CrossMark

ORIGINAL ARTICLE

Analyzing the effect of peptide-HLA-binding ability on the immunogenicity of potential CD8+ and CD4+ T cell epitopes in a large dataset

Shufeng Wang[1] · Jintao Li[1] · Xiaoling Chen[1] · Li Wang[1] · Wei Liu[1] · Yuzhang Wu[1]

Abstract Immunogenicity is a key factor that influences whether a peptide presented by major histocompatibility complex (MHC) can be a T cell epitope. However, peptide immunization experiments have shown that approximately half of MHC class I-binding peptides cannot elicit a T cell response, indicating the importance of analyzing the variables affecting the immunogenicity of MHC-binding peptides. In this study, we hierarchically investigated the contribution of the binding stability and affinity of peptide–MHC complexes to immunogenicity based on the available quantitative data. We found that the immunogenicity of peptides presented by human leukocyte antigen (HLA) class I molecules was still predictable using the experimental binding affinity, although approximately one-third of the peptides with a binding affinity stronger than 500 nM were non-immunogenic, whereas the immunogenicity of HLA-II-presented peptides was predicted well using the experimental affinity and even the predicted affinity. The positive correlation between the binding affinity and stability was only observed in peptide–HLA-I complexes with a binding affinity stronger than 500 nM, which suggested that the stability alone could not be used for the prediction of immunogenicity. A characterization and comparison of the 'holes' in the CD8+ and CD4+ T cell repertoire provided an explanation for the observed differences between the immunogenicity of peptides presented by HLA class I and II molecules. We also provided the optimal affinity threshold for the potential CD4+ and CD8+ T cell epitopes. Our results provide important insights into the cellular immune response and the accurate prediction of T cell epitopes.

Keywords Human leukocyte antigen (HLA) · Immunogenicity · Affinity · Stability · T cell repertoire

Introduction

In the vertebrate immune system, T cell receptors (TCRs) recognize peptides presented by major histocompatibility complexes (MHCs) on the surface of antigen-presenting cells. Peptides that bind to the MHC molecule and trigger a cellular immune response are referred to as T cell epitopes, which are of great importance in the development of epitope-based vaccines and immunotherapies against viral infections, tumors and autoimmune diseases [1–4]. The binding of a peptide to a MHC molecule is a prerequisite to eliciting a cellular immune response. In the past few years, immunologists have put great efforts into identifying peptides that bind to MHC molecules. The accumulation of these peptide data enables the development of highly accurate computational predictors of MHC-binding peptides. The predictions of MHC-binding peptides, especially the peptides that bind to MHC class I molecules, are well-studied problems in immunoinformatics [5–7].

Immunogenicity is the ability of a MHC-presented peptide to induce an immune response, which is an important factor that influences whether a MHC-presented peptide can be a T cell epitope. In the past, studies about

✉ Shufeng Wang
  wangshufeng81@hotmail.com

✉ Yuzhang Wu
  wuyuzhang@yahoo.com

[1] Institute of Immunology PLA, Third Military Medical University, Chongqing 400038, China

immunogenicity focused on immunodominance, which describes the dominant recognition of a relatively small number of potential epitopes by the immune system [8, 9]. Immunodominance is determined by many factors, such as the binding affinity of peptide–MHCs (pMHCs) [10], the stability of pMHCs [11, 12], the efficiency of MHC ligand processing [13–15], immunoregulatory phenomena [16] and competitive exclusion of the T cell repertoire [17].

Peptide-immunization experiments have shown that approximately half of MHC class I-binding peptides cannot elicit a T cell response [18, 19], indicating the importance of analyzing the variables that affect the immunogenicity of MHC-binding peptides. Harndahl et al. [20] suggested that peptide-MHC class I stability is more important than peptide affinity in the prediction of CTL immunogenicity. Calis et al. [21] performed a detailed study on the properties of MHC class I-presented peptides that enhance immunogenicity and found that several important amino acids at specific positions are associated with immunogenicity. In these studies, the term immunogenicity was based on the definitions established by Sercarz et al. [22], where dominant, subdominant, and even cryptic epitopes were classified as immunogenic peptides. Therefore, the immunogenicity of MHC-presented peptides could be defined as the recognition of pMHCs by TCRs.

The immunogenicity of the MHC-presented peptide is determined by the interaction between the pMHC and TCR. The available quantitative data describing the interactions of this ternary complex include the binding affinity and stability of pMHCs. To date, numerous studies have explored the contributions of the binding stability and affinity of pMHCs to the immunogenicity in response to a host-pathogen interaction [11, 12, 20]. However, experiments performed on limited numbers of peptides binding to a specific MHC molecule might induce biased results; the relative importance of these two factors needs further investigation with a large dataset, which should considerably improve the prediction of T cell epitopes. The building and the continual updates of the epitope databank IEDB (http://www.iedb.org/) [23–25], which is the largest database of immune epitopes and contains large amounts of experimental data on the binding abilities and T cell activities of pMHCs, make it possible to perform a system analysis of the variables affecting the immunogenicity of MHC-presented peptides based on a large dataset.

In this study, we extracted the T cell activity data and quantitative binding data of peptides presented by HLA molecules from IEDB and investigated the effect of peptide–HLA binding ability on the immunogenicity of potential CD8+ and CD4+ T cell epitopes by analyzing the interrelations among the binding affinity, stability and immunogenicity of the peptides presented by HLA class I and class II molecules.

Methods and materials

Data source

The datasets used here were mainly obtained from IEDB [23] (http://www.iedb.org/) in February 2015. The binding affinity and stability data of the peptide–MHC complexes were directly extracted from the "MHC ligand Assays" in the IEDB, while the data for the T cell activities of MHC-presented peptides were acquired from the published dataset downloaded from "T cell assays," where each entry contained a specific PMID number. The human and mouse proteomes were downloaded from the UniProt [26] database (http://www.uniprot.org/) in February 2015.

Immunogenicity of pMHCs

The immunogenicity of each HLA-presented peptide was assigned a specific value based on the reported T cell activity in the IEDB as follows: 1 for positive and −1 for negative. If there were more than one published entry for a specific pMHC, the immunogenicity of the pMHC was determined by the sum of these values. pMHCs with a positive value were collected as immunogenic pMHCs, and pMHCs with a negative value were defined as non-immunogenic pMHCs. Other pMHCs with a value of zero were removed from the dataset.

Immunogenicity of non-self pMHCs and self pMHCs

Because some peptide-immunization experiments were performed using transgenic mouse models, all peptides contained in the immunogenic and non-immunogenic datasets were checked using the human and mouse proteomes. Peptides that could be found as a subsequence of any protein in these two proteomes were defined as self pMHCs, and the other peptides contained in the immunogenic and non-immunogenic datasets were defined as non-self pMHCs.

