bioRxiv preprint doi: https://doi.org/10.1101/480301; this version posted November 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A theory for how the antigen presentation profile influences the timing of T-cell detection

Alberto Carignano[1,2] **and Neil Dalchau**[1]

[1]Biological Computation Group, Microsoft Research, Cambridge, CB1 2FB, UK
[2]Department of Electrical Engineering, University of Washington, WA

T-cells are activated when their receptor molecules recognize complexes of MHC proteins bound to peptides on the surface of neighbouring cells. Each T-cell expresses one variant of many possible receptor molecules, which are generated through a partially random process that culminates in approximately 10^7 possible T-cell receptors. As the peptide sequence bound to an MHC molecule is also highly variable, the optimal strategy of an antigen-presenting cell for displaying peptide-MHC complexes is not obvious. A natural compromise arises between aggressive peptide filtering, displaying a few peptides with high stability MHC binding in high abundance and regularity, and promiscuous peptide binding, which can result in more diverse peptides being presented, but in lower abundance. To study this compromise, we have combined a model of MHC class I peptide filtering with a simple probabilistic description of the interactions between antigen presenting cells (APCs) and cytotoxic T-cells (CTLs). By asking how long it takes, on average, for an APC to encounter a circulating CTL that recognises one of the peptides being presented by its MHC molecules, we found that there often exists an optimal degree of peptide filtering, which minimises this expected time until first encounter. The optimal degree of filtering is often in the range of values that the chaperone molecule tapasin confers on peptide selection, but varies between MHC class I molecules that have different peptide binding properties. Our model-based analysis therefore helps to understand how variations in the antigen presentation profile might be exploited for vaccine design or immunotherapies.

antigen presentation | cytotoxic T-cell | probabilistic modelling | peptide filtering | computational immunology

Correspondence: ***ndalchau@microsoft.com***

DRAFT

Introduction

An adaptive immune system is somatically generated during development based on the ability of circulating thymocytes to recognise the protein contents of infected cells. All cells in vertebrates express class I molecules of the Major histocompatibility complex (MHC), comprising the mostly sequence-invariant beta-2-microglobulin (β_2m) and a highly variable heavy chain (HC). In the endoplasmic reticulum (ER), this MHC class I complex binds to protein fragments arising from proteasomal degradation, which are transported from the cytoplasm into the ER by Transporter associated with antigen processing (TAP) molecules. While the specificity of proteasomes and TAP transporters are moderately sequence-dependent, the sheer diversity of the protein sequences of self and viral proteins naturally translates into a large diversity of peptide sequences being transported into the ER. As such, for MHC class I molecules to effectively communicate their internal contents for immune system surveillance, their peptide-binding specificity must be at least moderately promiscuous. It has now recently been shown that the MHC alleles vary in their promiscuity, and that this has a measurable impact on immune response (1).

Cytotoxic T lymphocytes (CTL) detect and destroy infections by recognising antigenic peptides presented on MHC class I molecules at the surface of antigen presenting cells (APC). Each CTL is able to recognise pMHC complexes via their T-cell receptor (TCR), though possess only one TCR type, and so the specificity of the CTL is encoded by its TCR sequence. The TCR is generated in a process of random arrangement of gene segments, known as V(D)J recombination, which creates an enormous diversity of TCRs across T-cell populations. Ideally, there would be sufficiently many T-cells to cover all non-self antigenic determinants, however this would require 10^7 TCRs to cover 10^{13} antigenic determinants, which naturally requires TCRs to be highly crossreactive, recognising multiple peptides (2).

The formation of the immunological synapse enables TCR to be engaged by peptide-MHC complexes, which can enable CTL to be activated if there are sufficient peptides that the TCR recognise. APCs can present multiple copies of the same peptide sequence, and we term the overall composition of peptides presented as the antigen presentation profile (APP). Given the high degree of crossreactivity between TCRs and peptide-MHC complexes, a compromise might arise in terms of how an APC might optimally select peptides for presentation. For instance, an APP with a small number of distinct peptide sequences in high abundance would maximise the chance that corresponding TCRs will be activated during an interaction. However, an APP with many distinct peptide sequences in lower abundance would increase the number of potential TCR matches, but at the cost of lowering the probability of any one of them leading to activation. Highly relevant to this question is the chaperone molecule tapasin, which skews the APP towards containing more high affinity peptides overall (3–6). Loss of tapasin leads to poor peptide selection, resulting in overall lower cell surface abundance of pMHC. Consequently, tapasin-deficient organisms have altered responses to viral clearance (7–9). It is therefore not surprising that some viruses and tumours down-regulate tapasin to evade immune detection (10). Also of relevance to presentation strategy is the intrinsic binding properties of the MHC haplotype, which have been shown to vary considerably (1).

bioRxiv preprint doi: https://doi.org/10.1101/480301; this version posted November 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Theoretical work has already been instrumental in understanding how T-cell responses depend on the biochemical parameters of specific TCR-pMHC interactions. It is now well established that an intermediate rate of TCR-pMHC unbinding (k_{off}) can provide a sufficient duration for downstream signalling without preventing additional receptor interactions (11, 12). Computational models of broader scope have helped to interpret how CTL killing depends on cell density within lymph nodes (13, 14). Stochastic models have helped to formulate a probabilistic treatment of T-cell activation in terms of the APP: for instance, (15) modelled the idea that infections are detected when the sum of the stimuli received by a T-cell from the interactions between its TCR and the pMHCs on the APC, exceeds a given threshold. The ability of T-cells of distinguishing between self and non-self is a consequence of a higher copy number of foreign peptides in comparison with self on the APC surface (*probabilistic recognition*). In particular, increasing the copy number of foreign peptides on the APC does not increase the average number of stimuli per T-cell, but the variance, suggesting that exceeding the activation threshold is possible only for high value of foreign peptide copy number (16). However, these theories have not yet been applied to specific MHC haplotypes, and do not consider realistic APC surface presentations. Therefore, here we sought to apply probabilistic models to the analysis of APPs that arise in different MHC haplotypes, and determine how they affect the timing of the CTL response. In order to make this relevant to known MHC molecules, we have made use of the extensive literature on peptide-MHC binding prediction, but also a kinetic model of peptide-MHC loading and presentation (5) to generate plausible antigen presentation profiles.

Results

Predicting the cell surface abundance of multiple peptides. Measuring the APP is made challenging by the sheer diversity of peptides, their low copy numbers, a lack of sufficiently many peptide-specific probes, cell-cell variability, and possibly many other factors. Therefore, we base our analysis on the T-cell activation capability of different APPs on computational models, but focus on well characterised MHC class I molecules. As the rate of peptide-MHC dissociation is the major determining factor of cell surface abundance, we sought to create a model that simulates combinations of peptides that differ only in their MHC dissociation rates. This assumes that i) distinct peptide sequences are in equal intracellular abundance, and ii) the association rate of peptide-MHC is homogeneous. To assign dissociation rates to the peptides, we made use of the extensive literature on peptide binding predictors (see Methods).

