

Customized Predictions of Peptide–MHC Binding and T-Cell Epitopes Using EPIMHC

Magdalena Molero-Abraham, Esther M. Lafuente, and Pedro Reche

Abstract

Peptide binding to major histocompatibility complex (MHC) molecules is the most selective requisite for T-cell recognition. Therefore, prediction of peptide–MHC binding is the main basis for anticipating T-cell epitopes. A very popular and accurate method to predict peptide–MHC binding is based on motif-profiles and here we show how to make them using EPIMHC (<http://imed.med.ucm.es/epimhc/>). EPIMHC is a database of T-cell epitopes and MHC-binding peptides that unlike any related resource provides a framework for computational vaccinology. In this chapter, we describe how to derive peptide–MHC binding motif-profiles in EPIMHC and use them to predict peptide–MHC binding and T-cell epitopes. Moreover, we show evidence that customization of peptide–MHC binding predictors can lead to enhanced epitope predictions.

Key words MHC, HLA, PSSM, T-cell epitope, Prediction

Abbreviations

MHC Major histocompatibility complex
HLA Human leukocyte antigen
PSSM Position specific scoring matrices

1 Introduction

T cells play a key role in fighting infectious agents such as pathogenic viruses, bacteria and parasites, as well as in cancer immune surveillance, eliminating tumoral cells. T cells respond to foreign peptide antigens (T-cell epitopes) bound the cell surface expressed major histocompatibility complex (MHC) molecules [1–4]. There are two main classes of MHC molecules, MHC class I (MHCI) and MHC class II (MHCII) that are in turn recognized by CD8⁺ T and CD4⁺ T cells, respectively [4]. In humans, MHC molecules are known as Human Leukocyte Antigens (HLAs) and are

extremely polymorphic [5]. HLA polymorphisms are the basis for distinct peptide binding specificity of HLA allelic variants [5].

The relevance of T-cell epitopes for understanding disease pathology [6] and for epitope-based vaccines [7–10] has led to the identification of thousands of epitopes and MHC-peptide ligands from all kind of antigens. The availability of this vast amount of data has had two major intertwined consequences. On the one hand, it has given rise to comprehensive databases and resources to store the ever-increasing data. On the other hand, it has fueled the development of computational approaches for the prediction of T-cell epitopes.

Relevant examples of T-cell epitope and MHC-peptide ligand databases include SYFPEITHI [11], JenPep [12], and MHCBN [13], TEPIDAS [14], ImmuneEpitope Database [15], and EPIMHC [16]. These resources are instrumental for the analysis of peptide–MHC binding and T-cell epitope immunogenicity, primarily serving as source of data but also providing analytic and predictive tools. All these databases are based on relational databases and share a considerable amount of capabilities. Yet they also have unique features. Here we will work with EPIMHC [16], a highly curated database of T-cell epitopes and MHC-binding peptides that unlike any related resource enables tailored predictions of T-cell epitopes using custom-made peptide–MHC binding motif-profiles [17–19].

T-cell epitopes are determined by several molecular events [20–22], of which peptide–MHC binding is the most selective. Therefore, prediction of peptide–MHC binding is the main basis to anticipate T-cell epitopes [23]. Peptide–MHC binding predictions can be achieved through a great variety of methods [23], including peptide–MHC binding motif-profiles [17–19] which rank among the most successful and popular of them [24]. These motif-profiles consist of weighted position specific scoring matrices (PSSM) [25] created from sets of aligned peptide sequences known to bind to the relevant MHC molecules.

Prediction of T-cell epitopes using a large set of MHC-specific motif-profiles is readily available for free public use in at our RANKPEP site (<http://imed.med.ucm.es/Tools/rankpep.html>). We generated the peptide–MHC binding profiles available in RANKPEP from the largest non-redundant set of peptides that we could identify. In computational cross-validations, RANKPEP profiles exhibited a great performance [17]. However, they are not necessarily the best for all predictive matters. In fact there is no general consensus on what peptides should be included for peptide–MHC binding model building. Therefore, in this chapter, we illustrate how to use EPIMHC to derive custom-made peptide–MHC binding motif-profiles and produce tailored T-cell epitope predictions.

