

# *HLA*restrictor—a tool for patient-specific predictions of HLA restriction elements and optimal epitopes within peptides

Malene Erup Larsen · Henrik Kloverpris · Anette Stryhn · Catherine K. Koofhethile · Stuart Sims · Thumbi Ndung'u · Philip Goulder · Søren Buus · Morten Nielsen

Received: 25 August 2010 / Accepted: 1 November 2010 / Published online: 16 November 2010  
© Springer-Verlag 2010

**Abstract** Traditionally, T cell epitope discovery requires considerable amounts of tedious, slow, and costly experimental work. During the last decade, prediction tools have emerged as essential tools allowing researchers to select a manageable list of epitope candidates to test from a larger peptide, protein, or even proteome. However, no current tools address the complexity caused by the highly polymorphic nature of the restricting HLA molecules, which effectively individualizes T cell responses. To fill this gap, we here present an easy-to-use prediction tool named *HLA*restrictor (<http://www.cbs.dtu.dk/services/HLArestrictor>), which is based on the highly versatile and accurate *NetMHCpan* predictor, which here has been optimized for the identification of both the MHC restriction element and the

corresponding minimal epitope of a T cell response in a given individual. As input, it requires high-resolution (i.e., 4-digit) HLA typing of the individual. *HLA*restrictor then predicts all 8–11mer peptide binders within one or more larger peptides and provides an overview of the predicted HLA restrictions and minimal epitopes. The method was tested on a large dataset of HIV IFN $\gamma$  ELISpot peptide responses and was shown to identify HLA restrictions and minimal epitopes for about 90% of the positive peptide/patient pairs while rejecting more than 95% of the negative peptide-HLA pairs. Furthermore, for 18 peptide/HLA tetramer validated responses, *HLA*restrictor in all cases predicted both the HLA restriction element and minimal epitope. Thus, *HLA*restrictor should be a valuable tool in any T cell epitope discovery

---

M. Erup Larsen · M. Nielsen (✉)  
Center for Biological Sequence Analysis, DTU Systems Biology,  
Building 208, Technical University of Denmark,  
DK-2800, Lyngby, Denmark  
e-mail: mniel@cbs.dtu.dk

M. Erup Larsen  
e-mail: mel@cbs.dtu.dk

H. Kloverpris · C. K. Koofhethile · S. Sims · P. Goulder  
Peter Medawar Building for Pathogen Research,  
University of Oxford,  
South Parks Road,  
OX1 3SY, Oxford, England, UK

H. Kloverpris  
e-mail: henrik.kloverpris@paediatrics.ox.ac.uk

C. K. Koofhethile  
e-mail: khei79@yahoo.com

S. Sims  
e-mail: stuart.sims@ndm.ox.ac.uk

P. Goulder  
e-mail: Philip.Goulder@paediatrics.ox.ac.uk

A. Stryhn · S. Buus  
Laboratory of Experimental Immunology,  
Faculty of Health Sciences, University of Copenhagen,  
Copenhagen, Denmark

A. Stryhn  
e-mail: astryhn@sund.ku.dk

S. Buus  
e-mail: sbuus@sund.ku.dk

T. Ndung'u  
HIV Pathogenesis Programme, Doris Duke Medical Research  
Institute, University of KwaZulu-Natal,  
Durban, South Africa  
e-mail: Ndungu@ukzn.ac.za

T. Ndung'u  
Ragon Institute of Massachusetts General Hospital,  
Massachusetts Institute of Technology and Harvard University,  
Charlestown, MA 02129, USA

process aimed at identifying new epitopes from infectious diseases and other disease models.

**Keywords** HLA restriction · Epitope prediction · MHC class I · Peptide binding · T cell epitope validation · HLA tetramer validation

## Introduction

CD8 positive cytotoxic T lymphocytes (CTLs) identify and eradicate host cells that have been infected with intracellular pathogens. They recognize protein antigen-derived peptides presented in complex with major histocompatibility complex (MHC) class I molecules. Prior to presentation, protein antigens are processed in a series of events beginning with the degradation of intracellular proteins by the proteasome (Rock et al. 2002), followed by transporter associated with antigen processing (TAP) mediated peptide translocation into the endoplasmic reticulum (ER) (Townsend and Trowsdale 1993; Uebel and Tampe 1999), N-terminal shortening of longer peptides by ER-resident amino peptidases (Serwold et al. 2002), and eventually some of the resulting peptides are specifically bound to MHC class I molecules. Once a stable peptide/MHC complex has been formed, it is transported via the Golgi apparatus to the cell surface ready for inspection by circulating CTLs.

Considering the many different peptides that can be generated, even from a small target protein, and the extensive polymorphism of the presenting MHC molecules, identifying pathogen-specific, HLA-restricted T cell epitopes can be an immense experimental task. However, only a few percent of a random collection of peptides can bind with sufficient affinity to a particular MHC molecule, making this event the most selective step in the entire pathway of antigen presentation (Yewdell and Bennink 1999) and a suitable starting point for T cell epitope discovery. Indeed, predictions of peptide/MHC interactions are widely used as an aid to identify T cell epitopes, and large efforts have been invested in developing accurate prediction method for peptide–MHC binding (reviewed in (Lundegaard et al. 2010)). *NetMHCpan* (Hoof et al. 2009; Nielsen et al. 2007) is one of the most precise publicly available predictors of peptide binding to MHC class I molecules (Lin et al. 2008; Zhang et al. 2009), and it has the added advantage that it is capable of addressing the binding specificity of any HLA molecule with known protein sequence. This is achieved by including both the peptide and MHC sequence as input hereby allowing the method to leverage information from multiple MHC specificities and extrapolate this information to MHC molecules for which limited or no binding data exist.

As described above, T cell epitope discovery involves the concurrent identification of stimulating antigenic pep-

ptide and their restricting HLA elements. In human populations, several thousand allelic HLA-A, -B, and -C variants have already been registered (IMGT/HLA database <http://www.ebi.ac.uk/imgt/hla/> (Robinson et al. 2001)). For any given human individual, a complete CTL epitope discovery effort would have to consider up to six (three loci with two heterozygous alleles each) different restricting HLA class I molecules. Fortunately, current DNA sequencing-based technology allows high-resolution (i.e., 4-digit, e.g., HLA-A\*0201) typing of all HLA-A, -B, and -C alleles of any given individual. Thus, information on all the HLA class I types needed to perform *NetMHCpan* predictions for any given individual can readily be provided. The other piece of information needed for *NetMHCpan* predictions is the input proteome, protein, or peptide. Whereas the number of HLA class I molecules can be limited to six, the number of peptides under consideration may be truly staggering; a problem that is compounded by the ability of HLA class I molecules to bind peptides of different length (note, *NetMHCpan* can handle peptides of 8–11 amino acids in length). To reduce this complexity, one could conveniently exploit a commonly used approach of T cell epitope discovery: testing pools of overlapping peptides (OLP) with a length of 15–18 amino acids in IFN $\gamma$  ELIspot or flow cytometry assays.

We here present a new online tool, *HLArestrictor*, aimed at identifying optimal peptides within one or several input peptides and corresponding HLA class I restriction elements targeted by CTL in given individuals. For a given individual, who has been fully typed for HLA-A, -B and -C, *HLArestrictor* is designed to identify all potential epitopes of length 8–11 amino acids that are predicted to bind to at least one of the individual's HLA restriction elements. A number of different sorting options are available for providing a user-friendly output.

We have benchmarked *HLArestrictor* with an HIV dataset of 5,145 18mer peptide IFN $\gamma$  ELIspot responses from 694 treatment-naive HIV-infected individuals and could successfully predict about 90% of the T cell epitopes. Using peptide/MHC class I tetramers, we furthermore demonstrated that *HLArestrictor* is able to correctly identify both the HLA restriction element and the optimal peptide length of a T cell epitope. The latter suggest that *HLArestrictor* could be an ideal design tool of HLA tetramer for the T cell epitope discovery.