MHC class I alleles differ in their binding repertoires. The BIMAS algorithm (17) was used to approximate the distribution of dissociation rates of five different HLA alleles (HLA-A*02:01, HLA-B*08:01, HLA-B*27:05, HLA-B*44:03 and HLA-B*58:01). For each allele, we applied the algorithm to a large number of peptides of either 9 or 10 amino acids in length (see Methods for further details). We found that the HLA alleles differed considerably in their distributions (Figure 1). HLA-B*27:05 was found to be highly promiscuous, binding over 4% of peptides with at least moderate stability (Table 1). All other alleles had fewer than 1% of peptides in this category, with HLA-B*08:01 predicted to bind as few as 0.04% with moderate stability.

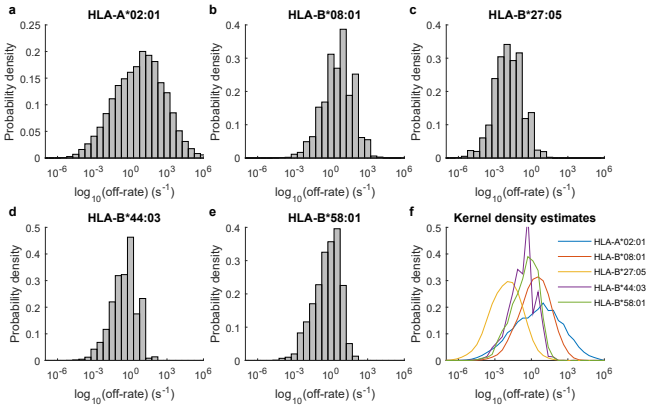

Fig. 1. Allelic variation in the distribution of peptide-off-rates from HLA molecules. (**a–e**) The BIMAS algorithm was used to predict the rate of dissociation for complexes of human protein-derived peptides and HLA molecules. Peptides of length 9 and 10 were included in the analysis, and HLA alleles considered are as indicated on each panel. (**f**) For each distribution, a kernel density estimate was obtained using the `kde` function on MatlabCentral (18), to facilitate comparison of each allele.

Table 1. Statistics of peptide off-rate distributions. Reported are the inferred means and standard deviations of the base-10 logarithm of peptide off-rate distributions for HLA molecules. The final column shows the percentage of off-rates that were more stable than 10^{-4} s^{-1}.

Allele	$\log_{10}$(Mean)	Std.	$< 10^{-4}$ s^{-1}
HLA-A*02:01	0.73	2.00	0.86%
HLA-B*08:01	0.51	1.22	0.04%
HLA-B*27:05	−1.88	1.19	4.35%
HLA-B*44:03	−0.55	1.05	0.25%
HLA-B*58:01	−0.35	1.07	0.16%

Simulation of the cell surface peptide distribution. Having established considerable variation in the distribution of peptide dissociation rates from different HLA alleles, we sought to characterise how this influences the shape of the distribution of peptides presented at the cell surface, the APP. To provide quantitative predictions of the cell surface abundance of each peptide, the peptide filtering model (5) was simulated for sets of peptides sampled from the distributions in Fig. 1. This allowed us to take into consideration both the peptide dissociation rate from a given MHC class I molecule and also the competition from other peptides for MHC class I loading. Furthermore, it was possible to simulate and therefore predict the consequences of perturbing the MHC class I system, such as in the case of tapasin-deficient cells.

Simulations were carried out for HLA-B*44:02 and HLA-B*27:05 molecules, as these were previously characterised as tapasin-dependent and tapasin-independent alleles in experimental studies (3). In (5), tapasin-dependency was predicted to be attributed to enhanced peptide-MHC association, and

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/480301; this version posted November 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in (6) the hypothesis has been refined to rely on a transition from a peptide-receptive to a peptide-nonreceptive conformation (i.e. closing). Since we had used the model in (5), simulations of each allele used different rates of peptide-MHC association (HLA-B*44:02 – 3.1773×10^{-11} molecule^{-1} s^{-1}; HLA-B*27:05 – 1.9446×10^{-9} molecule^{-1} s^{-1}). In each case, the off-rates of 10,000 peptides were sampled from log-normal distributions as parameterised in Table 1 (Fig. 2a,d). For HLA-B*44:02, we used off-rates for the closely related HLA-B*44:03 haplotype. Simulations of tapasin-competent and tapasin-deficient cells were carried out for both HLA molecules.

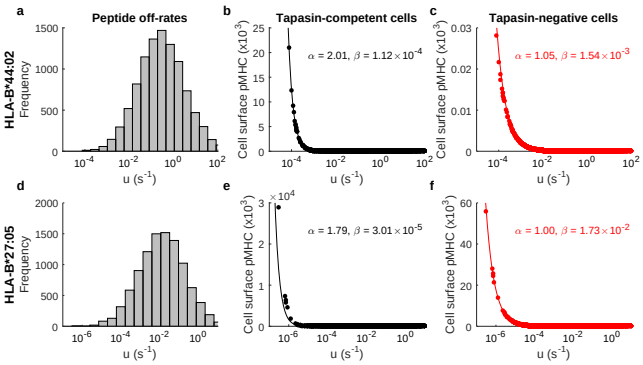

Fig. 2. Approximation of the MHC class I peptide filtering model with a power-law function. The peptide filtering model (5) was simulated for 10,000 peptides sampled from log-normal distributions corresponding to (**a-c**) HLA-B*44:02 and (d-f) HLA-B*27:05. (**a**,**d**) Histogram shows the distribution of peptides sampled for each MHC class I allele. (**b**,**e**) Simulations of tapasin-competent cells, for peptides as shown in panels a and d respectively. (**c**,**f**) Simulations of tapasin-negative cells. In panels b, c, e and f, the parameters of the power-law function $\beta u_i^{-\alpha}$ that best fit the simulation of the peptide filtering model are shown.

The effect of *peptide filtering* is clearly visible in Figure 2, as ER distributions of peptides with mode values between 10^{-2} and 10^{0} s^{-1} get mapped to cell surface distributions which are highly skewed towards peptides with low off-rates. The simulations revealed strongly tapasin-dependent presentation for HLA-B*44:02, with two peptides presented at approximately 10^4 copies each in tapasin-competent cells (Fig. 2b), and ≈ 60 copies each in tapasin-deficient cells (Fig. 2c), a fold change of $\approx 2,000$. For HLA-B*27:05, a strongly dominant peptide was presented at 3.7×10^7 copies in tapasin-competent cells (Fig. 2e) and at $\approx 6\times10^4$ copies in tapasin-deficient cells (Fig. 2f), a fold change of ≈ 500, indicating tapasin-dependency that is less than that of HLA-B*44:02.

