# CHAPTER 3
## Development of the T cell repertoire

### 3.1: Introduction

As noted in Chapter 2, VDJ recombination results in a diverse repertoire of T cell receptors. The large diversity enables us to mount T cell responses to a diverse and evolving world of microbes. That is, upon infection with a particular pathogen, with high likelihood at least a few of the pMHC molecules derived from its proteome will bind sufficiently strongly to some T cells in the organism. Sufficiently strong binding can lead to T cell activation, which is the necessary condition for a T cell response. The T cell repertoire exhibits several properties that cannot be explained by the diversity generated by VDJ recombination alone. T cell responses are extremely specific in that most point mutations to a peptide recognized by a particular T cell abrogate recognition. Such specificity is necessary for pathogen-specific T cell responses. At the same time, most T cells are cross-reactive to a few peptides. This seems necessary as the number of T cells in an organism cannot be as large as all possible pathogenic peptides that may emerge. Most cells display pMHC molecules with peptides derived from host proteins. The T cell repertoire is largely self-tolerant, in that most T cells do not respond to most such host-derived pMHC molecules. But some autoreactive T cells exist in every individual, and yet most individuals do not get autoimmune diseases. Such a largely self-tolerant, pathogen-specific, yet somewhat degenerate, T cell repertoire is the result of processes that occur during T cell development and other associated mechanisms. In this chapter, we will study some aspects of T cell development that provide insights into the mechanisms that result in the properties of the mature T cell repertoire noted above.

The thymus is an organ located behind the sternum (Fig. 3.1). Until the 1960s, the only known purpose of this organ was that the thymus of an animal is the main ingredient of sweetbreads, a dish that is considered a culinary delicacy in many countries. Jacques Miller discovered that the thymus played a critical role early in the life of an organism. He found that, if the thymus of a mouse is removed at birth, the animal suffers from immunodeficiency. This was indicated by these mice not rejecting transplanted skin from a genetically different mouse and being more prone to infections. After exiting the bone marrow immature T cells undergo development in the thymus to emerge as mature T cells. Removing the thymus prevented these developmental processes, and so there was no T cell response to transplanted skin or infections. After puberty, the thymus begins to wither and becomes fatty, and T cell output is dramatically reduced. Interestingly, without knowing any of this, chefs knew that sweetbreads should be prepared using the thymus of a young animal.

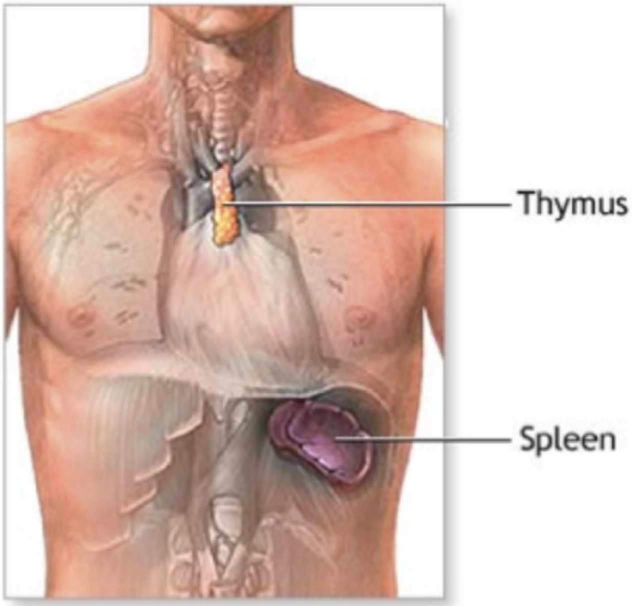

Fig. 3.1: Immature T cells migrate to the thymus, wherein they undergo various developmental processes and are exposed to pMHC molecules with peptides derived from diverse parts of an organism's own (self) proteome.

The processes that occur in the thymus during T cell development are complex, and detailed descriptions of the biology are available in standard immunology textbooks. In this chapter, we will first describe only some aspects of these processes that shape key features of the T cell repertoire. We will then focus on how insights into the consequences of these processes on the nature and function of the T cell repertoire can be obtained by complementary theoretical analyses, computation, experimental studies in mice, and clinical observations. Finally, we will end by noting some open questions that need to be studied in future.

## 3.2: Basic biology of thymic development

### *3.2.1: Positive and negative selection in the thymus*

The thymus is comprised of a network of cells called the thymic stroma, and it can roughly be partitioned into an outer region called the cortex, and an inner region called the medulla. In addition, the thymus also contains dendritic cells and macrophages. Precursors of T cells, called thymocytes, arrive in the thymus from the bone marrow. At this stage, the TCR genes have not rearranged, and these early stage thymocytes are called double negative thymocytes because they do not express the co-receptors, CD4 and CD8. Interactions with thymic stromal cells then result in various stages of thymocyte differentiation and development. During these steps, various surface molecules are expressed, the TCR $\beta$ chain genes rearrange, CD4 and CD8 are both expressed, and then the TCR $\alpha$ chain also rearranges. At this stage, thymocytes are called double positive because they express both types of co-receptors. Then, several stages of differentiation occur during which T cells become single positive (either CD4 or CD8 ceases to be expressed), and two processes called positive and negative selection occur. Ultimately, a fraction of single positive T cells that successfully undergo these processes exit the thymus and become part of the mature T cell repertoire that circulates in blood and peripheral tissues. In mice, an immature T cell precursor spends about 3 weeks in the thymus before it exits as a mature T cell.

Our focus in this chapter will be on the processes of positive and negative selection that occur in the thymus. Usually, cells in multicellular organisms express only a subset of genes in a tissue-specific manner. The expression of the autoimmune regulator (AIRE) gene causes promiscuous expression of diverse host genes in thymic cells. Thus, thymic cells express pMHC molecules with peptides derived from diverse parts of the host proteome that are normally expressed in different tissues. Thymocytes migrate through the thymus and interact with these self pMHC molecules. For a thymocyte to successfully mature into a peripheral T cell, it must pass two tests. It *must bind to at least one* of the self pMHC molecules it encounters with an affinity that exceeds a threshold value ($E_p$). This process is called positive selection, and there is some evidence that this is mediated largely by pMHC molecules presented on thymic epithelial cells. Thymocytes *must also not bind to any* of the encountered self pMHC molecules with an affinity that exceeds another affinity threshold ($E_n$), which corresponds to stronger binding than $E_p$. This process is called negative selection, and many cell types, including thymic dendritic cells and macrophages play a role in mediating negative selection.

A set of experiments in mice provided insights into the positive and negative selection thresholds, $E_p$ and $E_n$. Transgenic mice are genetically engineered animals whose T cells only express a single type of TCR. The T cells in such mice with a particular TCR, called OT-1, respond to a peptide (called OVA) from the chicken ovalbumin protein in complex with MHC class I molecules. The affinities and potencies of the OT-1 TCR for different variants of this so-called OVA peptide have been characterized. The potency can be characterized by pulsing APCs with the peptide so that the corresponding pMHCs are presented on its surface. Then T cells can be stimulated by these APCs, and a downstream marker of intracellular T cell signaling that serves as a proxy for activation can be measured. The variation of the extent of downstream signaling with peptide concentration exhibits a sigmoidal shape. A metric of the potency of a particular variant of the OVA peptide for stimulating OT-1 T cells is the peptide concentration that results in half the maximum amount of downstream signaling. This quantity is referred to as $EC_{50}$ (Fig. 3.2a). More potent peptides are characterized by smaller values of $EC_{50}$. The affinities, or half-lives of binding, of the variant peptides for the OT-1 TCR can be assessed using surface plasmon resonance experiments. Peptides with higher affinities, or usually, longer half-lives of binding to the TCR (see Chapter 6), are more potent.

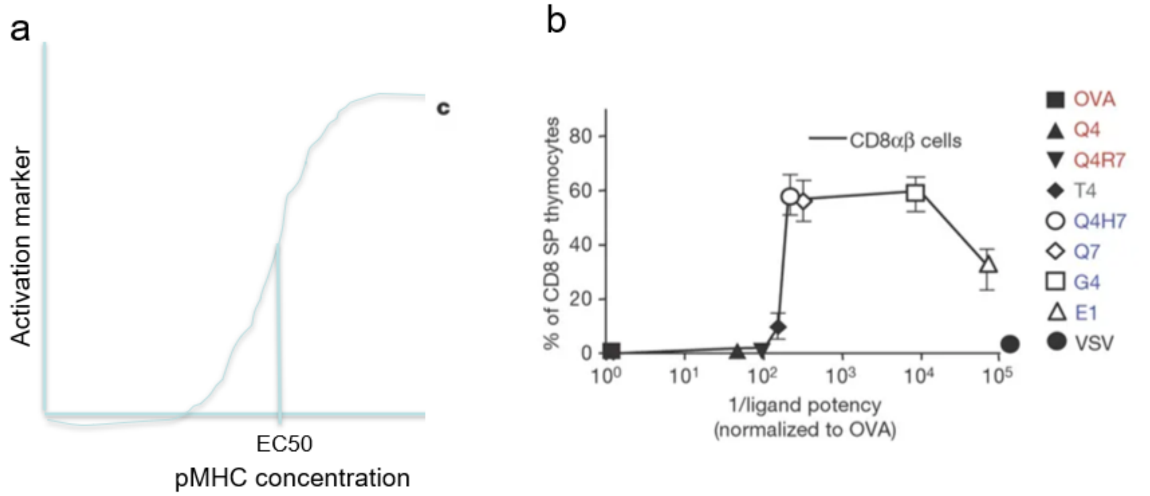

Fig 3.2: (a) A metric of the potency of a pMHC molecule in activating a T cell with a particular TCR is obtained by graphing the amount of a downstream signaling marker that is activated as a function of the pMHC concentration used to stimulate the T cell. The pMHC concentration that results in half the maximal amount of the stimulated activation marker is called the $EC_{50}$, and this value defines its potency. (b) Binding thresholds for positive and negative selection are defined by graphing the number of single positive thymocytes that successfully mature in FTOC pulsed with pMHC molecules of varying potency. The least potent pMHC (VSV) does not lead to positive selection, while the most potent pMHC (e.g., OVA) delete thymocytes by negative selection.