## Materials and methods

### *HLArestrictor* features

Using *NetMHCpan* version 2.3, *HLArestrictor* predicts all peptide binders of length 8–11 amino acids, which are

relevant for a given patient's HLA types within an input peptide. A detailed overview of features of the *HLArestrictor* method is given in Table 1, and the corresponding input fields of the online *HLArestrictor* predictor are marked on Fig. 1. Multiple input peptides can be given in the same file using the FASTA format. There is no principal limit neither for the length of the input peptide nor for the number of HLA types to predict for. Thus, *HLArestrictor* is not limited to its intended use: patient-specific predictions of epitopes within a peptide or a peptide pool but can be applied to complete protein sequences or proteomes. It is however important to stress that all benchmark calculations performed in this work have been limited to data sets of short peptides of length 15–18 amino acids, and that the high specificity of the method most likely will be compromised if complete protein sequences or proteomes are supplied as input. Three different sort modes are available: (1) *HLA oriented*, which groups predicted peptide/HLA pairs by HLA type and sorts by prediction score (see Fig. 2); (2) *Peptide oriented*, which groups predicted peptide/HLA pairs by peptide and sorts by prediction score (see Fig. 3); and (3) *Descending prediction scores*, which sorts predicted peptide/HLA pairs by prediction score without any grouping (see Fig. 4). Each of the three sort modes is available as a standard version, where predictions are shown separately for each input peptide, and as a pool version, where predictions are shown together. This latter mode is useful when predicting responses for a larger peptide pools. Finally, a “no-sort” option is available for optimal computational speed. Three binding thresholds are defined and labeled; *Strong binder*, *Weak binder*, and *Non-binder*. A fourth label *Combined binder* is used to indicate predictions, which do only

qualify as binders if both %rank and affinity thresholds are considered as described below. Non-binders are shown up to a user-defined factor (default=2) of the weak binding threshold. Additionally, the user can define a maximum number of predictions to show per input peptide, with the exception that all predicted strong or weak binders will always be shown. Binding thresholds and sorting are based on either the predicted binding affinity (in nM) or a predicted %rank score, while both values are always shown in the output. The %rank score is defined as the rank score of a given candidate peptide relative to a set of one million random natural peptides for a given allele, such that a 2 %rank score means that only 2% of random peptides bind the allele with a predicted affinity stronger than the candidate epitope. If the default sort method *Rank OR Affinity* is chosen, a prediction need to fulfill either the affinity or %rank at the strong or weak binding thresholds to be labeled as *Strong binder* or *Weak binder*, respectively. This is a practical feature when the user wants to be alerted by predictions meeting either threshold and would consider subsequently testing all suggested epitopes. In contrast, if the option *Rank AND Affinity* is chosen, a prediction need to fulfill both the affinity and %rank at the strong or weak binding thresholds to be labeled as *Strong binder* or *Weak binder*, respectively.

Note that all benchmark calculations in this work were performed with *HLArestrictor-1.0*, using the *NetMHCpan-2.2* method for peptide–MHC binding predictions. The default version of *HLArestrictor* is 1.1, and this method uses *NetMHCpan-2.3*. This update has allowed the use of the new HLA nomenclature and most recent HLA allelic coverage. None of the conclusions drawn in this work have been altered by the update.

**Table 1** *HLArestrictor* features

Input parameter	Default value	Description	Letter in Fig. 1
Input/file	–	Input peptides in FASTA format/file	A
List of HLAs	HLA-A02:01	Comma separated list of HLA alleles	B
Peptide lengths	8–11	Length(s) of submers to extract from input peptide	C
Sort mode	HLA oriented	Sort mode to use for showing the output: <i>HLA oriented</i> , <i>peptide oriented</i> , or <i>Descending prediction scores</i> . Each mode is also available as pool version	D
Sort method	Percent	Specifies if sorting should be done using %rank or affinity values, and whether the binding criteria should be based on thresholds for %rank, affinity, %rank OR affinity, or %rank AND affinity	E
Strong threshold	0.5 %rank or 50 nM	Threshold for strong binding in %rank or affinity	F
Weak threshold	2 %rank or 500 nM	Threshold for weak binding in %rank or affinity	F
Number of predictions per peptide	Off	Optional user defined maximal number of predictions to show per peptide. Predictions below weak binding threshold will, however, always be shown	G
Non-binders factor	2	Show non-binders with a score of factor*weak binding threshold	H

## HLArestrictor Server

HLArestrictor is a tool for patient-specific predictions of HLA restriction elements and optimal epitopes within peptides or proteins. Given the high-resolution (i.e. 4-digit) HLA typing of the individual, HLArestrictor predicts all 8–11mer peptide-binders within one or more larger peptides and provides an overview of the predicted HLA restrictions and minimal epitopes.

**Note, the HLA-restrictor 1.1 has been updated to include the newest MHC allele releases from the IMGT/HLA and IPD-MHC databases (for non-human primates and pig). These updates include adaptation of the new nomenclature for HLA and Rhesus macaque (Mamu) alleles. The update also includes a shift from NetMHCpan-2.2 to the more recent NetMHCpan-2.3 method for peptide MHC binding predictions.**

The project is a collaboration between CBS, and [IMMI](#)

View the [version history](#) of this server. All the previous versions are available on line, for comparison and reference.

Instructions	Output format	Article abstract
--------------	---------------	------------------

**SUBMISSION**

Paste a single sequence or several sequences in **FASTA** format into the field below:

```
>N067_IGSEELRSLYNTVATLY
IGSEELRSLYNTVATLY
>N067_ELAENREILKEPVHGVVY
ELAENREILKEPVHGVVY
```

Submit a file in **FASTA** format directly from your local disk:  
 no file selected

Select allele subtype (species)

Select Allele (max 20 per submission)

- HLA-C\*16:01
- HLA-C\*16:02
- HLA-C\*16:04
- HLA-C\*16:06
- HLA-C\*16:07
- HLA-C\*16:08
- HLA-C\*16:09
- HLA-C\*16:10
- HLA-C\*16:11
- HLA-C\*16:12
- HLA-C\*16:13
- HLA-C\*16:14

or type host allele names (ie HLA-A0101 or A0101) separated by commas.

For list of allowed allele names click here [List of MHC allele names.](#)

Peptide length

Sort mode:

Sort method

Sort predictions based on %rank score. Prediction are labeled **Weak binder** and **Strong binder** according to the thresholds defined below. Peptides are labeled **Combined binder** if %rank is less than the %rank weak binders threshold and the IC50 value (if defined) is stronger than the affinity weak binding threshold.

Threshold for strong binder (% Rank)  Threshold for strong binder (IC50)

Threshold for weak binder (% Rank)  Threshold for weak binder (IC50)

Number of predictions to show per peptide (0:report down to 2 x weak-binding cutoff.

Show non-binders with a score of factor\*weak binding threshold, 1 means no non-binders are shown.

**Fig. 1** Screenshot of *HLArestrictor*. The input fields are marked in *blue letters* and described in Table 1 and in the text. Note that multiple FASTA sequences can be given as input (see *A*)

### HIV benchmark set

We used a cohort of 1,000 antiretroviral naive HIV-infected adults, of whom 864 were recruited from Durban, KwaZulu-Natal, South Africa (Kiepiela et al. 2007) and 136 recruited from Thames Valley, Oxfordshire, England. Informed consent was obtained from all participating individuals and institutional review boards at the University of KwaZulu-Natal, Massachusetts General Hospital, and the University of Oxford approved the study. Four-digit high-

resolution typing of HLA-A, HLA-B, and HLA-C alleles was performed using the Dynal RELTIM reverse sequence-specific oligonucleotide kits as previously described (Kiepiela et al. 2007). Patients (694) were successfully typed on all alleles and were included in the benchmark set.