It is important to note that the simulated numbers of peptide-MHC complexes at the cell surface are considerably larger than what has been observed experimentally. For instance, it is reported in (19) that the cell surface abundances of specific peptides are in the range 0–100, whereas our simulations indicate abundances as high as $O(10^7)$ (Figure 2). This discrepancy most likely arises because the original model was not calibrated against absolute numbers of molecules, but instead against fluorescence measurements that were assumed proportional to cell surface abundance (5). Accordingly, while the dynamics of the model reproduced experimental observations well, the absolute scale of the model remains uncertain, and simulations should be considered as accurate only up to a multiplicative factor. While the output of the model could be rescaled for its usage in this study, this would not influence the qualitative nature of the results presented. Therefore, for consistency, we leave the peptide filtering model in its original scale.

Characterising the cell surface repertoire using a power-law function. While the peptide filtering model (5) provides a biochemically-based simulation of peptide selection, it does not provide a succinct description of the APP as a function of pMHC off-rate. However, we found that the equilibrium cell surface abundance of pMHC could be well approximated by the power-law function

$$[MP_i]_{\text{cs}} = \beta u_i^{-\alpha} \tag{1}$$

where u_i is the dissociation rate of peptide i from the MHC class I molecule (Fig. 2). The major advantage of using this simplified model was that it enabled us to determine the effect of the essential features of the APP on T-cell activation. The parameter α quantifies the degree of peptide filtering, which describes the extent to which peptides are selected according to their rate of dissociation from MHC class I. A value of 0 would indicate no dependence on u_i, with higher values indicating stronger selection for more stable peptides. Previous theoretical analysis (5) suggested that peptide filtering is at best proportional to u_i^{-3} in tapasin-competent cells ($\alpha = 3$), and proportional to u_i^{-2} ($\alpha = 2$) in tapasin-negative cells. The parameter β is a scale factor, and corresponds to the average cell surface abundance of a peptide with off-rate 1 s^{-1}. It incorporates the supply of peptide into the ER but also incorporates the overall competition between peptides. In this study, we ignore the effect of differential peptide supply as mediated by protein abundance, proteasomal degradation and TAP binding, and so a peptide non-specific value of β can be applied to all peptides.

To quantify how the APP differs between alleles and between tapasin-competent vs. tapasin-deficient cells, we inferred values of α and β in each of the four scenarios of Figure 2 (see Methods for details). The degree of peptide filtering was 2.01 for tapasin-competent cells expressing HLA-B*44:02, though only 1.79 for tapasin-competent cells expressing HLA-B*27:05, emphasising the tapasin-dependency of HLA-B*44:02. In tapasin-negative cells, the degree of filtering was low for both alleles (HLA-B*44:02 – 1.05; HLA-B*27:05 – 1.00). In all cases, the power-law function provided a good approximation of the peptide filtering model for all peptides that had abundance ≥ 1. Crucially, the change in the shape of the distributions was well described by the different values of α. In tapasin-competent cells, the large α was consistent with high presentation of the most highly stable peptides, at the cost of filtering out lower stability peptides (Fig. 2b,e). In tapasin-negative cells, low α was consistent with overall lower presentation, but also a greater number of peptides with moderate presentation.

Since the number of distinct peptide sequences that might be present in the ER at any one time is very large, and beyond what is feasible to simulate with this dynamical model, we

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/480301; this version posted November 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

performed a number of simulations in which the number of distinct peptides sampled from the off-rate distributions was varied. In doing so, we found that if more than approximately 10,000 peptides were used, the expected deviation from the mean, representative of a larger sample, was below 1% for both α and β, and for both tapasin-competent and tapasin-negative cells (Fig. S1).

Probabilistic model of APC-CTL interactions. To interpret how the distribution of peptides presented by MHC molecules on antigen presenting cells influences T-cell recognition, we developed a probabilistic model. The model describes interactions between CTLs and a single APC to determine how differences in the distribution of cell surface peptide-MHC might affect the timing of T-cell activation. For clarity, the parameters of the model are summarised in Table 2, but we introduce many of these quantities as we describe the model. Dendritic cells are thought to interact with as many as 5,000 T cells per hour *in vivo*, which is approximately 1 per second (19). Therefore, the time unit of this model can be conveniently thought of as 1 second. The same study (19) measured the average contact duration to be 3.4 minutes, suggesting that on average there are approximately 200 T-cells in contact with an APC at any one time. We may therefore formulate a simple model for the interactions between a single APC and multiple CTLs as

$$\text{APC} + \text{CTL}_j \underset{1-a(x)}{\overset{1/n_T}{\rightleftharpoons}} \text{APC–CTL}_j \xrightarrow{a(x)} \text{APC–CTL}^* \quad \textbf{(2)}$$

where n_T is the assumed number of CTLs and APC–CTL* represents an activated T-cell bound to the APC (Fig. 3). The activation rate $a(x)$ is a function that depends on the current surface distribution of the APC. We use the variable x to denote a random sample of the cell surface distribution, corresponding to the peptide-MHC molecules that are scanned in an immunological synpase. The activated state is considered to be *absorbing* in this model, which means that it is the final state considered. Moreover, the probability of T-cell activation is always positive here, which assumes that there is always a peptide on the surface for which at least one T-cell can recognise. Related to this, we note that negative selection, the deletion of T-cells that react with self peptide, is not explicitly considered here.

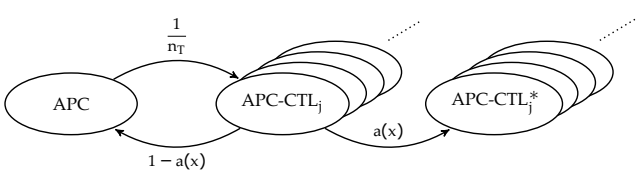

Fig. 3. Markov process model of APC-CTL interactions. The binding of an APC to a CTL is considered to take place once per time unit. Therefore, we normalise against the assumed number of CTL, n_T. Upon formulation of the immunological synpapse, the CTL is either activated with probability $a(x)$, which is a function of the APP x, or dissociates with probability $1-a(x)$. Multiple APC-CTL complexes are shown to illustrate that there are multiple CTLs that might bind, each occurring with frequency $\frac{1}{n_T}$.

