercise, not only to understand to origin of the above distribution, but also to rethink experimental design of current TCR repertoire studies that allow to test different possibilities.

8.1 On the role of allelic exclusion on peripheral tolerance

In Chapters 2 and 3, we argued that the γ locus operates in adult mice under a tradeoff between the yield of the process and allelic and isotypic exclusion. In other words, $\gamma\delta$ precursors undergo recombination of the γ locus in order to maximize the probability of a precursors surviving the recombination process but, at the same time, avoiding the generation of many cells that exhibit either two productive alleles of the same isotype or two productive in-frame rearrangements of different isotypes. This was achieved by a process occurring within a time window in which there is a first period with only one of the alleles available for recombination, followed by a second period where both alleles are accessible and, thus, allowed to recombine. During this time window, the occurrence of a productive rearrangement would lead to the shutting down of all recombination machinery. Since the closing of the process is not instantaneous, further rearrangements may occur in between.

The most common rationale for allelic and isotypic exclusion is minimizing the risk of pathologic autoimmune. In this line of reasoning, our results suggest that $\gamma\delta$ T cells might have an autoimmune potential, which should be avoided during V(D)J recombination. Yet, the link between $\gamma\delta$ T cells and autoimmunity is still unclear as discussed in these Chapters.

Notwithstanding, as opposed to β and γ chains, α locus is not operating under allelic and isotypic exclusion, even under healthy wildtype animals (Aifantis et al., 1997; Malissen et al., 1992). Therefore, there is an apparent paradox between this observation and the above reason for the allelic and isotypic exclusion. There might be several explanation for the lack of allelic exclusion of the TCR α chain. A simple explanation is that α chain is not important to antigen recognition. Other view is related to the so-called "phenotypic allelic exclusion" mechanism in which only one of the α chains will be effectively expressed by downregulation of TCRs exhibiting the second α chain (Gascoigne and Alam, 1999). One can also think that the presence of

two functional TCR α chains decrease the chance of TCR signaling by a lower quantity of each TCR variant available to bind to its ligand.

Recently, a study provided a different view on this phenomenon. Available experimental evidence suggest that human T_R cells are highly prone to express two functional TCRs exhibiting different α chains (Tuovinen et al., 2006), usually referred to as a dual TCR. It is still unclear the role of dual receptors on T_R-cell generation and function. However, the authors of the above study speculate that the expression of two functional α chains might open a selection window in the thymus towards autoreactive T_R cells. In their view, T_R cells might escape negative selection due to a putative downregulation of the autoreactive TCR, being selected by the other TCR. This is in agreement with current view of T_R-cell ontogeny that T_R cells seem more resistant to negative selection (van Santen et al., 2004). Notwithstanding, dual receptors on T_R cells seems to have beneficial role in the control of autoimmunity, since studies on transgenic mice expressing specific TCR for basic myelin protein on which only T cells expressing both the endogenous and the transgenic TCR were capable of protecting the host against autoimmune encephalitis (Hori et al., 2002), and not the endogeneous or the transgenic TCR alone. Yet, a word of caution should be given, because the above observations were taken in transgenic mice and, thus, the link between dual receptors and T_R-cell function is still elusive.

If two functional TCR α chains in the same cell are indeed relevant in the context of antigen recognition and response, one might think that this would imply an increase in the crossreactivity of the T cells. In this line of thought, T_R cells would be intrinsically highly crossreactivity due to allelic and isotypic inclusion. The results of Crossregulation model studied in Chapters 4 and 5 provide then a rationale for the occurrence and role of dual receptors in T_R cells. This model predicts that highly crossreactive T_R cells tend to be selectively maintained in the periphery and, therefore, one should expect that T_R-cell pool would be enriched on dual expressors. On a more functional rationale, the allelic and isotypic inclusion of α locus could be a mean to ensure that T_R cells generated in the thymus would have a higher chance to survive in the periphery. If this latter rationale is true, one can think that the parameters of the V(D)J recombination process on the α chain would be setup to optimize the yield of the process, and not much the probability of having double in-frame rearrangement, since the presence of multiple TCRs in the same cell would not cause damage to body due to the generation of T_R cells.

At this point, one should ask why then the β locus operates under stringent allelic and isotypic exclusion as opposed to the α locus. This might be related to $\alpha\beta:\gamma\delta$ lineage decision itself, and not much to a putative control of autoimmunity. Since β , γ and δ loci rearrange approximately at the same developmental time, one might expect that the occurrence of a given recombination event would determine to which T-cell lineage a precursor would divert to. It is like having a internal "race" running within a precursor in order to achieve a certain recombination event. In this regard, a functional TCR β chain associated to a surrogate non-productive TCR α chain seems enough to divert a cell to the $\alpha\beta$ lineage (Dudley et al., 1994; Fehling et al., 1995; von Boehmer et al., 1999) while, for $\gamma\delta$ T cells, the production of a full operational $\gamma\delta$ TCR seems required (Kang et al., 1998; Passoni et al., 1997; Pereira and Boucontet, 2004). A way to divert "quicker" a precursor cell to the $\alpha\beta$ lineage is to operate under allelic and isotypic exclusion on the β locus. Therefore, it is plausible that V(D)J recombination process

was shaped throughout evolution in order to optimize different aspects of thymic T-cell ontogeny.

It is worth noting that the study of the γ gene recombination process was only possible due to the fact that there were essentially two isotypes at the same locus that could be rearranged in adult mice. In this simple scenario, we were able to build up a stochastic model where the recombination status of each isotype in a cell could be traced throughout the recombination time window. This type of approach seems unfeasible when the number of gene segments (or isotypes) that can rearrange is very large, as happens in the α and β chains (see Table 1.1 of Chapter 1). This would require to develop appropriate modelling tools to overcome this problem properly.

In Chapter 6, we analyzed the TCR repertoires of a mouse model with limited diversity. In this mouse line, there is an endogeneous and a transgenic TCR α locus. The expression of the endogeneous TCR α gene is prevented by the insertion of an appropriate mutation at the endogeneous α locus. The transgenic TCR α locus is composed of a single V α region and two J α segments, and a TCR β transgene (Correia-Neves et al., 2001; Wong et al., 2007). A significant number of rearrangements on the endogeneous TCR β locus is not expected due to allelic exclusion. Similar mouse line was engineered by Pacholczyk et al. (2006) but using different J segments and β transgene. This scenario seems appropriate to apply similar stochastic approach in the study the rearrangement of the α locus. This would require to breed these mice in order to generate homozygous animals for the above TCR α minilocus with similar restriction on the TCR β gene. If one could assess the rearrangement status of the α locus in mature cells, one could detect the presence of a time window and/or of a feedback mechanism by developing similar stochastic models, as done in the study of the γ chain. Since there are no two distinct isotypes in the same locus, one could not infer the possibility of differential allelic accessibility during the recombination process. This limitation notwithstanding, we could study whether there is any optimization principle operating in the recombination process of the α locus. This would be a good way to test whether there is any selection towards the maximization of the yield of process, as hypothesized above.

