Published in final edited form as:

J Immunol Methods. 2015 October ; 425: 10–20. doi:10.1016/j.jim.2015.06.003.

Poor correlation between T-cell activation assays and HLA-DR binding prediction algorithms in an immunogenic fragment of *Pseudomonas* exotoxin A

Ronit Mazor¹, Chin-Hsien Tai¹, Byungkook Lee, and Ira Pastan^{*}

Laboratory of Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

Abstract

The ability to identify immunogenic determinants that activate T-cells is important for the development of new vaccines, allergy therapy and protein therapeutics. *In silico* MHC-II binding prediction algorithms are often used for T-cell epitope identification. To understand how well those programs predict immunogenicity, we computed HLA binding to peptides spanning the sequence of PE38, a fragment of an anti-cancer immunotoxin, and compared the predicted and experimentally identified T-cell epitopes. We found that the prediction for individual donors did not correlate well with the experimental data. Furthermore, prediction of T-cell epitopes in an HLA heterogenic population revealed that the two strongest epitopes were predicted at multiple cutoffs but the third epitope was predicted negative at all cutoffs and overall 4/9 epitopes were missed at several cutoffs. We conclude that MHC class-II binding predictions are not sufficient to predict the T-cell epitopes in PE38 and should be supplemented by experimental work.

Keywords

T-cell epitope; MHC alleles; T-cell activation; Immunogenicity; HLA binding prediction; HLA binding algorithm

1. Introduction

Helper T (T_H)² cells play a critical role in modulating the immune response. Activation of T_H cells requires the peptide-MHC complex and co-stimulatory proteins on the antigen

^{*}Corresponding author: Dr. Ira Pastan, Laboratory of Molecular Biology, National Cancer Institute, 37 Convent Drive, Room 5106, Bethesda, MD 20892-4264, Tel: 1-301-496-4797; Fax: 1-301-402-1344; pastani@mail.nih.gov.

¹Co-first authors

mazorr@mail.nih.gov (R. Mazor), TaiC@mail.nih.gov (C.H. Tai); BKLee@mail.nih.gov (B. Lee)

²Abbreviations: AUC, area-under-the curve; FN, false negative; FNR, FN rate; FP, false positive; FPR, FP rate; IEDB, immune epitope database; NIH, National Institutes of Health; PE38, *Pseudomonas* exotoxin; RIT, recombinant immunotoxins; ROC, receiver operating characteristic; T_H , helper T cells; TN, true negative; TP, true positive; TPR, TP rate

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of interest: The authors declare no financial or commercial conflict of interest.

presenting cells. An exogenous peptide is considered a T_H epitope if it is generated by protein-processing, binds to MHC class II molecules with high affinity, is transported to the cell surface, binds to a T-cell receptor and elicits an immune response.

The role of T_H cells and their epitopes in controlling the presence and magnitude of the immune response is well established (Baker et al., 2010; Paul and Zhu, 2010). Upon activation, T_H cells secrete cytokines, which provide the proper environment in the lymph nodes or other organs for B-cell proliferation, antibody class switching, and an increase in antibody production (Zhu and Paul, 2008; Swain et al., 2012). T_H cells also promote activation and growth of cytotoxic T-cells, and effect the innate immune system (Paul and Zhu, 2010).

Due to the clinical relevance of T_H cell epitopes in the immunogenicity of vaccines, allergens and protein therapeutics, there is an increasing interest in predicting T_H cell epitopes. Those predictions can be used to identify attractive immunogenic sequences for vaccine therapies (Sbai et al., 2001; Stevanovic, 2002; Sette and Rappuoli, 2010), to develop specific immunotherapy tolerance treatments against allergens (Anderson and Jabri, 2013), to achieve better understanding of autoimmune disorders and to help with novel protein therapeutic development in aspects of risk assessment, and in some instances to modify the therapeutic protein to avoid immunogenicity (Harding et al., 2005; Cantor et al., 2011; Moise et al., 2012; Mazor et al., 2014). There are many *in silico* algorithms that predict possible T-cell epitopes based on estimation of the affinity of a peptide to different MHC molecules (Jawa et al., 2013). The ambiguous value of these algorithms was recently demonstrated by Schwaiger et al. (Schwaiger et al., 2014) who found a striking low correlation between predicted and experimental T_H cell epitopes in transmembrane domains of viral envelope proteins.

We have been developing an experimental approach to predict and mitigate immunogenicity of recombinant immunotoxins (RITs). These are chimeric proteins designed to kill cancer cells and consist of a targeting portion (Fv or Fab) linked to a 38kDa portion of *Pseudomonas* exotoxin (PE38). RITs have produced striking tumor regressions in humans with hairy cell leukemia (Kreitman et al., 2000; Kreitman et al., 2012). However in patients with intact immune systems they exhibit high levels of immunogenicity (Hassan et al., 2007; Kreitman et al., 2009) that limit the number of treatment cycles that can be given. In a recent clinical trial, a RIT was administrated in combination with an immunosuppressive regimen that allowed increased number of treatment cycles to be given resulting in major tumor responses in some patients (Hassan et al., 2013). This indicates that the full potential of RITs to cause cancer regressions is limited by their immunogenicity and that decreasing their immunogenicity would be beneficial to patients.

To mitigate the immunogenicity of PE38, we have mapped the location of its T-cell epitopes. To do this we exposed PBMCs from 50 naïve donors and a few patients to RIT and mapped the T-cell epitopes by re-stimulation of the expanded cells with 15 mer peptides spanning the sequence of PE38. T cell activation was detected using IL-2 ELISpot. Functional cytokine secretion assays are now commonly used to detect T-cell epitopes (McMurry et al., 2007; Cohen et al., 2010; Oseroff et al., 2010). We used IL-2 because it

supports T-cell activation, differentiation, and memory and is a less specialized cytokine than IL-4 or IFN- γ (Tassignon et al., 2005). We identified one immuno-dominant and extremely promiscuous epitope in domain II in 21/50 of donors (Mazor et al., 2012) and seven additional major epitopes that promoted responses in four or more donors with various HLA haplotypes (Mazor et al., 2014). These epitopes were confirmed using samples from 16 cancer patients previously treated with PE38 containing RITs and who had mounted an immune response to the protein (Mazor et al., 2014).

To understand how well widely used *in silico* programs predict immunogenicity, we have carried out a comprehensive comparison between the *in silico* predictions and the experimentally identified T-cell epitopes in PE38.

2. Results

2.1 Normalization of experimental data

Previous analysis of the experimental data (Mazor et al., 2014) showed that many peptides do not cause T-cell responses among many donors. The large number of such unresponsive peptide-donor pairs provides a firm background level, which should help in detecting and accurately quantifying real responses. To take advantage of the large number of non-responsive data, we used the Z-score to normalize the ELISpot counts in this study (see Materials and methods).

The peptide that had a Z-score above a cutoff value for a donor was considered an epitope to the donor. We chose 6.0 as the cutoff value for this purpose because the epitopes that resulted from this choice agreed most closely with those published (Mazor et al., 2014). Figure 1A shows the normalized donor responses and Figure 1B shows the binarized data using the Z-score cutoff value of 6.0. The horizontal bars above the chart indicate the peptides that are epitopes to four or more donors.