The model assumes that, on average, an APC encounters T-cells at a frequency of 1 per time unit. As this is the only time-dependent process, the model can be arbitrarily scaled, and as such we leave the model in terms of this time unit. To further simplify this description so as to enable useful calculations to be carried out, we remove the probabilistic continuous-time nature of T-cell encounters and consider the discrete-time version of the model

$$P_{k+1} = AP_k \quad \textbf{(3)}$$

where the state space P_k records the number of copies of each state at time $t = k$ ($k = 1, 2, \ldots$) and the transition matrix A describes the transition propensities between each state. We write these as

$$P_k = \begin{pmatrix} \text{APC} \\ \text{APC–CTL}_1 \\ \vdots \\ \text{APC–CTL}_{n_T} \\ \text{APC–CTL}^* \end{pmatrix}, \quad A = \begin{pmatrix} 0 & \frac{1}{n_T} & \ldots & \frac{1}{n_T} & 0 \\ 1-a & 0 & \ldots & 0 & a \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-a & 0 & \ldots & 0 & a \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix} \quad \textbf{(4)}$$

Calculating the expected time until first detection. Using our model of APC-CTL encounters, it is possible to derive the expected number of T-cell encounters that are required until a T-cell is activated, which can be used as a proxy for the efficiency of an APC to stimulate a CTL response. We use the fact that the expected number of steps required of a discrete-time Markov chain to reach an absorbing state is given by the sum of the first row of

$$\sum_{i=0}^{\infty} Q^i = (I-Q)^{-1}$$

where Q is the transient part of the transition matrix, which in this case is given by the first $T_n + 1$ rows of A. As such,

$$(I-Q)^{-1} = \frac{1}{a(x)}(I+R)$$

Table 2. Definitions of modelling quantities used throughout.

Variable	Description	Value(s)
u_i	Dissociation rate of a peptide-MHC complex	Variable ($> 10^{-6}$ s^{-1})
β	Relative supply of each peptide	Inferred (HLA/Tpn-dependent)
α	Skewness of peptide selection	Inferred (HLA/Tpn-dependent)
$[MP_i]_{\text{cs}}$	Number of peptide-MHC complexes with peptide i	Eq. 1
P_T	Number of peptide-MHC complexes on the cell surface	$\sum_i [\text{MP}_i]_{\text{cs}}$
T_T	Number of TCRs on the surface of a CTL	1000
K_D	Half saturation constant	10
N_P	Number of distinct peptide sequences that can be presented on MHC molecules	
n_{cog}	Number of distinct peptide sequences that are cognate for a TCR	
n_{scan}	Number of pMHC-TCR bonds	Eq. 5
$n_{\text{diff}}(x)$	Number of distinct peptide sequences in a sample (x) of pMHC molecules presented on the APC surface	($\leq n_{\text{scan}}$)

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/480301; this version posted November 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where

$$R = \begin{pmatrix} 0 & \frac{1}{n_T} & \frac{1}{n_T} & \cdots & \frac{1}{n_T} \\ 1-a & \frac{(a-1)(n_T-1)}{n_T} & \frac{1-a}{n_T} & \cdots & \frac{1-a}{n_T} \\ 1-a & \frac{1-a}{n_T} & \frac{(a-1)(n_T-1)}{n_T} & \cdots & \frac{1-a}{n_T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1-a & \frac{1-a}{n_T} & \frac{1-a}{n_T} & \cdots & \frac{(a-1)(n_T-1)}{n_T} \end{pmatrix}$$

The sum of the top row of $(I-Q)^{-1}$ is therefore equal to $2/a(x)$, which is independent of n_T, the number of CTLs. Therefore, in this model, the probability of T-cell activation is determined purely by the probability of activation $a(x)$.

Calculating the probability of T-cell activation. In the model presented here, every time there is a contact between an APC and a CTL, the cell surface distribution is sampled so as to mimic the formation of the immunological synapse and binding of TCRs to surface pMHC complexes. The quantity $a(x)$ is then used to describe the probability that the CTL is activated from a sample x of pMHC complexes at the cell surface. The size of the sample is taken to be a function of the total number of pMHC molecules available for TCR binding ($P_T = \sum_i [MP_i]_{\text{cs}}$) and an assumed number of TCRs (T_T), as described in (12), as

$$n_{\text{scan}} = \frac{1}{2}\left(K_D + P_T + T_T - \sqrt{(P_T + T_T + K_D)^2 - 4P_T T_T}\right) \quad \textbf{(5)}$$

where K_D is the bond dissociation constant, which is the ratio of the off- and on- rates of pMHC-TCR binding. By using this model, we can incorporate the effect of TCR saturation, which will occur if there are very high cell surface levels of peptide-MHC. This imposes a limit on the number of peptide-MHC complexes that are scanned during an APC-CTL contact. As such, simply presenting more peptide-MHC complexes will not lead to more TCRs being engaged.

We use Eq. 5 to define the number of peptides that are sampled by a T-cell that comes into contact with the APC. This means that we have reinterpreted the K_D parameter to be an *averaged* dissociation constant over the different peptide sequences that are bound to the MHC molecule. As such, K_D is the value that equates

$$[\text{TCR}].\sum_i [MP_i]_{\text{cs}} = K_D.\sum_j [\text{TCR-}MP_i] \quad \textbf{(6)}$$

While the value of K_D is likely to be inaccurate for any specific pMHC-TCR interaction, the use of the equation represents the simplest way to describe the potential limitation of TCRs in the immunological synapse, which will become important as total APC presentation of pMHC increases.

To calculate the activation probability from a sample of n_{scan} pMHC complexes, we make some important simplifying assumptions. First, we assume that each T-cell may be activated by a fixed number of *cognate* peptides n_{cog}. Second, in a sample x of the APP, $n_{\text{diff}}(x)$ different peptides are seen by the T-cell. In this way, the activation probability $a(x)$ follows a hypergeometric distribution over N_P, the total number of peptides that might be presented. Since encountering one cognate peptide is thought to be sufficient to activate a T-cell (20), albeit via multiple TCR engagements (21), we formulate an approximate activation probability on the event that at least one cognate peptide is encountered. As such the probability of activation is 1− the probability of not encountering a cognate peptide, i.e.

$$a(x) = 1 - Pr(X = 0) \quad \textbf{(7)}$$

where X is a hypergeometric distribution with parameters $(N_P, n_{\text{cog}}, n_{\text{diff}}(x))$.

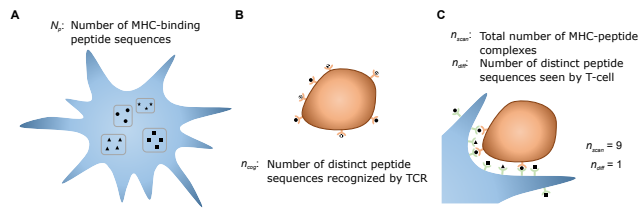

Fig. 4. Quantifying pMHC-TCR interactions. Shown are a series of diagrams that illustrate how model parameters are defined. An antigen-presenting cell is shown in blue, and a T-cell in orange.