8.2 On the shape of regulatory and effector T-cell repertoires

8.2.1 Distinct classes of peripheral T-cell repertoires predicted by the Crossregulation model

In Chapters 4 and 5, we studied the shape of peripheral T $_R$ and T $_E$ repertoire through two variants of the Crossregulation model. In the first version, we assume that T-cell clones recognize independently a mutually exclusive set of APCs. This simple model predicts that peripheral repertoires should be partitioned in three subsets. A first small subset refers to T-cell clones that recognize too few APCs and go extinct in the periphery. A second subset refers to a diverse pool of T-cell clones devoided of T $_R$ cells that recognize intermediate levels of APCs. Since this subset is devoided of T $_R$ cells, the respective T-cell clones are free to respond to infections. A third subset is composed of autoreactive T-cell clones with both cell types that recognize APCs at high densities. Because T $_R$ cells are present in these clones, autoimmunity

is avoided. The relative contribution of each subset to peripheral repertoire depends on the thymic export. Several predictions for the structure of the repertoire have been put forward.

A limitation of this first model was that none of the above subsets were just composed of T_R cells. In other words, the above model could not generate by itself a T_R -cell exclusive repertoire. This was achieved with a second Crossregulation model, where T-cell clones recognize non-exclusive subset of APCs. Simulations with this model under a constant thymic export revealed that, even for the same parameter set, different shapes of T-cell repertoire could emerge. These shapes were classified in three distinct classes, each one distinguished by the structure of the resident population (i.e., the oldest clone in the repertoire). On the one extreme, there is a class of repertoires characterized by an overdominance of T_E cells in the resident population. We argue that this class might represent repertoires of Foxp3-deficient mice, which exhibit massive T_E -cell proliferation and severe autoimmune diseases. On the other extreme, there is a class that shows an overdominance of T_R cells in the residence population. This scenario seems unlikely to describe a repertoire of an healthy individual, because an overdominance of T_R cells might imply a global regulation of the immune responses, either against self or foreign antigens. The third class is in between these two extremes, showing a repertoire structure of the repertoire similar to the one predicted by the first model. This class appear the most plausible to represent a repertoire of an healthy individual.

8.2.2 *Classifying peripheral T-cell repertoires of available limited TCR diversity mice*

After the analysis of these two Crossregulation models, available experimental data were assessed in order to investigate the repertoire structure present in healthy animal models. As stated above, the best way to distinguish the above repertoire class is through the resident population. A first difficulty then arose, because current TCR repertoire studies are just a snapshot survey of the repertoire, which do not allow to determine which TCR sequences could actually represent the resident population. Therefore, the subsequent statistical analysis proved that available data are not sufficient to resolve this issue.

In Chapter 7, we could not rule out the hypothesis of equal diversity between peripheral T_R and T_E repertoires in the animal model of Hsieh et al (2004,2006) Hsieh et al. (2004, 2006). At this point, we cannot say whether the acceptance of the hypothesis of equal diversity is just a statistical artifact due to small sample sizes. Moreover, even if this hypothesis is true, this does not necessary mean that T_R and T_E share 100% of their TCR diversity. The analysis of the intersection repertoire done in Chapter 7 might shed some light on this issue. Albeit scarce data, our results are compatible with the hypothesis that thymic T_R TCR diversity could be fully contained in their thymic T_E counterpart. This suggests that if a T_R cells is generated, there always a T_E cell exhibiting a similar TCR specificity. Yet, since thymic T_E repertoire seems more diverse than the T_R one, it is expected that thymus exports a large fraction of T_E -exclusive diversity. In this line of thought, it is not expected that the TCR specificities contained peripheral T_R and T_E repertoires would intersect totally with each other. Notwithstanding, the analysis of the intersection between peripheral T_R and T_E -cell repertoires revealed to be unconvulsive, because of the wide range of estimates for the fraction of shared diversity

that are compatible with the data (from 10% to 100%). On the extreme hand, a low fraction of shared TCR diversity would imply a positive correlation between T_R and T_E -cell densities of clones in the intersection repertoire. In its turn, this positive correlation might just reflect the contribution of recent thymic emigrant population, as argued in Chapter 5. In agreement with this interpretation, we showed that only a strong positive correlation between thymic T_R and T_E -cell density was compatible with the observed data (Chapter 7). Under the scenario of equal diversity, a small fraction of shared overlap would imply a large contribution of T_R and T_E exclusive diversity to the overall diversity. This seems close to the predictions for the repertoire class related to a healthy individual. On the other hand, approximately 100% shared diversity seems only possible when thymus exports preferentially T-cell clones with similar TCR specificity and crossreactivity of the clone is high enough to sustain both cell types together. In this line of reasoning, the underlying repertoire class is related to a widespread overdominance of T_R cells. This seems unlikely, since the fraction of T_R cells in these animals is more or less the same as in the wildtype (Hsieh et al., 2004). Between these two extremes lie repertoires with a diversity composition similar to the above-mentioned repertoire structure predicted for a healthy individual. Thus, available data agree with the prediction for the shape of repertoires in healthy individuals. Nonetheless, it is worth emphasizing again that the above interpretation is just based on a snapshot survey of the repertoire, which is not the best way to determine the type of structure underlying the peripheral T_R and T_E -cell repertoire. Therefore, a word of caution should be put forward.

The Crossregulation model seems also useful to disentangle the differences between repertoires of other two animal models available in the literature (Pacholczyk et al., 2006; Wong et al., 2007). LTD mice of Wong et al. (2007) show a high fraction of peripheral T_R cells than in wildtype animals, which prompts us to classify its repertoire in the repertoire structure class related to an overdominance of T_R cells. If our interpretation for this type of repertoire is correct, we expect that LTD mice would be more susceptible to infections than wildtype animals. A possible explanation for this type of repertoire structure can be explained by a preferential thymic export of T_R -cell clones recognizing APCs in quantity that can sustain them in the periphery. This view is in agreement with a higher fraction of thymic T_R cells taking together current models that proposes T_R cells are highly crossreactive (see, for example, Modigliani et al. 1996)

In the case of TCR^{mini} mice of Pacholczyk et al. (2006), peripheral T_R cells are in a smaller fraction than in wildtype animals, which suggests a repertoire structure related to an overdominance of T_E cells. Since these animals do not show any symptoms of autoimmune pathologies, as expected for such repertoires, we speculate that T_E -cell clones coming from the thymus recognize low to intermediate levels of APCs, setting the peripheral repertoire mainly in the first and second subsets referred in the first Crossregulation model. This is somehow in agreement with the observation that T-cell hybridomas derived from these animals could not be activated *in vitro* by autologous APCs (Pacholczyk et al., 2007). Yet, no study on the crossreactivity distribution is available to confirm the above interpretation.

