Defining CD8⁺ T Cell Determinants during Human Viral Infection in Populations of Asian Ethnicity

Laura Rivino, Anthony T. Tan, Adeline Chia, Emmanuelle A. P. Kumaran, Gijsbert M. Grotenbreg, Paul A. MacAry and Antonio Bertoletti

*J Immunol* 2013; 191:4010-4019; Prepublished online 20 September 2013;
doi: 10.4049/jimmunol.1301507
http://www.jimmunol.org/content/191/8/4010

Supplementary Material
http://www.jimmunol.org/content/suppl/2013/09/24/jimmunol.1301507.DC1

References
This article cites 28 articles, 13 of which you can access for free at:
http://www.jimmunol.org/content/191/8/4010.full#ref-list-1

Why *The JI*? Submit online.

- **Rapid Reviews!** 30 days* from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

*average

Subscription
Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

Permissions
Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts

*The Journal of Immunology* is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852
Copyright © 2013 by The American Association of Immunologists, Inc. All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.
The identification of virus-specific CD8+ T cell determinants is a fundamental requirement for our understanding of viral disease pathogenesis. T cell epitope mapping strategies increasingly rely on algorithms that predict the binding of peptides to MHC molecules. There is, however, little information on the reliability of predictive algorithms in the context of human populations, in particular, for those expressing HLA class I molecules for which there are limited experimental data available. In this study, we evaluate the ability of NetMHCpan to predict antiviral CD8+ T cell epitopes that we identified with a traditional approach in patients of Asian ethnicity infected with Dengue virus, hepatitis B virus, or severe acute respiratory syndrome coronavirus. We experimentally demonstrate that the predictive power of algorithms defining peptide–MHC interaction directly correlates with the amount of training data on which the predictive algorithm has been constructed. These results highlight the limited applicability of the NetMHCpan algorithm for populations expressing HLA molecules for which there are little or no experimental binding data, such as those of Asian ethnicity. The Journal of Immunology, 2013, 191: 4010–4019.
most frequently expressed Caucasian molecules (e.g., HLA-A*0201), although it is scarce or not present at all for many HLA class I alleles present in individuals of different ethnicities. This is the case for Asian populations for whom definitions of CD8+ T cell epitopes have thus far been limited (9). The NetMHCpan algorithm represents an important attempt to overcome the limited HLA coverage of predictive algorithms. It is a pan-specific method that is able to predict peptide binding to HLA molecules for which there are no experimental binding data available. The MHC class I molecule is represented by a "pseudosequence" of 34 aa that was extracted from the analyses of the polymorphic amino acids that are involved in the MHC-peptide contact of all known HLA-A, -B, and -C molecules. This pseudosequence paired with a given peptide sequence with known affinity for the MHC is used as input in the training system (5, 10, 11).

To date, few studies have addressed the reliability of predictive algorithms in terms of coverage and accuracy of human T cell epitope identification, particularly in populations that express HLA class I molecules in which HLA-peptide binding motifs have been inferred rather than derived directly from experimental data. In this study, we evaluate the ability of the NetMHCpan algorithm to predict antiviral CD8+ T cell epitopes that we identified with traditional methods in 22 patients of Asian ethnicity who were either infected with Dengue virus (DENV) or who have cleared infection with hepatitis B virus (HBV) or severe acute respiratory syndrome coronavirus (SARS-CoV). We experimentally demonstrate that the predictive power to define biochemical properties relating to the peptide–MHC interaction directly correlates with the amount of training data on which the predictive algorithm has been constructed. To our knowledge, this has never been explicitly demonstrated by experimental means. Our study therefore highlights the limited predictive value of bioinformatics tools in the context of Asian populations with poorly characterized HLA molecules.

Materials and Methods

Ethics statement

Human blood samples were obtained after written informed consent from all participants. The study was conducted in accordance with the Declaration of Helsinki and approved by the Singapore National Healthcare Group ethical review board and the Centralized Institutional Review Board of the Singapore Health Services Pte.

Patient samples

The patients included in this study were with DENV, HBV or SARS based on the criteria listed below.

DENV Dengue diagnosis was confirmed on the basis of either detection of DENV RNA by RT-PCR or of NS1 Ag by ELISA, as described previously (3). Some patients with positive IgM and IgG acute serology (Panbio Dengue Duo Cassette) were also included if they fulfilled the World Health Organization criteria for probable dengue (12).

Hepatitis B virus. Patients with resolved HBV infection were diagnosed based on the detection of IgG Abs against HBV core and the absence of HBsAg in the serum.

SARS Recovered SARS individuals (6 y postinfection) were diagnosed with SARS based on clinical examination during the period from March to May 2003, according to the World Health Organization’s definition of SARS (13). The diagnosis was further confirmed by serological detection of SARS-CoV–specific Abs by ELISA and/or RT-PCR for SARS-CoV mRNA, as described previously (14).