Based on this definition for $a(x)$, it remains for us to determine n_{diff}, the number of different pMHC seen by the TCRs. Since several copies of each peptide are present on the surface of the APC, n_{diff} will be less than the total number of peptides scanned by the T-cell (n_{scan}). For example, peptides with high affinity for one of the expressed MHC molecules will likely have higher copy numbers than lower affinity ligands. Moreover, the number of different peptides is huge ($N_P > 10^{15}$) (22). Hence, an exhaustive search of all the possible outcomes n_{diff} given the number of scanned peptides by the T-cell is computationally intractable and no analytic formula has been derived to our knowledge. However, one can compute its expected value as a function of the surface multiplicity of each peptide n_i (defined by Eq. 1) and n_{scan}. To obtain a probabilistic expression for n_{diff}, we introduce the index variable I defined as:

$$I(i) = \begin{cases} 1 & \text{if peptide } i \text{ is scanned by the T-cell} \\ 0 & \text{otherwise.} \end{cases} \quad \textbf{(8)}$$

Using the index variable I, we can find an expression for the expected number of n_{diff} as:

$$E[n_{\text{diff}}] = E[\sum_{i=1}^{N_P} I(i)] = \sum_{i=1}^{N_P} E[I(i)] \quad \textbf{(9)}$$

The expected value of $I(i)$ is by definition equal to 1− the probability that peptide i was not scanned. This is equal to:

$$P(\text{peptide } i \text{ was not one of the } n_{\text{scan}} \text{ scanned}) = \prod_{j=1}^{n_i}\left(\frac{P_T - j - n_{\text{scan}}}{P_T - j}\right) = \prod_{j=1}^{n_i}\left(1 - \frac{n_{\text{scan}}}{P_T - j}\right) \quad \textbf{(10)}$$

Hence, combining Eq. 9 and Eq. 10, we obtain the expected value of n_{diff} as:

$$E[n_{\text{diff}}] = \sum_{i=1}^{N_P}\left(1 - \prod_{j=1}^{n_i}\left(1 - \frac{n_{\text{scan}}}{P_T - j}\right)\right) \quad \textbf{(11)}$$

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/480301; this version posted November 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The probability of activation $a(x)$ can also be obtained by using $E[n_{\text{diff}}]$ instead of n_{diff} in Eq. 7.

To make the calculation of the expected time until first detection more computationally efficient, we grouped peptides together according to their cell surface multiplicity. By defining k_i to be the number of peptides with cell surface multiplicity i, and as such $\sum_i i.k_i = n_{\text{scan}}$, we can rewrite Eq. 11 as

$$E[n_{\text{diff}}] = \sum_{i=1}^{n_{\text{max}}} k_i \left(1 - \prod_{j=1}^{i}\left(1 - \frac{n_{\text{scan}}}{P_T - j}\right)\right) \qquad \textbf{(12)}$$

where n_{max} is the highest multiplicity observed. To summarise the procedure for calculating the expected time until first detection, we present the steps in Algorithm 1.

Algorithm 1 Calculating the expected time until first detection

1: Sample peptides with $\log_{10}(u_i) \sim \mathcal{N}(\mu, \sigma^2)$ for $i = 1, \ldots, N_P$.
2: Quantify their cell surface abundance bound to MHC class I, using the peptide filtering model (5), or the power-law function (Eq. 1).
3: Round each abundance value to the nearest integer, producing a set $\{P_i \in \mathbb{Z}^* : i = 1, \ldots, N_P\}$.
4: Determine the number of pMHC-TCR contacts, n_{scan}, using Eq. 5.
5: Determine the expected number of different peptides presented using Eq. 11.
6: Evaluate the activation probability $a(x)$ using Eq. 7 with $n_{\text{diff}} = E(n_{\text{diff}})$.
7: Evaluate the expected time as $\frac{2}{a(x)}$.

The distribution of peptides on an APC modulates the expected time until first productive T-cell encounter.. To determine how the timing of T-cell activation depends on the distribution of peptide off-rates for an MHC class I molecule, we used the power-law description of cell surface peptide-MHC abundance within the probabilistic model of T-cell detection (Fig. 3). Samples of peptides were drawn from log-normal distributions with varying mean (μ) and standard deviation (σ) parameter values (Fig. 5a). For each pair of values, the cell surface abundance was then calculated for a varying degree of filtering (α) and peptide supply (β), and then used to calculate the expected time until first detection using the Algorithm 1.

We found that the time until first detection routinely showed a compromise in the degree of filtering, with fastest detection reached with an intermediate value of α (Fig. 5b). Our interpretation of these results is that higher filtering raises overall cell surface abundance of high affinity peptides, but at the cost of diversity, with medium and low affinity peptides being filtered out. To understand this more clearly, we consider the extremes. Perfect filtering ($\alpha \to \infty$) would produce a very large abundance of only a single peptide, which would lead to only one class of T-cells being activated (assuming there exists a circulating CTL that reacts with this peptide). Whereas an absence of filtering ($\alpha = 0$) would lead to an equivalent number of each peptide being presented, but each at very low abundance. As total abundance drops with lower α, we find that only a small number of total peptides are presented at all when $\alpha = 0$, leading also to a low number of activatable T-cell classes. By increasing the filtering to an intermediate level, total surface abundance increases, but maintains some diversity, leading to a larger number of activatable T-cell classes.

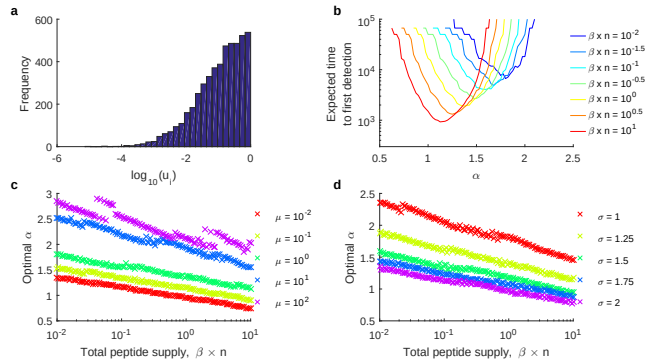

Fig. 5. The distribution of peptides on an APC modulates the expected time until first productive T-cell encounter. The expected time until first detection was calculated for peptides sampled from a range of log-normal distributions. (**a**) Shown is an example sample of 10,000 peptide off-rates drawn from a log-normal distribution with mean $\mu = 1$ and standard deviation $\sigma = 1.5$. Note the distribution is truncated to peptides with their off-rates no larger than the mean, for efficiency of the expected time calculation. (**b**) Expected time to first detection as a function of α and β, specifically for the case in (a). (**c**) Shown is the value of α that minimises the time until first detection, as a function of peptide supply (quantified as $\beta \times n$, where n is the number of peptides sampled), but also for varying μ (σ fixed at 1.25). (**d**) As in (c), but this time varying σ (μ fixed at 1).

As the supply of peptide was decreased, the optimal degree of filtering increased, but the expected time until first detection lengthened (Fig. 5b). Our interpretation is that in low supply scenarios, additional filtering can compensate for a lower total number of cell surface complexes, but will naturally result in fewer distinct peptides, and therefore fewer numbers of activatable T-cell classes.

We found that the optimal degree of filtering also varied with the underlying distribution of peptide off-rates. As the mean of the off-rates became lower (peptides bind MHC more stably), the optimal α decreased (Fig. 5c). This is because lower off-rates are presented in higher abundance, and so less filtering is required to maintain a sufficiently high total abundance to nearly saturate the TCRs. Similarly, increasing the standard deviation of the off-rates decreased the optimal α (Fig. 5d), as this increases the number of high stability peptides in the sample.