The detailed data of the immunogenicity of the self and non-self pMHCs can be found at http://www.immunoinformatics.net/supplementary/immunogenicity.xlsx.

Datasets used to analyze the variables affecting the immunogenicity of HLA-presented peptides

The immunogenic/non-immunogenic pMHCs with quantitative binding affinities (IC50 or EC50) or stabilities in the MHC ligand assays were collected to analyze the relationship between affinity and immunogenicity or the relationship between stability and immunogenicity. HLA–

peptide complexes with identified binding affinity and stability were collected to detect the correlation between affinity and stability. The stability dataset submitted by Weinhold et al. (2009) (Reference ID in IEDB: 1014192 and 1014194), which had an obvious error in the quantitative unit, was not used in this study.

An overview of the quantitative data used here is shown in Table S1 (see supplementary data). HLA-A*02:01 and HLA-DRB1*0101 are the best studied HLA class I and II alleles, respectively, and the datasets of these two alleles contained the largest number of peptides with verified binding ability and immunogenicity in the respective HLA class. The analysis of the relationship between the binding ability and immunogenicity was mainly performed based on the peptides presented by these two well-studied alleles. The detailed data used here are now available at http://www.immunoinformatics.net/supplementary/Quantitative.zip.

Datasets used to evaluate the performance of the affinity-based prediction of the immunogenicity of the HLA-presented peptides

HLA class I

HLA-A*02:01: The binding affinities of the HLA-A*02:01-restricted 9-mers in the immunogenicity dataset of non-self pMHCs were predicted using NetMHC-3.4, the updated version of the best performance predictor according to the benchmark study [27]. Peptides with a predicted affinity stronger than 500 nM were collected and defined as Dataset_pred. The HLA-A*02:01-restricted 9-mer peptides with an experimental binding affinity stronger than 500 nM were categorized as Dataset_exp. Redundant peptides were removed from the two datasets to ensure that all remaining peptides differed from each other by at least two amino acids and that the subsequences of P3–P8 differed from each other by at least one amino acid. The 9-mer peptides in Assarsson's dataset [18] (vaccinia-derived peptides presented in an HLA-A*02 transgenic mouse model) were also used in this study. The immunogenicity of the peptides from Assarsson's dataset was defined by Calis et al. [21] as follows: The dominant, subdominant, and cryptic epitopes were classified as immunogenic, and the peptides with a negative response were classified as non-immunogenic.

HLA-A*24:02, B*07:02 and B*35:01: These were the other three alleles with more than 100 peptides in the immunogenicity dataset of non-self pMHCs. Due to the lack of quantitative binding data for these allele-restricted peptides with experimental immunogenicity, only peptides with a predicted affinity stronger than 500 nM were used to evaluate the performance of the affinity-based prediction of immunogenicity.

HLA class II

HLA-DRB1*01:01: The HLA-DRB1*01:01-restricted peptides in the immunogenicity dataset of the non-self pMHCs were predicted using NetMHCII2.2 [28, 29] (http://www.cbs.dtu.dk/services/NetMHCII/). Peptides with a predicted affinity stronger than 500 nM were included and defined as HLA-DRB1*01:01_pred. HLA-DRB1*01:01_exp was the set of HLA-DRB1*01:01-restricted peptides with an experimental affinity stronger than 500 nM in the immunogenicity dataset of the non-self pMHCs.

We also evaluated the performance of the affinity-based prediction of immunogenicity for the other HLA class II alleles with enough peptide data. After consideration of the ratio of immunogenic and non-immunogenic peptides and the total number of peptides with an experimental or predicted affinity stronger than 500 nM, only HLA-DRB1*03:01 and DRB1*07:01-restricted peptides with a predicted affinity stronger than 500 nM were available to perform the evaluation.

Evaluation

The overall performance of the prediction model was assessed by receiver operating characteristic (ROC) curve analysis [30], which is a widely used nonparametric performance measure. A ROC analysis tests the ability of a model to separate positive data from negative data without the need for selecting a threshold. The area under the ROC curve (AUC) provided a useful measure of the prediction quality as follows: a value of 0.5 indicates random prediction, and a value of 1.0 indicates perfect prediction. The efficacy of the binding affinity for distinguishing the immunogenic peptides from the non-immunogenic peptides was assessed using the ROC curves generated using the experimental affinity or predicted affinity. The performance of the affinity-based prediction of the immunogenicity of HLA-A*0201-restricted peptides was also compared to that of the prediction tool developed by Calis et al. [21] (http://tools.immuneepitope.org/immunogenicity/).

Results

Interrelations among the affinity, stability, and immunogenicity of HLA-I–peptide complexes

The binding between a peptide and a MHC molecule is required to trigger a cellular immune response. Binding affinity and stability are two general factors used to evaluate the peptide-MHC-binding ability. We investigated the effects of these two factors on the immunogenicity of HLA
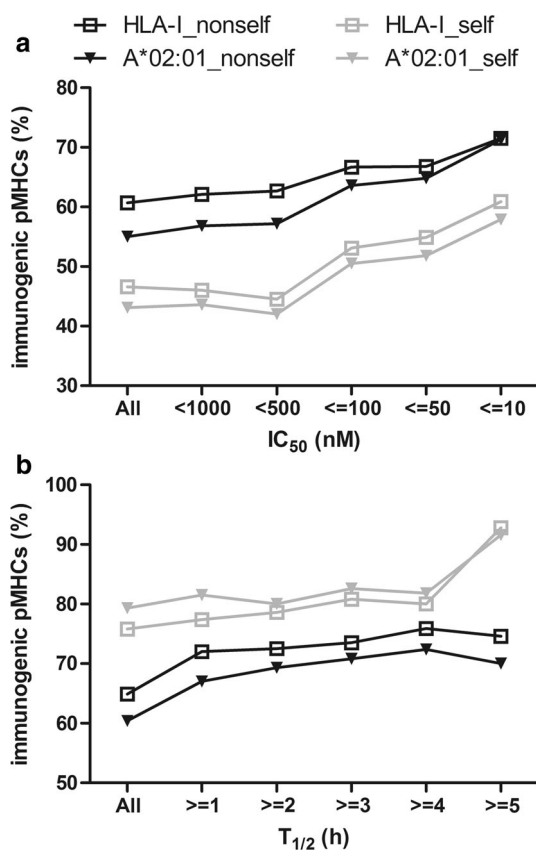
class I-presented peptides. As shown in Fig. 1, the percentage of immunogenic peptides in the non-self dataset increased as the binding affinity increased; this trend was observed in both the entire HLA-I-restricted dataset and the well-studied allele HLA-A*02:01-restricted dataset (Fig. 1a). The relationship between the stability and immunogenicity of HLA-I-peptide complexes showed a similar result to that observed between affinity and immunogenicity (Fig. 1b). According to the conventional standard for the definition of MHC binders, the stability of peptide–HLA-I complexes was more suitable than the affinity for distinguishing immunogenic peptides from non-immunogenic peptides. For example, 72.0 % of the HLA-I-presented non-self peptides with a high binding stability ($T_{1/2} \geq 1$ h) were immunogenic, whereas only 62.7 % of peptides with a binding affinity stronger than 500 nM (IC50 < 500 nM) were immunogenic; For HLA-A*02:01-restricted non-self peptides, 67.0 % of the peptides with high binding stability ($T_{1/2} \geq 1$ h) were immunogenic, while only 57.2 % of the peptides with a binding affinity stronger than 500 nM were immunogenic.