2.2. HLA allele coverage in the donor cohort

Since the prediction programs use specific MHC allele, we typed the donor DRB molecules (DRB1, 3, 4 and 5) with high resolution (see results in Supporting Information Table 1). We concentrated on DR because prediction programs use very few DP or DQ alleles. There are four different MHC DRB loci in a human, one pair from each parent, each of which is highly polymorphic. Not counting those that could not be typed for various reasons, there were 36 different DRB alleles in our cohort of 50 donors.

The DRA alleles were not considered because the DRA gene is conserved and allelic variations do not affect the peptide-binding site (Trowsdale et al., 1985; Marshall et al., 1994).

2.3. Evaluation of prediction for individual donors

Each prediction method works for only a given set of HLA alleles. The immune epitope database (IEDB) “recommended” method covers 33 of the 36 donor alleles and IEDB “consensus” method covers 21. All other methods have much lower coverage. We,

therefore, used only the IEDB “recommended” and IEDB “consensus” methods for comparison with the experimental data.

The prediction score for a peptide-allele pair is given in terms of the percentile rank of the predicted binding strength of the given peptide among a large number of randomly selected 15mer peptides; the lower the rank the better the binding. For each of the 111 PE38 peptides, each donor can have up to four different covered alleles and the corresponding number of different predictions. To assign one score per each donor for each peptide, we recorded either the minimum or the average of the percentile ranks for the donor. Figure 2A shows the heat map of the minimum percentile rank of each peptide for each donor predicted by the IEDB “recommended” method. Several regions such as peptides 13-15, 67-68 and 75-80 (peptide ID is indicated on the X-axis) are predicted to be very immunogenic, against which almost every donor has a strong predicted response (black in the heat map). On the other hand, the same method predicted that no donor would respond to peptides 22-24 and 35-40. This is consistent with experimental observation.

Figure 2B shows the predicted donor responses by the IEDB “recommended” method, if one considers only those for which the percentile rank is better than 2.84% as responses. Percentile rank cutoff value of 2.84% gives the False Positive Rate (FPR) of 10% when compared with binarized experimental Z-scores. Figure 2C shows the false positives in white and false negatives in black when the predictions using this particular cutoff value is compared with the experimental T-cell responses given in Figure 1B. The FPR (the white area in Fig. 2C as a fraction of the white area in Fig. 1B) using this cutoff value is 10%.

Figure 2D shows the receiver operating characteristic (ROC) curves, which are the plots of True Positive Rate (TPR) versus FPR as the percentile rank cutoff value is varied from low to high, for both the minimum and the average percentile ranks predicted by the two IEDB methods. Even though the IEDB “recommended” method covers more MHC alleles in our donor cohort than the IEDB “consensus” method, the two methods perform similarly. All have the AUC between 0.67 and 0.69. Table 1 gives the TPRs and the corresponding percentile rank cutoff values that yield a few samples of FPRs. It can be seen from Table 1 and the ROC curves that the TPR is below 50%, i.e. more than half of the epitopes are missed by the predictions, unless the percentile rank cutoff value is set so large that more than 10% of the many non-epitopes are predicted as epitopes. We note in passing that, for a given FPR, using minimum and average percentile ranks give similar TPRs, but the percentile rank cutoff values are much lower when the minimum is used than when the average is used.

2.4. Evaluation of prediction for the population using the 33 HLA-DRB alleles in our donor cohort

Generally, a donor was considered to be a responder to a peptide if the donor has the corresponding ELISpot Z-score higher than 6 and as a predicted responder if the donor has an HLA-DRB allele that is predicted by the IEDB “recommended” method to bind the peptide better than a certain threshold value, as measured by the binding percentile rank. The numbers of predicted responding donors for each peptide at the percentile rank threshold of 1%, 6% and 20% are shown in Figure 3 A, B and C. As expected, the more

stringent the threshold, the fewer predicted number of responding donors. Peptides 13-15, and 75-79 were ranked as the top two epitope regions in our previous studies. Interestingly, they are also predicted to be epitopes to many donors at the very stringent 1% binding threshold (Fig. 3A). On the other hand, many experimentally observed frequent epitopes are not predicted at this binding threshold. More peptides are predicted as the percentile rank cutoff value is raised (Fig. 3B) but some frequent peptides such as peptides 4, 8, 9, 10, and 93 remain undetected (predicted to produce no or very few responses) even when the percentile rank cutoff value is raised to 20% (Fig. 3C and Table 2).

To quantitatively evaluate the population-wide prediction of the immunogenicity of a peptide, we concentrate on those peptides that produce a response in a large number of people. To this end, we define “frequent epitope” as the peptide that produces a response in at least a certain number of donors. After testing values 3, 4, and 5 using the experimental Z-score, this donor frequency cutoff value was chosen as 4, which gave the results closest to our published data (Mazor et al., 2014). This cutoff value was used both for the experimental Z-score data and also for the predicted percentile rank data. Figure 4A shows the number of peptides that are correctly or incorrectly predicted as frequent epitopes at binding threshold of 1% to 50%. Among the total 111 peptides, 27 produced responses from four or more donors experimentally and are colored in red (FN) or green (TP). The remaining 84 experimental non-frequent epitope peptides are colored orange (FP) or white (TN). At 2% percentile rank cut off, there are 26 peptides that are predicted to be frequent epitope (TP and FP combined), of which only 12, less than half, are correctly predicted, shown in green. At 6%, users would have to test 56 predicted frequent epitope peptides, but only 1/3, or 19 of them, are truly frequent epitopes. In order not to miss any frequent epitope, the binding percentile rank threshold would need to be at 35%, at which threshold, all 111 peptides are predicted to be frequent epitopes.

The TPR (TP/(FN+TP), FNR (FN/(FN+TP), FPR (FP/(FP+TN), Precision (TP/(TP+FP), and Accuracy (TP+TN)/(FN+TP+FP+TN) at thresholds of 1% to 20% percentile rank are plotted in Figure 4B. Although the TPR increases as the binding threshold is loosened, the Accuracy decreases because the number of false positives increases more rapidly. The Precision, true among predicted epitopes, drops below 50% when the binding cutoff is 2% or larger. This means that if we wanted to use the prediction algorithm to narrow down the number of peptides to test *in vitro*, more than 50% of the peptides would be FPs, even at a stringent threshold of 2%.

2.5. Evaluation of prediction for the population using HLA-DRB alleles in the worldwide population

To evaluate epitope prediction in the case where no cohort is selected, we predicted percentile rank from IEDB “recommended” method for the 15 reference HLA-DRB alleles that represent the world’s population (Greenbaum et al., 2011). Figure 5A shows the predicted percentile rank from IEDB “recommended” method for the 15 reference HLA-DRB alleles as a heat map. Two alleles, DRB1*0901 and DRB4*0101, are not in our cohort although it does include DRB4*0101.

To compare these predictions with the experimental data of our cohorts, we computed the correlation coefficient between the fractions of experimentally determined responding donors and the fractions of predicted responding alleles for all the frequent epitope peptides. This calculation requires two parameters; the percentile rank cutoff value and the allele frequency cutoff value (see Materials and methods). The correlation coefficient was calculated for each combination of these two parameter values. The results for the percentile rank cutoff values of 1% - 20% and the allele frequency cutoff value of 1-15 are given in Supporting Information Table 2. The correlation is generally poor, reaching the maximum value of 0.60 when the percentile rank and the allele frequency cutoff values are 6% and 5, respectively.