The thymus of a fetal mouse can be cultured *in vitro* (fetal thymic organ culture or FTOC), and the development of thymocytes can be studied in FTOC. To characterize the positive and negative selection thresholds, double positive thymocytes were taken from transgenic mice expressing the OT-1 TCR and the number of single positive T cells that develop in FTOC pulsed with variant OVA peptides was assessed. The dependence of this quantity on the potency of the peptide (measured as described above) shows that the negative selection threshold is very sharply defined, while the positive selection threshold is somewhat softer (Fig. 3.2b). These data also indicate that the thresholds of negative and positive selection, $E_p$ and $E_n$, are separated by a relatively small difference (a few thermal energy units, $k_BT$, where $k_B$ is Boltzmann's constant and T is the absolute temperature at physiological conditions). Of course, these numbers are not known for humans as similar experiments would be difficult to do.

By now, it must be obvious to the reader that the process of negative selection aims to delete T cells with TCRs created by VDJ recombination that bind too strongly to self pMHC molecules. Such T cells could potentially initiate autoimmune responses. The process of positive selection likely ensures that a mature T cell is at least capable of binding to host MHC molecules and bind to peptides presented in the MHC groove. But T cells that successfully exit the thymus do not encounter all self pMHC molecules presented in the thymus, and so T cells that could be activated by self pMHC must be present in the mature repertoire; indeed, data from humans show that this is the case.

### *3.2.2: The effect of thymic selection on T cell specificity for pathogenic peptides*

Experiments by Huseby and co-workers revealed that developmental processes in the thymus influence the exquisite specificity with which T cells respond to antigenic peptides. These experiments were carried out with two groups of mice: 1] Mice that display a normal diverse array of self pMHC molecules in the thymus, and 2] Mice that were engineered to display only a single type of self pMHC molecule in the thymus but were otherwise genetically identical to those in the first group. For each of a few mature T cells isolated from both groups of mice, a stimulatory peptide was identified. The amino acids at different sites of each stimulatory peptide were mutated to other amino acids. For each mutant peptide, the response of the T cell specific for the original peptide was re-assessed. The data showed that T cells that develop in mice that express the usual diversity of self pMHC molecules in the thymus exhibit a significantly higher degree of specificity compared to those that developed in mice engineered to display one type of self pMHC (Fig. 3.3). That is, point mutations to a stimulatory peptide were more likely to abrogate activation of mature T cells in the normal mice. If half the mutations at a particular site in a stimulatory peptide abrogated a stimulatory response, the site was labelled a "hot spot". The greater the number of hot spots in the peptide, the more sensitive T cell recognition is to point mutations. The peptides that stimulated T cells derived from mice in group 1 exhibited more hot spots. These results raise intriguing questions regarding how thymic development shapes the antigen specificity of the mature T cell repertoire.

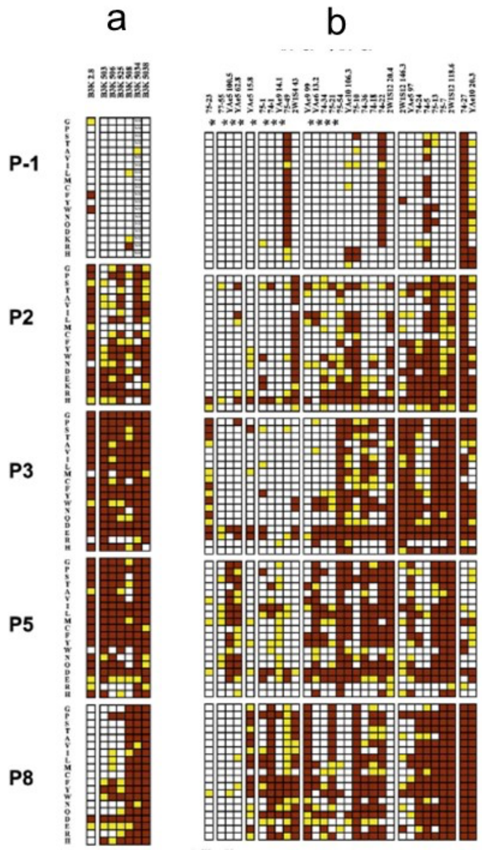

Fig. 3.3: T cell activation sensitivity to amino acid mutations in a stimulatory peptide. A few mature T cells isolated from mice that reacted to a particular peptide-MHC complex were identified. Each column in panels (a) and (b) of the figure corresponds to one such T cell. The proliferation of the same T cells to the peptide upon making amino acid changes at 5 of its residues (P1, P2, P3, P5, P8) was re-assessed. If T cell proliferation in response to the mutated peptide was greater than 50 % the rate of proliferation when stimulated by the original peptide, the mutated peptide was considered fully stimulatory (white pixels). Less than 5 % proliferation rate and between 5 and 50 % proliferation rate are shown with brown and yellow pixels, respectively. (a) Response of T cells isolated from a mouse displaying diverse self peptides in the thymus. (b) Response of T cells isolated from a genetically identical mouse as in (a) that was engineered to display only a single self peptide in the thymus. There are many more white pixels in panel (b) compared to panel (a). Figure adapted from Huseby et al, Cell (2005).

## 3.3: Exploring the mechanism of how thymic selection influences the antigen specificity of T cells

Let us first consider how positive and negative selection in the thymus shapes the characteristics of the T cell repertoire in terms of very simple models. Our goal is to obtain mechanistic insights into the origin of the experimental results described above. We begin with a primer on extreme value statistics.

### 3.3.1: A primer on extreme value statistics

Consider N independent samples $\{r_1, r_2, \ldots r_N\}$ drawn from identical independent Gaussian distributions. Let $P_N$ (x) be the probability that the maximum value of these numbers is less than or equal to x; i.e.,

$$P_N(x) = [1 - \bar{P}(x)]^N;\ \bar{P}(x) = \int_x^\infty dr\, p(r) \qquad (1)$$

where p (r) is the Gaussian probability distribution for each random variable. The probability distribution, $p_N(x)$, that the largest number lies between x and x+ dx is obtained by differentiating Eq. 1, and equals:

$$p_N\ (x) = N\ [1 - \ \bar{P}\ (x)]^{N-1}\ \left(- \frac{d\bar{P}\ (x)}{dx}\right) = \ N\ p(x)\ [1 - \ \bar{P}\ (x)]^{N-1} \qquad (2)$$

Differentiating Eq 2 with respect to x and setting the resulting expression to zero provides the most likely value of the largest number, x*. Specifically, x* must obey

$$p'\ (x^*) - \ p'\ (x^*)\ \bar{P}\ (x^*) + N\ p^2\ (x^*) - \ p^2\ (x^*) = 0 \qquad (3)$$

If N is large, we expect x* to be a large number. Therefore, the values of p (x*), $p'\ (x^*)$, and $\bar{P}\ (x^*)$ are sampled from the tails of the corresponding probability distributions, and so will be small. Thus, we consider the following leading order approximation to Eq. 3:

$$p'\ (x^*) + N\ p^2\ (x^*) = 0 \qquad (4)$$

We can solve Eq. 4 for x* by noting that p (x) is Gaussian; i.e., $p\ (x) = \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{x^2}{2\sigma^2}}$. This obtains:

$$\ln x^* + \frac{x^{*2}}{2\sigma^2} = \ln N + \ln \frac{\sigma}{\sqrt{2\pi}} \qquad (5)$$

For large N, and hence x*, we obtain:

$$x^* = \ \pm \sqrt{2\sigma^2 \ln N} \qquad (6)$$

We can also compute the entire probability distribution, $p_N$ (x). Toward this end, start with Eq. 1. As x will be large for large N, the value of x in the integral, $\int_x^\infty dr\ p\ (r)$, in Eq. 1 will be small. So, for large N, we can rewrite Eq. 1 as follows:

$$P_N\ (x) \sim \exp\left[- N\ \int_x^\infty dr\ p\ (r)\right] \qquad (7)$$

If the tail of the probability distribution, p (r), falls of exponentially as $a \exp(- \lambda r)$, then Eq. 7 becomes

$$P_N\ (x) \sim \exp\left[- \frac{Na}{\lambda}\ e^{-\lambda x}\right] \qquad (8)$$

So, $p_N$ (x) is

$$p_N\,(x) = \frac{dP_N\,(x)}{dx} = Na\exp\left[-\lambda x - \frac{Na}{\lambda}\,e^{-\lambda x}\right] \qquad (9)$$

Now $p_N$ (x) will be maximized when $\lambda = N\,a\,e^{-\lambda x^*}$, and the optimum value of x is x* = $\frac{1}{\lambda}\ln\left(\frac{Na}{\lambda}\right)$. Using this result, we can rewrite Eq. 9 as follows:

$$p_N\,(x) = \lambda\exp\left[-\lambda\,(x - x\,*) - e^{-\lambda(x - x*)}\right] \qquad (10)$$

The probability distribution represented by Eq. 10 is called the Gumbell distribution. For x > x*, $p_N\ (x)$ has an exponential tail. If p (r) is Gaussian, x* is given by Eq. 6.