Synthetic peptide libraries

Peptides were designed based on the full proteome sequence of DENV 2 virus (accession number: NC_001474, http://www.ncbi.nlm.nih.gov/nucleotide/158976983) and HBV (accession number: genotype B: AF121243, http://www.ncbi.nlm.nih.gov/nucleotide/AF121243; genotype C: AB112063, http://www.ncbi.nlm.nih.gov/nucleotide/AB112063), and the accessory 3a protein and nucleoprotein of SARS-CoV (Isolate Sin2774; accession number:AY283798, http://www.ncbi.nlm.nih.gov/nuccore/373691915). All peptides were purchased from Mimotopes (Australia). The peptide libraries consist of 15-mer peptides overlapping by 10 aa. The purity of the peptides was >80%, and their composition was confirmed by mass spectrometry analysis. Peptides were pooled in a two-dimensional matrix so that each individual peptide was present in two different pools as described previously (15).

PBMC isolation and T cell culture

PBMCs were isolated from peripheral blood by Ficoll gradient and cryo-preserved. For the in vitro assays, T cell lines were generated as described previously (15). In brief, 20% of PBMCs were pulsed with 10 μg/ml of the overlapping peptide libraries for 1 h at 37°C, subsequently washed, and cocultured with the remaining cells in AIM-V media (Life Technologies) supplemented with 2% AB human serum. T cell lines were cultured for 10 d in the presence of 20 U/ml IL-2 (R&D Systems).

IFN-γ ELISPOT

ELISPOT assays for the detection of IFN-γ-producing cells were performed as described previously (15). Ex vivo isolated PBMCs from DENV patients or in vitro expanded T cells from HBV and SARS resolved patients were used. For all patients, T cell lines were used to determine the minimal epitopes. To validate the HLA-A*1101– and A*2402–restricted predicted peptides, we used individual peptides to generate T cell lines. Testing by ELISPOT was performed with or without the peptide used to generate the T cell lines or with PMA/ionomycin as positive control (PMA 10 ng/ml, ionomycin 100 ng/ml). Positive values were calculated by subtracting the number of spots detected in the unstimulated control wells. Values were considered positive if they were >5 spot-forming cells (SPC) and at least two times the mean of the unstimulated control wells.

Intracellular staining and HLA restriction

Intracellular staining (ICS) and HLA restrictions were performed as described previously (15). In brief, T cell lines were stimulated with or without the peptide of interest (5 μg/ml) for 5 h at 37°C in the presence of brefeldin A (10 μg/ml). To assess degranulation, we added CD107a FITC Ab to the cells at the beginning of the stimulation. Cells were surface stained with anti-CD3 PerCP CY5.5 and anti-CD8 PE CY7, then fixed and permeabilized using the Cytofix/Cytoperm kit (BD Pharmingen). Cells were then stained with anti-IFN-γ PE and analyzed on a BD Canto or Fortessa FACS Scan. HLA restriction was determined by using EBV-transformed B cell lines matched for one or more HLA molecules with the patient. EBV-B cells were cultured with or without the peptide of interest (1 μg/ml) for 1 h at 37°C, washed, and then cocultured with the T cell lines for 5 h at 37°C in the presence of CD107a FITC Abs and brefeldin A. ICS was performed as described earlier.

Bioinformatics analyses

Viral proteomes were scanned for 8- to 11-mer peptides with HLA binding motifs for the different HLA molecules by using the NetMHCpan 2.2 (Table I), NetMHCpan 2.3 (Fig. 3), or NetMHCpan 2.4 (only nine-mers were selected, Supplemental Table III) prediction tool available on the Immune epitope database (http://www.immuneepitope.org/; or available at: http://www.cbs.dtu.dk). Peptides were ranked either by total scores (Table I) or by MHC IC50 values (Fig. 3, Supplemental Table III), setting a cutoff at IC50 < 1000 nM.

MHC stability ELISA

The 384-well polystyrene microtiter plates (Corning) were precoated with or without the peptide of interest (5 μg/ml) for 30 min at a distance of 5 cm from the UV lamps. After UV irradiation, the peptide exchanged MHC samples were incubated at 37°C for 1 h during which the streptavidin-coated plate was rinsed four times with 100 μl washing buffer (0.05% Tween 20 in PBS) per well and dried by vigorous tapping. Subsequently, 100 μl blocking buffer (0.005% polyvinyl alcohol alcohol in PBS) was added to the wells and then discarded after the incubation at room temperature for 30 min. Samples were further diluted in blocking...
buffer to a final concentration of 20 nM and then added to the wells (25 μl) in quadruplicate and incubated on ice for 1 h. The plate was rinsed four times with washing buffer, after which 25 μl HRP-conjugated anti-human β2-microglobulin (diluted 1:16,000 in blocking buffer; clone B2M-01; Novus Biologicals) was added and incubated on ice for 1 h. The plate was washed a further four times with washing buffer and cleared thoroughly. A total of 25 μl ABTS solution (Life Technologies) was added per well and incubated for 15 min at room temperature. The reaction was stopped by addition of 12.5 μl stop solution (0.01% sodium azide in 0.1 M citric acid solution), and the absorbance at 415 nm was measured using a spectrophotometer (SpectraMax M2; Molecular Devices).