We chose the allele frequency cutoff value of 5, which gives the best correlation, to define the predicted frequent epitopes, which were then compared with the experimental frequent epitopes in the same manner as that used to evaluate the predictions using our own cohort HLA-DRB alleles. Figure 5B gives the numbers of correctly and incorrectly predicted frequent epitopes using the IEDB “recommended” method on the 15 reference HLA-DRB alleles. Compared to the predictions using our own 33 cohort HLA-DRB alleles (Fig. 4A), there are noticeably fewer peptides predicted to be frequent epitopes. FNR is still 30% (8 missed out of 27 frequent epitopes) at a very loose binding threshold of 20% (Fig 5E). At a very stringent, 1%, binding threshold, there are no peptides predicted to be a frequent epitope (Fig 5C). On the other hand, at binding cutoff of 6%, 9 out of 11 predicted frequent epitopes are correct which gives the best Precision of 82%, but the Miss Rate (FNR) is still high at 67%. The nine TP frequent epitopes at this cutoff level are peptides 12-15, 67-68, 76-78 (Fig. 5D and Table 2), which are the previously identified epitopes 1, 2A, 2B and 5 (Mazor et al., 2014). This indicates that highly immunogenic peptides can be correctly predicted at this 6% binding percentile rank cutoff and predicted responding allele frequency cutoff of 5 out of 15 reference alleles. However, the remaining 18 peptides that were found to be frequent epitopes experimentally would not have been identified.

2.6. Number of donors and alleles predicted to respond to the experimentally determined frequent epitopes

Table 2 gives the 27 experimentally determined frequent epitopes, their peptide sequences and the number of responding donors, along with the predicted number of responding donors and reference alleles at different percentile rank cutoffs for these peptides. Since the peptides overlap by 12 amino acids, one 9-mer binding core is included in two or three overlapping peptides. In some cases the epitope is spread over a number of neighboring peptides representing different cores in the same region of sequence. For these reasons, frequent epitopes often occur over several neighboring peptides. In Table 2 the peptides are grouped into 9 regions of consecutive frequent epitopes centered around a peak number of donors, of which the regions are sorted in descending order. The shaded entries indicate the number of responding donors and alleles that are above the cutoff values (4 and 5, respectively) to make the peptide a frequent epitope. Overall, 5 or 6 of the 9 regions are predicted to be epitopes independent of percentile rank between 6% and 20%. Four peptide regions I, II, V and IX are predicted to be frequent epitopes even when the binding percentile cutoff is as stringent as 1% if our cohort is used. On the other hand, epitope

regions III, IV, VI and VII would be missed. Interestingly, peptide 8 has 10 responding donors (20% of the 50 in the cohort) indicating it is an important epitope. However, the analysis does not predict a single response even at 20% cutoff. Examination of the HLA alleles of the 10 responding donors for this peptide, given in Supporting Information Table 3, reveals that all of their DRB1 are covered by IEDB “recommended” method and that 9/10 donors have at least one allele included in the 15 reference HLA most common worldwide.

IEDB “recommended” method includes more than 200 DRB alleles, and only 5 and 6 DP and DQ alleles, respectively. This low selection did not cover most of our DP and DQ alleles in our cohort. Nevertheless, in order to examine if additional prediction for those alleles can improve the inability of the algorithm to predict epitope 3, we computed the binding of peptide 8 to the 5 DP and 6 DQ alleles covered by IEDB consensus. We found that both peptides got high percentile rank (>30) indicating that incorporation of DP and DQ binding predictions to the analysis would not have improved the prediction.

3. Discussion

In this study we report the similarities and discrepancies between experimental and algorithm predicted T-cell epitope mapping with a highly immunogenic bacterial protein, PE38. We evaluated seven prediction algorithms provided by IEDB and after identifying the algorithm that offers the best MHC II binding prediction and the best allelic coverage, we employed three different analysis strategies that cover different possible T-cell prediction needs. We found that five of the nine epitope regions were successfully predicted; however, high rate of false negatives in all analysis strategies is a major issue.

3.1. Comparison strategies

In this study we compared the promiscuous epitopes, which are those that produce a response in many donors that were identified experimentally to predicted epitopes using two separate DRB HLA populations. The first population includes the DRB alleles in our cohort that provide a good coverage of the world population (Mazor et al., 2012). The second population includes 15 reference HLA alleles that were shown to offer optimal coverage (Greenbaum et al., 2011) of the world’s population.

The use of cohort alleles will be useful to predict responses of specific patient populations that have unique HLA alleles or bias to certain alleles like DRB1*11 in well differentiated thyroid carcinoma and hairy cell leukemia patients (Juhasz et al., 2005; Arons et al., 2014), DRB1*0301 and DRB1*1001 in ovarian carcinoma patient (Kubler et al., 2006). On the other hand, the use of a limited number of reference alleles to predict T-cell epitopes is commonly used in the field of *in silico* T-cell epitope predictions (Oseroff et al., 2012; Schulten et al., 2013; Li et al., 2014; Salvat et al., 2014) and would be useful for researchers that do not have any cohort. Some investigators use a stringent threshold (low percentile rank), and identify the top binders as predicted epitopes (Cantor et al., 2011; Li et al., 2014). Others use a more forgiving threshold that provides a large number of positive candidate peptides, which can then be further analyzed using experimental methods (Iwai et al., 2003; Fonseca et al., 2006; Oseroff et al., 2012; Schulten et al., 2013).

While we assume that most end-users do not have a known cohort and would use a 15-allele list for epitope promiscuity prediction, we found that this approach gives a lower correlation to the experimental data and a much higher rate of FN compared with the HLA alleles in our cohort. Two of the experimentally positive peptides could not be identified even with a threshold cutoff of 50% percentile rank that accepts 90% of all peptides as positive. This could be attributed to the fact that some of the HLA alleles that contribute to the experimental response are not included in the list of 15 alleles.

Our analysis uses a list of 15 reference alleles and treats each of them with equal importance following a prediction strategy that is considered to be state of the art (Iwai et al., 2003; Tangri et al., 2005; Koren et al., 2007; Cantor et al., 2011; Abdel-Hady et al., 2014; King et al., 2014; Li et al., 2014; Salvat et al., 2014; Salvat et al., 2015); however, it may introduce a bias because those alleles actually have different frequencies in the population. For example, if a peptide is predicted positive to DRB1*0701 (which is the most frequent in the world population), but negative to DRB1*0401 (which is much less frequent in the world population (Gonzalez-Galarza et al., 2011)), those predictions are considered equally important in the in silico analysis. In fact, their impact on the immunogenicity of the end population will not be the same, and experimental T-cell epitope mapping with a cohort that represents the world population will indicate that the peptide is more immunogenic than the prediction.

3.2. Possible reasons for FP and FN predictions

We found a high rate of FP epitopes. This finding is not surprising because the MHC binding algorithms only cover a small fraction of the factors that affect the immune response. It does not consider the processing of the protein in APC, and therefore the algorithm may predict peptides as epitopes that are not created naturally. It also does not consider the transport of the peptide to the APC surface and most importantly, the lack of ability to discriminate between tolerated and non-tolerated proteins, namely the availability of a T-cell receptor to recognize a bound peptide. There are other factors that add additional complexity like regulatory components of the immune system that can down regulate the immune response, though these aspects are also difficult to predict using experimental methods. Alternatively, we cannot rule out the possibility that the experimental method “missed” some epitopes due to low frequency of antigen-specific naïve T cells (Delluc et al., 2011) or due to a stringent threshold (of 4/50 donors), and those epitopes can be mistakenly classified as “FP” predictions. However, we do not suspect those are important epitopes, because the epitopes in this study were confirmed by epitope mapping of samples from immunized patients that had epitope spreading and clonal expansion of T cells and we did not observe new major epitopes.