Therefore, we could explain the differences between peripheral repertoires of the two animal models by differences in the distribution of crossreactivity or APC density recognized by

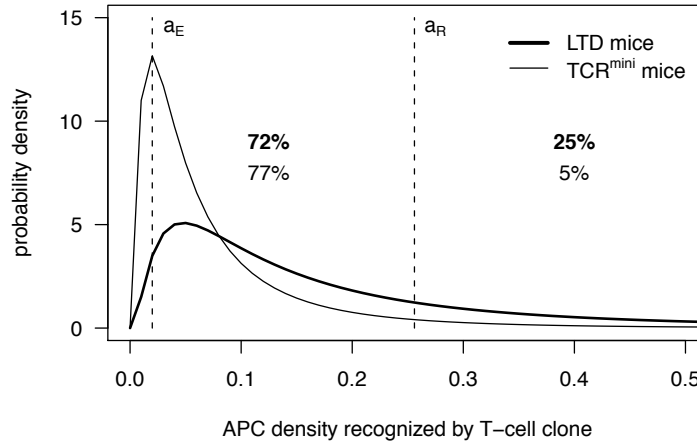


Fig. 8.1: Illustration of putative differences between the distribution of APC density recognized by T-cell clones generated in the thymus of LTD and TCR^{mini} mice. The critical points a_E and a_R are associated to the bifurcation diagram in Figure 5.2A of Chapter 5. Thick lines and bold text refer to LTD mice, while thin and normal text refer to TCR^{mini} mice.

each T-cell generated from the thymus. In the case of LTD mice, there is a higher probability of generating clones recognizing an APC density higher than the critical point a_R for T_R-cell maintenance in the periphery, while in TCR^{mini} mice the opposite happens (Figure 8.1).

It is worth noting that, in the same line of Chapters 6 and 7, further statistical analysis on data from these two animal models could have been performed. Yet, this task revealed to be full of limitations. Beside of being a snapshot survey of the repertoires, available data refer to a pooled sample of different animals. For us, these two reasons undermined any possibility of getting accurate results for the repertoire of each individual.

In summary, the Crossregulation model and its two variants studied here seem to provide a good interpretative framework in the study of peripheral TCR repertoires. As shown above, they can also give a hint on the thymic output, specially, in terms of the distributions of APC density "seen" by each T-cell clone.

8.2.3 Similar shape for the peripheral repertoire predicted by a model of peripheral T_R-cell generation

Recently, Fouchet and Regoes (2008) studied a mathematical model based on a mechanism of peripheral T_R-cell generation proposed by Powrie and Maloy (2003). The key players of the model are APCs and adaptive T cells, obeying to a complex interaction network between them. In this new model, APCs can be either resting or activated upon pathogen exposure. APCs are or enter a resting state either if they have not been previously conjugated with T_E cells or if activated APCs have been in contact with T_R cells, while activation of APCs is promoted by T_E-cell contact. Resting APCs promote naive T cells to become T_R cells upon cellular conjugation, while activated APCs would differentiate naive T cells into T_E cells also upon cell contact. In this model, T_R cells are able to inhibit the proliferation of T_E cells by two routes, one

upon "direct" interaction with T_E cells and other by decreasing the number of activated APCs that promote T_E -cell differentiation.

Fouchet and Regoes (2008) demonstrated that, depending on the parameter set, their model can reach two possible steady states. The first steady state refers to a state of weak regulation, in which T_E cells are highly abundant as opposed to T_R cells that are at extremely low density (eventually zero). The second steady state is associated with a strongly regulated state, in which T_R cells are in higher levels than their T_E counterparts. These two possible steady states are in close agreement with those of the Crossregulation model (illustrated in Figure 4.2 of Chapter 4).

Interestingly, this model exhibits a bistable behavior in certain parameter regions. For example, it was shown a bifurcation diagram as function of antigen stimulation with similar qualitative behavior to that for Crossregulation model given in Figure 4.6, which is in the core of the whole analysis given in Chapters 4 and 5. Since antigen stimulation is related to the rate of antigen uptake by APCs, which critically determines that the density of APCs in a resting or an activated state, we can argue that this parameter assumes similar role of APC density recognized by each T cell in the Crossregulation model. In this line of thought, this new model might predict similar qualitative results for the shape of repertoires as the Crossregulation model. Therefore, the results obtained in the thesis do not seem model dependent, and might be general properties of models that aim to capture the most relevant observations of T_R -cell immunobiology, such as the bistable behavior of many *in vitro* and *in vivo* experiments (reviewed in Chapter 4).

8.3 On the experimental design and analysis of current TCR repertoires studies

In Chapters 5, 6, and 7, we made different considerations for the design and analysis of TCR repertoire studies. We now summarize them and put them in a more general context (Table 8.1).

The first comment is related to the way that TCR sequences are obtained experimentally. Currently, there are two main experimental approaches, one uses an pre-amplified cDNA extract from a pool of T cells (Hsieh et al., 2004, 2006; Lathrop et al., 2008), and another uses single-cell RT-PCR (Pacholczyk et al., 2006, 2007; Wong et al., 2007). The first approach is useful to directly estimate TCR diversity. Yet, the same is not true to estimate the underlying clonal size distributions and overlap. As mentioned in Chapter 7, when pre-amplified cDNA extracts are used, one does not know whether the repeated TCR sequences effectively represent distinct T cells. With a series of assumptions on the efficiency of cDNA amplification and on the average number of cDNA copies per T cell, one can in theory obtain an estimate for clonal size distribution. Since many assumptions are made that might not completely true, we do not recommend this approach if the goal of the analysis is focused on the clonal size distribution itself.