**Results**

**Experimental validation of predicted CD8+ T cell epitopes restricted to the Asian HLA subtypes A*1101 and A*2402**

Results from a previous study suggested that bioinformatics tools are not accurate in predicting experimentally defined CD8+ T cell epitopes in the context of commonly expressed Asian HLA molecules (16). The lack of sufficient experimental data for these populations needed to adequately train the predictive algorithms is likely to contribute to this scenario (9). However, we do not exclude the possibility that algorithms could accurately predict epitopes that may have been missed by experimental screens. These are generally performed with 15-mer overlapping peptides that require further processing before they can bind to HLA class I.

To address this, we experimentally tested the recognition of the top 10 predicted CD8+ T cell epitopes from HBV Envelope protein (E) restricted to two of the most common HLA subtypes in the Southeast Asian population, HLA-A*1101 and -A*2402 (20.7 and 29.9%, respectively; Table I). Peptide predictions were performed using the NetMHCpan algorithm, which was the only algorithm that was able to predict epitope restrictors to the less characterized Asian HLA alleles (11). Immunogenicity of the predicted peptides was defined based on “total scores” that integrate the predictions of classical MHC class I binding affinity with proteasomal cleavage and TAP translocations. This integrated approach can outperform all methods based on the individual parameters (17).

Peptides were validated experimentally for their ability to bind to their restriction element and to elicit a T cell response in PBMCs from patients with resolved HBV (Fig. 1A–D). The binding capacity of the predicted peptides was assessed by using a microtiter plate–based peptide-MHC binding assay (18). This assay is based on the measurement by ELISA of HLA stability after irradiation of a UV-sensitive peptide bound to the HLA of interest (indicated as KB = known binder) and exchanged with our peptide of choice. By comparing HLA stability before and after the UV-epitope exchange, the assay also allows an approximation of the peptide’s affinity for the restriction element. As a negative control, HLA stability was measured after UV exchange with a peptide unable to bind the HLA of interest (NB = nonbinder). Experimental binding values for HLA-A*1101- and -A*2402-restricted peptides are shown Fig. 1A and 1B, respectively. Background is set according to NB values (black dotted line), and cutoff values for positive HLA-peptide binding are set at 0, 10, or 20% above the background value (black, red, or blue dotted line, respectively). Of the top 10 HLA-A*1101 predicted epitopes, few bound the restriction element (4/10 with 0% cutoff and 3/10 with 10 or 20% cutoff), whereas the majority of HLA-A*2402-restricted peptides could stabilize the restriction element (8/10 with 0% cutoff and 6/10 with 20% cutoff).

The immunogenicity of the predicted peptides was tested in a total of 29 Asian patients (20 HLA-A*1101* and in 9 HLA-A*2402; Supplemental Table I) who had resolved HBV infection. IFN-γ production by T cells was assessed by ELISPOT after 10 d of in vitro expansion of T cell lines with the relevant peptide, as described previously (19–21). In parallel for each patient, T cell lines were generated against a previously described A*2402-restricted HBV epitope (22) and against A*1101- (23, 24) or A*2402-restricted CMV epitopes (Fig. 1C, 1D). Overall, the predicted peptides elicited a CD8+ T cell response in 2 of 20 A*1101+ patients and in 3 of 9 A*2402+ HBV resolved patients. However, the responses observed in these patients were of an extremely low magnitude considering the cells had been expanded in vitro. Two of these HLA-A*1101-restricted peptides (7 and 8) appeared to bind poorly to the restriction element (Fig. 1A).

Thus, by using the NetMHCpan predictive algorithm, we were unable to identify relevant CD8+ T cell epitopes in the context of two common Asian HLA subtypes in 29 patients tested. Predicted peptides that were experimentally shown to efficiently bind HLA-A*1101 (peptides 2–4) or -A*2402 (peptides 1–3, 7–9) molecules also failed to induce a robust T cell response.