### IFN $\gamma$ ELISpot

Comprehensive IFN $\gamma$  ELISpot responses to a set of 420 overlapping peptides (OLPs) based on the 2004 C-clade

```
# HLArestrictor with NetMHCpan version 2.3
# HLA types used: HLA-A02:01, HLA-A02:05, HLA-B51:01, HLA-B58:01, HLA-C07:01, HLA-C16:02
# Peptide lengths: 8, 9, 10, 11
# Sort-method: OR. Sort-mode: HLA-oriented
# %rank threshold for strong binding peptides: 0.5%rank
# %rank threshold for weak binding peptides: 2.0%rank
# Affinity threshold for strong binding peptides: 50.0nM
# Affinity threshold for weak binding peptides: 500.0nM
# Number of predictions per peptide: Not specified
# Non-binders shown up to a prediction score of 2.0*(weak binding threshold)
```

Input options

Results for Peptide N067\_TGSEELRSLYNTVATLY: TGSEELRSLYNTVATLY

Pos	Length	Peptide	HLA	1-log50k(aff)	Affinity(nM)	%Rank	Label	Estimated accuracy
8	9	SLYNTVATL	HLA-A02:01	0.43	475	4.0	Combined binder	0.853
8	8	SLYNTVAT	HLA-A02:01	0.363	985	5.0	Non-binder	0.853
7	11	RSLYNTVATLY	HLA-B58:01	0.601	75	0.4	Strong binder	0.853
7	10	RSLYNTVATL	HLA-B58:01	0.512	196	0.8	Weak binder	0.853
7	10	RSLYNTVATL	HLA-C07:01	0.476	289	0.1	Strong binder	0.486
7	11	RSLYNTVATLY	HLA-C07:01	0.418	544	0.25	Strong binder	0.486
7	8	RSLYNTVA	HLA-C07:01	0.304	1867	1.5	Weak binder	0.486
7	9	RSLYNTVAT	HLA-C07:01	0.268	2747	3.0	Non-binder	0.486
7	11	RSLYNTVATLY	HLA-C16:02	0.112	NA	0.8	Weak binder	0.564
7	10	RSLYNTVATL	HLA-C16:02	0.106	NA	0.8	Weak binder	0.564

Results for Peptide N067\_ELAENREILKEPVHGVVY: ELAENREILKEPVHGVVY

Pos	Length	Peptide	HLA	1-log50k(aff)	Affinity(nM)	%Rank	Label	Estimated accuracy
8	9	ILKEPVHGV	HLA-A02:01	0.63	55	1.5	Weak binder	0.853
8	9	ILKEPVHGV	HLA-A02:05	0.566	109	2.0	Weak binder	0.833

**Fig. 2** HLArestrictor example output using sort-mode *hla-oriented*, by which the predictions are grouped by HLA-type, then ranked by prediction score. In this example, all predictions for HLA-C\*07:01 are listed in the second group marked (A). Ranking within the group is

marked with (B). Non-binders (C) are shown if their %rank score is less than two times the threshold for weak binding (2 %rank). All threshold values are user-defined

consensus were used in a matrix system with 11–12 peptides in each pool. Responses to matrix pools were subsequently confirmed by stimulating with individual peptides as previously described (Kiepiela et al. 2004). A total of 5,145 ELISpot responses to a total of 294 different 18mer peptides were observed when measured in the 694 treatment-naive HIV-infected individuals.

### Peripheral blood mononuclear cell isolation and generation of CTL lines

Purified peripheral blood mononuclear cells (PBMCs) were isolated from whole blood of HIV-infected individuals by Ficoll-Hypaque density gradient centrifugation and cryopreserved until used for tetramer staining and/or generation of peptide-specific CTL lines. Specific CD8+ T cell lines, also described as “CTL” lines, were generated from culturing of HLA class I peptide tetramer purified antigen-specific cells derived from HIV-infected human PBMCs as previously described (Payne et al. 2010). Briefly, tetramer-enriched

CD8+ T cell lines were cultured by sorting tetramer-positive cells from PBMCs by magnetic separation using anti-PE microbeads (Miltenyi Biotec). The flowthrough fraction was peptide pulsed (20 µg/ml) for 1 h at 37°C, irradiated (30 Gy), washed once in PBS, and cultured with the tetramer-positive fraction at a ratio of 100:1 in H10-10 medium (RMP1) medium supplemented with 10% human serum, 10% natural T cell growth factor [TCGF; Helvetica Healthcare], 2 mM L-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin. Every 10 days following the initial setup, the tetramer-positive cells were fed by using peptide-pulsed, irradiated, autologous PBMCs or BCLs at a ratio of 1:1 and irradiated feeder PBMCs from three healthy HIV-seronegative donors.

### Peptide/MHC class I tetramer synthesis and tetramer staining

Tetramers were generated in two different ways. The first method was described by (Altman et al. 1996). Briefly, HLA-B\*4201 heavy chain (HC) was expressed in Rosetta

```
# HLArestrictor with NetMHCpan version 2.3
# HLA types used: HLA-A02:01, HLA-A02:05, HLA-B51:01, HLA-B58:01, HLA-C07:01, HLA-C16:02
# Peptide lengths: 8, 9, 10, 11
# Sort-method: OR. Sort-mode: Peptide-oriented
# %rank threshold for strong binding peptides: 0.5%rank
# %rank threshold for weak binding peptides: 2.0%rank
# Affinity threshold for strong binding peptides: 50.0nM
# Affinity threshold for weak binding peptides: 500.0nM
# Number of predictions per peptide: Not specified
# Non-binders shown up to a prediction score of 2.0*(weak binding threshold)
```

} Input options

Results for Peptide N067\_TGSEELRSLYNTVATLY: TGSEELRSLYNTVATLY

Pos	Length	Peptide	HLA	1-log50k(aff)	Affinity(nM)	%Rank	Label	Estimated accuracy
7	10	RSLYNTVATL	HLA-C07:01	0.476	289	0.1	Strong binder	0.486
7	10	RSLYNTVATL	HLA-B58:01	0.112	196	0.8	Weak binder	0.853
7	10	RSLYNTVATL	HLA-C16:02	0.106	NA	0.8	Weak binder	0.564
7	11	RSLYNTVATLY	HLA-C07:01	0.418	544	0.25	Strong binder	0.486
7	11	RSLYNTVATLY	HLA-B58:01	0.601	75	0.4	Strong binder	0.853
7	11	RSLYNTVATLY	HLA-C16:02	0.112	NA	0.8	Weak binder	0.564
7	8	RSLYNTVA	HLA-C07:01	0.304	1867	1.5	Weak binder	0.486
7	9	RSLYNTVAT	HLA-C07:01	0.268	2747	3.0	Non-binder	0.486
8	9	SLYNTVATL	HLA-A02:01	0.43	475	4.0	Combined binder	0.853
8	8	SLYNTVAT	HLA-A02:01	0.363	985	5.0	Non-binder	0.853

Results for Peptide N067\_ELAENREILKEPVHGVVY: ELAENREILKEPVHGVVY

Pos	Length	Peptide	HLA	1-log50k(aff)	Affinity(nM)	%Rank	Label	Estimated accuracy
8	9	ILKEPVHGV	HLA-A02:01	0.63	55	1.5	Weak binder	0.853
8	9	ILKEPVHGV	HLA-A02:05	0.566	109	2.0	Weak binder	0.833