3.3.2: Analysis of thymic selection effects in terms of extreme value statistics

Recall that if a T cell has to successfully mature, it must bind at least one self-peptide it encounters in the thymus with an affinity that exceeds $E_p$, and to none with an affinity that exceeds $E_n$. Let us use a convention wherein $E_p$ and $E_n$ correspond to binding free energies and higher values correspond to higher affinity. Then the conditions for a thymocyte to successfully pass both positive and negative selection are met if the maximum value of the interaction free energies of the T cell with the N self-peptides encountered in the thymus lies between $E_p$ and $E_n$. Therefore, the probability that a T cell with a sequence of amino acids that contact peptide amino acids, $\vec{t}$, will successfully mature, $P^{mat}\,(\vec{t})$, is:

$$P^{mat}\,(\vec{t}) = \int_{E_p}^{E_n} p_N\,(x,\vec{t})\,dx \qquad (11)$$

where $p_N\,(x,\vec{t})$ is the probability that the largest value of the binding free energy of the T cell with the sequence, $\vec{t}$, with N types of self-peptides in the thymus lies between x and x + dx. We also know that $P^{mat}\,(\vec{t})$ is the probability of not being negatively selected minus the probability of not being positively selected; i.e.,

$$P^{mat}\,(\vec{t}) = \left[1 - \bar{P}(E_n,\vec{t})\,\right]^N - \left[1 - \bar{P}\,(E_p,\vec{t})\,\right]^N \qquad (12)$$

$\bar{P}$ has the same meaning above as in Eq. 1. Elementary calculus tells us that Eqs 11 and 12 imply that $p_N\,(x,\vec{t})$ is the exact differential of $\left[1 - \bar{P}(x,\vec{t})\,\right]^N$. Therefore, it is given by Eq. 2; i.e.,

$$p_N\,(x,\vec{t}) = N\,[1 - \ \bar{P}\,(x,\vec{t})]^{N-1}\,\left(-\,\frac{d\bar{P}\,(x,\vec{t})}{dx}\right) = \ N\,p(x,\vec{t})\,[1 - \ \bar{P}\,(x,\vec{t})]^{N-1} \qquad (13)$$

The distribution, $p_N\,(x,\vec{t})$, is the Gumbell distribution (Eq. 10). But we can also deduce the shape of $p_N\,(x,\vec{t})$ based on physical arguments. The probability that a T cell will bind very strongly to an arbitrary peptide is small. This is because a molecule is unlikely to bind to most other molecules avidly. So, $[1 - \ \bar{P}\,(x,\vec{t})]$ has the shape shown in Fig. 3.4a. If we assume that many interactions between T cells and peptides are sampled, p (x) is Gaussian, and so $p_N\,(x,\vec{t})$ has the shape shown in Fig. 3.4b.

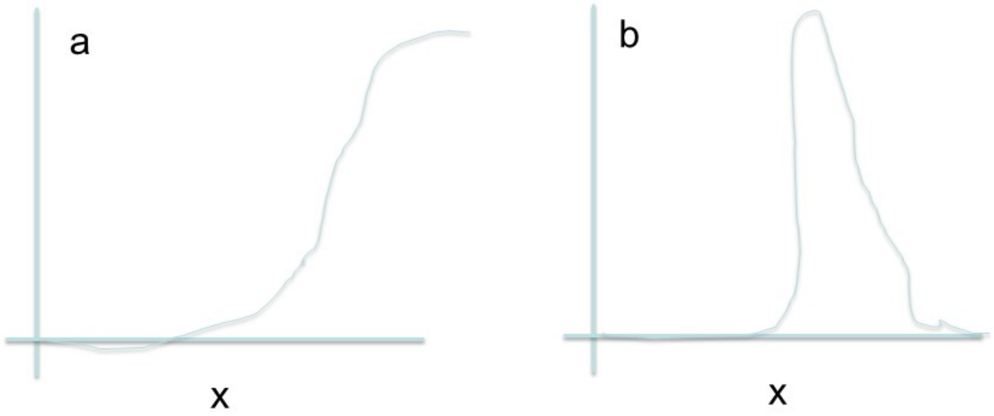

Fig. 3.4: (a) Schematic depiction of the shape of $[1 - P\,(x,t)]$ based on physical reasons in the text following Eq. 13. (b) Multiplying the function depicted in (a) by a Gaussian results in the curve shown.

Notice that the shape of $p_N\,(x,\vec{t})$ we have deduced based on physical arguments has a long tail for x > x*, as the Gumbell distribution does regardless of whether p (x) is Gaussian (Eq. 10). As N increases, $[1 - \ \bar{P}\ \ (x,\vec{t})]$ is raised to a high power in Eq. 13, which makes $p_N\,(x,\vec{t})$ become more sharply peaked. This is consistent with the fact that the Gumbell distribution is more sharply peaked as λ increases which, in turn, increases with N. For large enough N then, a good approximate condition for a T cell to successfully mature is that the most likely value of the largest of its N interaction free energies with self-peptides encountered in the thymus (E*) lies between $E_p$ and $E_n$.

The simplest estimate of the interaction free energy between a TCR with a sequence of amino acids that contact peptide amino acids, $\vec{t}$, and a peptide with amino acid sequence, $\vec{p}$ can be represented as:

$$E^t(\vec{t},\ \vec{p}) = E_c^t + \sum_{i=1}^{L} U\ (t_{i,}\ p_i) \qquad (14)$$

where $E_c^t$ is the interaction free energy between the TCR and the MHC, and the function, U, describes the interaction between the amino acids at the L peptide contact sites of the TCR ($t_i$ ) and the corresponding peptide amino acids ($p_i$). Most of the TCR sites that make contacts with a peptide are in the so called Complementary Determining Region 3 (CDR3) region. This is the region in which TCRs are most different from each other. We have assumed a very simple model in which the amino acids of a TCR that make contacts with peptide amino acids and the peptides are represented as strings of sites, and each TCR amino acid in the contact region interacts with a single amino acid on the peptide. In this representation, we are ignoring the conformations of the CDR loops and the peptide. As we are focused on short peptides and just the peptide contact sites of the TCR, this Occam's razor like approximation may not be so bad. The experiments carried out by Huseby and co-workers (see section 3.2.2) focused on T cells restricted by a single MHC allele, but still its interaction free energy with different TCRs will vary, and so $E_c^t$ should be picked from a distribution of values. If the value of $E_c^t$ is too close to the positive selection threshold (and hence also the negative selection threshold as $E_p$ and $E_n$ are not that far apart), then the TCR is likely to be negatively selected because interactions with the peptide are likely to make the total interaction free energy exceed $E_n$. Conversely, if the value of $E_c^t$ is too low, the TCR is unlikely to be positively selected. Therefore, the values of $E_c^t$ characterizing mature T cells are most likely to be drawn from a relatively narrow distribution. To obtain qualitative mechanistic insights let us take it to be a constant.

We can rewrite Eq. 14 as follows:

$$E^t(\vec{t},\ \vec{p}) = E_c^t + \sum_{i=1}^{L}\langle U\ (t_i)\rangle + \sum_{i=1}^{L} \delta\ U\ (t_i, p_i) \qquad (15)$$

where $< U\ (t_i) >$ is the average value of the interaction free energy of the amino acid at the i$^{th}$ peptide contact site of the TCR under consideration with all other amino acids; $\delta\ U\ (t_i, p_i)$ is the fluctuation around this average for interactions with a particular amino acid in the peptide sequence, $p_i$. If $\delta\ U\ (t_i, p_i)$ is distributed in a Gaussian fashion, we can use our previous result for x* (Eq. 6) to obtain the most likely value of the maximum interaction free energy of a TCR with N self-peptides in the thymus (E*) to be:

$$E^*(\vec{t}) = E_c^t + \sum_{i=1}^{L} \langle U\,(t_i)\rangle + \sqrt{2\sigma^2 \ln N} \qquad (16)$$

where $\sigma^2$ is the sum of the variances of the interaction free energies of the individual amino acids, $t_i$, that make up the peptide contact sites of the TCR under consideration. Eq. 16 makes a clear prediction. Different thymocytes will encounter different numbers of types of peptides in the thymus (N), but as more peptides are expressed in the thymus, each thymocyte will encounter more peptides. So, although how fluctuations in the value of N for each thymocyte affect repertoire-scale properties are worth exploring, we can learn a lot by just exploring the consequences of increasing the average number of types of peptides (N) that thymocytes encounter in the thymus.

As N increases, the value of E* increases. So, to avoid being negatively selected, the TCRs expressed by thymocytes that successfully mature will be characterized by smaller values of $\sum_{i=1}^{L} \langle U\,(t_i)\rangle$. That is, if N is large, mature T cells will express TCRs with peptide contact sites comprised of amino acids that do not bind to other amino acids very strongly. They cannot bind to other amino acids too weakly either as that would prevent thymocytes with such TCRs from being positively selected. So, Eq. 16 predicts that, when a sufficiently large diversity of self-peptides is expressed in the thymus, mature T cells are statistically likely to express TCRs whose peptide contact residues are comprised of amino acids that bind moderately to other amino acids. In contrast, when N is small, as for the engineered mouse in the experiments described in section 3.2.2 that displays a single type of self-peptide in the thymus, T cells with peptide contact sites comprised of amino acids that tend to bind to other amino acids more strongly are more likely to successfully mature into peripheral T cells.

The biophysical reason underlying these predictions is easy to understand. If selection of thymocytes is mediated by only one type of self peptide-MHC molecule, for a thymocyte to mature, its TCR must bind to this single molecule with a free energy that lies between $E_p$ and $E_n$. Two common ways in which this can happen are shown in Fig. 3.5. One way is for the TCR to bind relatively strongly to the MHC, and weakly to the peptide. Another is to bind moderately to the MHC, but the binding free energy for the single type of peptide is dominated by one or two very strong contacts. Both these two types of TCRs would be negatively selected if the corresponding thymocytes encountered a large diversity of self peptides in the thymus. For the first type of TCR, an encounter with any peptide with which its peptide contact amino acids interacted moderately with the peptide's amino acids would likely push its binding free energy past the negative selection threshold. Similarly, additional moderate contacts with amino acids of other peptides would lead to negative selection for the second type of TCRs noted above. The point to recognize is that avoiding being negatively selected by a diversity of self peptides encountered in the thymus imposes a strong constraint that is most likely to be satisfied by thymocytes with TCRs that bind moderately to the MHC molecule that restricts it, and whose peptide contact residues are comprised of amino acids that bind moderately to other amino acids.