Tapasin can accelerate or decelerate T-cell detection, depending on HLA allele. To determine how enhanced peptide filtering, conferred by the chaperone molecule tapasin, might influence the timing of T-cell detection, we applied our probabilistic model to two HLA alleles. Differences in the tapasin-dependency of HLA-B*27:05 and HLA-B*44:02 was previously modelled by changing the in-

bioRxiv preprint doi: https://doi.org/10.1101/480301; this version posted November 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

trinsic association rate of peptide-MHC binding (parameter b in (5)). A low b (HLA-B*44:02) translates into strong tapasin-dependency, with few peptide-MHC molecules being detectable at the cell surface in tapasin-negative cells (3, 5). A high b (HLA-B*27:05) permits sufficient peptide loading in tapasin-negative cells for detectable cell surface presentation, but in tapasin-competent cells leads to some peptides being loaded on tapasin-unbound molecules, circumventing the potential benefits of tapasin, and thereby reducing tapasin-dependency overall (3, 5).

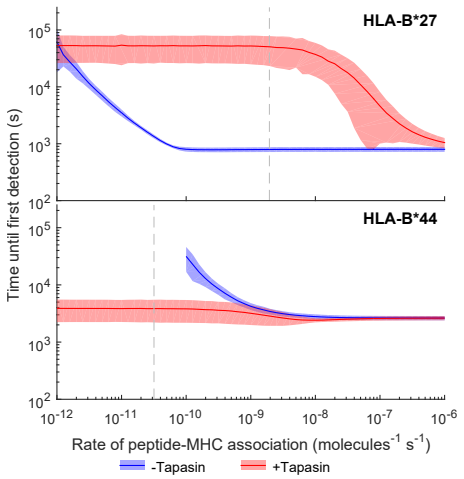

Fig. 6. Tapasin can accelerate or decelerate T-cell detection, depending on HLA allele. The rate of peptide-MHC association (b) was varied in simulations of HLA-B*27 (top panel) and HLA-B*44 (bottom panel). Compared are simulations of tapasin-competent (red) and tapasin-negative (blue) cells. The expected time until first detection is quantified according to Algorithm 1. Simulated cell surface abundance incorporated the allele-specific peptide-MHC on-rates established in (5), with $b = 1.9446 \times 10^{-9}$ molecule^{-1} s^{-1} for HLA-B*27 and $b = 3.1773 \times 10^{-11}$ molecule^{-1} s^{-1} for HLA-B*44. These rates are indicated by the dashed lines.

We carried out simulations for HLA-B*27 and HLA-B*44, using the off-rate distributions quantified in Table 1, varying the on-rate b over a large range, and comparing tapasin-competent with tapasin-negative cells. In tapasin-negative cells, increasing the on-rate b always sped up T-cell detection (Figure 6). At the inferred value for b (from (5)), HLA-B*27 had a surface distribution which produced efficient T-cell detection, such that increases in this rate provided no additional benefit. In contrast, the inferred value of b for HLA-B*44 did not lead to any peptide-MHC complexes at the surface of tapasin-negative cells. In tapasin-competent cells, we found that the expected time to first detection strongly differed between the HLA alleles. We found that tapasin expression led to less efficient T-cell detection for HLA-B*27, though more efficient T-cell detection for HLA-B*44 (Figure 6). Since tapasin expression increases peptide filtering (parameter α), we can use the analysis of Figure 5b to explain how this arises. HLA-B*27 presumably corresponds to the situation exemplified by the red trace in Figure 5b, with an optimal α near 1, and declining detection with increased α. On the other hand, HLA-B*44 presumably corresponds more to the situation exemplified by the blue trace, where the optimal α is closer to 2, and less peptide filtering leads to very slow T-cell detection. We suggest that the shift in the optimal degree of filtering of these two HLA alleles results from a shift in the means of the off-rate distributions. HLA-B*27 has a $\log_{10}$ mean of -1.88 and HLA-B*44 has a $\log_{10}$ mean of -0.55 (Table 1), which we would expect to increase the optimal degree of filtering (Figure 5c).

Discussion

Due to the combinatorial nature of antigen presentation, it has been difficult to assay for how T-cell activation depends on the genetic variations in MHC class I. By combining current quantitative understanding of intracellular peptide-MHC binding and cell surface presentation with pMHC-TCR binding in the immunological synpase, the theory presented illustrates how changes in the antigen presentation profile can influence the timing of T-cell responses. Accordingly, the theory provides a perspective on how changes in antigen presentation, as encountered during viral infection, might affect an immune response. During viral infection, it is known that interferon-γ (IFNγ) release stimulates enhanced expression in the MHC locus, leading to enhanced expression of MHC class I/II and other components of the PLC such as tapasin and TAP (23), which all increase overall antigen presentation (24). In contrast, many viruses interfere with the antigen presentation pathway, reducing or at least altering the composition of peptide-MHC molecules presented at the cell surface (25). The HIV protein Nef is known to target pMHC complexes in post-ER compartments (26), while human cytomegalovirus impairs formation of the PLC and reduces tapasin expression (10).

We found that changes in the antigen presentation profile (APP) of antigen presenting cells can lead to a non-intuitive impact on the expected time until first productive T-cell encounter. Depending on the supply of peptide-MHC molecules intracellularly (parameter β), efficient T-cell detection could be achieved optimally by either aggressive or weak peptide filtering, depending on the MHC alleles. When peptide supply is low, more aggressive filtering is needed to overcome low numbers of peptide-MHC molecules arriving at the cell surface. Whereas, when peptide supply is higher, the cell surface becomes saturated. In this case, weaker peptide filtering results in greater diversity in the APP, and therefore leads to a greater number of T-cells having the potential to be activated by the APC. This observation could help to explain why MHC class I alleles have evolved with variable dependency on the chaperone protein tapasin (3), which enhances peptide filtering (4–6).

A natural extension of our results could be obtained by incorporating some of the ideas developed in (16), which considers the more general scenario of *probabilistic recognition* of infections. Whereas our analysis computes the probability that a T-cell encounters at least one cognate peptide, the framework introduced by (16) considers partial activation by non-cognate peptides. Our model can be compared with their results, for T-cell clonotype j, by assigning $W_j = 1$ and $W_i = 0$ for $i \neq j$, and setting an activation threshold of 1. Our analysis could be extended to account for higher activation thresholds, since Eq. 7 follows a hypergeometric dis-

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/480301; this version posted November 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tribution, though this would not change our conclusions. In fact, our analysis corresponds to the case where T-cell *mistakes* are negligible. And according to (16), this could always be made true for large populations of cognate peptides on the APP (z_f in their study) by adjusting the threshold of activation accordingly.