**Table I.** Top 10 epitopes from HBV E protein predicted to bind HLA-A*1101 and -A*2402 molecules

<table>
<thead>
<tr>
<th>HLA aa</th>
<th>Peptide Length</th>
<th>Sequence</th>
<th>Proteasome Score</th>
<th>TAP Score</th>
<th>MHC Score</th>
<th>Processing Score</th>
<th>Total Score</th>
<th>MHC IC50 (nM)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>A*1101</td>
<td>254–263</td>
<td>10</td>
<td>LIFLLVLVDY</td>
<td>1.09</td>
<td>1.37</td>
<td>-2.53</td>
<td>2.46</td>
<td>-0.08</td>
<td>342.4</td>
</tr>
<tr>
<td></td>
<td>147–156</td>
<td>10</td>
<td>ASPISSIFSR</td>
<td>0.96</td>
<td>0.72</td>
<td>-1.97</td>
<td>1.69</td>
<td>-0.28</td>
<td>92.4</td>
</tr>
<tr>
<td></td>
<td>277–285</td>
<td>9</td>
<td>SSTSGHKCK</td>
<td>0.89</td>
<td>0.2</td>
<td>-1.41</td>
<td>1.1</td>
<td>-0.31</td>
<td>25.4</td>
</tr>
<tr>
<td></td>
<td>226–235</td>
<td>10</td>
<td>TSCPPICPGY</td>
<td>1.36</td>
<td>1.22</td>
<td>-2.92</td>
<td>2.58</td>
<td>-0.34</td>
<td>825.5</td>
</tr>
<tr>
<td></td>
<td>360–369</td>
<td>10</td>
<td>MMYVRGRSLY</td>
<td>1.46</td>
<td>1.39</td>
<td>-3.22</td>
<td>2.86</td>
<td>-0.36</td>
<td>1661.4</td>
</tr>
<tr>
<td></td>
<td>178–187</td>
<td>10</td>
<td>LQAGFFFLTR</td>
<td>1.04</td>
<td>0.71</td>
<td>-2.36</td>
<td>1.75</td>
<td>-0.62</td>
<td>231.6</td>
</tr>
<tr>
<td></td>
<td>354–363</td>
<td>10</td>
<td>LSIVVMDWNYY</td>
<td>1.19</td>
<td>1.28</td>
<td>-3.1</td>
<td>2.48</td>
<td>-0.63</td>
<td>1263.3</td>
</tr>
<tr>
<td></td>
<td>357–366</td>
<td>10</td>
<td>IVIMMMGYGR</td>
<td>1.11</td>
<td>0.82</td>
<td>-2.59</td>
<td>1.93</td>
<td>-0.66</td>
<td>389.6</td>
</tr>
<tr>
<td></td>
<td>276–285</td>
<td>10</td>
<td>TSISGMPCK</td>
<td>0.89</td>
<td>0.18</td>
<td>-1.81</td>
<td>1.08</td>
<td>-0.73</td>
<td>63.8</td>
</tr>
<tr>
<td></td>
<td>355–363</td>
<td>9</td>
<td>LSVIVMMMDY</td>
<td>1.19</td>
<td>3.32</td>
<td>-2.48</td>
<td>2.49</td>
<td>-0.79</td>
<td>1899.8</td>
</tr>
<tr>
<td>A*2402</td>
<td>320–328</td>
<td>10</td>
<td>LWEWNASVRF</td>
<td>1.56</td>
<td>1.12</td>
<td>-2.47</td>
<td>2.69</td>
<td>0.22</td>
<td>291.8</td>
</tr>
<tr>
<td></td>
<td>375–383</td>
<td>9</td>
<td>LWDSASVRF</td>
<td>1.56</td>
<td>1.12</td>
<td>-2.47</td>
<td>2.69</td>
<td>0.22</td>
<td>291.8</td>
</tr>
<tr>
<td></td>
<td>374–382</td>
<td>9</td>
<td>FPPHPLPPIF</td>
<td>1.49</td>
<td>1.02</td>
<td>-2.34</td>
<td>2.51</td>
<td>0.17</td>
<td>216.5</td>
</tr>
<tr>
<td></td>
<td>234–243</td>
<td>10</td>
<td>GYRWNCRLRF</td>
<td>1.32</td>
<td>1.09</td>
<td>-2.54</td>
<td>2.42</td>
<td>-0.13</td>
<td>350.6</td>
</tr>
<tr>
<td></td>
<td>334–342</td>
<td>9</td>
<td>SWLSLVPFF</td>
<td>1.14</td>
<td>1.19</td>
<td>-2.35</td>
<td>2.2</td>
<td>-0.15</td>
<td>223.5</td>
</tr>
<tr>
<td></td>
<td>108–116</td>
<td>9</td>
<td>AGMQNVTIFP</td>
<td>1.37</td>
<td>1.18</td>
<td>-2.74</td>
<td>2.55</td>
<td>-0.19</td>
<td>551.1</td>
</tr>
<tr>
<td></td>
<td>375–383</td>
<td>9</td>
<td>FPPMPLLPIF</td>
<td>1.49</td>
<td>1.04</td>
<td>-2.76</td>
<td>2.53</td>
<td>-0.22</td>
<td>570.4</td>
</tr>
<tr>
<td></td>
<td>318–326</td>
<td>9</td>
<td>SWAFAKYLN</td>
<td>1.28</td>
<td>0.54</td>
<td>-2.11</td>
<td>1.82</td>
<td>-0.29</td>
<td>129.8</td>
</tr>
<tr>
<td></td>
<td>320–328</td>
<td>9</td>
<td>AFAYKLYN</td>
<td>1.44</td>
<td>0.5</td>
<td>-2.28</td>
<td>1.95</td>
<td>-0.33</td>
<td>189.3</td>
</tr>
<tr>
<td></td>
<td>353–362</td>
<td>10</td>
<td>WVLSSVQMDM</td>
<td>1.39</td>
<td>0.53</td>
<td>-2.28</td>
<td>1.92</td>
<td>-0.36</td>
<td>191.6</td>
</tr>
</tbody>
</table>

CD8+ T cell epitope predictions were performed using the NetMHCpan 2.2 bioinformatics tool. Peptides were ranked based on total scores that combine MHC I binding, predictions of proteasomal cleavage, and TAP translocations.
Ability to predict experimentally determined antiviral CD8+ T cell epitopes by using NetMHCpan