It is noteworthy that the two strongest experimental epitopes (epitopes I and II in table 2) were consistently predicted correctly and were predicted to be in the top 5 epitopes throughout most of the analyses, whereas weaker epitopes (epitopes III and IV in Table 2) were considered FN in most of the analyses. A similar observation was reported by Schwaiger et al. (2014) who found very good agreement between the strong epitopes and their predictions and a poorer correlation for the weaker epitopes.

There are also surprising examples of FNs, which in hindsight does not seem so surprising. In this comparison we did not include predictions for DP or DQ molecules because there is not yet sufficient prediction data for those molecules (Wang et al., 2010). IEDB “recommended” method includes more than 200 DRB alleles, and only 5 and 6 DP and DQ alleles, respectively. This low selection did not cover most of our DP and DQ alleles in our cohort. Furthermore, previous restriction assays we have done for epitope 1 (peptide 15) using 10 donors that had a positive response to that epitope showed that this epitope is restricted by DR alone (Mazor et al., 2012). Moreover, focus on DR binding prediction is considered highly common in the art of immunogenicity prediction (Iwai et al., 2003; Tangri et al., 2005; Koren et al., 2007; Cantor et al., 2011; Abdel-Hady et al., 2014; King et al., 2014; Li et al., 2014; Salvat et al., 2014; Salvat et al., 2015) and it has been recently shown that prediction of DP and DQ had very low correlation with experimental data. (Paul et al., 2015) Therefore, we chose to focus only on DR. In our comparison study we found four epitopes that were missed by the prediction algorithm (FN) in most thresholds. It is plausible that the FN peptides are ones that cause immune response through presentation by DP or DQ (or both) and not by DR and thus could not be predicted without a prediction of DP and DQ. The inability to handle most of DP and DQ alleles is potentially a serious drawback of current prediction software.

3.3. Individual donor prediction

One of the possible applications for HLA binding predictions is the ability to individually predict immunogenicity responses for each patient based on his HLA. For example, whether a certain protein therapeutic will be immunogenic to an individual, a certain vaccine will work well or even develop specific immunotherapy tolerance for treating allergy patients. For a given person, it is unknown which HLA allele contributes to the presentation of a specific peptide or how much each different allele contributes to the activation of the T-cell population recognizing a specific peptide. Therefore, we included two types of analyses to evaluate the ability of the binding algorithm to predict the immunogenicity response of individuals. The minimum percentile rank represents the case where a single high affinity peptide-HLA allele complex is sufficient to elicit a response, while the average predicted percentile rank assumes that four presentation molecules, each with only a moderate binding, can elicit a strong immunogenicity response in combination. Interestingly, both types of analyses gave very similar TPRs at a given FPR and the AUC, except that a lower binding percentile cut off should be used if minimum percentile rank is used for analysis. Overall, the best AUC achieved for individual prediction was only 0.69.

Examination of the heat map reveals a good correlation between the experimental and the predicted responses in the most immunogenic peptides and the least immunogenic ones. The majority of the FN and FP are in the moderate responding epitopes. The presence of FP is expected as prediction algorithms are compelled to a high rate of false positives because peptide binding to an HLA molecule represents only a small and critical step in T cell activation, however the presence of FN as apparent in peptides 8-9 and others is surprising and indicates the shortcomings of immunogenicity prediction by current software (see above).

3.4. Previous efforts to map PE38 T-cell epitopes

A previous study to map the T-cell epitopes in PE38 used activation of PBMCs from donors using a similar peptide library. Peptide presentation was done using dendritic cells that were loaded with peptides and there was no stimulation with the whole protein. (Harding, 2008). This study identified three epitopes; the major one corresponding to peptides 75 and 76, and two others corresponding to peptides 15 and 65. Peptides 75, 76, and 15 were identified in our experimental assay (ranked 1st and 2nd). As shown in Table 2, they were predicted to be among the most frequent epitopes using all thresholds and analysis strategies. This supports the observation that the prediction algorithm predicts the strongest and most promiscuous epitopes. The third epitope identified by Harding (Harding, 2008) is peptide 65. This peptide was not found positive in our experimental results. However it was predicted to be relatively immunogenic (ranked 20th, 24th and 40th using 1%, 6% and 20% thresholds, respectively). We consider that the discrepancy between the experimental methods is a result of antigen processing and presentation of the full protein, such that peptide 65 is non-immunogenic in a native setting. This is supported by an analysis of the sequence of peptide 65 using protein processing algorithms PeptideCutter (Wilkins et al., 1999), which predicted this peptide to be cleaved by common proteases at position 8 of the peptide.

3.5. Current uses for algorithms

One of the common uses of algorithm based T-cell epitope prediction is to narrow the peptide population to be examined *in vitro* (Fonseca et al., 2006; Oseroff et al., 2012; Paul et al., 2013). This approach assumes that the peptides that have a low binding (high percentile score) will not bind to HLA molecules and thus there is no chance for those peptides to be epitopes. This approach is based on the assumption that algorithms always over-predict and rarely miss. We found a high rate of FNs in our analysis and 5 peptides that would not have been considered positive unless the entire 111-peptide library was considered positive. Furthermore, a promiscuity prediction holds the risk of missing peptides that bind at high affinity to only one or two HLA molecules and not be considered a promiscuous epitope.

When we had completed our study, an analysis comparing the computed and experimental responses of more than 95 donors and peptides derived from 30 proteins was published (Paul et al., 2015). This work focused solely on the false positive rate of the predictions, mainly due to the nature of the experimental work that did not allow identification of false negative predicted peptides. Nevertheless, the correlation and conclusions from this work strongly agree with the high rate of false positive responses observed in our study.

In conclusion, T-cell epitope prediction using HLA binding algorithms provides a convenient and easy to use tool. Current prediction tools offer very good binding data for DRB molecules but are hindered by the lack of knowledge of DP and DQ binding. Based on the rates of FN peptides in this analysis we conclude that current HLA binding algorithms are not yet reliable as a standalone strategy and should be accompanied with experimental work.

4. Materials and methods

4.1. Description of the experimental procedure

The experimental method that was used included a step of *in vitro* expansion of PBMC from 50 donors with the whole protein for 14 days, followed by re-stimulation with a library of 15-mer peptides spanning the whole sequence of PE38 with 12 residues overlap between neighboring peptides. Assessment of the T-cell activation response was done by counting the number of activated cells using IL-2 ELISpot. The detailed steps were previously described (Mazor et al., 2012). The epitopes identified in this study were further corroborated using samples of clinical patients that were previously treated with the protein and had a neutralization response to it. Epitope mapping using those samples gave the same epitopes (Mazor et al., 2014). The initial stimulation with the whole protein and *in vitro* expansion step reduces the rate of experimental FP (compared to other assays that only stimulate the cells with peptides) because it eliminates positive response from peptides that are not naturally processed. It also improves the sensitivity and reproducibility of the T-cell epitope mapping.