We also investigated the relationships between the binding abilities and the immunogenicity of self peptides

Fig. 1 Percentages of immunogenic peptides in the collected dataset at various binding affinity and stability levels. **a** Relationship between affinity and immunogenicity. **b** Relationship between stability and immunogenicity

presented by HLA class I molecules. After the self-reactive T cells were removed during the negative selection, most of self peptides were non-immunogenic. Thus, compared to non-self peptides, self peptides showed much lower percentages of immunogenic peptides at the corresponding affinity cutoffs. Nevertheless, the percentage of immunogenic peptides also increased with the binding affinity for the self peptides with a binding affinity stronger than 500 nM (Fig. 1a). The relationship between the stability and immunogenicity of self peptides showed was similar to that observed in non-self peptides, but the percentages at the corresponding stability cutoffs were higher than those of the non-self peptides (Fig. 1b), which might suggest that the binding stability of the HLA-I–peptide complex is a key factor for identifying self-reactive epitopes.

We then extracted the HLA-I-presented peptides with a binding affinity stronger than 500 nM or with a high stability ($T_{1/2} \geq 1$ h) and compared the affinities and stabilities between the immunogenic and non-immunogenic groups. As shown in Fig. 2a, the immunogenic peptides presented a significantly higher affinity (low IC50 values) than the non-immunogenic peptides; this difference was observed in both the HLA-A*02:01-restricted non-self and self datasets (non-self $p < 0.0001$, self: $p = 0.0203$, unpaired two-tailed t tests). However, significant differences in the stability were not observed between these two groups (Fig. 2b). We also collected peptides in a HLA-A*02:01-restricted peptide dataset in which the affinity, stability, and immunogenicity had all been experimentally determined, and we performed the comparison described above; the same results were observed (Fig. S1, see supplementary data). Due to the lack of quantitative peptide data, the stabilities could not be compared between the immunogenic and non-immunogenic groups for the other HLA class I alleles. We compared the affinities between the immunogenic and non-immunogenic groups for the peptide data of HLA-A*01:01 and B*07:02, the two alleles other than HLA-A*02:01 that contained the largest number of peptides with verified binding affinity and immunogenicity. However, significant differences were not observed. This result might be due the limited number of binding peptides with experimentally verified immunogenicity (A*0101: 44 and B*0702:131) and the unbalanced distribution of the binding affinities (e.g., about 80 % of the non-immunogenic peptides shared a binding affinity stronger than 100 nM in the collected HLA-B*0702-restricted dataset).

Furthermore, we investigated the correlation between the stability and affinity of peptide–MHC-I based on pMHCs whose binding affinity and stability were both experimentally determined. As shown in Fig. 3a, the stability of HLA-A*02:01-restricted peptides showed a significant tendency to increase with the affinity

Fig. 2 Comparison of the immunogenic and non-immunogenic peptides presented by HLA-A*02:01. **a** Comparison of the affinity of the immunogenic and non-immunogenic peptides with a binding affinity stronger than 500 nM. **b** Comparison of the stability of the immunogenic and non-immunogenic peptides with a high binding stability ($T_{1/2} \geq 1$ h). *P* values were obtained using unpaired two-tailed *t* tests
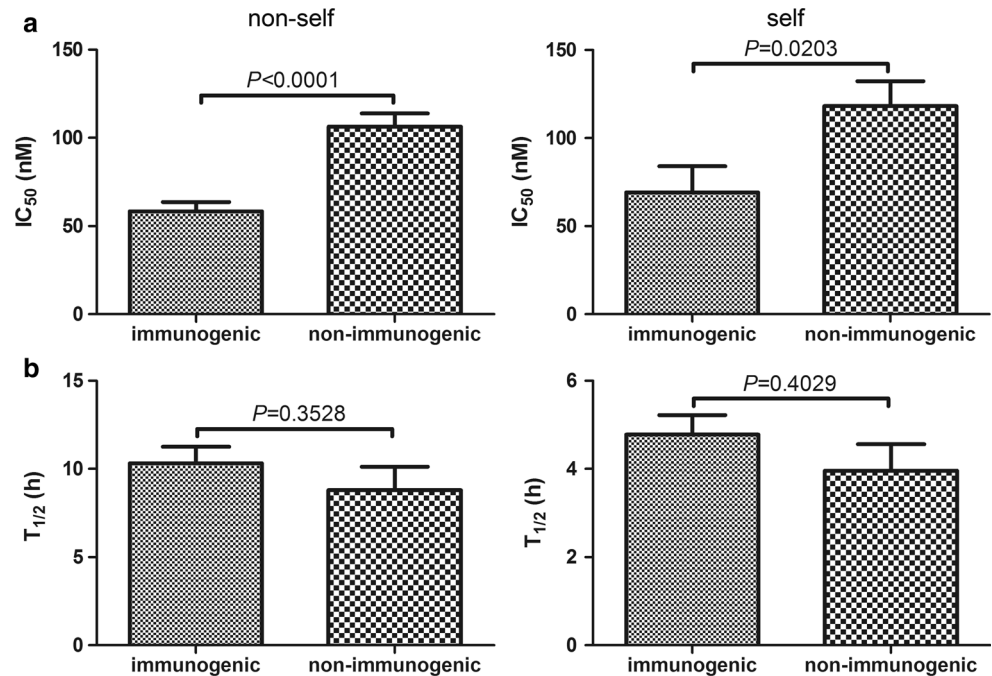
($R^2 = 0.2951$, $p < 0.0001$), and this increase was only observed in the peptides with a binding affinity stronger than 500 nM. The correlation between stability and affinity was also detected for HLA-A*30:02-restricted peptides that contained the second largest number of peptides with verified binding affinity and stability, and similar results were observed (Fig. 3b). Combining with the result that significant difference between the immunogenic and non-immunogenic peptides was only observed in pMHCs with a binding affinity stronger than 500 nM, we proposed that the stability of a pMHC could not be used solely as a predictor to distinguish immunogenic pMHCs from non-immunogenic pMHCs. For the peptides with high binding affinities, the high binding stability was in favor of the immunogenicity. In our collected datasets, more than 70 % of HLA-A*02:01-restricted peptides with a binding affinity stronger than 500 nM and a high stability ($T_{1/2} \geq 1$ h) were found to be immunogenic. This percentage was higher than the values based on the two factors alone at the corresponding thresholds.