To further evaluate the performance of the algorithm, we used an alternative approach where we analyzed the ability of the NetMHCpan algorithm to predict CD8+ T cell epitopes that we had previously identified experimentally with traditional methods of T cell epitope mapping. Antiviral CD8 T cell epitopes were identified and characterized in Asian patient cohorts infected with DENV (n = 10) or who had cleared infections with HBV (n = 7) or SARS-CoV (n = 5). We used 15-mer peptide libraries spanning the entire proteome of DENV and HBV or selected proteins from SARS-CoV (3a and NP), as described previously (4, 15, 16). The peptide libraries were screened directly against ex vivo isolated PBMCs (DENV; Fig. 2A) or they were used to generate T cell lines that were subsequently screened for peptide specificity (HBV and SARS). Peptide-specific CD8+ T cell responses were detected by IFN-γ ELISPOT and positive responses confirmed by ICS for peptide-induced production of IFN-γ and CD107a (Fig. 2B). The HLA restriction of the identified responses was characterized by evaluating the capacity of HLA-matched EBV transformed B cell lines to present the peptide of interest to CD8+ T cells (Fig. 2C). The optimal epitope was determined by assessing the capacity of truncated variants of the 15-mer to elicit IFN-γ production by a specific T cell line (Fig. 2D). A total of 61 CD8+ T cell peptides were identified (32 peptides from DENV, 22 from HBV, and 7 from SARS). The HLA restriction was identified for 22 of these peptides, and the optimal epitope was further defined for 7 of these 15-mers.

For CD8+ T cell epitope predictions, the same peptide library sequences were input into the NetMHCpan 2.3 software, and 8- to 11-mer epitopes were predicted for each of the six patient alleles (Fig. 2E). CD8+ T cell epitopes are generally assumed to bind MHC molecules with an affinity, 500 nM (27–29). Thus, we included only peptides with predictive IC50 values below 1000 nM. The performance of the algorithm was then evaluated in terms of “coverage” (i.e., the percentage of experimentally defined responses that were successfully predicted by NetMHCpan) and “accuracy” (i.e., the comparison between the predicted and experimentally determined HLA restriction and minimal epitope sequence).

The “coverage” of the peptide predictions was assessed by analyzing the data in two different ways (strategies I and II, Fig. 2F, 2G). For strategy I, prediction results for each HLA molecule were merged and peptides were ranked according to IC50 values, regardless of their HLA restriction. The percentage of experimentally defined responses that were successfully predicted among the top 10, 50, 100, or 200 binders was evaluated for each patient. For the HBV and DENV patient cohorts, 100 predicted peptides per patient (with a total of 1000 and 700 peptides for the 10 DENV and 7 HBV patients, respectively) were needed to obtain at least 50% coverage of the experimentally identified peptides.
Because the comprehensive 15-mer peptide libraries consist of 313 HBV and 660 DENV peptides, this bioinformatics approach appears neither cost nor time effective. Predictions were more successful for the SARS cohort for whom 50 predicted peptides (10 per patient) would give 70% coverage of experimentally identified peptides from a library of 136 fifteen-mers. Nevertheless, 250 peptides would need to be screened for coverage of 80% (Fig. 3A, 3C).

Strategy I takes into account the competition that may exist between HLA molecules, but it implies that peptide libraries will need to be tailored individually for each patient, a scenario that is impractical for large patient cohorts. As an alternative approach, we assumed that each HLA molecule was equally contributing to peptide binding and predicted epitopes were ranked separately by IC50 values of individual HLA molecules (strategy II, Fig. 3B). Peptide libraries could therefore be designed for each HLA allele.
FIGURE 3. Evaluation of the coverage of NetMHCpan. Performance of the NetMHCpan 2.3 predictive algorithm was evaluated for strategies I (A) and II (B) in 10 DENV, 7 HBV, and 5 SARS patients. The percentage of coverage was determined as the percentage of experimentally identified peptides that were successfully predicted among the top 2, 5, 10, 50, or 200 binders for strategy I or among the top 12, 30, 60, or 180 binders for strategy II. The total number of predicted peptides that would need to be synthesized for each condition for the total patient cohort is summarized (C).

Table II. Experimental and predicted HLA restrictions of DENV, HBV, and SARS-CoV–specific CD8+ T cell peptides