**Fig. 3** *HLArestrictor* example output using sort-mode *peptide oriented*, by which the predictions are ordered such that multiple predictions of the same sub-peptide are grouped together, then ranked according to prediction score. In this example, all predictions for

peptide RSLYNTVATL are listed in the group marked (A). Ranking within the group is marked with (B). Ranking between groups is marked with (C) and is based on the best scoring values marked with blue bullets

(DE3)pLysS (Novagen), purified, and refolded around the peptide of interest in the presence of human  $\beta$ 2M light chain. Unrefolded HC and peptide were separated from refolded peptide/MHC monomer complexes using FPLC prior to tetramerization of monomers and conjugation to R-phycoerythrin (Extravidin PE, Sigma) to obtain PE-labeled HLA-B\*4201 tetramers. The second method was recently described by Leisner et al. (2008). Briefly, HLA class I heavy chain were expressed in *Escherichia coli* BL21 (DE3), which had been co-transfected with a vector encoding the BirA holoenzyme, leading to the expression of biotinylated HLA class I heavy chain when the proteins were induced in the presence of biotin. Pre-oxidized, pre-biotinylated isomers were purified by column chromatography and stored at  $-20^{\circ}\text{C}$  until use. Peptide/HLA monomers were made by incubating these highly active isomers in a refolding buffer with excess b2m and peptide. Peptide/HLA tetramers were then generated by the addition of PE-labeled streptavidin. PBMCs or CTL lines were thawed and stained with PE-conjugated tetramer for 20 min, then washed

and stained with the following extracellular antibodies, CD3 Alexa Flour 700 (BD) or CD3 Pacific Orange (Invitrogen), CD8 Qdot 605 (invitrogen) or CD8 Alexa Flour 750 (eBioscience), and Live/Dead marker Violet (Invitrogen) for another 20 min. Cells were washed, fixed, and samples were acquired within 24 h on a BD LSR II flow cytometer. Cells were gated on singlets, lymphocytes, live cells, CD3, and then evaluated for CD8<sup>+</sup> T cells binding the peptide/MHC tetramer. Data were analyzed using FlowJo version 8.8.6.

## Results

### Benchmarking *HLArestrictor* on HIV data

The performance of the *HLArestrictor* at different %rank thresholds was benchmarked on a HIV dataset of 5,145 18mer peptide IFN $\gamma$  ELISpot responses representing a total of 294 different 18mer peptides being recognized in one or more of 694 treatment-naive HIV-infected individuals. To

```
# HLArestrictor with NetMHCpan version 2.3
# HLA types used: HLA-A02:01, HLA-A02:05, HLA-B51:01, HLA-B58:01, HLA-C07:01, HLA-C16:02
# Peptide lengths: 8, 9, 10, 11
# Sort-method: OR. Sort-mode: Descending prediction scores
# %rank threshold for strong binding peptides: 0.5%rank
# %rank threshold for weak binding peptides: 2.0%rank
# Affinity threshold for strong binding peptides: 50.0nM
# Affinity threshold for weak binding peptides: 500.0nM
# Number of predictions per peptide: Not specified
# Non-binders shown up to a prediction score of 2.0*(weak binding threshold)
```

Results for Peptide N067\_TGSEELRSLYNTVATLY: TGSEELRSLYNTVATLY

Pos	Length	Peptide	HLA	1-log50k(aff)	Affinity(nM)	%Rank	Label	Estimated accuracy
7	10	RSLYNTVATL	HLA-C07:01	0.476	289	0.1	Strong binder	0.486
7	11	RSLYNTVATLY	HLA-C07:01	0.418	544	0.25	Strong binder	0.486
7	11	RSLYNTVATLY	HLA-B58:01	0.601	75	0.4	Strong binder	0.853
7	10	RSLYNTVATL	HLA-B58:01	0.512	196	0.8	Weak binder	0.853
7	10	RSLYNTVATL	HLA-C16:02	0.106	NA	0.8	Weak binder	0.564
7	11	RSLYNTVATLY	HLA-C16:02	0.112	NA	0.8	Weak binder	0.564
7	8	RSLYNTVA	HLA-C07:01	0.304	1867	1.5	Weak binder	0.486
7	9	RSLYNTVAT	HLA-C07:01	0.268	2747	3.0	Non-binder	0.486
8	9	SLYNTVATL	HLA-A02:01	0.43	475	4.0	Combined binder	0.853
8	8	SLYNTVAT	HLA-A02:01	0.363	985	5.0	Non-binder	0.853

Results for Peptide N067\_ELAENREILKEPVHGVY: ELAENREILKEPVHGVY

Pos	Length	Peptide	HLA	1-log50k(aff)	Affinity(nM)	%Rank	Label	Estimated accuracy
8	9	ILKEPVHGV	HLA-A02:01	0.63	55	1.5	Weak binder	0.853
8	9	ILKEPVHGV	HLA-A02:05	0.566	109	2.0	Weak binder	0.833

**Fig. 4** *HLArestrictor* example output using sort-mode *descending prediction score*, by which the predictions are ordered solely by their prediction score, in this case, their %rank score. Note that sorting by

affinity is different than sorting by %rank, since the affinity distribution of ranked peptides is specific for each allele

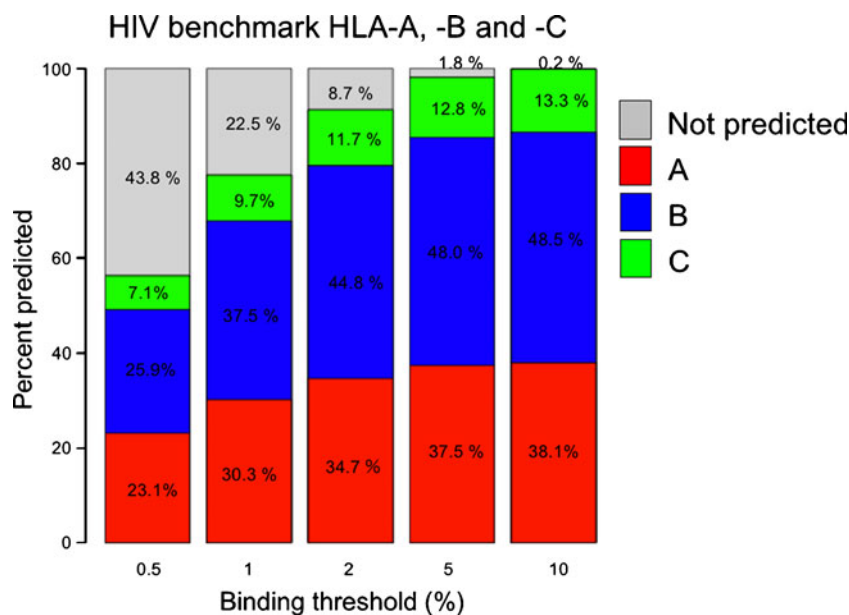
calculate the fraction of validated responses, which could be predicted at a given %rank threshold, predictions were carried out for each validated patient/peptide response for all six patient HLA molecules typed (Fig. 5). A response was considered correctly predicted if at least one of the patients' HLA types was predicted to bind an 8, 9, 10 or 11mer (a “submer”) within the 18mer peptide with a binding strength stronger than the given threshold. At a 2 %rank threshold for instance, the method suggested at least one such epitope for 91% of the 5,145 responses observed. With an average of 5.3 predicted epitopes per response, *HLArestrictor* thus rejected 98% of the negative peptide-HLA combinations. Likewise, at a 2 %rank threshold, the method suggested at least one epitope for 78% of the responses, with an average of 3.5 predicted epitopes per response. Furthermore, the figure shows the fraction of responses with predicted HLA-A, HLA-B, and HLA-C restrictions at each threshold. At the 2 %rank threshold, approximately 50% of the responses were predicted to be HLA-B restricted.