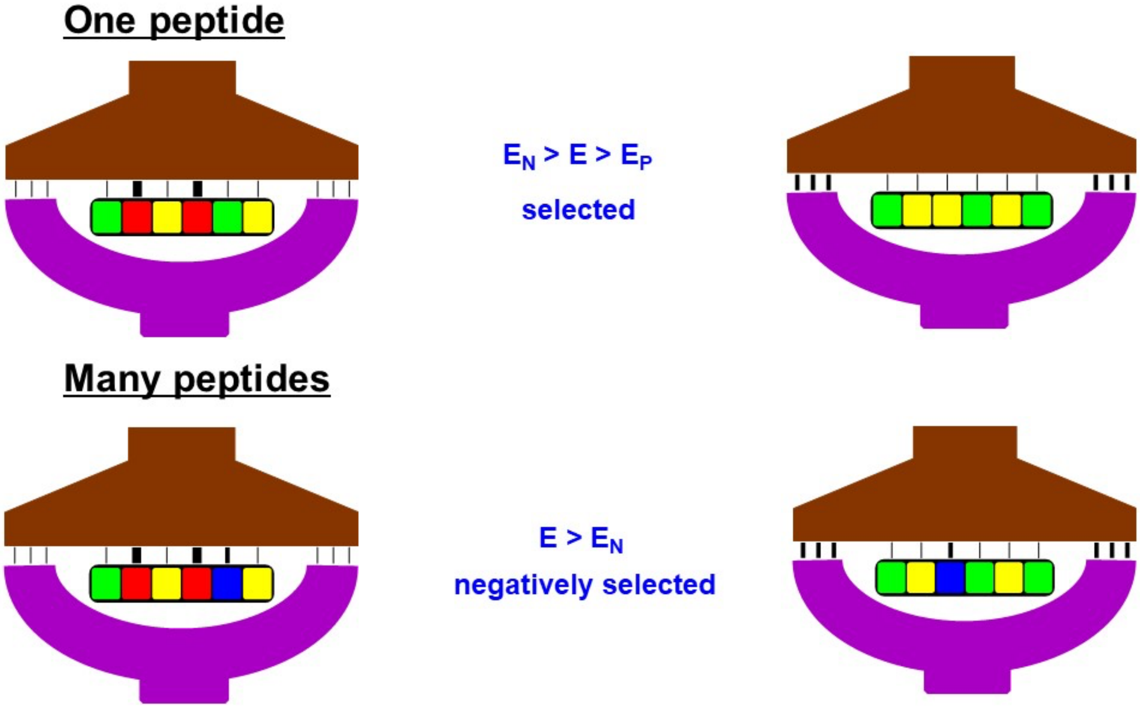

Fig. 3.5: The top row is a schematic depiction of two common ways in which thymocytes expressing a certain TCR could develop successfully in a thymus expressing only one type of self pMHC molecule. The TCRs are shown in brown, the MHC in magenta, and different possible single peptides are shown with differently colored amino acids. Wider and bolder bars show stronger interactions. The bottom row shows how both the types of TCRs that develop successfully when there is a single self pMHC molecule would likely be negatively selected due to additional moderate to strong interactions with other self peptides when many self peptides are expressed in the thymus.

The predictions emerging from Eq. 16 can explain why T cells that mature in a normal thymus expressing a diversity of self peptides exhibit high peptide antigen specificity (many hotspots). Experiments show that the threshold binding free energy for mature T cells to be activated by a pathogenic peptide is close to $E_n$. Eq. 16 predicts that T cells that mature in a normal thymus are likely to express TCRs with peptide contact sites comprised of moderately interacting amino acids. For such a T cell to recognize a pathogenic peptide, the amino acids of the peptide must be among the more strongly binding complements of the amino acids that comprise this TCR's peptide contact residues. Otherwise, it would be difficult to achieve a binding free energy of the order of $E_n$ which was avoided in the thymus when interacting with randomly selected self peptides. Furthermore, each contact with the peptide amino acids is likely to contribute a significant fraction of the total binding free energy; i.e., the threshold binding free energy required to recognize a pathogenic peptide is achieved by multiple moderate interactions acting collectively (Fig. 3.6a). Because the recognized peptide's amino acids are likely to be among the stronger binding complements of the TCR's peptide contact residues, if one of the peptide amino acids is mutated, it will likely be to an amino acid that binds less strongly to the corresponding TCR peptide contact site. This will result in a decrease in binding free energy for the mutant peptide. Since each contact contributes a significant fraction of the overall binding free energy, this effect will likely result in a significant percentage reduction in the binding affinity. The binding free energy threshold that leads to recognition in the periphery is sharply defined (see Chapter 6), and so a significant reduction in binding free energy is likely to lower it below this threshold. Taken together, the statements above argue that, for T cells that mature in a normal thymus, most point mutations to a pathogenic peptide that activates it are likely to abrogate recognition. That is, there will be many hot spots as observed in experiments.

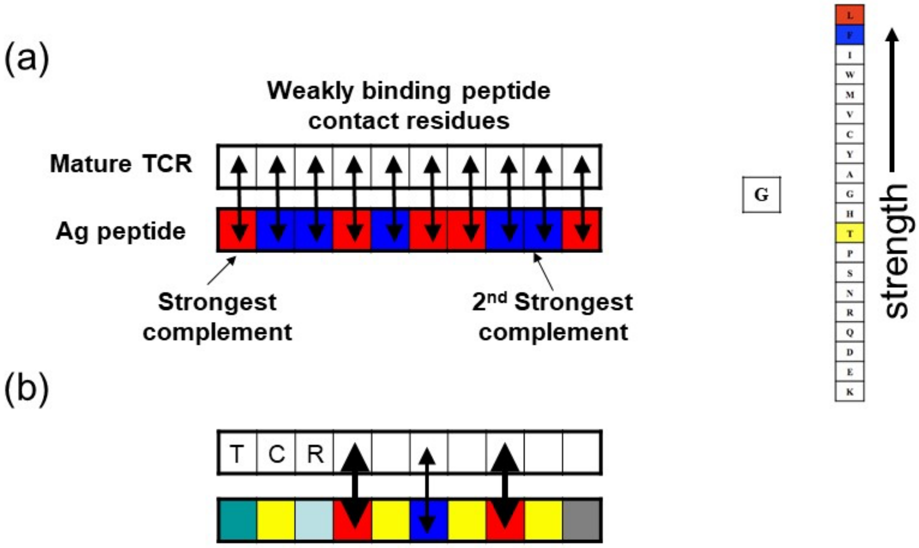

Fig. 3.6: (a) For T cells that undergo selection against many self peptide-MHC molecules,the "moderate" peptide contact residues on the TCR must bind a sufficient number of its stronger complementary amino acids for recognition via multiple moderate interactions. The schematic on the right illustrates that any amino acid interacts with others with different strengths. For the case shown, L is the strongest complement. (b) For T cells that undergo selection in a thymus with only one type of self peptide-MHC molecule, strongly interacting amino acids on the TCR will require only a few important contacts to mediate recognition.

Eq. 16 also predicts that T cells that mature in a thymus with a small diversity of expressed self peptides, such as an engineered mouse that displays a single type of self peptide, will exhibit few hot spots. This is because some of the peptide contact sites of these T cells are more likely to be comprised of amino acids that bind strongly to others. So, strong interactions between a small number of TCR amino acids with a peptide's corresponding amino acids may be sufficient to exceed $E_n$ (Fig. 3.6b) So, only mutations to these few peptide amino acids are likely to abrogate recognition. Thus, there will be fewer hot spots, as observed in experiments.

The implications of Eq. 16 also enable us to understand why T cells that mature in a normal thymus can simultaneously exhibit high specificity for certain peptides (sensitivity to point mutations), while simultaneously exhibiting some degree of cross-reactivity. This is because shuffling the peptide amino acid sequence in a way that maintains the same number of moderate interactions with the TCR's peptide contact sites can result in activation.

Our analyses suggest that the mechanism underlying T cell recognition of specific peptide-MHC molecules is not like Emil-Fisher's lock and key metaphor for the specificity of enzyme-substrate reactions. Rather, recognition is the consequence of statistical matching of the patterns of the amino acids of a TCR's peptide contact sites and the amino acids of the peptide. This is perhaps not surprising, because unlike enzyme-substrate pairs, T cells did not co-evolve with the diverse microorganisms whose peptides they recognize with specificity. Indeed, our T cells can specifically recognize peptides derived from microorganisms that evolved after we were born. The mechanistic picture that emerges from our consideration of highly simplified models is undoubtedly further embellished by many molecular details that were ignored. These details matter for individual TCRs and the peptide-MHC molecules that they interact with. It is important, however, to test whether the essence of the proposed mechanism is statistically correct as the statistical collective properties of the T cell repertoire help clear infections. Below, we first briefly note tests that can be carried out using numerical simulations and then focus on experimental tests in mice.

### 3.3.3: Numerical tests of model predictions

Results of numerical studies of the thymic selection process are consistent with insights regarding the origin of the antigen specificity of TCRs that follow from eq. 16. We can generate a repertoire of TCRs expressed by thymocytes by picking amino acids for each peptide contact site in individual TCRs according to the frequencies of occurrence of amino acids in the human proteome. Similarly, we can generate an ensemble of peptides that are displayed in the thymus. One can make peptide generation more realistic by using one of a number of publicly available computer programs to select self peptides based on whether they bind to particular MHC alleles. We can use Eq. 14 to calculate each TCR's interaction free energies with N randomly selected self peptides. The simplest way to do this (Fig. 3.7) is to choose a value for $E_c^t$ that is neither too high or too low (see arguments in the preceding section), and then use a parametrized potential function to compute the free energy of interactions between the amino acids in the TCR's peptide contact sites and those that comprise the peptide ($\sum_{i=1}^{L} U\ (t_{i,}\ p_i)$). Only thymocytes that bind to at least one self peptide-MHC complex with a free energy that exceeds $E_p$ and none of the N self peptide-MHC complexes with a free energy that exceeds $E_n$ are allowed to successfully mature. Thus, we obtain a mature T cell repertoire for a given choice of the number of self peptides (N) expressed in the thymus.