<table>
<thead>
<tr>
<th>Virus</th>
<th>Peptide</th>
<th>aa</th>
<th>Sequence</th>
<th>Experimental HLA Restriction</th>
<th>Predicted Epitope</th>
<th>Predicted HLA Restriction</th>
<th>MHC IC50 (nM)</th>
<th>Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>DENV 1</td>
<td>Env 41</td>
<td>201–215</td>
<td>MENKAVLHVRQ0FWD</td>
<td>B*3701</td>
<td>KAVLHVRFQ</td>
<td>B*5701</td>
<td>15.61</td>
<td>0.89</td>
</tr>
<tr>
<td>DENV 2</td>
<td>NS2A 11</td>
<td>51–65</td>
<td>RDLGLRVWYGVATMT</td>
<td>A*0101</td>
<td>RVNVVGA</td>
<td>A*3001</td>
<td>59.09</td>
<td>0.65</td>
</tr>
<tr>
<td>DENV 4</td>
<td>NS3 41</td>
<td>201–215</td>
<td>KRYLPAIVREA1IKRG</td>
<td>A*1101</td>
<td>GTSGSPIIDK</td>
<td>A*1101</td>
<td>37.11</td>
<td>0.84</td>
</tr>
<tr>
<td>DENV 5</td>
<td>NS3 41</td>
<td>201–215</td>
<td>KRYLPAIVREA1IKRG</td>
<td>A*3101</td>
<td>RVGLPVR</td>
<td>A*2402</td>
<td>29.89</td>
<td>0.14</td>
</tr>
<tr>
<td>SARS-CoV 1</td>
<td>Core 1</td>
<td>6–16</td>
<td>IGLFQDSGDGQVSLYN</td>
<td>B*5801 (5701)</td>
<td>RGSNSHSHW</td>
<td>B*5801 (5701)</td>
<td>4.26</td>
<td>1.32</td>
</tr>
<tr>
<td>SARS-CoV 2</td>
<td>NP65</td>
<td>120–135</td>
<td>DQFTETSPSPTWLY</td>
<td>B*5801</td>
<td>SPDTSVSPPQPSW</td>
<td>B*5801</td>
<td>625.01</td>
<td>1.81</td>
</tr>
<tr>
<td>SARS-CoV 3</td>
<td>NP 65</td>
<td>120–135</td>
<td>DQFTETSPSPTWLY</td>
<td>B*5801</td>
<td>SPDTSVSPPQPSW</td>
<td>B*5801</td>
<td>625.01</td>
<td>1.81</td>
</tr>
</tbody>
</table>

Experimental and predicted HLA restrictions for a given peptide are shown in bold when identical.

*Novel epitopes.

and then combined to protect the patients’ unique HLA composition. The number of unique HLA alleles within the patient cohorts was 38 and 26 for DENV or HBV and SARS, respectively (Supplemental Table II). The percentage of coverage was then calculated for the top 2, 5, 10, 20, or 30 predicted epitopes per HLA (a total of 12, 30, 60, 120, or 180 epitopes per patient). Similar results were obtained for the HBV and DENV cohorts for whom 760 or 780 predicted peptides would need to be screened to obtain 50–70% coverage of experimentally identified peptides. Predictions were more successful for the SARS cohort as the top 130 predicted peptides (top 5 per HLA) provided 90% coverage.

HLA–peptide interactions rely on few amino acids critically positioned along the peptide, and disruption or addition of a single residue may interfere with peptide binding. An accurate bioinformatics prediction should therefore encompass both HLA restriction and minimal epitope sequence. Predicted HLA restrictions were compared with those defined experimentally for 22 CD8+ T cell peptides and were in large part identical (16/22 epitopes = 73%, indicated in bold in Table II). In contrast, prediction of the minimal epitope sequence was not as accurate. Three of seven predicted epitopes coincided with the minimal epitopes shown to experimentally elicit optimal IFN-γ production by specific CD8+ T cells (DENV NS3 27 and NS3 68, SARS NP 65, indicated in bold in Table III). These were also identical in terms of HLA restriction. Three epitopes differed in only one residue that was either missing from the C terminus (HBV core 1 and SARS NP 44) or was added at the N terminus (HBV Env 72). We have recently shown for SARS NP44 that, although the absence of the C-terminal l-residue (GETALALLL-L) in the predicted epitope did not affect HLA binding, it significantly impacted the ability of specific CD8+ T cells to produce IFN-γ (16). Predicted and experimental HBV Env 35 epitope differed for both sequence and HLA restriction, and thus identify two overlapping epitopes within the same 15-mer.

In conclusion, for the HBV and DENV cohort, CD8+ T cell epitope predictions by NetMHCpan did not provide a satisfac-
NetMHCpan 2.4 and ranked according to MHC IC50 values expressed in the Asian population. Peptides were predicted using the HBV proteome restricted to nine HLA molecules commonly derived CD8+ T cell responses were absent from the peptide profound. Regardless of the approach taken, five experimentally some predicted epitope sequences differed by only single amino available for the given HLA subtype (Fig. 4D, 4E).