4.2. HLA-DRB alleles used within our cohort

Human PBMCs from naive donors were collected under research protocols approved by the National Institutes of Health (NIH) Review Board (99-CC-016). HLA typing was performed either by the HLA typing unit at NIH or by Thermo Fisher transplant diagnostic service using sequence specific primer extension and sequence based typing (Bunce et al., 1995; McGinnis et al., 1995).

4.3. Analysis of ELISpot experimental data

To distinguish the positive signals from noise for each donor, the ELISpot count of each peptide of each donor was normalized by calculating the Z-score as

$$Z_i = \frac{E_i - \bar{E}}{S}$$

where E_i is the average ELISpot count of four replicas of each peptide i for each donor, $\bar{E}$ is the average ELISpot count of that donor toward all PE38 peptides, and S is the standard deviation about the average. The ELISpot counts from the negative control, no peptide added, were not included in the mean and standard deviation calculations. The histogram of the Z-score does not show normal distribution, with the peak shifted to the left and a long tail extended to the right. There would still be many signals hidden in the background within 3 sigmas. To better differentiate the signals from noise, Z-scores were calculated twice: The first Z-score calculation used responses toward all 111 PE38 peptides to calculate the average and the standard deviation. In the second calculation, the responses that had the first Z-score greater than or equal to 1 were omitted when calculating the mean and standard deviation.

The Z-scores were then “binarized” using a cutoff score, i.e. turned into 1 or 0 depending on whether the value was greater or less than the Z-score cutoff value, respectively. A peptide

was considered immunogenic, or to be an epitope, to a donor if the corresponding binarized Z-score was 1. The Z-score cutoff value chosen was 6, which produced epitopes that matched best with the epitopes previously identified (Mazor et al., 2014).

4.4. Prediction programs and analysis of prediction data

The prediction programs used in this study are from the IEDB MHC-II binding predictions tool (Zhang et al., 2008) which includes the following prediction methods: IEDB “recommended”, IEDB “consensus” (Wang et al., 2008), Combinatorial library (Comlib) (Sturniolo et al., 1999), NN-align (netMHCII-2.2) (Nielsen and Lund, 2009), SMM-align (netMHCII-1.1) (Nielsen et al., 2007), Sturniolo (Sturniolo et al., 1999) and NetMHCIIpan (Nielsen et al., 2008). The sequences for the 111 peptides spanning PE38 and a set of HLA-alleles (see below) were input to the website. The prediction results are given in terms of the percentile rank of the score of each input peptide-allele pair against the scores of five million random 15mer peptides from the SWISSPROT database for each allele. The smaller the percentile rank, the higher the affinity.

The prediction results were also “binarized” using a percentile rank cutoff value. When the prediction result was considered for a donor, who may have up to four different HLA-alleles (two DRB1 and two DRB3/4/5), the binarization was done either using the average percentile rank or on the basis of the best (the smallest) percentile rank over all the alleles of the donor (see below).

4.5. Evaluation by counting correctly predicted donor-peptide pairs; ROC curves

We predicted the epitopes for individual donors by computing the binding scores of 111 peptides to four HLA DRB alleles of each donor (Supporting Information Table 1). The responses were binarized using a percentile rank cutoff value.

To evaluate the performance of the predictions, we calculated the rate of TPs and FPs using the binarized data. At a given percentile rank threshold, TP for example, is the peptide-donor pair for which the predicted response is below the percentile rank cutoff (the lower the percentile rank, the stronger the binding) and the ELISpot Z-score is above the Z-score cutoff value of 6. The TPR is the fraction of TPs among all experimentally positive donor-peptide pairs. The ROC curve (Hanley and McNeil, 1982) was made by plotting the TPR against the FPR as the percentile rank cutoff value was varied and the AUC was calculated.

4.6. Evaluation by counting the correctly predicted “frequent epitopes”

A peptide was considered to be a “frequent epitope” if it was an epitope for four or more donors among our cohort of 50 donors. Since the IEDB “recommended” method has greater allele coverage and the performance is similar to IEDB “consensus” method, only the predictions from IEDB “recommended” method were analyzed. Also, for this evaluation, we used the best binding allele among up to four different alleles to represent each donor because the previous ROC curve analysis indicated that prediction accuracy was slightly better by using the best prediction score (lowest percentile rank) rather than the average.

The TPs, FPs, TNs, and FNs were computed by comparing the predicted frequent epitopes to the experimentally determined frequent epitopes. Following quantities were calculated to measure the performance:

Sensitivity = TPR = $TP/(TP+FN)$ = fraction correctly predicted as frequent epitope among all experimentally determined frequent epitopes.

Precision = positive predictive value = $TP/(TP+FP)$ = fraction of experimentally determined frequent epitopes among all that are predicted as frequent epitopes.

Accuracy = $(TP+TN)/(TP+FN+FP+TN)$ = fraction correctly predicted, both frequent and not frequent, among all peptides.

Miss rate = FNR = $FN/(FN+TP)$ = fraction of peptides not predicted as frequent epitope among all experimentally determined frequent epitopes.

Fall-out = FPR = $FP/(FP+TN)$ = fraction of peptides predicted to be frequent epitopes among those non-epitope peptides determined experimentally.

4.7. Evaluation using HLA-DRB alleles in the worldwide population

To evaluate the ability of the prediction programs to predict epitopes without donor cohort information, 15 reference alleles that provide maximal population coverage (Greenbaum et al., 2011) were submitted to IEDB “recommended” method to obtain the predicted binding percentile rank for each allele-peptide pair. These would be useful in cases where the HLA sequences of the human subjects are not known.

As before, a peptide is a frequent epitope experimentally if four or more donors have Z-scores above 6 for the peptide. A predicted (as opposed to experimentally determined) frequent epitope in this case is the peptide that is an epitope (percentile rank below a cutoff value) to more than a certain number, the allele frequency cutoff, of alleles. The frequent epitopes predicted and experimentally determined in this manner were compared by calculating the TPs, FPs, TNs and FNs in the same way as described before. In addition, for each percentile rank cutoff value and each allele frequency cutoff value, the Pearson Correlation Coefficient was computed between the fractions of predicted responding alleles and the fractions of the experimentally determined responding donors for all peptides.

4.8. Statistical analysis

All the ROC curves, HEAT maps and Pearson correlation coefficients were calculated and generated with MATLAB software.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Peters Bjoern at La Jolla Institute for Allergy & Immunology and Chris Bailey-Kellogg at Dartmouth University for helpful discussions.