These difficulties are not present in single-cell PCR experiments, which is the most correct way to obtain sample of TCR sequences because the sampling unit would be effectively each T cell. In this scenario, the sampling scheme is conceptually equivalent to sampling individ-

uals from an ecological community, which allows to extend the results of the samples to the level of the whole repertoire, as illustrated in Chapter 6. It is worth noting that the single-cell PCR approach has its own drawbacks. Since sequencing might not work well in all sorted T cells, one should compensate the putative failures in sequencing with a higher sample size, increasing the cost and time of such experiments. In this regard, only one of the above studies provided the proportion of positive RT-PCR reactions on different 96-well plate, ranging from 30% to 80% (Correia-Neves et al., 2001). Moreover, single-cell RT-PCR might also show some bias in the TCR sequences that fail PCR amplification. This issue is not mentioned in the above-mentioned studies, except in the original reference of one of the limited diversity mice (Correia-Neves et al., 2001). The authors of this study claim no bias on sequencing, because TCR samples were diverse and show no bias towards rare TCR sequences, as many repeats of the same sequence were present. This argument seems insufficient. Interestingly, comparing the two above extremes for the proportion of positive RT-PCR reactions, they are statistically different ($p < 10^{-11}$, Fisher's exact test), but not reported in the original reference. On the one hand, this result can be simply explained by differences in TCR composition of each plate. In fact, this composition is critically dependent on the underlying TCR diversity and respective shape of the clonal size distribution. On the other hand, it does not formally rule out the possibility of a putative bias on the TCR variants that could actually be sequenced. Therefore, we conclude that, for the single-cell RT-PCR approach be considered as a gold standard procedure for a detailed study of T-cell repertoire, one should first investigate how TCR sequence bias can be inferred from the data.

Hitherto, all TCR repertoire studies were made by taking a snapshot of the repertoire (Hsieh et al., 2004, 2006; Lathrop et al., 2008; Pacholczyk et al., 2006, 2007; Wong et al., 2007). Yet, in Chapter 5, we demonstrate with simulations of the Crossregulation model that different repertoire classes are better distinguished by their underlying resident population. Since the above-mentioned studies are just snapshot surveys of the repertoires, it was a matter of speculation to classify the repertoire structure of the available mouse models with limited diversity. In theory, this could be avoided by performing longitudinal studies. In this scenario, one would expect that resident population would consistently appear in samples of different time points. Yet, in mice with TCR diversity, such as those analyzed throughout the thesis, one can also expect that some TCR sequences could appear in different time points just due to preferential V(D)J recombination and thymic selection towards them. Therefore, one should possess powerful analytic and experimental tools in helping to decide which TCR sequences could actually represent the resident population. The bivariate Poisson-Lognormal model presented in Chapter 7 could help in this task when there are only two time points. One can try to extend this model to more than two samples, but the respective estimation would be cumbersome. In this regard, estimation based on massive stochastic simulation, such as the one employed in Chapter 7 or the recently-proposed Approximate Bayesian Computation methods (Beaumont et al., 2002), might be a good solution to this problem in an era of great computer power.

Another advantage of longitudinal studies in Ecology was shown in a recent study of an estuarine community (Magurran and Henderson, 2003). By recording a 21-year data, the authors of the study showed that the estuarine community could be divided in core species that per-

sist longer in the population and occasional species that occur infrequently in the population. By doing this partition of the data, species-abundance distribution revealed to be a mixture of two distributions, one for core species and another for occasional species. Similar type of distribution might happen in the peripheral TCR samples, since peripheral T cell repertoire is divided in resident and recent thymic emigrant populations, as illustrated by the simulations of the Crossregulation model studied in Chapter 5. In this line of thought, snapshot surveys obscure underlying complexities in the clonal size distribution, which might be unraveled by collecting samples from the repertoire at different time points.

At this point, we recommend that, before performing a TCR repertoire study, the pros and cons of each experimental design should be weighted in order to achieve the best result. In this stage, one should also think the best available analytical tools for the data to be collected. To avoid any "post-mortem" examination of the data (Fisher, 1962) that ultimately lead to nowhere, we set out some final recommendations. First, experimental protocols should be standardized across different studies. This would allow to directly compare the results of different experiments. Second, TCR studies should report data from each individual and not pooled data of different animals, as often done (Hsieh et al., 2004; Pacholczyk et al., 2006, 2007; Wong et al., 2007). This would allow to estimate the effective TCR diversity of each individual. As shown in Chapter 6, TCR estimation on pooled data tends to overestimate the diversity of each animal, and the extent of overestimation is directly correlated with the number of pooled data sets from different individuals. By doing the analysis discriminated by individuals, it avoids distortions on the clonal size distributions by an overrepresentation of TCR sequences shared across different animals. Third, adequate sample sizes should be applied in order to provide accurate TCR diversity estimates. Since the determination of sample size is critically dependent on the underlying shape of the repertoire, we recommend to monitor the probability of obtaining a new TCR sequence while gathering data, as done in subsection and also in Wong et al. (2007) but monitoring the abundance-coverage estimates. When this probability reaches a plateau, it is reasonable to assume that samples possess all information to accurately estimate TCR diversity. Similar idea can be applied when studying other aspects of TCR repertoires, such as the size of the intersection repertoire. Following these recommendations, we hope that more accurate and comparable TCR studies could be achieved.

8.4 On the good fit of Poisson-Lognormal distribution for experimental data

8.4.1 Poisson-Lognormal model should be applied with caution

In Chapters 6 and 7, we concluded that the Poisson-Lognormal was a good model to dissect different structural properties of TCR repertoires. In spite of its general good fit for the available experimental data, the model proved to be sensitive to the presence of highly abundant TCR sequences in the samples when estimating TCR diversity. In fact, since available data show invariantly a great proportion of singletons and few highly abundant TCR sequences, the respective diversity estimate usually decreases when the highly abundant TCR sequences are discarded from the analysis. Therefore, this model should be used with extreme caution.

Tab. 8.1: Main advantages and disadvantages of different experimental procedures and analysis.

Single-cell PCR sequencing	
<i>Advantages</i>	• Direct estimation of TCR diversity, clonal size distribution, and intersection repertoire
<i>Disadvantages</i>	• Higher sample sizes to account for putative sequencing failure • Putative bias in TCR sequences that fail PCR amplification
PCR sequencing based on a pre-amplified cDNA pool	
<i>Advantages</i>	• Direct estimation of TCR diversity
<i>Disadvantages</i>	• Indirect estimation of clonal size distribution, and intersection repertoire
Longitudinal surveys	
<i>Advantages</i>	• Partitioning clonal size distribution into the resident and recent thymic emigrant populations • Putative classification of repertoire structure according to Crossregulation model
<i>Disadvantages</i>	• Difficult to define the resident population in mice with limited TCR diversity • "High" time consuming and cost • Limited set of statistical tools to assess changes between different time points
Snapshot surveys	
<i>Advantages</i>	• Availability of a vast statistical tools to the respective analysis • "Low" time consuming and cost
<i>Disadvantages</i>	• Obscure underlying complexities in the clonal size distribution
Analysis of data sets from each animal	
<i>Advantages</i>	• Estimation of the effective TCR diversity of each individual and respective clonal size distribution
<i>Disadvantages</i>	• Large sample sizes to have accurate TCR diversity estimates for each animal • Current experimental procedures do not allow to large sample sizes for each animal
Analysis of pooled data sets from distinct animals	
<i>Advantages</i>	• Estimation of the overall TCR diversity of pooled data
<i>Disadvantages</i>	• Distortion of clonal size distribution by the overrepresentation of TCR sequences shared across different individuals • Overestimation of TCR diversity of each individual • Difficult to compare results from different studies due to differential number of pooled data sets