Predicting the immunogenicity of HLA-I-presented peptides based on binding affinity

Since the immunogenic peptides presented a significantly higher binding affinity than did the non-immunogenic peptides, we evaluated the ability of the binding affinity to distinguish immunogenic peptides from non-immunogenic peptides using the following three HLA-A*02:01-restricted 9-mer datasets: Dataset_pred, Dataset_exp, and Assarsson's dataset. The performance was assessed by the ROC

curves generated using the experimental affinity or the predicted affinity.

As shown in Fig. 4, the predicted affinity was a poor predictor for separation of the immunogenic peptides from the non-immunogenic peptides, and the area under the ROC curve (AUC) that was generated based on the predicted IC50 values was only 0.517 (Fig. 4a). However, the immunogenicity of pMHCs was predictable when using the experimental binding affinities (AUC of 0.650) (Fig. 4b). The prediction tool of immunogenicity provided by IEDB was useful when experimental binding affinity data were lacking, but the evaluation on Dataset_exp indicated that the experimental binding affinity was better at distinguishing immunogenic peptides from non-immunogenic peptides. Furthermore, we evaluated the overall performance on Assarsson's 9-mer peptide data, where each peptide was bound to HLA-A*02:01 with an affinity stronger than 100 nM. The areas under the ROC curves generated using the experimental IC50 values and the predicted immunogenicity of the IEDB tool were 0.739 and 0.663, respectively (Fig. 4c), which further proved that the experimental binding affinity could be useful for defining immunogenic pMHCs.

The effect of binding ability on the immunogenicity of HLA-II-presented peptides

During a cellular immune response, the TCRs on the surface of Th cells recognize the peptides presented by MHC class II. Therefore, the variables affecting the immunogenicity of HLA-II-presented peptides were also

Fig. 3 Correlation between the stability and affinity of HLA-I-restricted peptides. **a** HLA-A*02:01-restricted peptides; **b** HLA-A*30:02-restricted peptides
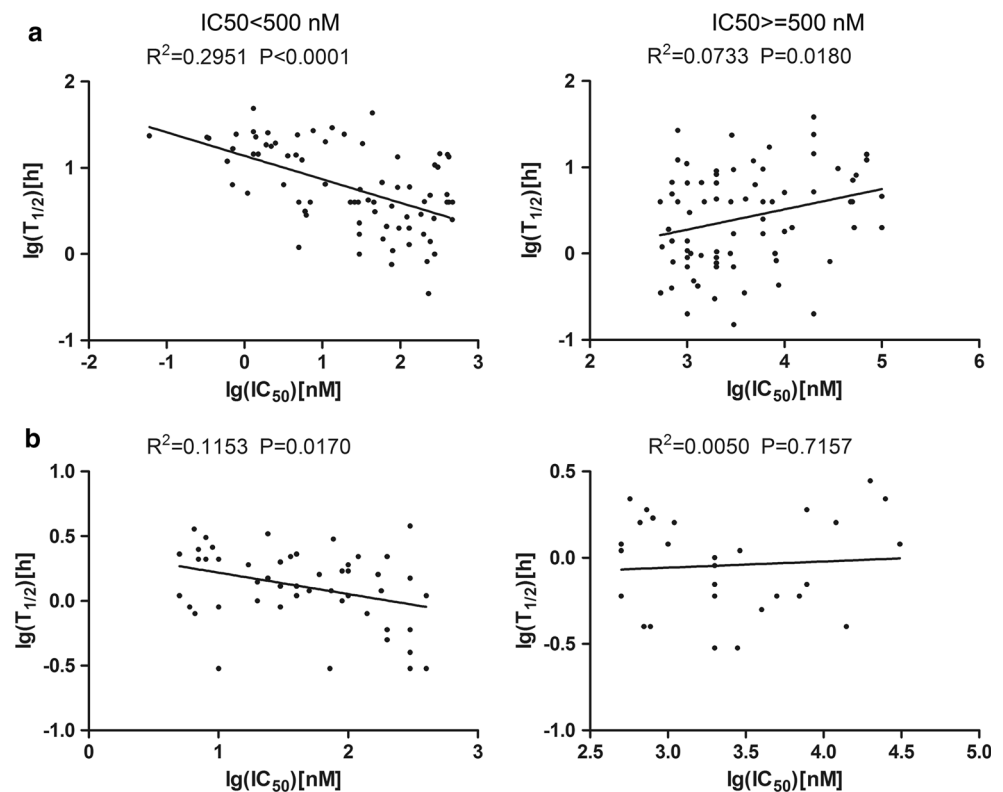

investigated in this study. Because of the biased composition of the collected self peptides presented by HLA-II molecules (more than 75 % of these peptides were immunogenic) and the relative low number of self peptides presented by a specific HLA-II molecule, the peptides used here were non-self peptides. The percentage of immunogenic peptides increased with the binding affinity, and this was observed in the entire HLA-II-restricted dataset and the well-studied allele HLA-DRB1*01:01-restricted dataset (Fig. 5a); this finding was similar to the result for the peptide–HLA-I complexes. Significant differences in the binding affinities were also observed between the immunogenic and non-immunogenic HLA-II-restricted peptides with a strong binding affinity (IC50 < 500 nM, $p = 0.0003$) (Fig. 5b). Most of the HLA-DRB1*01:01-restricted peptides with a strong binding affinity were immunogenic (79/90), and significant differences between these two groups were not found (data not shown). In the collected stability dataset of peptide–HLA-II complexes, only 15 in 94 of the peptide–HLA-II complexes with a high stability ($T_{1/2} \geq 1$ h) were immunogenic. Each of the 15 immunogenic peptides had a binding affinity stronger than 1000 nM, and 12 of the 15 immunogenic peptides had an affinity stronger than 500 nM. Some of the peptides that showed binding to a specific HLA-II molecule with a high stability shared a low binding affinity, which further proved that the stability of pMHC could not be solely used as a

predictor to distinguish the immunogenic from non-immunogenic peptides. However, the high percentage of immunogenic peptides at the threshold of 500 nM indicated that the binding affinity is sufficient for distinguishing immunogenic from non-immunogenic HLA-II-restricted peptides.