Funding: This work was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References

- Abdel-Hady KM, Gutierrez AH, Terry F, Desrosiers J, De Groot AS, Azzazy HM. Identification and retrospective validation of T-cell epitopes in the hepatitis C virus genotype 4 proteome: an accelerated approach toward epitope-driven vaccine development. *Hum. Vaccin. Immunother.* 2014; 10:2366–77. [PubMed: 25424944]
- Anderson RP, Jabri B. Vaccine against autoimmune disease: antigen-specific immunotherapy. *Curr. Opin. Immunol.* 2013; 25:410–7. [PubMed: 23478068]
- Arons E, Adams S, Venzon DJ, Pastan I, Kreitman RJ. Class II human leucocyte antigen DRB1*11 in hairy cell leukaemia patients with and without haemolytic uraemic syndrome. *Br. J. Haematol.* 2014; 166:729–38. [PubMed: 24931452]
- Baker MP, Reynolds HM, Lumicisi B, Bryson CJ. Immunogenicity of protein therapeutics: The key causes, consequences and challenges. *Self/nonself.* 2010; 1:314–322. [PubMed: 21487506]
- Bunce M, O'Neill CM, Barnardo MC, Krausa P, Browning MJ, Morris PJ, Welsh KI. Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens.* 1995; 46:355–67. [PubMed: 8838344]
- Cantor JR, Yoo TH, Dixit A, Iverson BL, Forsthuber TG, Georgiou G. Therapeutic enzyme deimmunization by combinatorial T-cell epitope removal using neutral drift. *Proc. Natl. Acad. Sci. U.S.A.* 2011; 108:1272–7. [PubMed: 21209329]
- Cohen T, Moise L, Ardito M, Martin W, De Groot AS. A method for individualizing the prediction of immunogenicity of protein vaccines and biologic therapeutics: individualized T cell epitope measure (iTEM). *J. Biomed. Biotechnol.* 2010; 2010:961752. doi:10.1155/2010/961752. [PubMed: 20706613]
- Delluc S, Ravot G, Maillere B. Quantitative analysis of the CD4 T-cell repertoire specific to therapeutic antibodies in healthy donors. *FASEB J.* 2011; 25:2040–8. [PubMed: 21368101]
- Fonseca SG, Coutinho-Silva A, Fonseca LA, Segurado AC, Moraes SL, Rodrigues H, Hammer J, Kallas EG, Sidney J, Sette A, Kalil J, Cunha-Neto E. Identification of novel consensus CD4 T-cell epitopes from clade B HIV-1 whole genome that are frequently recognized by HIV-1 infected patients. *Aids.* 2006; 20:2263–73. [PubMed: 17117012]
- Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res.* 2011; 39:D913–9. [PubMed: 21062830]
- Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics.* 2011; 63:325–35. [PubMed: 21305276]
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982; 143:29–36. [PubMed: 7063747]
- Harding, FA. Pseudomonas exotoxin A CD4+ T-cell epitopes. MedImmune, L., editor. United States: 2008. p. 56
- Harding FA, Liu AD, Stickler M, Razo OJ, Chin R, Faravashi N, Viola W, Graycar T, Yeung VP, Aehle W, Meijer D, Wong S, Rashid MH, Valdes AM, Schellenberger V. A beta-lactamase with reduced immunogenicity for the targeted delivery of chemotherapeutics using antibody-directed enzyme prodrug therapy. *Mol. Cancer Ther.* 2005; 4:1791–800. [PubMed: 16276001]
- Hassan R, Bullock S, Premkumar A, Kreitman RJ, Kindler H, Willingham MC, Pastan I. Phase I study of SS1P, a recombinant anti-mesothelin immunotoxin given as a bolus I.V. infusion to patients with mesothelin-expressing mesothelioma, ovarian, and pancreatic cancers. *Clin. Cancer Res.* 2007; 13:5144–9. [PubMed: 17785569]
- Hassan R, Miller AC, Sharon E, Thomas A, Reynolds JC, Ling A, Kreitman RJ, Miettinen MM, Steinberg SM, Fowler DH, Pastan I. Major cancer regressions in mesothelioma after treatment

- with an anti-mesothelin immunotoxin and immune suppression. *Sci. Transl. Med.* 2013; 5:208ra147.
- Iwai LK, Yoshida M, Sidney J, Shikanai-Yasuda MA, Goldberg AC, Juliano MA, Hammer J, Juliano L, Sette A, Kalil J, Travassos LR, Cunha-Neto E. In silico prediction of peptides binding to multiple HLA-DR molecules accurately identifies immunodominant epitopes from gp43 of *Paracoccidioides brasiliensis* frequently recognized in primary peripheral blood mononuclear cell responses from sensitized individuals. *Mol. Med.* 2003; 9:209–19. [PubMed: 15208742]
- Jawa V, Cousens LP, Awwad M, Wakshull E, Kropshofer H, De Groot AS. T-cell dependent immunogenicity of protein therapeutics: Preclinical assessment and mitigation. *Clin. Immunol.* 2013; 149:534–55. [PubMed: 24263283]
- Juhász F, Kozma L, Stenszky V, Gyory F, Luckas G, Farid NR. Well differentiated thyroid carcinoma is associated with human lymphocyte antigen D-related 11 in Eastern Hungarians: a case of changing circumstances. *Cancer*. 2005; 104:1603–8. [PubMed: 16121403]
- King C, Garza EN, Mazor R, Linehan JL, Pastan I, Pepper M, Baker D. Removing T-cell epitopes with computational protein design. *Proc. Natl. Acad. Sci. U.S.A.* 2014; 111:8577–82. [PubMed: 24843166]
- Koren E, De Groot AS, Jawa V, Beck KD, Boone T, Rivera D, Li L, Mytych D, Koscec M, Weeraratne D, Swanson S, Martin W. Clinical validation of the “in silico” prediction of immunogenicity of a human recombinant therapeutic protein. *Clin. Immunol.* 2007; 124:26–32. [PubMed: 17490912]
- Kreitman RJ, Hassan R, Fitzgerald DJ, Pastan I. Phase I trial of continuous infusion anti-mesothelin recombinant immunotoxin SS1P. *Clin. Cancer Res.* 2009; 15:5274–9. [PubMed: 19671873]
- Kreitman RJ, Tallman MS, Robak T, Coutre S, Wilson WH, Stetler-Stevenson M, Fitzgerald DJ, Lechleider R, Pastan I. Phase I trial of anti-CD22 recombinant immunotoxin moxetumomab pasudotox (CAT-8015 or HA22) in patients with hairy cell leukemia. *J. Clin. Oncol.* 2012; 30:1822–8. [PubMed: 22355053]
- Kreitman RJ, Wilson WH, White JD, Stetler-Stevenson M, Jaffe ES, Giardina S, Waldmann TA, Pastan I. Phase I trial of recombinant immunotoxin anti-Tac(Fv)-PE38 (LMB-2) in patients with hematologic malignancies. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2000; 18:1622–36. [PubMed: 10764422]
- Kubler K, Arndt PF, Wardelmann E, Krebs D, Kuhn W, van der Ven K. HLA-class II haplotype associations with ovarian cancer. *Int. J. Cancer*. 2006; 119:2980–5. [PubMed: 17016821]
- Li X, Yang HW, Chen H, Wu J, Liu Y, Wei JF. In Silico Prediction of T and B Cell Epitopes of Der f 25 in *Dermatophagoides farinae*. *Int. J. Genomics*. 2014; 2014:483905. [PubMed: 24895543]
- Marshall KW, Liu AF, Canales J, Perahia B, Jorgensen B, Gantz RD, Aguilar B, Devaux B, Rothbard JB. Role of the polymorphic residues in HLA-DR molecules in allele-specific binding of peptide ligands. *J. Immunol.* 1994; 152:4946–57. [PubMed: 8176213]
- Mazor R, Eberle JA, Hu X, Vassall AN, Onda M, Beers R, Lee EC, Kreitman RJ, Lee B, Baker D, King C, Hassan R, Benhar I, Pastan I. Recombinant immunotoxin for cancer treatment with low immunogenicity by identification and silencing of human T-cell epitopes. *Proc. Natl. Acad. Sci. U.S.A.* 2014; 111:8571–6. [PubMed: 24799704]
- Mazor R, Vassall AN, Eberle JA, Beers R, Weldon JE, Venzon DJ, Tsang KY, Benhar I, Pastan I. Identification and elimination of an immunodominant T-cell epitope in recombinant immunotoxins based on *Pseudomonas* exotoxin A. *Proc. Natl. Acad. Sci. U.S.A.* 2012; 109:E3597–603. [PubMed: 23213206]
- McGinnis MD, Conrad MP, Bouwens AG, Tilanus MG, Kronick MN. Automated, solid-phase sequencing of DRB region genes using T7 sequencing chemistry and dye-labeled primers. *Tissue Antigens*. 1995; 46:173–9. [PubMed: 8525476]
- McMurry JA, Kimball S, Lee JH, Rivera D, Martin W, Weiner DB, Kutzler M, Sherman DR, Kornfeld H, De Groot AS. Epitope-driven TB vaccine development: a streamlined approach using immunoinformatics, ELISpot assays, and HLA transgenic mice. *Curr. Mol. Med.* 2007; 7:351–68. [PubMed: 17584075]