As illustrated in Chapters 6 and 7, the robustness of TCR diversity estimates derived from Poisson-Lognormal model can be evaluated by a procedure based on splitting the data set into two parts, one referring to the most abundant species and other to the rarer ones (Chao et al., 1993). This procedure brings into the analysis some degree of subjectivity, because the threshold for data splitting has to be determined by the analyst. The usual criterion to do such is to determine the threshold that provides stable diversity estimates and a good fit for the data. On the one hand, a good fit can be precisely defined by performing an appropriate goodness-of-fit test with a given significance level (e.g., 5%). On the other hand, how can one exactly define stable diversity estimates? In this regard, the stability of the estimates — or the reasonableness of the estimates — is assessed by the prior beliefs that the analyst has to the problem at hand. This "feeling" usually comes from previous experience on the subject. In this case, Bayesian analysis should be a good alternative to the frequentist methods that we employed here. Prior beliefs could then be coherently included in the analysis framework, and help to decide the best TCR diversity estimate for the data at hand. This is an ongoing work and we expect to have an answer in a near future.

8.4.2 Immunological and ecological communities: the recurrent observation of a Lognormal pattern for species-abundance distributions

In Chapter 6, the analysis of the TCR repertoire based on the Poisson abundance models indicates that the clonal size distribution follows a Lognormal distribution. If one knows the number of T cells in the compartment under analysis, the parameters of the respective clonal size distribution for the whole repertoire of an individual can be derived from the corresponding estimates for the observed data. Yet, for this inference to be correct, each TCR sequence in a data set should represent unequivocally a cell, as in single-cell RT-PCR experiments. For experiments based on pre-amplified cDNA extracts, the clonal size distribution can be indirectly inferred but this would involve a series of assumptions for cDNA amplification, as discussed in Chapter 6.

Interestingly, the Lognormal distribution has been found to be particularly adequate to describe the observed species-abundance distributions in different ecological contexts, even under complete and incomplete enumeration of species (Bulmer, 1974; McGill, 2003; Preston, 1948). Because of this apparent generality of the Lognormal distribution, some authors suggested that this model should be regarded as a good null model for the species-abundance distribution (McGill, 2003). Yet, there is no general consensus in this regard, because different mechanisms can generate a Lognormal species-abundance distribution (reviewed in Williamson and Gaston 2005). Moreover, one of the reasons for attractiveness of the Lognormal distribution is to invoke the Central Limit Theorem in a situation where species grow exponentially subject to multiplicative random effects. Yet, as pointed out by different authors (Pielou, 1974; Williamson and Gaston, 2005), a Lognormal distribution can only be expected by the Central Limit Theorem for the samples of same species under equilibrium or under the strong assumption that all species have the same demographics parameters.

With respect to null models of Ecology, Hubbell (2001) proposed a neutral model for ecological communities, inspired by the neutral theory for evolution (reviewed in Alonso et al. 2006).

This model involves a minimal set of assumptions for the community dynamics, such migration, growth, and speciation. Under this model, the expected species-abundance distribution follows the so-called zero-sum Multinomial distribution, which has a shape resembling a Lognormal distribution (Williamson and Gaston, 2005) but has no close form expression. Interestingly, using different fitting strategies, a recent study demonstrates that the Lognormal distribution outperforms the zero-sum Multinomial distribution when fitted to many distinct ecological data sets (Magurran and Henderson, 2003). This result shows that even a neutral theory for community dynamics was not enough to dethrone the Lognormal distribution as a general good model for the species-abundance distribution. Yet, the cause of these empirical observations remains a mystery and a matter of constant debate (see Williamson and Gaston (2005) and the references therein).

This thesis brings to Immunology the interest and also the controversies of the Lognormal species-abundance distribution. Even that the mechanisms shaping the T-cell repertoires seem different from those structuring an ecological community, the (Poisson-)Lognormal distribution emerged as a good candidate model for the most data sets. This is specially intriguing for thymocyte populations analyzed in Chapters 6 and 7, which result from a complex process of positive and negative selection. In this case, the simple explanation for this result is just a matter of small sample size that do not allow to reject the model.

In the case of peripheral populations, a possible explanation might be related to an interesting observation of Magurran and Henderson when studying an estuarine community (Magurran and Henderson, 2003). These authors demonstrated that the community could be partitioned in persistent and occasional species. Typically, persistent species form the core of the community and are highly abundant, while occasional species occur infrequently in the community and are low abundant due to different habitat requirements. In this scenario, species abundance distribution appear to be a mixture of a Lognormal distribution for persistent species with a Log-series distribution for occasional species. More importantly, it was only when this partition was done that the Lognormal distribution drowns out the Log series distribution from the data. Based on this result, we speculate that the good fit of the Poisson-Lognormal of peripheral populations might be just due to the fact it was not possible distinguish persistent TCR variants from occasional ones.

It is worth noting that peripheral T-cell repertoires seem more close to a "natural" ecological community than their thymic counterparts in terms of the forces governing their organization. First, peripheral T cells compete with each other for their resources (Freitas and Rocha, 2000). In Chapters 4 and 5, we consider APCs, which promote T-cell activation and proliferation, as the main resources of T cells. Second, peripheral T cells migrate across different anatomical locations. Third, periphery might be structured in different niches (Carneiro et al., 1995; Freitas and Rocha, 2000), which are often invoked in Ecology (e.g., Begon et al. 2006). Immunological niches can be defined as distinct anatomical sites or as APC foci in which T cells compete for their cognate APCs (Carneiro et al., 2007, 1995). Since a Lognormal distribution is often a good model to describe the species-abundance distributions in many ecological communities, we may extend this empirical argument to peripheral T-cell repertoires. Yet, this rationale might be blurred by the fact that peripheral T_E cells can be converted to T_R pool. In

theory, this process poses no problem when studying the overall $CD4^+$ T cell repertoire, because peripherally-converted T_R cells would not affect the overall abundance of a given TCR species. In agreement with this observation is the good fit of the Poisson-Lognormal model for the lymph node $CD4^+$ T cells analyzed in Chapter 6. When studying T_R and T_E cell repertoires individually, a peripherally-converted T_R cell would be interpreted either as a result of proliferation of a T-cell clone exhibiting an existing TCR species or as an arrival of a new T-cell clone if its respective TCR was not already present in the community. In this line of thought, peripheral conversion would be regarded as a "natural" process occurring in an ecological community and, thus, the above heuristic argument seem to hold in this case. The good fit of the Poisson-Lognormal model for peripheral T_R and T_E TCR samples presented in Chapter 7 seems to favor this interpretation.