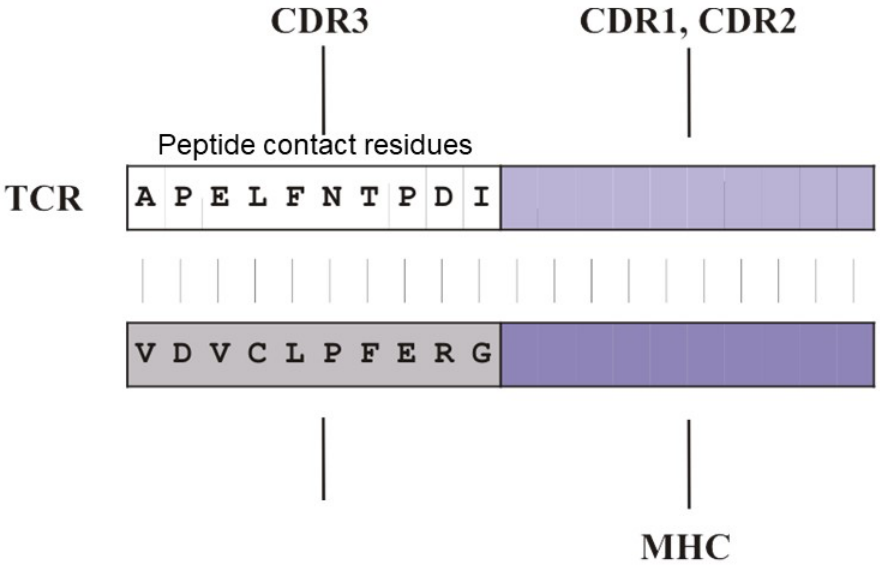

Fig. 3.7: A simple "string model" to estimate free energies of interaction between a TCR and a peptide-MHC complex. The CDR1 and CDR2 regions of a TCR make most contacts with the MHC, while the CDR3 region usually contains the peptide contact residues. Eq. 14 is used to estimate the free energy of interaction as described in the text.

We can numerically generate a panel of pathogenic peptides by picking their amino acids according to the frequency of occurrence of amino acids in a model pathogen, for example, *Listeria Monocytogenes*. One can then carry out the numerical analog of the experiments described in section 3.2.2. The interaction free energy of a mature TCR with several pathogenic peptides is determined. A TCR is considered to recognize a pathogenic peptide if the corresponding interaction free energy exceeds $E_n$ (for reasons noted earlier). For peptides recognized by a mature TCR, each of its amino acids can be mutated to its nineteen possible mutants, and the binding free energy reassessed using Eq. 14. Peptide residues wherein half the mutations result in a binding free energy below $E_n$ are hot spots (as in experiments). These numerical calculations must be carried out with many realizations of thymocyte sequences, the peptides displayed in the thymus, and pathogenic peptides in order to obtain statistically meaningful measures of the specificity of the mature TCR repertoire.

Eq. 16's predictions hold true for any potential function, $U\ (t_i, p_i)$, in Eq. 14. To simulate the thymic selection process numerically as described above, however, we have to choose a particular form for this interaction free energy as well as the size of the gap between $E_p$ and $E_n$.

One possible choice for $U\ (t_i, p_i)$ is the Miyazawa-Jernigan potential that describes interactions between amino acids, and which has been used to study protein folding. This potential tends to overemphasize hydrophobic effects because of their importance in forming the core of folded proteins. Hydrophobic effects should also be important for forming the TCR-peptide-MHC interface, and we anticipate that hydrophobicity of an amino acid may be a reasonable zeroth order proxy for its strength of binding to other amino acids at the TCR-pMHC interface. Based on experiments in mice described earlier, the gap between $E_p$ and $E_n$ can be chosen. Fig 3.8 shows numerical results of the calculations outlined above with these choices of parameters and moderate values of $E_c^t$ (for reasons noted earlier). These results mirror the conclusions from Eq. 16 and experimental findings. Mature T cells that develop in a thymus with a larger diversity of expressed self peptides exhibit a higher degree of specificity for recognition of pathogenic peptides (more hot spots).

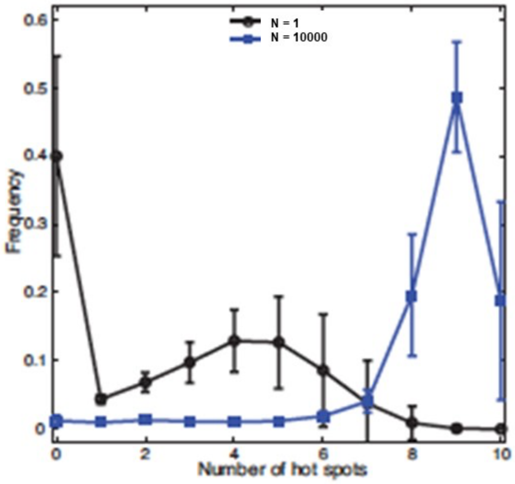

Fig 3.8: Results of numerical calculations for the frequency of occurrence of different numbers of "hot spots" (defined in text) characterizing T cell recognition of pathogenic peptides is shown for a large diversity of self peptide-MHC molecules (N = 10000) expressed in the thymus, and for a case when only one self peptide-MHC is expressed.

### 3.3.4: Experimental tests of the model predictions

The veracity of the model described in section 3.3.2 and its conclusions can only be established by experimentally testing its main predictions. The main predictions are: 1] Mature T cells that develop in a normal thymus expressing diverse self peptides express TCRs whose peptide contact sites are statistically enriched in moderately hydrophobic amino acids; 2] Negative selection constraints result in the depletion of TCRs with highly hydrophobic amino acids at the peptide contact sites. These predictions are challenging to test for two reasons. First, they are statistical statements and so one must examine properties of a very large number of TCRs. Second, they pertain to the peptide contact sites of the TCRs, and so these few sites in the much longer CDR3 region of the TCR have to be identified. Huseby and co-workers carried out a set of experiments in mouse models that overcame these challenges.

The investigations began by considering two TCRs, called YAe62 and B3K506, both of which are restricted by a common MHC class II molecule. In some, but not all, instances, the TCR $\beta$ chain is dominant in determining interactions with peptide-MHC molecules. Therefore, transgenic mice with either the $\beta$ chain of YAe62 or B3K506 were bred. The $\beta$ chains randomly pair with a diversity of $\alpha$ chains in these mice. The following differences were noted between thymocytes of these TCRs:

1] The cells with YAe62 TCRs displayed higher levels of downstream signaling markers, suggesting that they were more self-reactive. However, the total number of cells expressing either YAe62 or B3K506 that matured into single positive T cells were the same. These results suggest that greater numbers of YAe62 T cells are deleted by negative selection.

2] By breeding transgenic mice that do not express MHC molecules, and thus cannot carry out positive or negative selection, pre-selection thymocytes expressing either YAe62 or B3K506 TCR were isolated. *In vitro* assays showed that these YAe62 thymocytes were more reactive to self peptides. These results also suggest that pre-selection TCRs with the YAe62 $\beta$ chain paired with diverse $\alpha$ chains exhibit a high degree of self-reactivity.

3] The CDR3 regions of the $\beta$ chains of the two TCRs were found to differ primarily at two sites. Analyses of crystal structures revealed that these two sites are located at the center of the TCR-pMHC interface, and since they contact the sixth and seventh amino acids of peptides they were labeled, P6 and P7. The P6 and P7 amino acids for YAe62 are Phenylalanine and Tryptophan, while they are Serines for B3K506. Swapping the amino acids at the P6 and P7 sites of the two TCRs reversed the order of their self-reactivity. As

Phenylalanine and Tryptophan are more hydrophobic than Serine, these results indicated that the hydrophobicity of the peptide contact sites of TCRs play a role in determining the strength of reactivity to self peptide-MHC molecules.

4] Analyses of crystal structures of 53 different human and mouse TCR with different $\beta$ chains of varying lengths complexed with peptide-MHC molecules revealed that all use either the P6 or P7 sites to contact the peptide-MHC molecule, and in 43 cases both are used. These results suggested that sequencing many TCRs that develop in mice and relating the frequency of different types of amino acids at the P6 and P7 sites to their functional properties and fates during thymic development should be informative and enable testing the predictions emerging from Eq. 16.

5] A scale of hydrophobicity based on partitioning in octane-water mixtures was used to quantify the degree of hydrophobicity of the P6 and P7 doublet of sites for TCRs with different $V_\beta$ chains. Large scale sequencing studies revealed that pre-selection TCRs with more hydrophobic amino acids at the P6 and P7 sites promote self-reactivity. Decreasing hydrophobicity was found to correlate statistically with being neutral or reducing self-reactivity.

Based on the results noted above, one way to test the predictions that follow from Eq. 16 is to examine the enrichment or depletion of amino acids of different hydrophobicity at P6 and P7 in the mature T cell repertoire compared to the pre-selection repertoire (see point # 2 above to see how the latter class of TCRs can be isolated). Sequencing pre- and post-selection TCRs in mice shows that, for both CD4 and CD8 T cells, only moderately hydrophobic amino acids are enriched in the peptide contact sites of mature T cells compared to the preselection repertoire (Fig. 3.9 shows illustrative results for CD4 T cells). These results are concordant with the qualitative predictions emerging from Eq. 16.

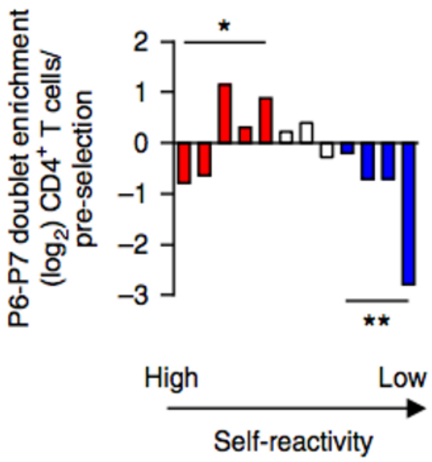

Fig. 3.9: Comparison of relative hydrophobicity of the P6-P7 doublet in mature and pre-selection TCRs is shown for CD4 T cells. Less hydrophobic amino acids are shown in blue. Similar results are seen for CD8 T cells.