- Moise L, Song C, Martin WD, Tassone R, De Groot AS, Scott DW. Effect of HLA DR epitope de-immunization of Factor VIII in vitro and in vivo. *Clin. Immunol.* 2012; 142:320–31. [PubMed: 22222093]
- Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics.* 2009; 10:296. [PubMed: 19765293]
- Nielsen M, Lundegaard C, Blicher T, Peters B, Sette A, Justesen S, Buus S, Lund O. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput. Biol.* 2008; 4:e1000107. [PubMed: 18604266]
- Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics.* 2007; 8:238. [PubMed: 17608956]
- Oseroff C, Sidney J, Kotturi MF, Kolla R, Alam R, Broide DH, Wasserman SI, Weiskopf D, McKinney DM, Chung JL, Petersen A, Grey H, Peters B, Sette A. Molecular determinants of T cell epitope recognition to the common Timothy grass allergen. *J. Immunol.* 2010; 185:943–55. [PubMed: 20554959]
- Oseroff C, Sidney J, Vita R, Tripple V, McKinney DM, Southwood S, Brodie TM, Sallusto F, Grey H, Alam R, Broide D, Greenbaum JA, Kolla R, Peters B, Sette A. T cell responses to known allergen proteins are differently polarized and account for a variable fraction of total response to allergen extracts. *J. Immunol.* 2012; 189:1800–11. [PubMed: 22786768]
- Paul S, Kolla RV, Sidney J, Weiskopf D, Fleri W, Kim Y, Peters B, Sette A. Evaluating the immunogenicity of protein drugs by applying in vitro MHC binding data and the immune epitope database and analysis resource. *Clin. Dev. Immunol.* 2013; 2013:467852. [PubMed: 24222776]
- Paul S, Lindestam Arlehamn CS, Scriba TJ, Dillon MB, Oseroff C, Hinz D, McKinney DM, Carrasco Pro S, Sidney J, Peters B, Sette A. Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *J. Immunol. Methods.* 2015 doi: 10.1016/j.jim.2015.03.0022.
- Paul WE, Zhu J. How are T(H)2-type immune responses initiated and amplified? *Nat. Rev. Immunol.* 2010; 10:225–35. [PubMed: 20336151]
- Salvat RS, Choi Y, Bishop A, Bailey-Kellogg C, Griswold KE. Protein deimmunization via structure-based design enables efficient epitope deletion at high mutational loads. *Biotech. Bioeng.* 2015 doi: 10.1002/bit.25554.
- Salvat RS, Parker AS, Williams A, Choi Y, Bailey-Kellogg C, Griswold KE. Computationally driven deletion of broadly distributed T cell epitopes in a biotherapeutic candidate. *Cell. Mol. Life Sci.* 2014; 71:4869–80. [PubMed: 24880662]
- Sbai H, Mehta A, DeGroot AS. Use of T cell epitopes for vaccine development. *Current drug targets. Infect. Disord.* 2001; 1:303–13.
- Schulten V, Greenbaum JA, Hauser M, McKinney DM, Sidney J, Kolla R, Lindestam Arlehamn CS, Oseroff C, Alam R, Broide DH, Ferreira F, Grey HM, Sette A, Peters B. Previously undescribed grass pollen antigens are the major inducers of T helper 2 cytokine-producing T cells in allergic individuals. *Proc. Natl. Acad. Sci. U.S.A.* 2013; 110:3459–64. [PubMed: 23401558]
- Schwaiger J, Aberle JH, Stiasny K, Knapp B, Schreiner W, Fae I, Fischer G, Scheinost O, Chmelik V, Heinz FX. *J. Virol.* 2014; 88:7828–42. [PubMed: 24789782]
- Sette A, Rappuoli R. Reverse vaccinology: developing vaccines in the era of genomics. *Immunity.* 2010; 33:530–41. [PubMed: 21029963]
- Stevanovic S. Identification of tumour-associated T-cell epitopes for vaccine development. *Nat. Rev. Cancer.* 2002; 2:514–20. [PubMed: 12094237]
- Sturniolo T, Bono E, Ding J, Radrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol.* 1999; 17:555–61. [PubMed: 10385319]
- Swain SL, McKinstry KK, Strutt TM. Expanding roles for CD4(+) T cells in immunity to viruses. *Nat. Rev. Immunol.* 2012; 12:136–48. [PubMed: 22266691]

- Tangri S, Mothe BR, Eisenbraun J, Sidney J, Southwood S, Briggs K, Zinckgraf J, Bilsel P, Newman M, Chesnut R, Licalsi C, Sette A. Rationally engineered therapeutic proteins with reduced immunogenicity. *J. Immunol.* 2005; 174:3187–96. [PubMed: 15749848]
- Tassignon J, Burny W, Dahmani S, Zhou L, Stordeur P, Byl B, De Groote D. Monitoring of cellular responses after vaccination against tetanus toxoid: comparison of the measurement of IFN-gamma production by ELISA, ELISPOT, flow cytometry and real-time PCR. *J. Immunol. Methods.* 2005; 305:188–98. [PubMed: 16157348]
- Trowsdale J, Young JA, Kelly AP, Austin PJ, Carson S, Meunier H, So A, Erlich HA, Spielman RS, Bodmer J, Bodmer FE. Structure, sequence and polymorphism in the HLA-D region. *Immunol. Rev.* 1985; 85:5–43. [PubMed: 2995241]
- Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput. Biol.* 2008; 4:e1000048. [PubMed: 18389056]
- Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics.* 2010; 11:568. [PubMed: 21092157]
- Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* 1999; 112:531–52. [PubMed: 10027275]
- Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, Bui HH, Buus S, Frankild S, Greenbaum J, Lund O, Lundegaard C, Nielsen M, Ponomarenko J, Sette A, Zhu Z, Peters B. Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res.* 2008; 36:W513–8. [PubMed: 18515843]
- Zhu J, Paul WE. CD4 T cells: fates, functions, and faults. *Blood.* 2008; 112:1557–69. [PubMed: 18725574]

Highlights

- Comparison of experimental data and in silico class II epitope predictions
- Use of binarization and multiple thresholds to analyze HLA class II prediction
- In silico methods failed to predict 4/9 epitopes
- In silico methods correctly predict 2 major epitope

Figure 1.