8.4.3 Can the Crossregulation model predict a Lognormal clonal size distribution?

In Ecology, the Lognormal distribution is usually ascribed to communities in equilibrium or to persistent species that form the core of the community (Magurran and Henderson, 2003). In this scenario, the Crossregulation model studied in Chapters 4 and 5 could be used as a rationale for the above model for the clonal size distribution.

We now take the chance to discuss with some detail the reasonableness of the Lognormal distribution. To this end, we recall the simple Crossregulation model in which every T-cell clone recognizes a mutually exclusive APC subset (Chapter 4). The shape of the peripheral repertoire follows closely the pattern of the bifurcation diagram illustrated in Figure 5.2A in Chapter 5. To calculate the clonal size distribution, we assume that all clones are generated with similar initial conditions and the peripheral repertoire is in equilibrium. In this scenario, the steady states of T_R and T_E cell densities are given as in Figure 5.2B. On top of this bifurcation diagram, we specify an distribution of APC density recognized by each T-cell clone entering the periphery (Figure 8.2A). In this scenario, the clonal size distribution is just the probabilistic transformation of that APC distribution according to the respective steady states for T_R and T_E cells (Figure 8.2A). This transformation was obtained via a Monte Carlo procedure because of the great complexity of the steady states leading to a coexistence of T_R and T_E cells in the same clone. We used the following simulation algorithm to obtain the underlying clonal size distribution: (i) generate a value of the APC density distribution (in our case, Exponential and Lognormal distributions), (ii) determine the respective steady state predicted by the simulated for APC density, (iii) repeat previous steps several times, and (iv) estimate the respective clonal size distribution with the simulated samples (with sample size of 5000 values). In this last step, we apply the popular Gaussian kernel estimation method implemented in the R software. It is worth noting that, when the simulated APC density did not allow the maintenance of either T_R and T_E cells in the system (that is, an APC density less than a_E ; see Figure 5.2), we did not include the value 0 in the generated samples. The same was done when estimating T_R clonal size distribution, but for simulated APC densities less than $\tilde{a}_R$.

As shown in Figures 8.2A and B, the underlying clonal size distribution of T_E cells is far from being a Lognormal distribution, when APC density follows either an Exponential or a Lognormal

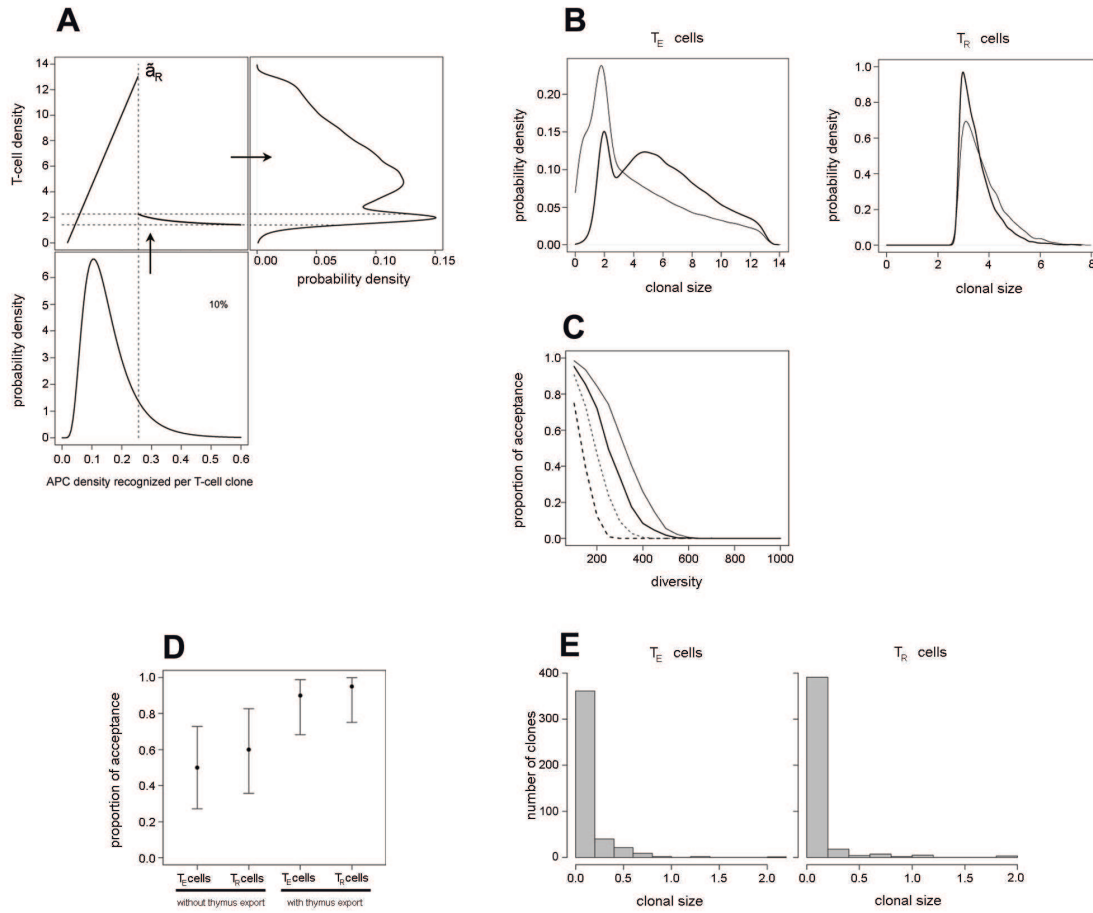


Fig. 8.2: Clonal size distributions predicted by the Crossregulation models of Chapters 4 (plots **A**, **B** and **C**) and 5 (plots **D** and **E**). **A**. Transformation of the distribution of APC density recognized per T-cell clone into the respective clonal distribution of T_E cells through the bifurcation diagram of the steady states, where $\tilde{a}_R$ is the critical point for T_R cell maintenance in the system (see Figure 5.2B in Chapter 5 for more details). **B**. Clonal size distributions of T_E and T_R cells predicted by an Exponential and a Lognormal distribution for APC density (thin and thick lines, respectively). **C**. Proportion of Kolmogorov-Smirnov tests that have accepted a Lognormal clonal size distribution in repertoires with a given diversity (Crossregulation model of Chapter 4). **D**. Proportion, and respective 95% confidence interval, of Kolmogorov-Smirnov tests that have accepted a Lognormal clonal size distribution in repertoires generated according to the Crossregulation model of Chapter 5, with and without thymic export (20 simulations in total for each condition). The respective parameters of the model are given in Figures 5.4 and 5.6 of Chapter 5. **E**. Clonal size distributions of T_R and T_E obtained in one simulation with thymic export. The significance level of above-mentioned Kolmogorov-Smirnov tests was 5%.

distribution. The clonal size distribution is then a mixture of two populations. The first one is associated with T-cell clones that exhibit similar steady states for T_E -cell density but lying two distinct parameter regions (Figure 8.2A). A lower or higher contribution of this population to the clonal distribution depends on the proportion of clones generated with an APC higher than $\tilde{a}_R$ (results not shown). It is worth noting that this population exhibits a "small" T_E -cell densities.