Bim is a molecule that plays an important role in mediating apoptosis (death) of thymocytes. Mice with the gene that codes for Bim removed at both loci (Bim knock out or Bim-/- mice) exhibit defects in negative selection. When the frequencies of amino acids in mature and preselection TCRs were studied in Bim-/- mice, strongly hydrophobic amino acids were statistically enriched in the mature T cell repertoire (Fig. 3.10). Taken together, the results shown in Figs. 3.9 and 3.10 demonstrate that negative selection promotes the deletion of thymocytes with TCRs whose peptide contact residues are comprised of strongly interacting, or highly hydrophobic, amino acids. These results suggest that the analyses of thymic selection developed in section 3.3.2 provides qualitatively accurate mechanistic insights into the statistical consequences of thymic selection on some important characteristics of the mature T cell repertoire.

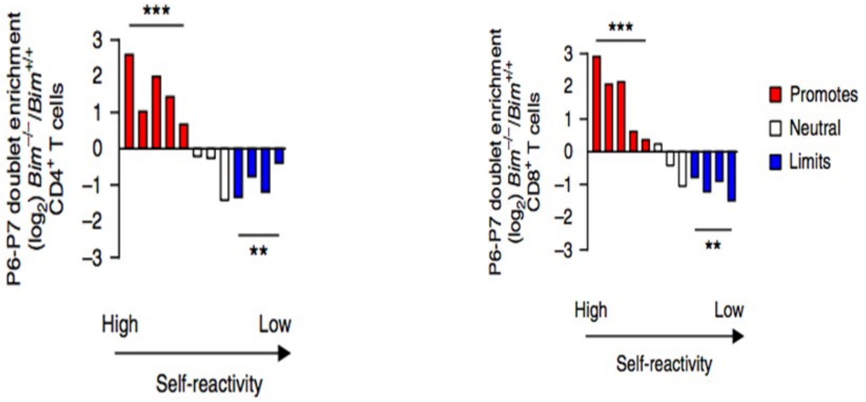

Fig. 3.10: Comparison of relative hydrophobicity of the P6-P7 doublet in mature and pre-selection TCRs in Bim-/- mice is shown for both CD4 and CD8 T cells. Less hydrophobic amino acids are shown in blue.

Additional experimental data are indicative, but not confirmatory, of the notion that a greater diversity of self peptides expressed in the thymus suppresses the frequency of highly hydrophobic amino in the peptide contact sites of mature TCRs. A common strain of mice that is employed in laboratory studies is the C57BL6 strain (hereon labeled B6), which was used in the studies that led to the results shown in Figs. 3.9 and 3.10. Another strain of mice, called Non-obese diabetic (NOD) mice, spontaneously develop diabetes, an autoimmune disease. Fig. 3.11 shows that mature CD4 T cells in NOD mice have a higher frequency of highly hydrophobic amino acids at the P6 and P7 sites compared to B6 mice. In contrast, this feature is not observed for CD8 T cells (Fig. 3.11). This result suggests that the higher frequency of hydrophobic amino acids at P6 and P7 sites observed for CD4 T cells is not because of any generic selection defect in NOD mice.

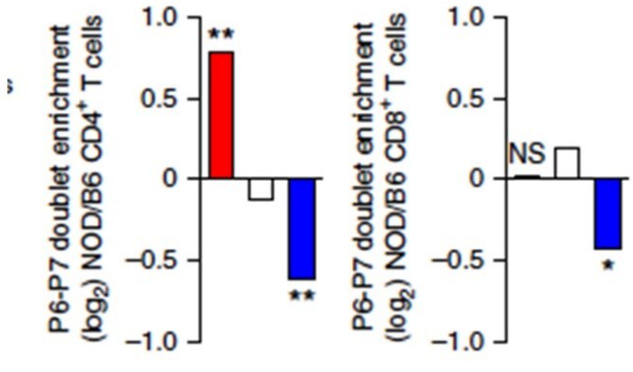

Fig 3.11: CD4 T cells of NOD mice show enrichment of hydrophobic residues (red) at P6 and P7 compared to B6 mice. This enrichment is not seen for CD8 T cells. Less hydrophobic residues are in blue.

The susceptibility of NOD mice to diabetes has been linked to their possessing a particular MHC class II allele which expresses the MHC class II molecule called $IA^g$. To explore whether the MHC class II alleles in NOD mice mediate the character of their CD4 repertoire (Fig. 3.11), Huseby and co-workers generated two sets of mice. One set, called NOD.b, had all the background genes of NOD mice, but the MHC genes were those of B6 mice. Another set, called $B6.H2^g$ had the background genes of the B6 strain, but the MHC genes were those of the NOD strain. For these strains of mice, the phenomenon observed in Fig. 3.11 was reversed in that the $B6.H2^g$ mice's TCR exhibited the characteristics of the NOD.b mice TCRs. Furthermore, the $B6.H2^g$ mouse strain exhibited a higher frequency of highly hydrophobic amino acids in the mature CD4 T cell repertoire compared to B6 mice, and similar frequencies as the NOD mice (Fig. 3.12). Other studies have shown that $IA^g$ binds peptides in a more unstable manner, suggesting that they may present a smaller diversity of self peptides in the thymus. Thus, the results shown in Figs 3.11 and 3.12 are consistent with Eq. 16; i.e., T cells that are selected against fewer self peptides are more likely to have strongly interacting amino acids at peptide contact sites.

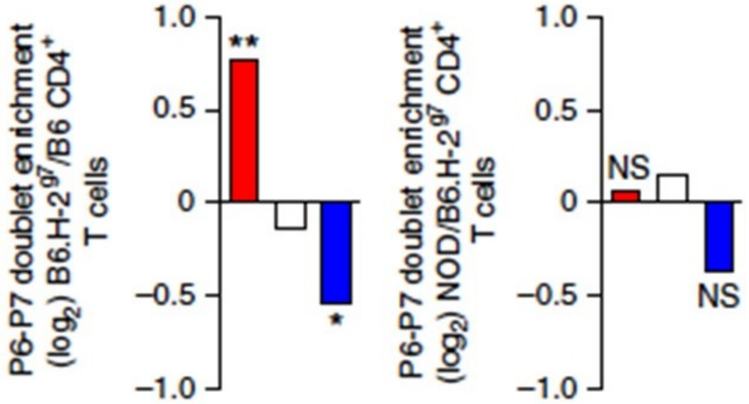

Fig 3.12: Compared to B6 mice, more hydrophobic residues (red) were enriched at P6 and P7 positions for mice that had the NOD MHC but were otherwise the same as B6 mice (left panel), and similar levels of more hydrophobic residues as NOD mice (right panel). Less hydrophobic residues are in blue.

Eq. 16 may also be consistent with some findings in humans. Human MHC genes are called Human Leukocyte Antigens (HLA). Two HLA class I alleles, HLA-B57 and HLA-B27, are statistically overrepresented in cohorts of patients who maintain low levels of HIV and HCV viruses without treatment. Interestingly, individuals with these HLA alleles are also statistically more likely to suffer from T cell-mediated autoimmune diseases, such as spondylitis and psoriasis. In the next chapter, we will see that the peptides presented by these HLA alleles are from regions of the HIV proteome that are highly vulnerable to mutations, and this is likely the principal reason why they are more likely to control HIV infections. However, as discussed below another factor could also play a role in this regard, which may also explain why individuals with these genes have a higher propensity for autoimmunity.

As noted in Chapter 2, machine learning approaches have been applied to experimental data on the equilibrium binding constants for diverse peptides binding to HLA molecules. The resulting predictive algorithms perform quite accurately for HLA class I molecules because they bind peptides whose lengths are fairly tightly constrained. The largest possible set of human peptides is all peptides in the proteome that overlap by one amino acid. A peptide binding prediction algorithm showed that HLA B-57 and HLA B-27 molecules bind to fewer human peptides thus constructed than most other HLAs. Given this result, Eq. 16 predicts that mature T cells with TCRs restricted by these HLA molecules will be more likely to have more strongly hydrophobic amino acids at their peptide contact sites. These T cells would be more likely to be cross-reactive to point mutants of the peptides that they recognize. Therefore, T cells in individuals with HLA-B57 and HLA-B27 alleles are more likely to be cross-reactive to point mutations of peptides that they recognize, and so would be more likely to inhibit mutational escape of highly mutable pathogens from the T cell response. This factor may contribute to their higher likelihood of being better at controlling infection with highly mutable pathogens, such as HIV and HCV. At the same time, T cells in these individuals would be more likely to exhibit autoimmune responses, since they were subjected to negative selection against a smaller diversity of self peptides compared to individuals with other HLA genes.

## 3.4: Collective effects enable the T cell repertoire to be relatively self-tolerant and pathogen sensitive

### 3.4.1: Quorum sensing by T cells

During the time spent in the thymus, a thymocyte that successfully exits as a mature T cell does not encounter the complete set of self peptides that can be presented by MHC molecules. Since the peptides presented by MHC molecules are short, it is not possible for a particular T cell to distinguish between a particular self peptide that it did not encounter in the thymus and a pathogen-derived peptide. So, the likelihood that such a T cell is activated upon interactions with this particular self peptide or a pathogen-derived peptide is similar. Indeed, autoreactive T cells are known to exist in humans. Why do these auto-reactive T cells cause autoimmune diseases only in a few individuals? The analysis below suggests that such rampant autoimmunity is prevented by collective effects that require multiple T cells to be activated in order to mediate a functional response, and because thymic selection endows the mature T cell repertoire with the statistical properties described in section 3.3.