Normalized donor response. Each normalized response represents a normalized average of two assays, each was run in quadruplicates for 50 PBMC samples (A) the Z-scores of the ELISpot counts of each donor for each peptide. The darker the color, the higher the Z-score. (B) Binarized Z-scores using cut off value of 6.0. The horizontal bars above the chart indicate epitope regions where four or more donors responded.

Figure 2.

Prediction of epitope peptides for each donor. (A) The minimum percentile rank of each peptide for each donor predicted by the IEDB "recommended" method. Peptide ID is indicated on the X-axis. The lower the percentile rank, the stronger the predicted binding and darker the color in the heat map. Percentile ranks more than 20% are considered non-binders and colored in white. The horizontal bars above the chart indicate epitope regions where four or more donors responded experimentally. (B) The data of panel A, binarized using the percentile rank cutoff value of 2.84%, which gives the FPR of 10% when compared with binarized experimental Z-scores. (C) The comparison between the binarized percentile rank data of panel B and binarized experimental Z-score data of Fig. 1B. White bars indicate the FN and black bars the FP predictions. The black area is 10% of the white

area in Fig. 1B. (D) ROC curves (plot of TPR versus FPR as the percentile rank cutoff value is varied from low to high) using the minimum and average percentile ranks predicted by the IEDB “recommended” (red solid and dotted, respectively) and “consensus” (black solid and dotted, respectively) methods. The area under the AUC is shown in the figure. The vertical dotted line is at the FPR of 10%, which is obtained by the IEDB “recommended” method when the cutoff value for the minimum percentile rank is 2.84%.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3.

The number of responding donors for each peptide predicted by IEDB “recommended” method. The number of responding donors according to the normalized experimental Z-scores (black bars) and as predicted by the IEDB “recommended” method (gray bars) using binding percentile rank cutoff values of 1% (A), 6% (B), and 20% (C). The horizontal solid line is the threshold for a peptide to be considered frequent epitope.

Figure 4.

The performance of IEDB “recommended” method evaluated by using 33 HLA-DRB alleles in donor cohort. (A) Each horizontal line gives the number of TP (red), FP (green), TN (orange) and FN (white) peptides at the percentile rank cutoff value indicated by the Y-axis. The three dotted lines indicate the binding cutoff values of 1%, 6% and 20%, respectively. (B) Various performance measures, including Sensitivity, Precision, Accuracy, Miss Rate and Fall-out Rate, as the percentile rank cutoff value varies from 1% to 20%.

Figure 5.

The percentile ranks predicted by IEDB “recommended” method for the 15 reference HLA-DRB alleles and the performance evaluated using the 15 reference HLA-DRB alleles. (A) Percentile ranks more than 20% were considered non-binders and colored in white; those below 20% are colored in black for the strongest binder and in beige for the weakest. The horizontal bars above the chart indicate epitope regions where four or more donors responded experimentally. Peptide ID is indicated on the X-axis. (B) The number of TP, FP, TN and FN frequent epitopes predicted at different percentile rank binding cutoff values. A peptide was predicted as a frequent epitope if it is predicted to cause 5 or more alleles to respond. The dotted blue lines indicate the binding cutoff at 1%, 6% and 20% ranks. (C, D, E) The fraction of responding donors according to the normalized experimental Z-scores (black bars, $n=50$) and as predicted by the IEDB “recommended” method (gray bars) using for the 15 reference HLA-DRB alleles at binding percentile rank cutoff values of 1% (B), 6% (C), and 20% (D). The horizontal lines are the thresholds for a peptide to be considered frequent epitope black for experimental epitopes ($4/50=8\%$) and the gray for predicted ($5/15=30\%$).

Table 1
The TPR and FPR at various percentile rank cut off values for the IEDB “recommended” and IEDB “consensus” methods

FPR	IEDB “consensus” method				IEDB “recommended” method			
	Minimum rank		Average rank		Minimum rank		Average rank	
	TPR (%)	Cutoff	TPR (%)	Cutoff	TPR (%)	Cutoff	TPR (%)	Cutoff
1%	5.3	0.2	8.8	1.1	5.3	0.2	7.6	1.5
5%	23.3	1.2	23.7	4.3	22.1	1.1	24.8	6.1
10%	35.5	3.3	32.8	7.7	38.9	2.8	34.7	9.9
1-TPR	65.7	12.2	62.6	21.6	66.0	10.9	64.5	26.3

TPR, true positive rate

FPR, false positive rate

Table 2

Prediction of epitope regions

Epitope region	Peptide sequence	Number of responded donors	Predicted number of responded donors				Predicted number of responded reference alleles			
			1% cut off	3% cut off	6% cut off	20% cut off	1% cut off	3% cut off	6% cut off	20% cut off
I	12 EQGYPVQR LVALYL	4	1	35	44	50	1	5	7	14
	13 GYPVQRLVA LYLAAR	6	15	43	48	50	1	4	7	14
	14 VQRLVALYL AARLSW	19	15	38	50	50	1	4	9	14
	15 LVALYLAARL SWNQV	22	7	36	44	50	1	3	6	12
II	75 RGRIRNGALL RVYVP	4	7	29	34	50	0	2	4	10
	76 IRNGALLRVY VPRSS	9	24	32	40	50	2	3	7	13
	77 GALLRVYVPR SSLPG	12	24	32	47	50	2	4	8	13
	78 LRVYVPRSSL PGFYR	11	18	22	43	50	0	1	5	10
	79 YVPRSSLPGF YRTSL	6	10	14	29	41	0	0	2	6
	80 RSSLPGFYRT SLTLA	4	0	7	7	45	0	0	0	9
III	7 ETFTRHQRPR GWEQL	6	0	0	0	24	0	0	0	3
	8 TRHRQPRGW EQLEQC	10	0	0	0	0	0	0	0	0
	9 QPRGWLEQL EQCGYP	8	0	0	0	7	0	0	0	0
	10 RGWEQLEQC GYPVQR	6	0	0	0	7	0	0	0	0
4	AHQACHLPL ETFTRH	4	0	0	0	0	0	0	0	1

Epitope region	Peptide sequence	Number of responders	Predicted number of donors						Predicted number of reference alleles			
			1% cut off	3% cut off	6% cut off	20% cut off	1% cut off	3% cut off	6% cut off	20% cut off		
IV	5 ACHLPLETFT RHRQP	7	0	1	15	35	0	0	1	4		
	6 LPLETFRHR QPRGW	8	0	1	15	43	0	0	1	6		
V	67 WRGFYIAGDP ALAYG	7	12	40	43	50	3	8	8	13		
	68 FYIAGDPALA YGYAQ	8	10	27	36	47	2	5	6	9		
VI	1 PEGGSLAALT AHQAC	6	0	0	3	41	0	0	1	7		
	2 SLAALTAHQ A CHLPL	5	0	0	4	48	0	0	1	10		
VII	3 ALTAHQACH LPLETF	4	0	0	0	16	0	0	0	2		
	93 GPEEGGRLE TILGW	4	0	0	0	0	0	0	0	0		
VIII	94 EEGRLETIL GWPLA	5	0	0	4	31	0	0	1	6		
	95 GRLETILGWP LAERT	5	0	3	13	46	0	0	2	10		
IX	51 TVERLLQAH RQLEER	4	1	10	33	48	0	1	2	6		
	57 FVGYHGTFL E AAQSI	4	12	12	29	47	0	0	2	9		

Responses that were considered positive in the analysis are highlighted in gray.