HLA restriction identification by association studies

We next investigated what prediction threshold should be applied to give the highest predictive performance of the

*HLArestrictor* method. As illustrated in Fig. 5, choosing a relative high %rank prediction score naturally led to a higher sensitivity of the predictions; however, this came at the price of a higher number of potential false-positive predictions that in real life would have to be analyzed in subsequent immuno-assay validations. To address the question of which prediction threshold would be optimal, we carried out a simple computational experiment. A commonly used method for assigning HLA restriction to immunogenic peptides is HLA association studies. In these studies, the HLA restriction of a given immunogenic peptide is assigned based on prevalence of an HLA allele in a large patient cohort that responds positive to the peptide. A set of 85 (35 HLA-A and 50 HLA-B) HLA-peptide associations was identified using a Fisher's exact test-based analysis that corrects for multiple comparisons. Briefly, a two-by-two contingency table is constructed for each peptide-HLA pair. *P* values are then computed using Fisher's exact test for each table, and exact *q* values (Storey and Tibshirani 2003) are computed by summing over the null space of all observed contingency tables, as previously described (Carlson et al. 2009). All associations had *P* values less than 0.05. We next applied the *HLArestrictor* method to validate these associations. We identified the patients in the HIV cohort data set expressing the HLA

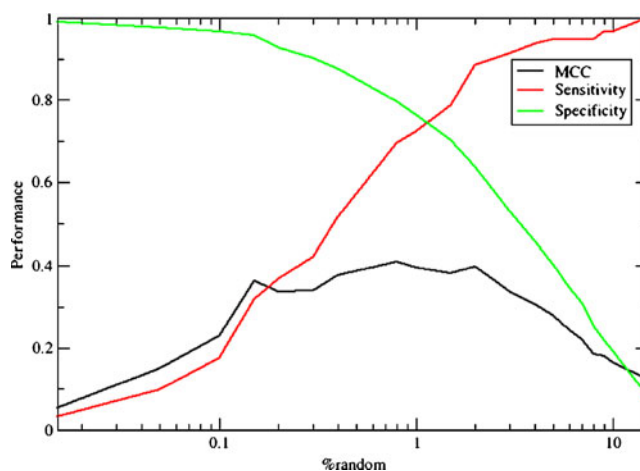
**Fig. 5** Benchmark on HIV peptide IFN $\gamma$  ELISpot responses. The *barplots* are divided into *colors* indicating the number of cases in which a HLA-A, -B, or -C type is predicted with the lowest %rank score



allele in question and that had responded positively to the given peptide. Next, a predicted binding value was assigned for the peptide to each of the patients' HLA types as the strongest %rank score for all 8–11submers contained within the peptide. All peptide/HLA pairs matching the restriction element identified from the association studies were taken as positive, and all other suggested restriction elements as negative. This led to a set consisting of 1,067 positive and 5,115 negative data points (on average, each peptide was tested in 12.5 patients expressing the given HLA allele). The predictive performance of *HLArestrictor* was finally evaluated in terms of the Matthews correlation coefficient (MCC), sensitivity, and specificity for different values of the %rank threshold. The results of this analysis are shown in Fig. 6 and demonstrate that the HLA restriction method achieved its highest predictive performance in terms of MCC for %rank threshold values in the range 0.5–2. If reducing the screening load is essential even at the expense of missing some of the epitopes (i.e., high specificity is a requirement and the concurrent loss in sensitivity is acceptable), then a threshold of 0.5 %rank is recommended. Here, the specificity is about 90%, and the sensitivity is about 50%. If, on the other hand, identifying as many epitopes as possible is essential even at the expense of an increased screening load (i.e., high sensitivity is a requirement and the concurrent loss in specificity is acceptable), then a threshold of 2 %rank is suggested. Here, the sensitivity is close to 90%, and the specificity is about 50%.

The Matthews correlation coefficient showed a highly significant ( $P < 0.005$ , exact permutation test) correlation between the physiological analysis of T cell responses and the prediction of the biochemical analysis of peptide/HLA interaction. In fact, we found that 73 of the 85 peptides (86%)

contained an 8–11submer peptide predicted to bind to the restriction element proposed by the association studies with a binding strength stronger than or equal to 2 %rank. In these cases, the two methods thus agree on the assignment of the HLA restriction element. However, for 12 of the peptides, the *HLArestrictor* method failed to confirm the proposed HLA restriction using the suggested 2 %rank threshold. For these 12 peptides, we identified the patients with the proposed HLA allele that had responded positively to the peptide and analyzed whether any of the other HLA alleles of these patients would predict alternative HLA restriction elements. This analysis allowed us to suggest alternative HLA restrictions for the majority of the positive patient/peptide pairs using the default threshold value of 2 %rank (see Table 2). For instance, the two peptides GKKHYMLKHLVWASREL and



**Fig. 6** The predictive performance of *HLArestrictor* evaluated in terms of the Matthews correlation coefficient (MCC), sensitivity, and specificity for different values of the %rank threshold



**Table 2** Alternative HLA restrictions

Allele	Peptide	<i>P</i> value	%rank	No. of patients	Fraction predicted
A01:01	<b>EVGFPVRPQVPLRPMTFK</b>	2.79E-03	15	26	1.00
A01:01	EWEFVNRPPVLKWLWYQL	6.78E-03	4	12	1.00
B35:01	DEALLQAVRIIKILY	7.41E-09	3	10	1.00
A36:01	<b>GKKHYMLKHLVWASREL</b>	3.92E-02	15	6	1.00
B39:10	PPIVAKEIVASCDKCQLK	2.29E-03	15	3	0.67
A32:01	FRLPIQKETWETWWTDYW	2.35E-04	3	3	0.67
B53:01	AGRWPVKVIHTDNGSNF	3.79E-49	4	14	0.64
A68:01	FWEVQLGIPHPAGLKKKK	7.13E-26	5	11	0.64
B81:01	PPIVAKEIVASCDKCQLK	2.16E-32	9	21	0.62
B39:10	EVNIVTDSQYALGII	4.94E-20	3	10	0.60
A03:01	LVSIVGGQIKEALL	1.57E-02	3	4	0.50
A11:01	IKKDKSTKWRKLVDFREL	2.46E-03	3	2	0.50

Allele gives the restriction element predicted from the large cohort population studies. *P* value gives the association study *P* value for the HLA restriction prediction. %rank gives the predicted binding score in terms of the %rank score for the peptide to the proposed HLA restricting allele. No. of patients gives the number of patients in our study matching the proposed HLA restriction and responding to the given peptide, and fraction predicted gives the fraction of the responding HLA-matched patients having alternative HLA restrictions identified by *HLArestrictor* with a %rank score less than or equal to 2%. In bold are highlighted the two examples explained in the text

EVGFPVRPQVPLRPMTFK were by association studies predicted to be restricted by the alleles HLA-A\*36:01 and HLA-A\*01:01, respectively. The predictions by the *HLArestrictor* indicated that these restrictions were highly unlikely with a %rank score of 15. Investigating the HLA types of each patients responding to these peptide and having the allele HLA-A\*36:01 or HLA-A\*01:01, we find in each case that the response could be explained in terms of an alternative HLA restriction with a prediction score below the suggested 2 %rank threshold. Note that HLA-A\*36:01 is in linkage disequilibrium with HLA-B\*53:01 ( $P=9.75E-08$ ), and five of the seven patients responding to the GKKHYMLKHLVWASREL peptide shared this allele. The peptide YMLKHLVW is predicted to bind HLA-B\*53:01 with a %rank score of 0.8, strongly suggesting that this HLA is a dominant restriction element for this peptide response. Likewise, 21 of the 26 patients responding to the EVGFPVRPQVPLRPMTFK peptide shared both the HLA-A01:01 and HLA-B\*81:01 alleles. The peptide FPVRPQVPL is predicted to bind the HLA-B\*81:01 allele with a %rank score of 0.01, strongly suggesting that this HLA is a dominant restriction element for this peptide response. Note that these suggested alternative HLA restrictions are based on predictions only, and that future experimental validation is needed to confirm both the HLA restriction and optimal epitope sequence.