Similarly to (16), we did not consider the mechanism of negative selection explicitly in our study. However, this effect could be easily added to our framework by adjusting the transition matrix A in Eq. 4. Rows corresponding to T-cells that cannot be activated by an APC only presenting self-peptides are set to $[1\,0\,0\cdots 0]$. The expected time for detection then becomes $\frac{n_T}{n_T - n_{inac}}\frac{2}{a(x)}$, where n_{inac} is the number of T-cells that do not respond to the APC, which amounts to a simple scaling factor. Therefore, consistent with (15), T-cell activation dynamics are not qualitatively affected by negative selection, although the activation threshold is reduced. Furthermore, our analysis into the effect of tapasin remains unaltered, since the quantity $a(x)$ is independent of negative selection.

By combining established theories of antigen presentation and T-cell stimulation, we have produced predictions of immune system functioning that we hope can be tested experimentally. As we have simulated established HLA alleles directly, those same alleles could be incorporated into experimental analyses. However, murine models offer more potential for obtaining quantitative estimates of APPs in infected or uninfected organisms, and downstream T-cell responses. New methods based on mass spectrometry are beginning to enable high throughput quantitation of cell surface peptide-MHC, sorted by peptide sequence (27), offering the means to characterise the APP in detail, and over time. By measuring the timing of T-cell responses in a corresponding range of MHC backgrounds, it will soon be possible to establish whether the composition of the APP, rather than simply the presence of a handful of immunodominant epitopes, is a more representative determinant of T-cell response times.

Materials and methods

Off-rate prediction using BIMAS. The dissociation rates of five HLA class I alleles were characterised by applying the BIMAS algorithm to a large set of peptides. The peptides were generated from human protein sequences obtained from UniProtKB (release 2010_08). All subsequences of length 9 and 10 were extracted from the sequences of the proteins across the 24 chromosomes, generating a total of 22,160,455 peptide sequences. The dissociation rates were computed in Matlab, following the methodology described in (17).

Inferring the power-law parameters from simulated cell surface pMHC abundance. For a given distribution of peptide off-rates according to $\log_{10}(u) \sim \mathcal{N}(\mu, \sigma^2)$, the values of α and β are inferred from simulations of the peptide filtering model (5). The power-law function is fitted to equilibrium simulated cell surface abundance of peptide-MHC complexes by first $\log_{10}$-transforming the simulated values. Given that $\log_{10}(\beta u^{-\alpha}) = \log_{10}(\beta) - \alpha \log_{10}(u)$, we can then identify α and $\log_{10}(\beta)$ using linear regression of the $\log_{10}$ transformed equilibrium simulations against $\log_{10}(u)$. We used the *polyfit* function in Matlab, applied to peptides with $\log_{10}(u) < \mu - \sigma$ to calibrate α and β in this way. This sub-sample was used to ensure high accuracy approximation of the high affinity peptides.

ACKNOWLEDGEMENTS
The authors wish to thank Marco Ferrarini for critical reading of the manuscript.

Bibliography

1. Paul Chappell, El Kahina Meziane, Michael Harrison, Łukasz Magiera, Clemens Hermann, Laura Mears, Antony G Wrobel, Charlotte Durant, Lise Lotte Nielsen, Søren Buus, Nicola Ternette, William Mwangi, Colin Butter, Venugopal Nair, Trudy Ahyee, Richard Duggleby, Alejandro Madrigal, Pietro Roversi, Susan M Lea, and Jim Kaufman. Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. *eLife*, 4:1–22, 2015. ISSN 2050-084X. doi: 10.7554/eLife.05345.
2. Don Mason. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunology Today*, 19(9):395–404, sep 1998. ISSN 01675699. doi: 10.1016/S0167-5699(98)01299-7.
3. Anthony P Williams, Chen Au Peh, Anthony W Purcell, James McCluskey, and Tim Elliott. Optimization of the MHC class I peptide cargo is dependent on tapasin. *Immunity*, 16(4): 509–520, apr 2002. ISSN 10747613. doi: 10.1016/S1074-7613(02)00304-7.
4. Mark Howarth, Anthony Williams, Anne B Tolstrup, and Tim Elliott. Tapasin enhances MHC class I peptide presentation according to peptide half-life. *Proceedings of the National Academy of Sciences of the United States of America*, 101(32):11737–42, aug 2004. ISSN 0027-8424. doi: 10.1073/pnas.0306294101.
5. Neil Dalchau, Andrew Phillips, Leonard D. Goldstein, Mark Howarth, Luca Cardelli, Stephen Emmott, Tim Elliott, and Joern M. Werner. A peptide filtering relation quantifies MHC class I peptide optimization. *PLoS Computational Biology*, 7(10), 2011.
6. Alistair Bailey, Neil Dalchau, Rachel Carter, Stephen Emmott, Andrew Phillips, Jörn M. Werner, and Tim Elliott. Selector function of MHC I molecules is determined by protein plasticity. *Scientific Reports*, 5:14928, 2015. ISSN 2045-2322. doi: 10.1038/srep14928.
7. N Garbi, P Tan, a D Diehl, B J Chambers, H G Ljunggren, F Momburg, and G J Hämmerling. Impaired immune responses and altered peptide repertoire in tapasin-deficient mice. *Nature immunology*, 1(3):234–8, sep 2000. ISSN 1529-2908. doi: 10.1038/79775.
8. Denise S M Boulanger, Roberta Oliveira, Lisa Ayers, Stephen H Prior, Edward James, Anthony P Williams, and Tim Elliott. Absence of tapasin alters immunodominance against a lymphocytic choriomeningitis virus polytope. *Journal of Immunology*, 184(1):73–83, 2010. ISSN 1550-6606. doi: 10.4049/jimmunol.0803489.
9. Shirin Ashraf, Katja Nitschke, Usama M. Warshow, Collin R. Brooks, Arthur Y. Kim, Georg M. Lauer, Theresa J. Hydes, Matthew E. Cramp, Graeme Alexander, Ann-Margaret Margaret Little, Robert Thimme, Christoph Neumann-Haefelin, and Salim I. Khakoo. Synergism of tapasin and human leukocyte antigens in resolving hepatitis C virus infection. *Hepatology*, 58(3):881–889, 2013. ISSN 02709139. doi: 10.1002/hep.26415.
10. Anne Halenius, Sebastian Hauka, Lars Dölken, Jan Stindt, Henrike Reinhard, Constanze Wiek, Helmut Hanenberg, Ulrich H Koszinowski, Frank Momburg, and Hartmut Hengel. Human cytomegalovirus disrupts the major histocompatibility complex class I peptide-loading complex and inhibits tapasin gene transcription. *Journal of virology*, 85(7):3473–3485, 2011. ISSN 0022-538X. doi: 10.1128/JVI.01923-10.
11. Alexis M. Kalergis, Nicole Boucheron, Marie-Agnés Doucey, Edith Palmieri, Earl C. Goyarts, Zsuzsanna Vegh, Immanuel F. Luescher, and Stanley G. Nathenson. Efficient T cell activation requires an optimal dwell-time of interaction between the TCR and the pMHC complex. *Nature Immunology*, 2(3):229–234, 2001. ISSN 1529-2908. doi: 10.1038/85286.
12. Omer Dushek, Milos Aleksic, Richard J Wheeler, Hao Zhang, Shaun-Paul P Cordoba, Yan-Chun C Peng, Ji-Li L Chen, Vincenzo Cerundolo, Tao Dong, Daniel Coombs, and Philip Anton van der Merwe. Antigen potency and maximal efficacy reveal a mechanism of efficient T cell activation. *Science signaling*, 4(176):ra39, 2011. ISSN 1945-0877. doi: 10.1126/scisignal.2001430.
13. Saikrishna Gadhamsetty, Athanasius F M Marée, Joost B Beltman, and Rob J de Boer. A general functional response of cytotoxic T lymphocyte-mediated killing of target cells. *Biophysical journal*, 106(8):1780–91, 2014. ISSN 1542-0086. doi: 10.1016/j.bpj.2014.01.048.
14. Sotiris Prokopiou, Loic Barbarroux, Samuel Bernard, Julien Mafille, Yann Leverrier, Christophe Arpin, Jacqueline Marvel, Olivier Gandrillon, and Fabien Crauste. Multiscale Modeling of the Early CD8 T-Cell Immune Response in Lymph Nodes: An Integrative Study. *Computation*, 2(4):159–181, 2014. ISSN 2079-3197. doi: 10.3390/computation2040159.
15. Natali Zint, Ellen Baake, and Frank Den Hollander. How T-cells use large deviations to recognize foreign antigens. *Journal of Mathematical Biology*, 57(6):841–861, 2008. ISSN 03036812. doi: 10.1007/s00285-008-0191-1.
16. Hannah Mayer and Anton Bovier. Stochastic modelling of T-cell activation. *Journal of Mathematical Biology*, 70(1-2):99–132, 2015. ISSN 0303-6812. doi: 10.1007/s00285-014-0759-x.
17. Kenneth C Parker, Maria A Bednarek, and John E Coligan. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *The Journal of Immunology*, 152(1):163–175, 1994. ISSN 0022-1767. doi: 0022-1767/94/$02.0.
18. Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion, 2010. ISSN 0090-5364.
19. Mark J Miller, Arsalan S Hejazi, Sindy H Wei, Michael D Cahalan, and Ian Parker. T cell repertoire scanning is promoted by dynamic dendritic cell behavior and random T cell motil-