We evaluated the efficacy of the binding affinity in predicting the immunogenicity of HLA-II-presented peptides with a binding affinity stronger than 500 nM as was assessed for peptide–HLA-I complexes. The performance was assessed in the following three datasets: all HLA-II-restricted peptides with an experimental affinity stronger than 500 nM, HLA-DRB1*01:01-restricted peptides with an experimental affinity stronger than 500 nM, and HLA-DRB1*01:01-restricted peptides with a predicted affinity stronger than 500 nM (Fig. 6). These areas under the three generated ROC curves were 0.713, 0.792 and 0.760, respectively, for the above-mentioned datasets. This result suggested that the immunogenicity of HLA-II-presented peptides could be well predicted using the binding affinity and even the predicted binding affinity.
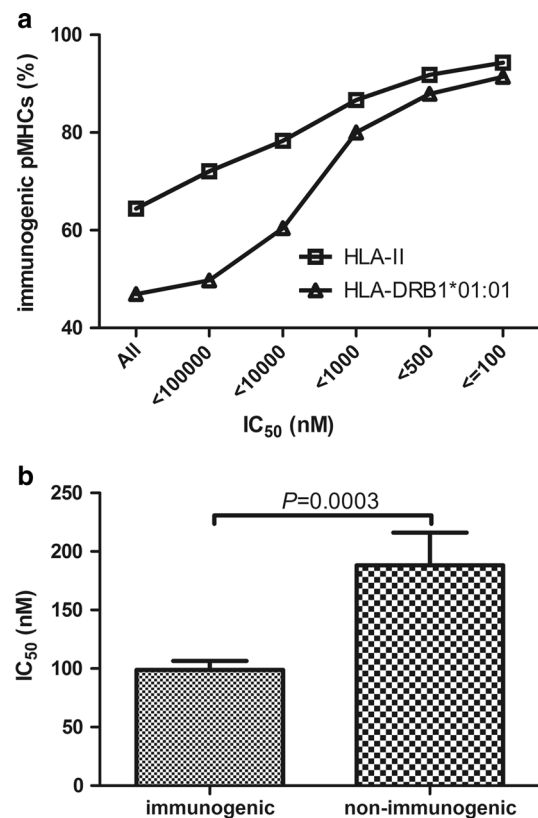
The 'holes' in the T cell repertoire

The above results indicated that binding affinity of pMHC could be a predictor of T cell immunogenicity, but its efficacy for HLA-I- and HLA-II-restricted complexes

a

b

c
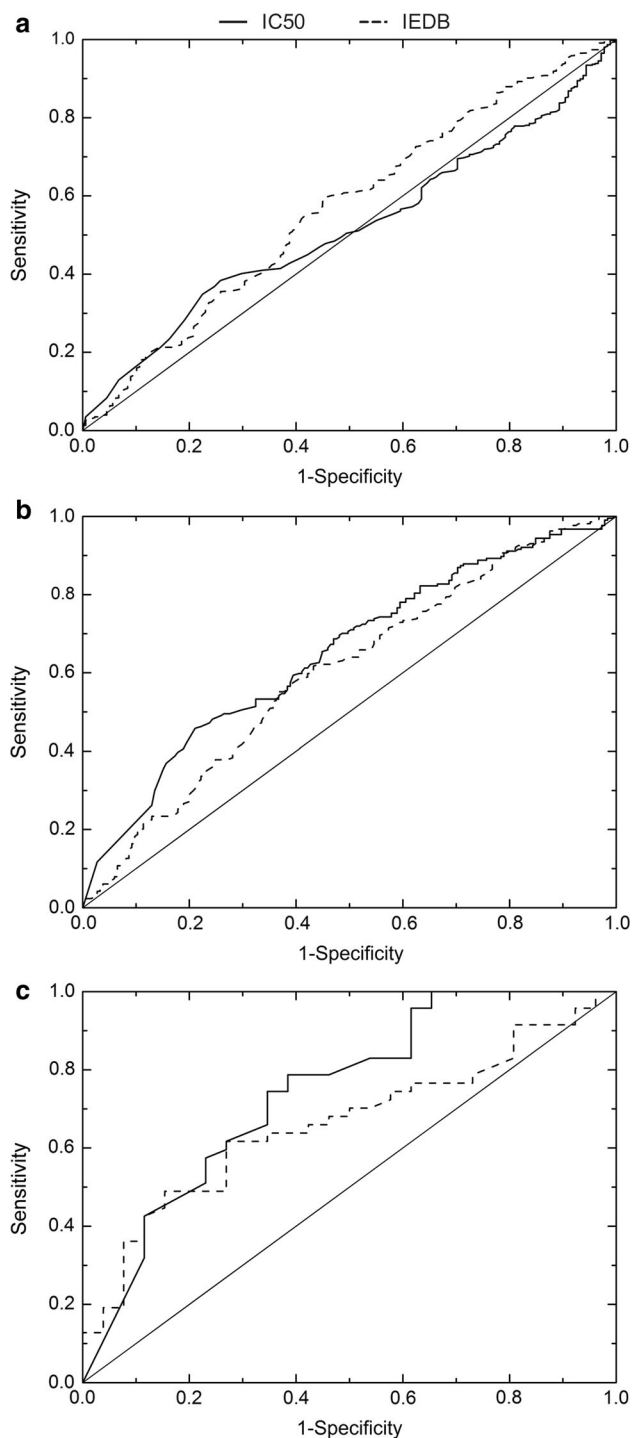
Fig. 4 Prediction of the immunogenicity of HLA-A*02:01-restricted peptides using the binding affinity and the IEDB tool. Performances were evaluated for the peptides with a predicted IC50 < 500 nM (**a**), peptides with an experimental IC50 < 500 nM (**b**) and Assarsson's dataset (**c**)

a

b

Fig. 5 Effect of binding affinity on the immunogenicity of HLA-II-restricted peptides. **a** Relationship between affinity and immunogenicity. **b** Comparison of the affinity of the immunogenic and non-immunogenic peptides with a binding affinity stronger than 500 nM. p values were obtained using an unpaired two-tailed t test
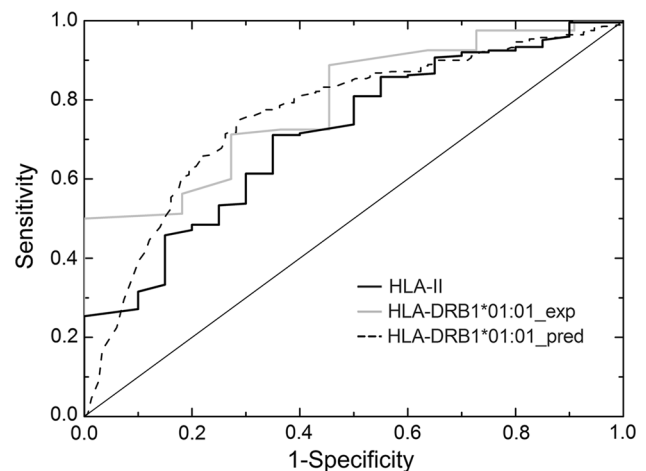
Fig. 6 Predicting the immunogenicity of peptide–HLA-II complexes based on the binding affinity. The performance was assessed using the following three datasets: all HLA-II-restricted peptides with an experimental affinity stronger than 500 nM (HLA-II); HLA-DRB1*01:01-restricted peptides with an experimental affinity stronger than 500 nM (HLA-DRB1*01:01_exp); and HLA-DRB1*01:01-restricted peptides with a predicted affinity stronger than 500 nM (HLA-DRB1*01:01_pred)

presented obvious differences. As shown in Table 1, a much better performance (AUC values) for distinguishing immunogenic from non-immunogenic peptides using the

experimental or predicted binding affinity was observed in the HLA-II-restricted datasets. The percentages of immunogenic peptides in the HLA-II-restricted datasets with an experimental binding affinity stronger than 500 nM were also much higher than those in the HLA-I-restricted datasets. The relatively low percentage of immunogenic peptides in the HLA-II-restricted dataset with a predicted binding affinity stronger than 500 nM could be due to the relatively low performance of prediction tools for HLA-II-binding peptides. Resetting the threshold, we found that almost 80 % of HLA-II-restricted peptides with a predicted binding affinity stronger than 50 nM were immunogenic. Thus, we investigated how these large differences between the immunogenicity of HLA class I- and class II-restricted peptides were generated.