Validation of CD8<sup>+</sup> T cell responses using peptide/MHC class I tetramers

To validate the optimal peptide and the corresponding HLA restriction element, we used a panel of 18 peptide/MHC

class I tetramers across eight different HLA alleles in ten HIV-infected individuals as shown in Table 3 and Fig. 7. The optimal epitope within the 18mer was selected based on previously described epitopes combined with information about the binding motif of the restriction element of interest. All 18 epitopes are listed in the Los Alamos HIV molecular immunology database (<http://www.hiv.lanl.gov/content/immunology>) (Korber et al. 2009). To confirm the optimal epitope and HLA class I restriction element, the corresponding peptide/MHC class I tetramer was produced and used to stain PBMCs from HIV-infected IFN $\gamma$  ELISpot responders or in vitro expanded CTLs as described in “Materials and methods”. In 16 of 18 cases, *HLArestrictor* successfully predicted the HLA restriction element and the minimal epitope with a %rank score below 2. In two cases, *HLArestrictor* failed to predict the validated HLA restriction and minimal epitope with a %rank score of 2. However, in the first case of VKVIEEKAF/HLA-B\*15:03, the predicted affinity was 155 nM and the predicted %rank score was 6.0. In the second case of SLYNTVATL/HLA-A\*02:01, the predicted affinity was 436 nM, and the predicted %rank score was 4.0. Thus, both epitopes were predicted to bind stronger than the commonly used threshold of 500 nM (Lundegaard et al. 2007; Moutaftsi et al. 2006; Sette et al. 1994). Further, the previously described HLA-B\*58 restricted epitope RSLYNTVATLY is predicted to bind HLA-B\*58:01 with an affinity of 62 nM and a 0.4 %rank. This example based on the peptide TGSEELRSLYNTVATLY, illustrates how the *HLArestrictor* often will predict multiple restrictions, in this case both correct, within a given positive peptide. Additional known epitopes not tested in our patients were predicted in several other cases as well. For example, within

**Table 3** Tetramer validations on selected patients as exemplified in Fig. 7

Patient	HIV 18mer with ELISpot response	Validated epitope	Validated allele	Pred. affinity (nM)	Pred. %rank
N080	PRTLNAW <b>VKVIEEKAF</b>	VKVIEEKAF	B15:03	155	6.0
N080	YHCLVC <b>FQTKGLGISY</b> GR	FQTKGLGISY	B15:03	8	0.8
N080	VKAACWWAG <b>IQQEFGIPY</b>	IQQEFGIPY	B15:03	4	0.1
N080	AVFIHNF <b>KRKGGIGGYS</b> A	KRKGGIGGY	B15:03	24	1.5
H044	WVKVIEE <b>KAFSPEVPMF</b>	KAFSPEVPMF	B57:03	NA	0.4
N021	ELKQEA <b>VRHFPRPWLHGL</b>	FPRPWLHGL	B42:01	49	0.05
N012	ACQGVG <b>GPSHKARVLA</b> EA	GPSHKARV <sup>a</sup>	B07:02	36	0.3
R050/N012	CRAIRN <b>IPRRIRQGL</b>	IPRRIRQGL <sup>a</sup>	B07:02	10	0.1
R050	NY <b>TPGPGVRYPL</b> TFGWCF	TPGPGVRYPL	B07:02	45	0.3
R050	QGWKGS <b>SPAIQSSM</b> TKIL	SPAIQSSM	B07:02	10	0.1
R050/R039	WVKVIEE <b>KAFSPEVPMF</b>	KAFSPEVPMF	B57:01	67	0.1
R039	PVGEIY <b>KRWIILGLNK</b> IV	KRWIILGLNK	B27:05	22	0.05
R039	AVFIHNF <b>KRKGGIGGYS</b> A	KRKGGIGGY	B27:05	289	1.0
R035	ELKNEA <b>VRHFPRIW</b> LHSL	VRHFPRIW	B27:05	357	1.0
R014	MASEFN <b>LPPIVAKEI</b> VA	LPPIVAKEI	B42:01	NA	1.5
N067	TGSEELR <b>SLYNTVATL</b> Y	SLYNTVATL	A02:01	436	4.0
N067/N096	ELAENRE <b>ILKEPVHGV</b> YY	ILKEPVHGV	A02:01	42	1.5
N096	SDIAGT <b>TSTLQEQIA</b> WM	TSTLQEQIAW	B58:01	32	0.2

The validated epitope within each 18mer is underlined. The predicted affinity is listed when available and the predicted %rank score is marked in bold if it is below or equal to a 2 %rank score threshold. 16 of 18 peptide/MHC class I tetramer-validated CD8+ T cell epitopes are successfully predicted by *HLArestrictor* at this threshold setting, while all 18 are predicted either below or equal to 2 %rank score OR below 500 nM

<sup>a</sup> In vitro expanded CTL line used for tetramer staining

the 18mer WVKVIEEKAFSPEVPMF, the known epitopes B\*45:01-EEKAFSPEV and B\*57:03-KAFSPEVI were predicted to bind at 0.3 %rank and 0.1 %rank, respectively.

## Discussion

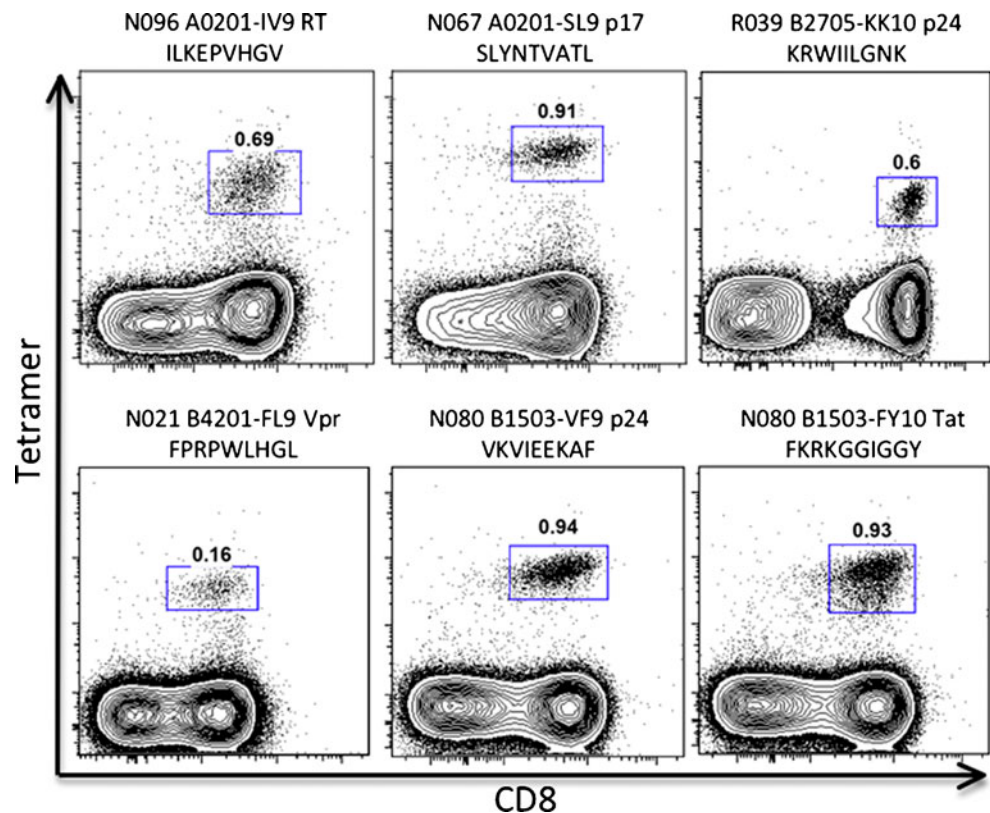
T cell epitopes consist of antigen-derived peptides presented in the context of HLA molecules; and the identification and validation of peptide/HLA complexes, which can stimulate T cell responses, is at the heart of any T cell epitope discovery process. Finding the stimulatory peptide and the presenting HLA restriction element is not a simple task. Here, we present an immunoinformatics method, *HLArestrictor*, which has been tailored to support T cell epitope discovery in individual subjects. As inputs, it needs the amino acid sequence of the target protein(s), and the HLA type of the individual in question (high-resolution HLA typing, e.g., HLA-A\*01:01, and preferably for all relevant loci, e.g., for HLA-A, -B, -C for HLA class I restricted CTL responses). Using these inputs, *HLArestrictor* creates all possible 8, 9, 10, and 11mer peptides from the target protein(s), predicts their binding to all the HLA molecules in question, and generates an output file consisting of the most likely peptide/HLA combination(s). Peptide/HLA tetramers is one of the most efficient means to validate