Correlation between the accuracy of peptide predictions and the amount of training data available for a given HLA molecule

The accuracy of HLA-I/peptide predictions depends on sufficient experimental data being available for the HLA in question. This implies that the outcome of CD8+ T cell predictions would be different for populations that express well-characterized HLA subtypes (e.g., Caucasian individuals) compared with those that express HLA molecules for which there are little training data available (e.g., Asians). To experimentally test this hypothesis, we selected the top 20 predicted CD8+ T cell nine-mer epitopes from the HBV proteome restricted to nine HLA molecules commonly expressed in the Asian population. Peptides were predicted using NetMHCpan 2.4 and ranked according to MHC IC50 values (Supplemental Table III). The ability of the predicted peptides to bind to their restriction element was tested in a microtiter-based peptide-MHC binding assay (Fig. 4A). The number of peptides capable of binding the correspondent HLA molecule was assessed by defining the cutoff for positive HLA-peptide binding at 0, 10, or 20% above background levels (Fig. 4A, black, red, or blue dotted lines, respectively). The number of experimental data points used to train the NetMHCpan 2.4 algorithm differed significantly between the HLA molecules (data from http://www.cbs.dtu.dk/services/NetMHCpan/; Fig. 4B) and ranged from 0 (HLA-C*0304, HLA-C*0702, and HLA-C*0801) to almost 8000 for HLA-A*0201. In general, when all known human HLA molecules are classified into those that are found more frequently in Asian or white populations (defined as “Asian” or “Caucasian”; distributions obtained from the National Center for Biotechnology Information Web site), the number of data points available for Asian HLA molecules is directly correlated with the amount of training data available for the given HLA subtype (Fig. 4D, 4E).

Discussion

The relevance of T cell epitope identification for vaccine design has fueled the development of computational methods that allow us to economically and efficiently map pathogen-specific T cell responses. These methods have proved to be extremely powerful in mouse studies (8), but it is still unclear whether the same is true for human populations.

Several variables distinguish a natural human infection from the well-controlled scenario of inbred mice infected with a known dose and strain of the pathogen of interest. First, the variability of the infectious virus that is naturally infecting the patient is an unavoidable reality in human studies. The sequence of the infecting agent can often be inferred only from a “consensus sequence,” and mutations may impact on the predictive ability of bioinformatics tools. We have, for example, demonstrated that mutations in HBV genotypes alter the immunogenicity of HLA-A2 supertype epitopes in Asian HBV patients (15). Second, because humans experience frequent encounters with other pathogens, T cell responses to individual viruses will be strongly influenced by cross-reactive T cell recognition (30, 31). This does not occur in laboratory mice that are rigorously maintained in controlled pathogen-free environments. Third and most important, humans display high levels of genetic polymorphisms at the MHC level (but also in other genes involved in Ag presentation and T cell activation), in contrast with laboratory mice that are often homozygous at the MHC-I loci. Only a small proportion of the >7000 known polymorphisms of the MHC-I loci have been directly characterized in their peptide-binding specificities, and even fewer of these were tested with a large set of peptide binders. This information is still largely unavailable for HLA molecules that are common in less studied ethnic groups such as Southeast Asian populations that represent a significant proportion of the world population and are often those more in need of vaccines. The NetMHCpan algorithm represents an important attempt to overcome the limited HLA coverage of predictive algorithms. This method is trained on the largest database of binding data and allows the prediction of binding peptides for any HLA specificity based on the “nearest neighbor” prediction (10). Experimental validations of this method showed that 86% of peptide binding predictions could be experimentally confirmed, with peptide binding affinities <500 nM. To date, however, few studies have addressed whether this also translates to an efficient prediction of relevant human CD8+ T cell determinants, in particular, for Asian populations.

Our previous studies suggested that bioinformatics tools may be inefficient for the prediction of CD8+ T cell epitopes in the context of Asian populations (9). In this study, we further confirm our observations in two different experimental settings. First, we show that the top 10 predicted peptides from HBV E protein restricted in HLA-A*1101 and -A*2402 do not represent relevant CD8+ T cell epitopes in 29 patients with resolved HBV (Fig. 1C, 1D). This analysis may not seem exhaustive because it was performed using 1% of the top ranking peptides restricted to one or two HLA molecules out of the complete repertoire expressed by each subject. However, the previously described HBV A*2402-restricted

Table III. Experimental and predicted minimal epitope sequences of CD8+ T cell epitopes specific for DENV, HBV, and SARS-CoV

<table>
<thead>
<tr>
<th>Virus</th>
<th>Peptide</th>
<th>15-Mer Peptide</th>
<th>Epitope Predicted Epitope</th>
<th>Experimental HLA Restriction</th>
<th>Predicted HLA Restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>DENV 1</td>
<td>NS3 27</td>
<td>SPGTSGSPIIDKKGK</td>
<td>GTSGSPIIDKK GTSGSPIIDKK</td>
<td>A*1101</td>
<td>A*1101</td>
</tr>
<tr>
<td>DENV 2</td>
<td>NS3 68</td>
<td>REIEIPERSWMSGWHE</td>
<td>RSWMSGWHE RSWMSGWHE</td>
<td>B*5801</td>
<td>B*5801</td>
</tr>
<tr>
<td>HBV 3</td>
<td>Env 72</td>
<td>SVIVNWGNWGQPSKLNY</td>
<td>SVIVNWGNWY LSIVWSGNYW</td>
<td>B*5801</td>
<td>B*5801</td>
</tr>
<tr>
<td>HBV 4</td>
<td>Core 1</td>
<td>DIPIDPEKPEGASVEL</td>
<td>KEPASVEL KEKASVEL</td>
<td>A*0201</td>
<td>B*4001</td>
</tr>
<tr>
<td>HBV 5</td>
<td>Env 35</td>
<td>FLPGLLVLQQPGFFLL</td>
<td>FLPGLLVLQA VLGAGFFLL</td>
<td>Cw*0801</td>
<td>A*0201</td>
</tr>
<tr>
<td>SARS 6</td>
<td>NP 44</td>
<td>GERALALLLLDLRNLQ</td>
<td>GERALALLL GERALALLL</td>
<td>B*4001</td>
<td>B*4001</td>
</tr>
<tr>
<td>SARS 7</td>
<td>NP 65</td>
<td>IGMVFPTSGTWLT Y</td>
<td>MEVTFTSGTWL MEVTFTSGTWL</td>
<td>B*4001</td>
<td>B*4001</td>
</tr>
</tbody>
</table>