The probability that a T cell with sequence, $\vec{t}$, is activated ($P\ (\vec{t})^{act}$) by a randomly chosen peptide is:

$$P\ (\vec{t})^{act} = \int_{E_a}^{\infty} p\ (x,\ \vec{t})\ dx \qquad (17)$$

where $E_a$ is the binding free energy threshold required for a mature T cell to be activated, and the other symbols have been previously defined. Earlier we found that the strongest interaction, $E^*\ (\vec{t})$, between self peptides encountered in the thymus by a thymocyte that successfully matured must lie between $E_p$ and $E_n$. Since the magnitude of the difference between $E_p$ and $E_n$ is small, and the activation threshold, $E_a$, for mature T cells is close to $E_n$, as an approximation, let us take $E^*\ (\vec{t})$ to be roughly equal to $E_a$ in Eq. 17. This obtains:

$$P\ (\vec{t})^{act} = \int_{E^*\ (t)}^{\infty\ (\vec{t})} p\ (x,\ \vec{t})\ dx = \bar{P}\ (E^*\ (\vec{t})) \qquad (18)$$

Now, E* obeys Eq. 3 with x* therein replaced by E*. If the thymocyte with TCR, $\vec{t}$, encountered a large diversity of self peptides in the thymus, as is normal, then the value of $E^*\ (\vec{t})$ is going to be large, and so as per previous arguments, the probabilities in Eq. 3 will be drawn from the tails of the corresponding distributions. Therefore, following Eq. 4 we can conclude that

$$-\frac{p^2\left(E^*\ (\vec{t})\right)}{p'\left(E^*\ (\vec{t})\right)} \sim \frac{1}{N} \qquad (19)$$

Note that the left-hand side of Eq. 19 is positive as p' (x) is negative for positive values of x. We can always write Eq. 18 as:

$$P\ (\vec{t})^{act} = \int_{E^*\ (\vec{t})}^{\infty} \exp[\ln p\ (x,\ \vec{t})]\ dx \qquad (20)$$

Because values of $p\ (x, \vec{t})$ are drawn from the tail of the distribution, we can expand the logarithm in Eq. 20 around $E^*\ (\vec{t})$ and carry out the integral to obtain:

$$P\ (\vec{t})^{act} = -\frac{p^2\left(E^*\ (\vec{t})\right)}{p'\left(E^*\ (\vec{t})\right)} \qquad (21)$$

Comparing Eqs. 19 and 21, we conclude that $P\ (\vec{t})^{act} \sim \frac{1}{N}$. This estimate is not quite correct because we know that $E^*(\vec{t})$ is smaller than $E_n$, and hence $E_a$. So, we expect that $P\ (\vec{t})^{act}$ will be smaller than 1/N as it was estimated with the range of the integral in Eq. 18 to be larger than reality. However, although $P\ (\vec{t})^{act} \sim \frac{1}{KN}$ with K > 1, the probability that a mature T cell is activated scales as 1/N and is small for large N.

To obtain Eq. 21, we used properties of E*, but we did not consider the statistical features conferred on the mature T cell repertoire by undergoing thymic selection. Let us now ask whether $P\ (\vec{t})^{act}$ for a randomly picked mature T cell is different depending on whether it interacts with a self or pathogenic peptide. On average, there should be a difference because every mature T cell interacted with some fraction of self peptides in the thymus, and was not negatively selected by these peptides. So, on average, a T cell is less likely to be activated by a randomly picked self peptide compared to a randomly picked pathogen-derived peptide. This argument suggests that, on average, the probability that a T cell is activated by a randomly chosen self peptide can be estimated as follows:

$$p^{act}\,(self) = \, p^{act}\,\left[1 - \frac{N}{M}\right] = \, p^{act}\,x \qquad (22)$$

For $p^{act}\,(pathogen)$, x in Eq. 22 would equal unity. Our estimates for the average value of $p^{act}\left(\vec{t}\,\right)$ for either self or pathogen-derived peptides apply to any T cell, and so as in Eq. 22, we will henceforth drop reference to a particular TCR sequence.

Given the arguments above, let us ask if the T cell repertoire can be completely self-tolerant while also being able to respond to diverse pathogens. Upon infection with a pathogen, only a small number of pathogen-derived peptides that are processed and presented by MHCs are also able to activate T cells. Let us denote this number of so-called immunogenic peptides by I. For the T cell repertoire to be able to respond to diverse pathogens, if the number of T cells is T, IT$p^{act}$ > 1. This ensures that the mean number of activation outcomes is greater than 1 for the immunogenic peptides. Since $p^{act}$ is small (~ 1/N) and I is not a large number, T must be large for this to be true, and indeed the average human has roughly $10^{11}$ T cells. Furthermore, as I is not a large number, we expect T$p^{act}$ > 1. We can also estimate the probability that the T cell repertoire is completely able to avoid autoimmune responses to every self peptide-MHC molecule. This probability is given by $[(1 - \, p^{act}\,x\,)^T]^M$. Since $p^{act}\,x$ is small, we can approximate $[(1 - \, p^{act}\,x\,)^T]^M$as $\exp[-\,MT\,p^{act}\,x]$. Even if every T cell encounters all, but one, self peptide during development (N = M-1), the probability of completely avoiding autoimmunity is given by $e^{-Tp^{act}}$, which is small if the repertoire is to be able to respond effectively to diverse pathogens (T $p^{act}$ > 1). These rough estimates lead us to the conclusion that a T cell repertoire that is effective at combating diverse pathogens cannot avoid having T cells that are activated by some self peptides unless every thymocyte encounters every possible self peptide-MHC molecule during thymic development. The latter condition is not true. So, what prevents the T cell repertoire from being rampantly autoreactive even though it has T cells that are autoreactive to some self peptides?

Activated T cells require various growth factors and cytokines to proliferate and differentiate into functional cells that mediate an immune response. Activated T cells can produce these cytokines. Regulatory T cells are known to suppress T cell responses, and they consume such cytokines, but do not produce them. If a sufficiently large number of T cells are activated, they could secrete enough cytokines and growth factors themselves to overcome the suppressive effects of regulatory T cells. Thus, a T cell response would be predicated upon activation of a threshold number of T cells. That is, collective effects, not activation of individual T cells alone, determine whether an effective T cell response occurs. This hypothesis is analogous to the concept of quorum sensing that underlies the response of bacterial populations to environmental stimuli. In this analogy, a quorum of T cells must be activated to mount an effective immune response.

If quorum sensing underlies an effective T cell response, we can calculate the probabilities associated with two kinds of errors that the T cell repertoire might make. The first type of error is that an effective T cell response is mounted against self peptide-MHC molecules, and the second is that an effective response is not mounted against pathogenic peptides. If the number of T cells that must be activated to establish a quorum is q, the probability that q or more T cells are activated by a self peptide (P ($E_1$)) is given by:

$$P\ (E_1) = \sum_{k=q}^{T} {}_{k}^{T}C\ (p^{act}\ x)^k\ (1 - p^{act}x)^{T-k} \qquad (23)$$

where ${}_{k}^{T}C$ is the number of ways of choosing k T cells from a total of T in the repertoire. Similarly, we can calculate the probability (P ($E_2$)) that less than q T cells are activated by a pathogen derived peptide:

$$P\ (E_2) = \sum_{k=0}^{q-1} {}_{k}^{T}C\ (p^{act})^k\ (1 - p^{act})^{T-k} \qquad (24)$$

Fig. 3.13 shows graphs of Eqs 23 and 24 as a function of q. If the threshold, q, is small, P ($E_1$) is large, and it decays to zero as q increases. P ($E_2$) is large if q is large and it decays to zero for small q. Note that the graph for P ($E_1$) decays to zero faster with q than [1 – P ($E_2$)] because x <1. Thus, as Fig. 3.13 indicates there can exist values of q = q* for which both errors $E_1$ and $E_2$ can be avoided.

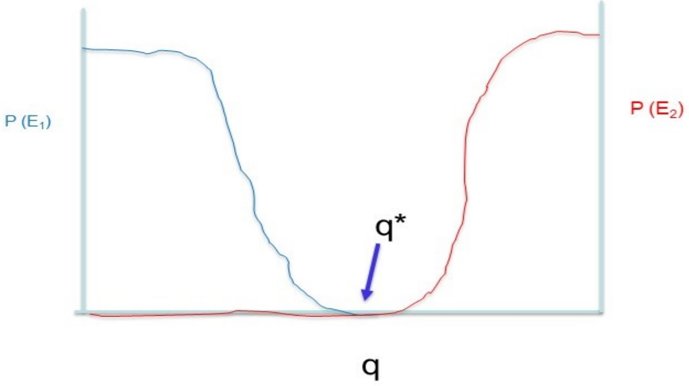

Fig. 3.13: Schematic depiction of graphs of Eqs 23 (blue) and 24 (red). The values of q* can lie in a range where both types of errors can be minimized.

While graphing Eqs. 23 and 24 as in Fig. 3.13a is conceptually instructive, if parameters were known, these equations are not sufficient for calculating the value of q*. The value of q* would need to ensure that there was no response to any of the self peptides displayed on APCs with which T cells may interact in a typical tissue. The value of q* would also have to ensure that a response was mounted to at least one of the pathogen's peptides displayed on APCs upon infection. If the typical number of self peptides encountered by T cells in a tissue is S, the probability of not mounting autoimmune responses, (P (~S), is:

$$P\ (\sim S) = \ (1 - \ P\ (E_1))^S \qquad (25)$$

Upon infection, if the number of pathogenic peptides typically expressed on APCs is I, the probability that a response to a typical pathogen will be mounted, P (I), is:

$$P\ (I) = (1 - \ P\ (E_2)^I) \qquad (26)$$

Eqs. 25 and 26 will result in a larger range of possible values of q* compared to that obtained using Eqs. 23 and 24 (Fig. 3.14). The controlling parameter that determines the range of values of q* is N/M, which is the average fraction of the total number of self peptide-MHC molecules that a typical thymocyte encounters during development. If N/M is sufficiently large, x in Eq. 22 is sufficiently small, which should allow for a range of possible values of q*. Thus, quorum sensing may be a mechanism that allows the T cell repertoire to be largely self-tolerant in spite of some autoreactive T cells, while also maintaining the ability to mount effective responses to diverse pathogens.