T cell epitopes, and *HLArestrictor* can also be viewed as a tool for efficient design of specific peptide/HLA tetramers.

We have successfully applied *HLArestrictor* to the search for patient-specific HLA restriction elements and optimal epitopes. To this end, we have re-analyzed a major study of T cell epitopes within the 2004 C-clade consensus HIV sequence (Kiepiela et al. 2007). In this study, the consensus sequence was represented by 420 overlapping 18mer peptides and tested in 694 treatment-naive HIV-infected individuals, which had been high-resolution typed for HLA-A, -B, and -C. A large set of HLA restrictions were identified from these data by association studies. Initially, we asked which %rank threshold should be used to generate T cell epitope predictions. At the 2 %rank and 1 %rank thresholds, the *HLArestrictor* method suggested HLA restriction elements and optimal peptides for 91% and 78%, respectively, of the 5,145 identified HIV-specific IFN $\gamma$  ELISpot peptide responses.

Next, we asked to what degree known T cell epitopes could be successfully predicted. A large set of HLA restrictions had been identified by association studies and could conveniently be used to further validate the predictive performance of the *HLArestrictor* method. Using this benchmark data set, the *HLArestrictor* was shown to achieve its optimal predictive performance for %rank score thresholds in the range from 0.5 to 2.0. At the 0.5 %rank

**Fig. 7** Examples of peptide/MHC class I tetramer stainings used to validate optimal epitopes and HLA restriction of CD8<sup>+</sup> T cell responses in HIV-infected individuals shown in Table 2. Cells are gated on single-ts→lymphocytes→live→CD3<sup>+</sup> T cells and CD8<sup>+</sup> T cells plotted against tetramer-positive cells with the *number* indicating the percentage of tetramer-positive cells in the CD3 gate. The patient ID, tetramer HLA, name of the peptide, and the peptide sequence is shown above each plot



threshold, *HLArestrictor* would correctly identify 50% of the known HLA restriction elements, while rejecting 90% of the non-restricting HLA elements; whereas at the 2% threshold, it would identify 90% of the known HLA restriction elements, while rejecting 50% of the non-restricting HLA elements. Note that this type of analysis is very crude and simple, and that in particular, the estimated specificity value of *HLArestrictor* should be interpreted with great caution. By way of example, we have only included the strongest association as the true positive HLA restriction element and assigned all other possible HLA restriction element as being negative. This is not always correct as some of the less strongly associated HLA restriction elements may well be bona fide HLA restriction elements. Indeed, as pointed out above, some of the HLA restriction elements that are rejected in this way are well-known and experimentally characterized HLA restrictions. When *HLArestrictor* correctly identifies them as possible HLA restriction elements, they inadvertently end up being considered false positives. Nonetheless, the calculation is simple and un-biased and clearly demonstrates that the HLA restriction method achieves its highest predictive performance for %rank threshold values in the range of 0.5–2. The benchmark demonstrated that here was a strong agreement between the HLA restriction identified by association study and the *HLArestrictor* predictions. However, in 12 of 85 cases, *HLArestrictor* failed to reproduce the restriction element suggested by the association study analysis. In these cases,

*HLArestrictor* suggested alternative HLA restriction elements and minimal peptides for the majority of patients responding to the peptide in question. These findings strongly suggest that *HLArestrictor* is capable of providing information beyond what is obtainable using association studies and that the method is highly sensitive and specific when predicting potential peptide/HLA restrictions.

Further supporting the strong predictive power of the method and demonstrating that it goes beyond identification of the most likely HLA restriction element and also identifies the minimal peptide, we used a panel of 18 peptide MHC class I tetramers across eight different HLA alleles in ten HIV-infected individuals to validate both the optimal epitope CD8<sup>+</sup> T cell response and the corresponding HLA restriction. In 16 out of 18 cases, *HLArestrictor* successfully predicted the HLA restriction and minimal epitope with a %rank score below or equal to 2. If the settings of the *HLArestrictor* were changed so that they also included any predicted binding affinity below 500 nM, then this figure changed to identifying 18 of 18 tetramer-validated epitopes. These observations illustrate the strength of *HLArestrictor*, as it does not only predict a patient's IFN $\gamma$  ELISpot response to an N-mer but also the correct HLA restriction and optimal epitope. The method thus provides a valuable guidance for researchers designing tetramers to validate HLA restriction elements and minimal epitopes corresponding to a given cellular response.

During the development of the *HLArestrictor*, we preferred the %rank measure rather than the affinity measure since the %rank measure lends itself to the needed comparisons across different HLA molecules and HLA isotypes. Predicting the affinity measure is a more demanding task, and not all HLA alleles are yet represented at a level that allows such quantitative predictions. It has further been suggested that not all MHC molecules present peptides at the same binding threshold (Rao et al. 2009; Stranzl et al. 2010). The %rank score removes this bias by placing binding scores for all MHC molecules on an equal scale. However, it has also been suggested that immunogenic peptide are characterized by an HLA binding affinity threshold of 500 nM (Assarsson et al. 2007; Sette et al. 1994). As more peptide/HLA binding data becomes available, it will eventually become possible to use affinity measurements for more and more HLA molecules when interpreting the results. It is even possible to estimate how reliable the prediction of affinity is for each HLA molecule (note that the output of the *HLArestrictor* includes this estimate of reliability). Whenever reliable, it is possible to include an affinity threshold in the interpretation of the output. For ease of operation, *HLArestrictor* includes an *Rank OR Affinity* setting (the default) that allows the selection of peptide/HLA combinations that meets either the %rank or the affinity thresholds. This allows that a predicted HLA restriction with, e.g., a borderline %rank score, might still be classified as an epitope due to a strong affinity. Indeed, this was the case for the two HLA tetramer-validated epitopes, which the *HLArestrictor* failed to recognize when running in the %rank only setting. Both these peptides were predicted to bind to one of the patient's HLA alleles stronger than the commonly used affinity threshold of 500 nM, even though they failed to be predicted below the 2 %rank threshold. If running the *HLArestrictor* in the %rank $\leq$ 2% OR affinity $\leq$ 500 nM threshold setting for the definition of a positive HLA restriction predictions and applied this setting to the 5,145 IFN $\gamma$  ELISpot peptide responses, as much as 94.0% of the responses were predicted with an average of 6.4 predicted epitope:HLA restrictions per peptide compared to 91.3% and 5.3 predicted epitope:HLA restrictions per peptide at a %rank $\leq$ 2%. This increase in sensitivity at a relative minor loss in specificity suggests that interpretations of minimal epitopes and HLA restrictions from the *HLArestrictor* predictions should be based on a combined evaluation of both the %rank and affinity prediction values.