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/480301; this version posted November 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ity in the lymph node. *Proceedings of the National Academy of Sciences of the United States of America*, 101(4):998–1003, 2004. ISSN 0027-8424. doi: 10.1073/pnas.0306407101.

20. Yuri Sykulev, Michael Joo, Irina Vturina, Theodore J. Tsomides, and Herman N. Eisen. Evidence that a single peptide-MHC complex on a target cell can elicit a cytolytic T cell response. *Immunity*, 4(6):565–571, 1996. ISSN 10747613. doi: 10.1016/S1074-7613(00)80483-5.
21. Salvatore Valitutti and Antonio Lanzavecchia. Serial triggering of TCRs: A basis for the sensitivity and specificity of antigen recognition, 1997. ISSN 01675699.
22. Linda Wooldridge, Julia Ekeruche-Makinde, Hugo a. Van Den Berg, Anna Skowera, John J. Miles, Mai Ping Tan, Garry Dolton, Mathew Clement, Sian Llewellyn-Lacey, David a. Price, Mark Peakman, and Andrew K. Sewell. A single autoimmune T cell receptor recognizes more than a million different peptides. *Journal of Biological Chemistry*, 287(2):1168–1177, 2012. ISSN 00219258. doi: 10.1074/jbc.M111.289488.
23. Alan Belicha-Villanueva, Sarah McEvoy, Kelly Cycon, Soldano Ferrone, Sandra O. Gollnick, and Naveen Bangia. Differential contribution of TAP and tapasin to HLA class I antigen expression. *Immunology*, 124(1):112–120, 2008. ISSN 00192805. doi: 10.1111/j.1365-2567.2007.02746.x.
24. Denise SM Boulanger, Ruth Charlotte Eccleston, Andrew Phillips, Peter V Coveney, Tim Elliott, and Neil Dalchau. A mechanistic model for predicting cell surface presentation of competing peptides by MHC class i molecules. *Frontiers in Immunology*, 9:1538, 2018.
25. Ted H Hansen and Marlene Bouvier. MHC class I antigen presentation: learning from viral evasion strategies. *Nature reviews. Immunology*, 9(7):503–13, 2009. ISSN 1474-1741. doi: 10.1038/nri2575.
26. O Schwartz, V Maréchal, S Le Gall, F Lemonnier, and J M Heard. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nature medicine*, 2(3):338–42, 1996. ISSN 1078-8956. doi: 10.1038/nm0396-338.
27. Nathan P. Croft, Stewart A. Smith, Yik Chun Wong, Chor Teck Tan, Nadine L. Dudek, Inge E A Flesch, Leon C W Lin, David C. Tscharke, and Anthony W. Purcell. Kinetics of Antigen Expression and Epitope Presentation during Virus Infection. *PLoS Pathogens*, 9(1), 2013. ISSN 15537366. doi: 10.1371/journal.ppat.1003129.

Supplementary Note A: Supplementary Figures

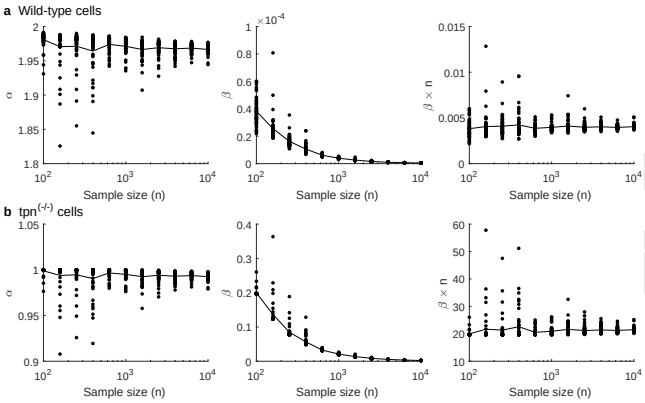

Fig. S1. Effect of sample size on relating the MHC class I peptide filtering model to power-law cell surface distributions. The number of distinct peptide sequences (n) was varied between 10^2 and 10^4, to determine the extent to which under-sampling peptides affects the distribution of peptides presented at the cell surface. This is quantified in terms of the best fit parameter values α and β of the power-law approximation for cell surface presentation of peptide-MHC complexes. The rightmost panels illustrates that variations in β can be compensated for by scaling out the sample size n.

DRAFT