Therefore, besides a severe undersampling, this result provides an alternative reasoning for the great representation of singletons and doubletons in TCR samples analyzed in Chapter 7. The second population refers to T-cell clones recognizing APC densities that lead to T_E steady states that are unique across all APC densities. This comprises T-cell clones with either very small or high clonal sizes.

Depending on the shape of the APC density distribution, the corresponding clonal size distribution of T_E cells can be more or less close to a Lognormal distribution. This is illustrated in Figure 8.2B with Exponential and Lognormal APC density distribution. Visually, the clonal size distribution generated with an Exponential APC density distribution seems more closely to a Lognormal distribution than that generated with a Lognormal APC density distribution. In this line of thought, one can argue that T_E -cell repertoires with limited diversity coming from an Exponential APC density distribution might exhibit a clonal size distribution that is statistically indistinguishable from a Lognormal distribution. To tackle this issue, we generate repertoires with different T_E -cell diversity and test the Lognormal distribution for the respective clonal size distribution by the Kolmogorov-Smirnov test (Figure 8.2C). The results agree with the above expectation. In fact, a Lognormal clonal size distributions distribution could only be rejected with 100% chance in repertoires coming from an Exponential APC density distribution with the diversity higher than 600. Similar result is obtained from the Lognormal APC density distribution but for repertoires higher than 500. However, when a more stringent tests were applied, such as the Shapiro-Wilk test or the Anderson-Darling test, both simulated repertoires could be readily disentangled from a Lognormal clonal size distribution (results not shown). This demonstrates that the stringency of the test might determine a higher or a lower chance of accepting the Lognormal distribution.

In contrast to T_E -cell clonal size distribution, the respective distribution for T_R cells is composed of a "single" population (Figure 8.2B), since the steady states of T_R -cell densities are all distinct across the APC density that allows to maintain those cells in the system. For Exponential or Lognormal APC density distributions, the generated clonal size distribution exhibits a shape similar to a Lognormal distribution. Yet, since the T_R -cell density has a minimal steady states that is higher than zero (see Figure 5.2B), the corresponding Lognormal distribution should be truncated at left in that point. Applying the Kolmogorov-Smirnov test to the generated samples for T_R -cell clonal size distribution, we could not indeed reject that truncated Lognormal distribution ($p = 0.64$ and $p = 0.10$ for Exponential and Lognormal distributions, respectively). It is worth noting that the Kolmogorov-Smirnov test could not reject an untruncated Lognormal clonal size distribution for T_R -cell repertoires with diversity less than 250 (Figure 8.2C). Yet, if one uses a more stringent test (the Shapiro-Wilk or the Anderson-Darling test), the hypothesis of an untruncated Lognormal distribution is unequivocally rejected for repertoires with T_R -cell diversity higher than 100 (results not shown).

We now examine the same issue with the Crossregulation model of Chapter 5. In this case, the respective clonal size distribution cannot be directly determined from the bifurcation diagram owing to the complex non-linearities introduced by interclonal competition. To this end, we just test through the Kolmogorov-Smirnov test the clonal size distribution in the generated repertoires, with and without thymic export. Considering an APC diversity of 100, Figure 8.2D

shows that the Lognormal distribution could not be rejected in many generated repertoires. This is particularly true for the simulations with thymic export, where the chance of accepting the Lognormal distribution is around 90%. Figure 8.2E shows an example of T_R and T_E clonal size distributions. The histograms exhibit a pattern similar to a Lognormal distribution, but this pattern can be decomposed as a mixture of the clonal distribution of recent emigrants and of the resident population. In fact, the peak at the small clonal sizes are mainly due to the recent thymic emigrants while the right tail of the histograms refers to resident population. Under the light of this interpretation, the Lognormal-like clonal size distribution seems a consequence of having a mixture of two populations, the recent thymic emigrants and the resident one. Therefore, the Crossregulation model cannot generate a Lognormal clonal size distribution.

In summary, this small study shows that a Lognormal distribution for clonal size distributions could be just an artifact due to inappropriate experimental designs or owing to study repertoires with small TCR diversity or severe undersampling. In this line, the consequences of wrongly accepting the Poisson-Lognormal distribution for the data remain to be assessed, namely, in estimating TCR diversity. Future studies should be then conducted to tackle this issue with more detail than done here in this final chapter. We hope that, in a near future, one could indeed explain the true nature of the clonal size distribution. For now, a Lognormal clonal size distribution is just based on a statistical argument.

8.5 Concluding remarks

We then conclude that mathematical modeling provided clues on how the T-cell repertoire is shaped. It is important to say that most of the above conclusions could only be reached because of our constant attempt to be in close contact with experimental data. In the case of the recombination of the TCR γ locus, we developed mathematical models with the aim of fitting them directly to the data, which allowed us to put out more realistic predictions for the outcome of the process. In the case of the Crossregulation model, the simulations that we performed seem an important basis for the interpretation of future T-cell repertoire studies. Moreover, the simulations done with this model were fundamental to understand the need of altering current experimental designs for the study of T-cell repertoire, such as the usage of longitudinal studies. Notwithstanding, combining our models with experimental data revealed to be a hard task as illustrated in Chapter 7, where insufficient sample sizes led to inconclusive results or the unavailability of the raw data did not permit a more profound statistical analysis. This shows the difficulties underlying the type of approach followed in this thesis. These difficulties could be partially overcome with a greater cooperation between "wet" and "dry" labs. This seems the only way to understand better complex biological phenomena, such as the shaping of the T-cell repertoire at its different stages.

Bibliography

Aifantis, I., Buer, J., von Bohemer, H., and Azugui, O. (1997). Essential role of the pre-T cell receptor in allelic exclusion of the T cell receptor β locus. *Immunity*, 7:601–607.