The *HLArestrictor* is specifically aimed at T cell epitope discovery in individual subjects. The use of overlapping peptides has emerged as a very powerful way to scan entire proteins for the presence of immunogenic epitopes. Conventionally, identifying HLA restriction elements and

minimal epitopes within this approach are done by presenting peptides on partially HLA-matched B cells and using more or less systematic peptide truncations, respectively. This requires considerable resources (high-resolution typing, extensive peptide synthesis, and extensive cellular testing). *HLArestrictor* automates the bioinformatic analysis and avoids any bias inherent to a manual “eye-balling” analysis and should relieve the experimenter of much tedious and costly work. Validating all potential 8–11mer peptides to all the patient HLA types is clearly a highly costly and inefficient brute force approach. An 18mer peptide will for instance contain up to 38 distinct 8–11mer peptides leading to a total of 228 peptide/HLA pairs when tested against six HLA types of a patient. Furthermore, *HLArestrictor* was developed with the rapid identification and design of peptide/HLA tetramers in mind. These have emerged as the most direct and efficient method to detect peptide-specific, HLA-restricted T cells. High-throughput methods are now available for peptide/HLA tetramer generation (Leisner et al. 2008; Toebes et al. 2006). It is therefore entirely feasible to use *HLArestrictor* as a rational guide to rapid and large-scale peptide/HLA tetramer formation for direct T cell epitope discovery.

In conclusion, *HLArestrictor* (<http://www.cbs.dtu.dk/services/HLArestrictor>) is a user-friendly tool for patient-specific epitope discovery within peptides. The user can adjust a number of parameters, predictions can be made for 8–11mers, and the different sort modes provide the user with a flexible overview of the predictions. The large-scale benchmarking on experimental data of the method makes it the best validated prediction tool of its kind to date and proves how the method can be a valuable tool to guide the rational identification of new epitopes from infectious diseases and other disease models. The method will be updated continuously as data becomes available for improving the underlying peptide/HLA predictors (e.g., the representation of binding data for HLA-C molecules is expected to improve significantly in the near future, thereby improving HLA-C predictors in particular from an affinity measurement perspective). Another area of future development will be to include HLA class II predictions.

**Acknowledgements** This work was supported by a NIH (National Institute of Health) grant (contract no. HHSN272200900045C) and The Danish Council for Independent Research Technology and Production Sciences (FTP 274-05-0582).

## References

- Altman JD, Moss PA, Goulder PJ, Barouch DH, McHeyzer-Williams MG, Bell JI, McMichael AJ, Davis MM (1996) Phenotypic analysis of antigen-specific T lymphocytes. *Science* 274:94–96

- Assarsson E, Sidney J, Oseroff C, Pasquetto V, Bui HH, Frahm N, Brander C, Peters B, Grey H, Sette A (2007) A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. *J Immunol* 178:7890–7901
- Carlson JM, Heckerman D, Shani G (2009) Estimating false discovery rates for contingency tables. Microsoft Corporation TechReport. Available at <http://research.microsoft.com/en-us/um/redmond/projects/MSCompBio/FalseDiscoveryRate/>
- Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M (2009) NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 61:1–13
- Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, Zimbwa P, Moore S, Allen T, Brander C, Addo MM, Altfeld M, James I, Mallal S, Bunce M, Barber LD, Szinger J, Day C, Klennerman P, Mullins J, Korber B, Coovadia HM, Walker BD, Goulder PJ (2004) Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* 432:769–775
- Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, Reddy S, de Pierres C, Mncube Z, Mkhwanazi N, Bishop K, van der Stok M, Nair K, Khan N, Crawford H, Payne R, Leslie A, Prado J, Prendergast A, Frater J, McCarthy N, Brander C, Learn GH, Nickle D, Rousseau C, Coovadia H, Mullins JI, Heckerman D, Walker BD, Goulder P (2007) CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat Med* 13:46–53
- Korber BTM, Brander C, Haynes BF, Koup R, Moore JP, Walker BD, Watkins DI (2009) HIV molecular immunology. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, LA-UR 09-05941
- Leisner C, Loeth N, Lamberth K, Justesen S, Sylvester-Hvid C, Schmidt EG, Claesson M, Buus S, Stryhn A (2008) One-pot, mix-and-read peptide–MHC tetramers. *PLoS ONE* 3:e1678
- Lin HH, Ray S, Tongchusak S, Reinherz EL, Brusic V (2008) Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunol* 9:8
- Lundegaard C, Lund O, Kesmir C, Brunak S, Nielsen M (2007) Modeling the adaptive immune system: predictions and simulations. *Bioinformatics* 23:3265–3275
- Lundegaard C, Lund O, Buus S, Nielsen M (2010) Major histocompatibility complex class I binding predictions as a tool in epitope discovery. *Immunology* 130:309–318
- Moutaftsi M, Peters B, Pasquetto V, Tscharke DC, Sidney J, Bui HH, Grey H, Sette A (2006) A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat Biotechnol* 24:817–819
- Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, Roder G, Peters B, Sette A, Lund O, Buus S (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE* 2:e796
- Payne RP, Klooverpris H, Sacha JB, Brumme Z, Brumme C, Buus S, Sims S, Hickling S, Riddell L, Chen F, Luzzi G, Edwards A, Phillips R, Prado JG, Goulder PJ (2010) Efficacious early antiviral activity of HIV Gag- and Pol-specific HLA-B 2705-restricted CD8+ T cells. *J Virol* 84:10543–10557
- Rao X, Costa AI, van Baarle D, Kesmir C (2009) A comparative study of HLA binding affinity and ligand diversity: implications for generating immunodominant CD8+ T cell responses. *J Immunol* 182:1526–1532
- Robinson J, Waller MJ, Parham P, Bodmer JG, Marsh SG (2001) IMGT/HLA database—a sequence database for the human major histocompatibility complex. *Nucleic Acids Res* 29:210–213
- Rock KL, York IA, Saric T, Goldberg AL (2002) Protein degradation and the generation of MHC class I-presented peptides. *Adv Immunol* 80:1–70
- Serwold T, Gonzalez F, Kim J, Jacob R, Shastri N (2002) ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* 419:480–483
- Sette A, Vitiello A, Rehman B, Fowler P, Nayarsina R, Kast WM, Melief CJ, Oseroff C, Yuan L, Ruppert J, Sidney J, del Guercio MF, Southwood S, Kubo RT, Chesnut RW, Grey HM, Chisari FV (1994) The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol* 153:5586–5592
- Storey JD, Tibshirani R (2003) Statistical significance for genome-wide studies. *Proc Natl Acad Sci USA* 100:9440–9445
- Stranzl T, Larsen MV, Lundegaard C, Nielsen M (2010) NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics* 62(6):357–368
- Toebes M, Coccoris M, Bins A, Rodenko B, Gomez R, Nieuwkoop NJ, van de Kastele W, Rimmelzwaan GF, Haanen JB, Ovaas H, Schumacher TN (2006) Design and use of conditional MHC class I ligands. *Nat Med* 12:246–251
- Townsend A, Trowsdale J (1993) The transporters associated with antigen presentation. *Semin Cell Biol* 4:53–61
- Uebel S, Tampe R (1999) Specificity of the proteasome and the TAP transporter. *Curr Opin Immunol* 11:203–208
- Yewdell JW, Bennink JR (1999) Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu Rev Immunol* 17:51–88
- Zhang H, Lundegaard C, Nielsen M (2009) Pan-specific MHC class I predictors: a benchmark of HLA class I pan-specific prediction methods. *Bioinformatics* 25:83–89