2 Materials

2.1 EPIMHC Database and Query Form

EPIMHC is a database with comprehensive information on MHC-restricted ligands and T-cell epitopes that are observed in real proteins from a great variety of sources including tumor antigens. Peptide data were collected from related databases [11, 26, 27] and the literature and was incorporated into the database upon computational curation (altered peptide ligands are not included). EPIMHC data is structured as a relational database with a set of related tables (Fig. 1). Entries in EPIMHC are unique with regard to the combination of two features: the sequence of the peptide and the MHC restriction element. Main annotations in EPIMHC include information on the ligand (sequence, length, MHC binding, T-cell activity, processing, protein source, protein name, and organisms), the MHC restriction element (CLASS, MHC molecule and MHC source), and publication reference. The “processing” field indicates whether the peptides are processed and presented from their protein sources in vivo (annotated as *natural*). “MHC binding” and “immunogenicity level” fields follow a qualitative annotation of four values (high, moderate, little, unknown). The immunogenicity level only applies to peptides with reported “T-cell activity.” Immunogenicity and MHC-binding binding levels were obtained from the literature and translated onto the indicated qualitative values as previously reported [26]. If no information on peptide–MHC binding and/or Immunogenicity level was found, then such fields were annotated as unknown.

EPIMHC database is accessible online at <http://imed.med.ucm.es/epimhc/> (Fig. 2) through an intuitive and user-friendly Web interface. This server allows for complex database queries,

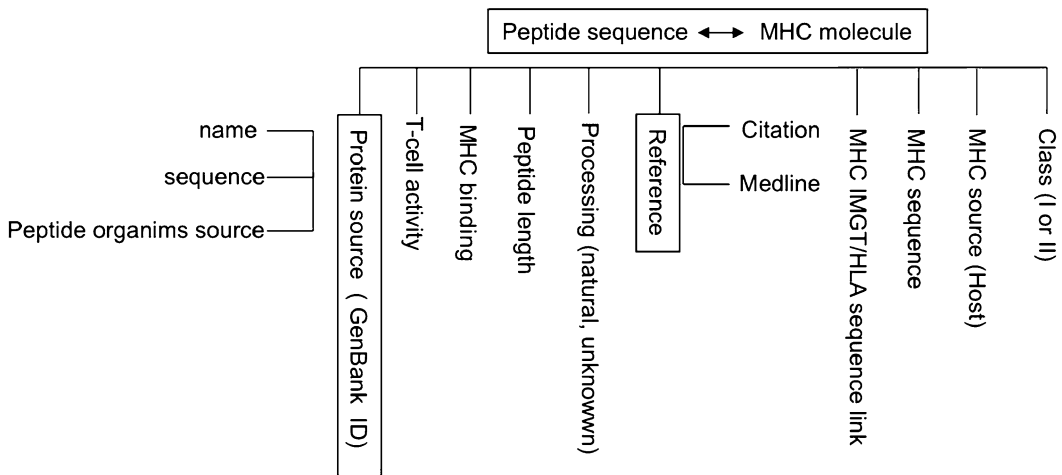


Fig. 1 EPIMHC database structure



Home Research Services Technology Contact

TOOLS

Tools >> EPIMHC Database

EPIMHC

A Database of Naturally Processed MHC-restricted Peptide Ligands and Epitopes for Customized Computational Vaccinology