Recent studies have demonstrated that the degenerate T cell recognition of peptides on MHC class I molecule created large holes in the CD8+ T cell (CTL) repertoire [31, 32]. Due to the deletion of self-reactive T cells caused by thymus selection and peripheral tolerance, foreign peptides that are highly similar to a self peptide do not trigger a T cell response. Foreign peptides that are never used in immune responses are referred as the "holes in the repertoire." The similarity of foreign peptides to self peptides was correlated to the length of the peptides. MHC class I molecules usually bind peptides of 8–10 residues in length in an extended conformation with the anchor residues (P2 and C-termini positions) buried in specificity pockets. The middle positions (P3–P8) of the binding peptide bulge out of the binding groove and play the main role in contacting the TCR [33]. A similarity calculation, based on the middle positions, indicated that almost 30 % of foreign peptides binding HLA class I molecules overlapped with the self peptides, which

strikingly agreed with the experimental results. The peptides binding to MHC-II molecules were much longer than MHC-I-binding peptides, and the TCR interacts with the binding core (9-core) and the N-terminal residues of the peptide [33]. According to the result of Calis et al. [31], in which the overlap of self and foreign peptides was lower than 1 % when every position of the MHC class I-binding peptide was taken into account, we inferred that the overlaps of the foreign and self peptides that bound to MHC class II molecules were no more than 1 %.

The TCR repertoire of CD4+ and CD8+ T cells can then be described using the MHC-presented peptides (Fig. 7). Assuming that 2 % of all possible 9-mer peptides actually bind to a particular MHC molecule, the number of peptides that can be presented by a given MHC molecule was calculated to be $20^9 \times 0.02 \approx 10^{10}$ (MHC-II-restricted peptides were calculated based on the binding core: 9-core). The estimated number of self peptides that can be presented by this MHC molecule is approximately 10^5 (the upper limit) [34]. Although the number of self-reactive peptides was much less than the total number of MHC-presented peptides, the deletion of the self-reactive T cell repertoire made some foreign peptides that highly overlapped with self peptides unable to evoke an immune response due to the degenerate T cell recognition of pMHCs (the gray regions in the diagrams of Fig. 7). Because the self/non-self overlaps in MHC-I-restricted peptides were much higher than that of MHC-II-restricted peptides, the holes in the CD8+ T cell repertoire (3×10^9) were much larger than those in the CD4+ T cell repertoire (10^8). The differences between the immunogenicity of MHC-I- and MHC-II-presented peptides can then be easily understood.

Table 1 Comparison of the effects of binding affinity on the immunogenicity of HLA class I and class II-restricted peptides

| HLA class	Affinity	Dataset	Immunogenic (%)	AUC
HLA-I	Experimental	HLA-I_all	62.7	0.594
		HLA-A*02:01[a]	57.0	0.653
	Predicted	HLA-A*02:01[a]	63.2	0.517
		HLA-A*24:02	37.9	0.654
		HLA-B*07:02	42.3	0.606
		HLA-B*35:01	53.6	0.646
HLA-II	Experimental	HLA-II_all	91.8	0.713
		HLA-DRB1*01:01	87.8	0.792
	Predicted	HLA-DRB1*01:01	65.2	0.760
		HLA-DRB1*03:01	65.3	0.760
		HLA-DRB1*07:01	74.4	0.752

Experimental: peptides with an experimental affinity stronger than 500 nM; Predicted: peptides with a predicted affinity stronger than 500 nM

[a] Only the 9-mer peptides were used here

Fig. 7 Diagram describing the
TCR repertoire of the CD8+
and CD4+ T cells using the
HLA-presented peptides.
a CD8+ T cell (or CTL);
b CD4+ T cell (or Th). *Black*
self peptides; *White* foreign
peptides; *Gray* foreign peptides
with high similarity to self
peptides
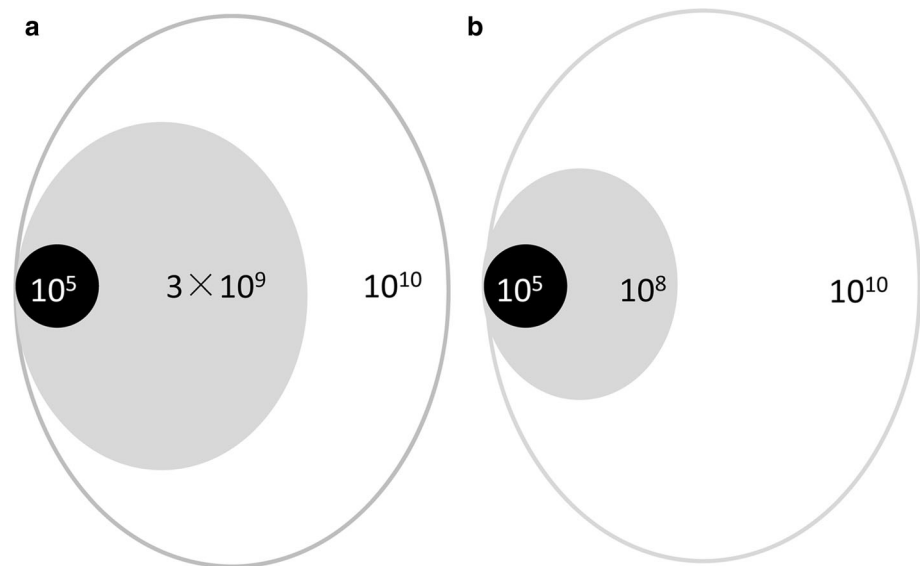
Discussion and conclusions

Immunogenicity is the ability of a MHC-presented peptide
to induce an immune response and is an important factor
that influences whether the peptide can be a T cell epitope.
Understanding the variables affecting the immunogenicity
of MHC-presented peptides will be useful for studying the
cellular immune response and improving the prediction of
T cell epitopes. The immunogenicity of a MHC-presented
peptide is determined by the interaction between the TCR
and the pMHC. The binding ability of a peptide to a MHC
molecule, which can be quantitatively measured by the
binding affinity or stability, is key to the interaction
between the TCR and pMHC. Our results showed that the
immunogenicity of HLA-I-presented peptides was still
predictable using the experimental binding affinity,
although approximately one-third of the peptides with a
binding affinity stronger than 500 nM were non-immuno-
genic; in contrast, the immunogenicity of HLA-II-pre-
sented peptides was well predicted using the experimental
affinity and even the predicted affinity. The positive cor-
relation between the binding affinity and stability was only
observed in peptide-HLA-I complexes with a binding
affinity stronger than 500 nM, which suggested that the
stability alone could not be used for the prediction of
immunogenicity.