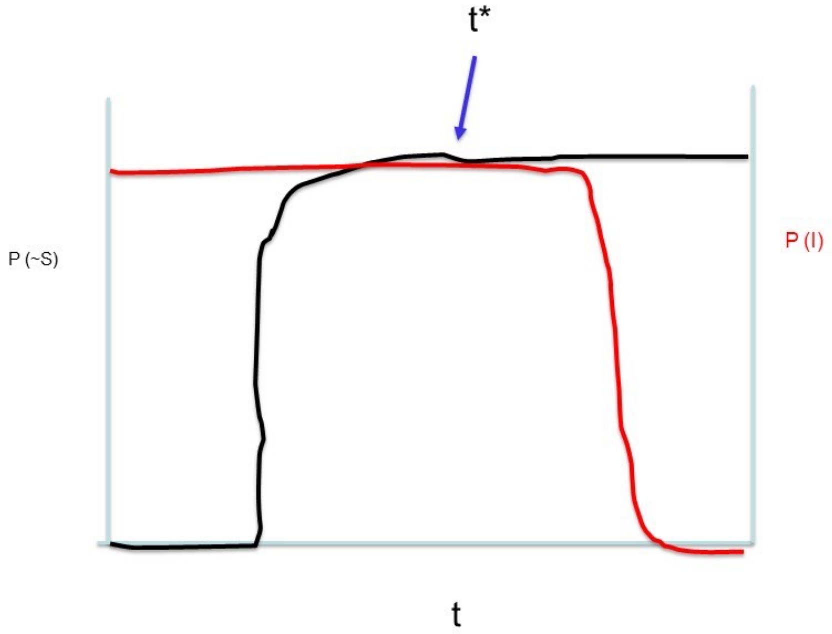

Fig. 3.14: Graphs of Eqs 25 (black) and 26 (red).

### 3.4.2: Experimental evidence for Quorum sensing by T cells

Several lines of experimental evidence suggest that a quorum of T cells needs to be activated for an effective response. It has long been observed that T cells form clusters around APCs. CD8 T cells adhere to APCs such as Dendritic Cells (DCs) aided by interactions between adhesion molecules; for example, a receptor called LFA-1 on T cells binds to a protein called ICAM-1 expressed on DCs. At about the same time as the theoretical proposal of the quorum model discussed in the previous section, two-photon microscopy experiments of mouse lymph nodes led to the observation that CD8 T cells formed clusters via interactions between adhesion molecules expressed on the T cells themselves. The formation of these clusters was found to make T cells respond more sensitively to cytokines that mediate differentiation of these cells to the memory phenotype.

Another study carried out *in vitro* showed that clustering of CD8 T cells around APCs is aided by the expression of ICAM-1 on activated T cells, which then allows them to bind to the LFA-1 proteins on other T cells. CD80 is a molecule expressed on APCs that binds to a molecule called CD28 that is expressed on T cells. It was found that, upon activation, CD80 expression levels increased on T cells themselves. Thus, T cells were able to bind to CD28 on neighboring T cells in the cluster. CD28 binding to CD80 initiates intracellular biochemical signaling that is known to promote the secretion of several cytokines, including one called IL2, which is an important growth factor mediating T cell proliferation. Experiments showed that clustered T cells that interact mutually through CD28 and CD80 expand in an IL2 dependent manner, and that the amount of IL2 produced scaled with the density of clustered activated T cells. These experiments also demonstrated that CD28 and IL2 mediated signaling lowers the expression of apoptotic factors in T cells. Thus, T cells mutually regulated their expansion and survival by forming a quorum of clustered activated T cells.

Another study in mice and *in vitro* also provides similar evidence for quorum sensing during the differentiation of activated CD4 T cells to memory cells. In mice, differentiation into memory T cells depended upon the precursor frequency of antigen specific CD4 T cells in the naïve repertoire. Motivated by this finding, studies were carried out in microwells containing different densities of antigen specific CD4 T cells. The differentiation into memory cells was shown to be density-dependent, and this was mediated by local interactions between cells present at sufficiently high density. Patches of the IL2 receptor were expressed on activated cells and the patches pointed toward other T cells. IL2 secreted by T cells is directed toward the junctions between T cells, and

IL2 concentrations were found to decline rapidly with distance from the secreting T cell. IL2 binding to IL2 receptors leads to intracellular signaling that results in cell differentiation. Thus, expression of the IL2 receptors was correlated with upregulation of downstream signaling markers important for differentiation. This finding is consistent with the quorum model which posits that when a sufficient number of T cells is activated at the same time, the suppressive effects that inhibit autoimmune responses by consuming growth factors like IL2 can be overcome. It was also found that IL2 blockade severely limited T cell numbers, but only at later time points and this may not be due to limitations in cell proliferation. This finding does not necessarily contradict the study noted earlier regarding T cell density and IL2-dependent expansion of CD8 T cells. Rather, taken together, the two studies suggest that complex signaling pathways involving various cytokines (IL2, IL6, etc) and other surface molecules that interact between clustered T cells regulate T cell proliferation and differentiation in a cell density-dependent way (i.e., establishment of a quorum).

As noted earlier, regulatory T cells ($T_{regs}$) consume IL2. Experiments showed that inhibiting the ability of $T_{regs}$ to bind to IL2 and signal intracellularly to mediate their suppressive functions, enabled activated T cells to overcome the suppressive effects of $T_{regs}$. Consistent with the quorum model, the results of these experiments also showed that the balance between $T_{reg}$ density and that of activated T cells determined whether the suppressive effects of $T_{regs}$ or IL2-mediated signaling by activated T cells prevailed.

Evidence for quorum sensing is also suggested by comparing the statistical properties of mouse T cell repertoires at different stages of thymic development. Upregulation of the gene encoding the Nur77 protein is an important marker of T cell stimulation. Mice that expressed a reporter linked to Nur77 upregulation and a marker of cell death were engineered. These reporters allowed determination of which thymocytes had been activated as well as which of them had subsequently died (negative selection), thus enabling TCR sequences at different stages of development to be distinguished. Three such mice were sacrificed and their T cell repertoires were sequenced. The repertoires at different stages of development could only be distinguished based on their statistical properties, not sequences of individual TCRs. For example, it was not possible to classify an individual TCR sequence to the mature or pre-selection T cell repertoire. However, T cell repertoires at different stages of development could be distinguished by the statistical properties of the combinations of their V and J genes and CDR3 amino acids. This finding is similar to observations described earlier that, in a statistical sense, the hydrophobicity of the TCR peptide contact residues could distinguish between the post and pre-selection T cell repertoires. Sequencing the entire CDR3 region, which contains many more amino acids than the peptide contact residues, also shows these statistical differences, but at a lower resolution compared to when peptide contact residues are known (see section 3.3.4). The fact that the post and pre-selection repertoires can only be distinguished based on their statistical properties, and not sequences of individual T cells, argues that the statistical features imprinted by thymic selection enables the mature T cell repertoire to discriminate between self and non-self via collective effects (embodied in the quorum model).

Much work remains to be done to comprehensively understand how collective effects mediate T cell-mediated immune responses and how its aberrant regulation leads to autoimmunity. For example, it will be important to study the dynamics of effector T cells and regulatory T cells that suppress autoimmunity to understand how an immune response against foreign antigens develops with time, but not against self-antigens. Such studies will also allow determination of the conditions that lead to "instabilities" in these dynamics resulting in the onset of autoimmunity. It will also be important to explicitly study how innate immunity regulates these processes.

**Suggested Readings**

1] H. Spits, "Development of αβ T cells in the human thymus", Nature Reviews Immunology, 2, 760 (2002).

2] K.M. Murphy, C. Weaver, L.J. Berg, "*Janeway's Immunobiology*", W.W. Norton & Company Inc. (2022).

3] A.K. Abbas, A.H. Lichtman, S. Pillai, "*Cellular and Molecular Immunology*", Elsevier (2021).

4] M. A. Daniels et al., "Thymic selection threshold defined by compartmentalization of Ras/MAPK signalling", *Nature*, **444**, 724 (2006).

5] E.S. Huseby et al., "How the T cell repertoire becomes peptide and MHC-specific", *Cell*, **122**, 247 (2005).

6] E.S. huseby et al., "Interface-disrupting amino acids establish specificity between T cell receptors and complexes of major histocompatibility complex and peptide", *Nature Immunology*, **7**, 1191 (2006).

7] A. Kosmrlj et al., "Thymic Selection of T-Cell Receptors as an Extreme Value Problem", *Physical Review Letters*, **103**, 068103 (2009).

8] A. Kosmrlj et al., “How the thymus designs antigen-specific and self-tolerant T cell receptor sequences”, *Proceedings of the National Academy of Sciences*, **105**, 16671 (2008).

9] B.D. Stadinski et al., “Hydrophobic CDR3 residues promote the development of self-reactive T cells”, *Nature Immunology*, **17**, 946 (2016).

10] S. Miyazawa, R. Jernigan, “Estimation of Effective Inter-residue Contact Energies from Protein Crystal Structures: Quasi-Chemical Approximation”, *Macromolecules*, **18**, 534 (1985).

11] A. Kosmrlj et al., “Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection”, *Nature*, **465**, 360 (2010).

12] L. Klein, et al. “Positive and negative selection of the T cell repertoire: What thymocytes see (and don’t see)”, *Nat. Rev. Immunol.*, **14**, 377 (2014).

13] W. Yu et al., “Clonal deletion prunes but does not eliminate self-specific CD8+ T lymphocytes”, *Immunity*, **42**, 929 (2015).

14] T. Butler et al., “Quorum Sensing Allows T Cells to Discriminate between Self and Nonself” *Proceedings of the National Academy of Sciences*, **110**, 11833 (2013).

15] G. Audrey et al., “Secondary T Cell–T Cell Synaptic Interactions Drive the Differentiation of Protective CD8+ T Cells” *Nature Immunology*, **4**, 356 (2013).

16] S. Zenke et al., “Quorum Regulation via Nested Antagonistic Feedback Circuits Mediated by the Receptors CD28 and CTLA-4 Confers Robustness to T Cell Population Dynamics” *Immunity* **52**, 313 (2020).

17] M. Polonsky et al., “Induction of CD4 T Cell Memory by Local Cellular Collectivity.” *Science*, **360**, eaaj1853 (2018).

18] H. Wong et al., “A Local Regulatory T Cell Feedback Circuit Maintains Immune Homeostasis by Pruning Self-Activated T Cells.” *Cell*, **184**, 3981(2021).

19] F. Camaglia et al., “Quantifying changes in the T cell receptor repertoire during thymic development”, *eLife*, **12**, e81622 (2023).