Experimental and predicted minimal epitopes and HLA restrictions are indicated in bold when identical.
FIGURE 4. The predictive power of HLA-peptide binding algorithms directly correlates with the amount of data points used to train the algorithm. (A) HLA-peptide binding assays for nine common Asian HLA molecules are shown. Results are expressed as relative absorbance. Peptides were ranked according to IC50 (1–20: from lowest to highest, gray bars). White bars represent KB; black bars represent NB. Background was set using NB values (black dotted line), and cutoff for positive HLA-peptide binding was set at 10 or 20% (red or blue dotted line, respectively) above the background value. (B) The number of experimental data points used to train the NetMHCpan 2.4 algorithm is shown for each HLA molecule (data from http://www.cbs.dtu.dk/services/NetMHCpan/). (C) The median number of data points is shown for HLA molecules that are found more frequently in Asian or Caucasian populations or in other ethnic groups (others). (D and E) Correlation between the number of experimental HLA-peptide binders from (A) and the number of data points used to train the predictive algorithm. The cutoff for HLA-peptide binding was set at either 10 (D) or 20% (E) above the background value. Linear regression analysis was performed using Prism.
epitope Pol 756 (22) elicited a T cell response in six of nine A*2402+ patients, confirming the presence of a detectable HBV-specific T cell memory response in at least some of these patients. Of note, the magnitude of the Pol 756-specific response was 5-fold higher than that of T cells specific for the predicted A*2402-restricted peptides 1, 2, and 5 (Fig. 1D).

In a second approach, we first identified virus-specific CD8+ T cell epitopes experimentally by using comprehensive 15-mer peptide libraries and subsequently addressed whether we were able to predict the same epitopes using NetMHCpan. A total of 22 Asian patients infected with DENV or with resolved HBV and SARS infections were analyzed. Results were similar for the HBV and DENV patients for whom the algorithm was not efficient in predicting the experimentally identified CD8+ T cell responses. The number of peptides that we would need to synthesize and screen to include these epitopes would be far larger than that of the comprehensive peptide libraries (Fig. 3A–C). This was true whether we took into account the competition that may occur between HLA molecules (Strategy I, Fig. 3A) or whether we assumed that each HLA molecule would equally contribute to peptide presentation (Strategy II, Fig. 3B). Predictions were significantly more efficient for the five SARS patients. The reason for this is not completely understood, although we note that the sample size was small (n = 5) and that the peptide library used for the analysis spans only two highly conserved SARS proteins (3a and NP), in contrast with the HBV and DENV libraries that span whole proteomes, which include highly variable proteins. It is thus possible that the DENV and HBV sequences used to construct the peptide library (and those used for the peptide predictions) do not always coincide with the sequence of the naturally infecting virus. This would imply that the genetic variability of the pathogen of interest plays an important role in the success of CD8+ T cell epitope predictions.

Based on our observations, we hypothesize that the poor predictive outcome of NetMHCpan in the context of the DENV and HBV cohorts may be because of insufficient experimental training data available for the HLA molecules in question. The number of experimental data points used to train the NetMHCpan algorithm differs significantly between HLA molecules ranging from 0 to almost 8000 for HLA-A*0201 for NetMHCpan 2.4 (training data differs significantly between HLA molecules ranging from 0 to 4018 ANTIVIRAL CD8+ T CELL EPITOPES IN ASIAN POPULATIONS [45x255]algorithm as it is continuously updated and improved as new experimental data become available. To address whether the differences in the number of data points used to train the NetMHCpan algorithm (version 2.4) could play a role in determining the accuracy of the peptide predictions, we experimentally tested the predictive efficiency (i.e., the number of binders out of the top 20 predicted peptides) for 9 common Asian HLA subtypes. We show that the number of peptide binders directly correlates with the amount of training data available for the given HLA subtype (Fig. 4D, 4E). Thus, our results suggest that the predictive power of algorithms may be limited in the context of Asian populations that express less characterized HLA molecules. Our findings have important implications for the immunomonitoring of Asian populations during infection and for the subsequent rational design of vaccines.

Acknowledgments
We thank Dr. Nattiya Kirankarn for kindly providing the patient samples of HLA-A*1101+ and -A*2402+ HBV resolved patients and Pei Yiang Lim for technical assistance.

Disclosures
The authors have no financial conflicts of interest.

References