- Alonso, D., Etienne, R. S., and McKane, A. J. (2006). The merits of neutral theory. *Trends Ecol. Evol.*, 21:451–457.
- Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in Population Genetics. *Genetics*, 162:2025–2035.
- Begon, M., Townsend, C. R., and Harper, J. L. (2006). *Ecology: from individuals to ecosystems*. Blackwell Publishing Ltd, Malden, USA, 4th edition.
- Bulmer, M. G. (1974). On fitting the Poisson Lognormal distribution to species-abundance data. *Biometrics*, 30:101–110.
- Carneiro, J., Léon, K., Caramalho, I., van den Dool, C., Gardner, R., Oliveira, V., Bergman, M.-L., Sepúlveda, N., Paixão, T., Faro, J., and Demengeot, J. (2007). When three is not a crowd: a Crossregulation model of the dynamics and repertoire selection of regulatory CD4+ T cells. *Immunol. Rev.*, 216:48–68.
- Carneiro, J., Stewart, J., Coutinho, A., and Coutinho, G. (1995). The ontogeny of class-regulation of CD4+ T lymphocyte populations. *Int. Immunol.*, 7:1265–1277.
- Chao, A., Ma, M.-C., and Yang, M. C. K. (1993). Stopping rule and estimation for recapture debugging with unequal detection rates. *Biometrika*, 80:193–201.
- Correia-Neves, M., Waltzinger, C., Mathis, D., and Benoist, C. (2001). The shaping of the T cell repertoire. *Immunity*, 14:21–32.
- Dudley, E. C., Petrie, H. T., Shah, L. M., Owen, M. J., and Hayday, A. C. (1994). T cell receptor β chain gene rearrangement and selection during thymocyte development in adult mice. *Immunity*, 1:83–93.
- Fehling, H. J., Krotkova, A., Saint-Ruf, C., and von Boehmer, H. (1995). Crucial role of the pre-T cell receptor $\alpha\beta$ gene in the development of $\alpha\beta$ but not $\gamma\delta$ T cells. *Nature*, 375:795–798.
- Fisher, R. A. (1962). The place of the design of experiments in the logic of scientific inference. *Colloq. Intern. Cent. Nat. Recherche Scientifique*, 110:13–19.
- Fouchet, D. and Regoes, R. (2008). A population dynamics analysis of the interaction between adaptive regulatory T cells and antigen presenting cells. *PLoS One*, 3:e2306.
- Freitas, A. A. and Rocha, B. (2000). Population biology of lymphocytes: the flight for survival. *Annu. Rev. Immunol.*, 18:83–111.
- Gascoigne, N. R. J. and Alam, S. M. (1999). Allelic exclusion of the T cell receptor α -chain: developmental regulation of a post-translational event. *Sem. Immunol.*, 11:337–347.
- Hori, S., Haury, M., Coutinho, A., and Demengeot, J. (2002). Specificity requirements for selection and effector functions of CD25+4+ regulatory T cells in anti-myelin basic protein T cell receptor transgenic mice. *Proc Natl Acad Sci U S A*, 99:8213–8218.

- Hsieh, C.-S., Liang, Y., Tyznik, A. J., Self, S. G., Liggitt, D., and Rudensky, A. Y. (2004). Recognition of the peripheral self by naturally arising CD25⁺ CD4⁺ T cell receptors. *Immunity*, 21:267–277.
- Hsieh, C.-S., Zheng, Y., Liang, Y., Fontenot, J. D., and Rudensky, A. Y. (2006). An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nat. Immunol.*, 7:401–410.
- Hubbell, S. P. (2001). *The unified neutral theory of Biodiversity and Biogeography*. Princeton University Press, Princeton.
- Kang, J., Coles, M., Cado, D., and Raulet, D. H. (1998). The developmental fate of T cells is critically influenced by TCR $\gamma\delta$ expression. *Immunity*, 8:427–438.
- Lathrop, S. K., Santacruz, N. A., Pham, D., Luo, J., and Hsieh, C.-S. (2008). Antigen-specific peripheral shaping of the natural regulatory T cell population. *J. Exp. Med.*, 205:3105–3117.
- Magurran, A. E. and Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, 422:714–716.
- Malissen, M., Trucy, J., Jouvin-Marche, E., Cazenave, P.-A., Scollay, R., and Malissen, B. (1992). Regulation of TCR α and β gene allelic exclusion during T-cell development. *Immunol. Today*, 13:315–322.
- McGill, B. J. (2003). A test of the unified neutral theory of biodiversity. *Nature*, 442:881–885.
- Modigliani, Y., Bandeira, A., and Coutinho, A. (1996). A model for developmentally acquired thymus-dependent tolerance to central and peripheral antigens. *Immunol. Rev.*, 149:155–200.
- Pacholczyk, R., Ignatowicz, H., Kraj, P., and Ignatowicz, L. (2006). Origin and T cell receptor diversity of Foxp3⁺CD4⁺CD25⁺ T cells. *Immunity*, 25:249–259.
- Pacholczyk, R., Kern, J., Singh, N., Iwashima, M., Kraj, P., and Ignatowicz, L. (2007). Nonself-antigens are the cognate specificities of Foxp3⁺ regulatory T cells. *Immunity*, 27:493–504.
- Passoni, L., Hoffman, E. S., Kim, S., Crompton, T., Pao, W., Dong, M.-Q., Owen, M. J., and Hayday, A. C. (1997). Intrathymic selection events in $\gamma\delta$ cell development. *Immunity*, 7:83–95.
- Pereira, P. and Boucontet, L. (2004). Rates of recombination and chain pair biases greatly influence the primary $\gamma\delta$ TCR repertoire in the thymus of adult mice. *J. Immunol.*, 173:3261–3270.
- Pielou, E. C. (1974). *Ecological Diversity*. Wiley-Interscience, New York.
- Powrie, F. and Maloy, K. J. (2003). Regulating the regulators. *Science*, 299:1030–1031.
- Preston, F. W. (1948). The commonness and rarity of species. *Ecology*, 29:254–283.

- Tuovinen, H., Salminen, J. T., and Arstila, T. P. (2006). Most human thymic and peripheral-blood CD4⁺ CD25⁺ regulatory T cells express 2 T-cell receptors. *Blood*, 108:4063–4070.
- van Santen, H.-M., Benoist, C., and Mathis, D. (2004). Number of T reg cells that differentiate does not increase upon encounter of agonist ligand on thymic epithelial cells. *J. Exp. Med.*, 200:1221–1230.
- von Boehmer, H., Aifantis, I., Feinberg, J., Lechner, O., Saint-Ruf, C., Walter, U., Buer, J., and Azogui, O. (1999). Pleiotropic changes controlled by the pre-T-cell receptor. *Curr. Opin. Immunol.*, 11:135–142.
- Williamson, M. and Gaston, K. J. (2005). The lognormal distribution is not an appropriate null hypothesis for the species-abundance distribution. *J. Animal Ecol.*, 74:409–422.
- Wong, J., Obst, R., Correia-Neves, M., Losyev, G., Mathis, D., and Benoist, C. (2007). Adaptation of TCR repertoires to self-peptides in regulatory and nonregulatory CD4⁺ T cells. *J. Immunol.*, 178:7032–7041.