Our results contradicted the results of the recent study
performed by Harndahl et al. showing that the stability is a
better predictor than the peptide affinity for CTL
immunogenicity [20]. In their study, the measured stability
was significantly correlated to the affinity of the selected
HLA-A*02:01-restricted peptides where all of the peptides
shared a relatively high binding affinity (IC50 < 1000 nM)
and more than 90 % were stronger than 500 nM. However,

the affinities of the collected peptides in our study ranged
from 10^{-1} to 10^{6} nM. The significant correlation between
the stability and affinity values was not observed in the
HLA-A*02:01-restricted peptides with a binding affinity
weaker than 500 nM (IC50 $\geq$ 500 nM). Combining with
the result that significant difference between the immuno-
genic and non-immunogenic peptides was only observed in
pMHCs with a binding affinity stronger than 500 nM, we
proposed that the stability of a pMHC could not be used
solely as a predictor to distinguish immunogenic pMHCs
from non-immunogenic pMHCs. Our results from the
HLA-II-presented peptides, which showed that only 15 of
the 94 collected HLA-DRB1*01:01-restricted peptides
with a high binding stability of $T_{1/2} \geq 1$ h were immuno-
genic, indirectly proves our argument. A high binding
stability favored the immunogenicity for peptides with high
binding affinities, which was consistent with the results
presented by Harndahl [20]. In our collected datasets, more
than 70 % of HLA-A*0201-restricted peptides with a
binding affinity stronger than 500 nM and a high stability
($T_{1/2} \geq 1$ h) were found to be immunogenic. This per-
centage was higher than the values based on the two factors
alone at the corresponding thresholds. The improved
accuracy for T cell epitope discovery was also achieved by
integrating the predictions for pMHC binding affinity and
stability [35].

The immunogenicity of MHC-presented peptides could
be predicted using the experimental binding affinity of
pMHCs. For peptide–MHC-II, approximately 90 % of
MHC-II-restricted peptides with a binding affinity stronger
than 500 nM were immunogenic. The predicted affinity
was also useful in distinguishing the immunogenic peptides
from the non-immunogenic peptides. For HLA-
DRB1*01:01-restricted peptides with a predicted affinity

stronger than 500 nM, the area under the ROC curve generated using the predicted affinity values was 0.760, which almost corresponds to a good performance [30]. The immunogenicity model developed by Calis et al. [21] was built by analyzing the residue properties that enhanced the immunogenicity of MHC class I presented peptides with a predicted affinity stronger than 500 nM. The cross-validation and blind tests of the prediction model showed that the immunogenicity is predictable with an average AUC of 0.65 [21]. In this study, our result suggested that the immunogenicity of MHC-I-restricted peptides was also predictable using the experimental binding affinity (AUC = 0.650 for 9-mer peptides binding to HLA-A*02:01 with an affinity stronger than 500 nM). An evaluation of Assarsson's 9-mer peptides, each of which showed a binding affinity to HLA-A*02:01 stronger than 100 nM, confirmed our results.

Since the immunogenicity of MHC-binding peptides with an affinity stronger than 500 nM could be predicted using their experimental binding affinity, there should be an optimal affinity threshold that determines the capacity of a MHC-binding peptide to elicit a T cell response. For MHC class I-restricted peptides, the threshold was approximately 50 nM, which agreed with Sette's early result [10]. In the HLA-A*0201-restricted epitopes of vaccinia virus identified by Assarsson et al. [18], each dominant epitope shared an experimental binding affinity stronger than 50 nM and a predicted affinity stronger than 50 nM was also observed in 12 of the 15 dominant epitopes. For MHC class II-restricted peptides, an experimental affinity stronger than 100 nM or a predicted affinity stronger than 50 nM was the best choice in which more than 80 % peptides were immunogenic. This result is important for the identification of T cell epitopes. Although peptides with high binding affinity could be more likely to evoke T cell responses, it should be noted that these epitopes might also lead to the development of immune escape variants. It has been suggested that mutations within CTL epitopes may be exploited by viruses, especially viruses with high mutagenicity, such as HIV and influenza A virus, to evade protective immune responses [36–38]. Therefore, conserved epitopes are likely the best choice for the development of vaccines against these viruses [39].

The degenerate T cell recognition of the pMHCs creates large holes in the T cell repertoire that accounted for the T cell tolerance to a large fraction of the presented foreign peptides [31]. A similarity calculation indicated that the overlaps of the foreign and self peptides presented by MHC-I and MHC-II molecules were 30 and 1 %, respectively. After self-reactive T cells were removed during the negative selection, the 'holes' in CD4+ T cell repertoire were much smaller than those in CD8+ T cell repertoire. A recent study on the TCRβ repertoire of CD4+ and CD8+ T cells indicated that the estimated species richness of TCRβ is approximately five times greater in CD4+ than in CD8+ T cells with the same number of cells [40]. This result might imply that CD4+ TCRs recognize greater numbers of antigenic epitopes than do CD8+ T cells, which is consistent with our findings that the number of foreign peptides recognized by CD4+ TCRs was greater than those for CD8+ TCRs (the white regions in the diagrams of Fig. 7). The diagram describing the TCR repertoire using MHC-presented peptides should facilitate the understanding of the cellular immune response.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Sette A, Fikes J. Epitope-based vaccines: an update on epitope identification, vaccine design and delivery. Curr Opin Immunol. 2003;15(4):461–70.
2. Ferrari G, Kostyu DD, Cox J, Dawson DV, Flores J, Weinhold KJ, et al. Identification of highly conserved and broadly cross-reactive HIV type 1 cytotoxic T lymphocyte epitopes as candidate immunogens for inclusion in Mycobacterium bovis BCG-vectored HIV vaccines. AIDS Res Hum Retroviruses. 2000; 16(14):1433–43. doi:10.1089/08892220050140982.
3. Knutson KL, Schiffman K, Disis ML. Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients. J Clin Invest. 2001;107(4):477–84. doi:10.1172/JCI11752.
4. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discov. 2007;6(5):404–14.
5. Trolle T, Metushi IG, Greenbaum JA, Kim Y, Sidney J, Lund O, et al. Automated benchmarking of peptide-MHC class I binding predictions. Bioinformatics. 2015;. doi:10.1093/bioinformatics/btv123.
6. Liao WW, Arthur JW. Predicting peptide binding to major histocompatibility complex molecules. Autoimmun Rev. 2011;10(8): 469–73. doi:10.1016/j.autrev.2011.02.003.
7. Lundegaard C, Lund O, Nielsen M. Prediction of epitopes using neural network based methods. J Immunol Methods. 2011; 374(1–2):26–34. doi:10.1016/j.jim.2010.10.011.
8. Adorini L, Appella E, Doria G, Nagy ZA. Mechanisms influencing the immunodominance of T cell determinants. J Exp Med. 1988;168(6):2091–104.
9. Berzofsky JA. Immunodominance in T lymphocyte recognition. Immunol Lett. 1988;18(2):83–92. doi:10.1016/0165-2478(88) 90046-6.
10. Sette A, Vitiello A, Reherman B, Fowler P, Nayersina R, Kast WM, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. J Immunol. 1994;153(12):5586–92.
11. Lazarski CA, Chaves FA, Jenks SA, Wu S, Richards KA, Weaver JM, et al. The kinetic stability of MHC class II: peptide complexes

is a key parameter that dictates immunodominance. Immunity. 2005;23(1):29–40. doi:10.1016/j.immuni.2005.05.009.

12. Busch DH, Pamer EG. MHC class I/peptide stability: implications for immunodominance, in vitro proliferation, and diversity of responding CTL. J Immunol. 1998;160(9):4441–8.

13. Pang KC, Sanders MT, Monaco JJ, Doherty PC, Turner SJ, Chen W. Immunoproteasome subunit deficiencies impact differentially on two immunodominant influenza virus-specific CD8+ T cell responses. J Immunol. 2006;177(11):7680–8.

14. Tenzer S, Wee E, Burgevin A, Stewart-Jones G, Friis L, Lamberth K, et al. Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance. Nat Immunol. 2009;10(6):636–46. doi:10.1038/ni.1728.

15. Chen W, Norbury CC, Cho Y, Yewdell JW, Bennink JR. Immunoproteasomes shape immunodominance hierarchies of antiviral CD8(+) T cells at the levels of T cell repertoire and presentation of viral antigens. J Exp Med. 2001;193(11):1319–26.

16. Haeryfar SM, DiPaolo RJ, Tscharke DC, Bennink JR, Yewdell JW. Regulatory T cells suppress CD8 + T cell responses induced by direct priming and cross-priming and moderate immunodominance disparities. J Immunol. 2005;174(6):3344–51.

17. De Boer RJ, Perelson AS. T cell repertoires and competitive exclusion. J Theor Biol. 1994;169(4):375–90. doi:10.1006/jtbi.1994.1160.

18. Assarsson E, Sidney J, Oseroff C, Pasquetto V, Bui HH, Frahm N, et al. A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. J Immunol. 2007;178(12):7890–901.

19. Kotturi MF, Botten J, Sidney J, Bui HH, Giancola L, Maybeno M, et al. A multivalent and cross-protective vaccine strategy against arenaviruses associated with human disease. PLoS Pathog. 2009;5(12):e1000695. doi:10.1371/journal.ppat.1000695.

20. Harndahl M, Rasmussen M, Roder G, Dalgaard Pedersen I, Sorensen M, Nielsen M, et al. Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. Eur J Immunol. 2012;42(6):1405–16. doi:10.1002/eji.201141774.

21. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Comput Biol. 2013;9(10):e1003266. doi:10.1371/journal.pcbi.1003266.

22. Sercarz EE, Lehmann PV, Ametani A, Benichou G, Miller A, Moudgil K. Dominance and crypticity of T cell antigenic determinants. Annu Rev Immunol. 1993;11:729–66. doi:10.1146/annurev.iy.11.040193.003501.

23. Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, et al. The immune epitope database 2.0. Nucleic Acids Res. 2010;38:D854–62. doi:10.1093/nar/gkp1004.

24. Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, et al. The design and implementation of the immune epitope database and analysis resource. Immunogenetics. 2005;57(5):326–36. doi:10.1007/s00251-005-0803-5.

25. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. Nucleic Acids Res. 2015;43(Database issue):D405–12. doi:10.1093/nar/gku938.

26. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 2004;32(Database issue):D115–9. doi:10.1093/nar/gkh131.

27. Lin HH, Ray S, Tongchusak S, Reinherz EL, Brusic V. Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. BMC Immunol. 2008;9:8. doi:10.1186/1471-2172-9-8.

28. Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinform. 2009;10:296. doi:10.1186/1471-2105-10-296.

29. Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinform. 2007;8:238. doi:10.1186/1471-2105-8-238.

30. Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988;240(4857):1285–93.

31. Calis JJ, de Boer RJ, Kesmir C. Degenerate T-cell recognition of peptides on MHC molecules creates large holes in the T-cell repertoire. PLoS Comput Biol. 2012;8(3):e1002412. doi:10.1371/journal.pcbi.1002412.

32. Frankild S, de Boer RJ, Lund O, Nielsen M, Kesmir C. Amino acid similarity accounts for T cell cross-reactivity and for "holes" in the T cell repertoire. PLoS ONE. 2008;3(3):e1831. doi:10.1371/journal.pone.0001831.

33. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. Annu Rev Immunol. 2006;24:419–66. doi:10.1146/annurev.immunol.23.021704.115658.

34. Yates AJ. Theories and quantification of thymic selection. Front Immunol. 2014;5:13. doi:10.3389/fimmu.2014.00013.

35. Trolle T, Nielsen M. NetTepi: an integrated method for the prediction of T cell epitopes. Immunogenetics. 2014;66(7–8):449–56. doi:10.1007/s00251-014-0779-0.

36. Pircher H, Moskophidis D, Rohrer U, Burki K, Hengartner H, Zinkernagel RM. Viral escape by selection of cytotoxic T cell-resistant virus variants in vivo. Nature. 1990;346(6285):629–33. doi:10.1038/346629a0.

37. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature. 1991;354(6353):453–9. doi:10.1038/354453a0.

38. Price GE, Ou R, Jiang H, Huang L, Moskophidis D. Viral escape by selection of cytotoxic T cell-resistant variants in influenza A virus pneumonia. J Exp Med. 2000;191(11):1853–67.

39. Stanekova Z, Vareckova E. Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. Virol J. 2010;7:351. doi:10.1186/1743-422X-7-351.

40. Li HM, Hiroi T, Zhang Y, Shi A, Chen G, De S, et al. TCRbeta repertoire of CD4+ and CD8+ T cells is distinct in richness, distribution, and CDR3 amino acid composition. J Leukoc Biol. 2015;. doi:10.1189/jlb.6A0215-071RR.