MHC SELECTION

- All
- BOLA-A11
- DRS2
- ELA-A5
- ELA-A9
- H2-AB
- H2-AD
- H2-AC7
- H2-AK
- H2-AS
- H2-AU
- H2-DB
- H2-DD
- H2-DK
- H2-DQ
- H2-EB
- H2-ED
- H2-EG7
- H2-EK
- H2-ES
- H2-K8
- H2-KD
- H2-KK
- H2-LD
- H2-M3
- H2-M3WT
- H2-QA-1A
- H2-QA-1B
- H2-QA-2
- H2-QA-2A
- H2-A
- H2-B

AND [v] SEQ []

AND [v] LENGTH [v] All 6 7 8

AND [v] CLASS [v] All

MHC SOURCE [v] All BONOBO CHIMPANZEE CHIMPANZEE COTTON-TOP TAMARIN

PEPTIDE SOURCE ORGANISM [v] All Chlamydia trachomatis Clostridium botulinum Dengue virus

AND [v] PEPTIDE BINDING LEVEL [v] All HIGH LITTLE MODERATE

AND [v] T-CELL ACTIVITY/IMMUNOGENIC PEPTIDES (EPITOPES) [v] All

AND [v] IMMUNOGENICITY LEVEL [v] All HIGH LITTLE MODERATE

AND [v] PROCESSING: [v] All

RESULTS DISPLAY

Select Fields to be Displayed [v] DEFAULT MHC MOLECULE SEQUENCE MHC Source CLASS SEQUENCE LENGTH

Select field to order results by: [v] MHC [v] Ascending

Reset Search

RELATED RESOURCES

Fig. 2 EPIMHC database Web interface. EPIMHC resource is available for free public use at <http://imed.med.ucm.es/epimhc/>

combining any annotation field, thanks to the underlying SQL language. For example, users can search for peptide ligands and/or epitopes that are restricted by one, various or all MHC molecules (left side of the screen), and restrict the search according to various criteria like length and source of the peptide (right of the screen). Also, any field of interest can be included in the search output. The EPIMHC search output will be described in detail in the Method section in the context of the generation of custom-made profiles.

2.2 Prediction of a Peptide-MHC Binding and T-Cell Epitopes Using Profiles

As mentioned, motif-profiles consist of weighted PSSM [25] created from a set of aligned peptide sequences known to bind to a given MHC molecule. In order to predict peptide-MHC binding and T-cell epitope using motif-profiles, we used a search algorithm known as RANKPEP [17, 18]. The algorithm uses the profile coefficients to score all possible peptide fragments in a protein with the width of the PSSM and ranks them by score. The width of a PSSM is given by the number of residue sites in a multiple sequence alignment. Although rank per se is insufficient to assess whether a peptide is a potential binder, we have shown that T-cell epitopes score among the top 2 % ranking peptides [17, 18]. Motif-profiles assume that peptide residues contribute independently to MHC binding. This assumption is well supported by experimental data, although there are reported instances in which the contribution of peptide residues to MHC-binding is influenced by neighboring residues [28].

RANKPEP is accessible online for public use at the site <http://imed.med.ucm.es/Tools/rankpep.html> (Fig. 3). Currently, 88 and 50 different MHCI and MHCII molecules, respectively, can be targeted for peptide binding predictions in RANKPEP using the relevant motif-profiles. The profiles available in RANKPEP have been derived from large sets of non-redundant peptide-MHC binders, without taking in consideration their T-cell activity and source. These sets can include self-peptides eluted from MHC molecules. The RANKPEP Web server is flexible, intuitive and combines several interesting features. Notably, users can upload their own motif-profiles, such as those generated using EPIMHC (*see* Subheading 3). A simplified version of the RANKPEP input form can also be launched from EPIMHC to facilitate tailored prediction of T-cell epitopes using custom-made profiles (*see* Subheading 3).

3 Methods

In this section, we show a step-by-step guide to derive a specific peptide-MHC binding motif-profile in EPIMHC and produce tailored T-cell epitope predictions. In particular, we will target the prediction of A*0201-restricted CD8 T-cell epitopes from SARS coronavirus nucleoprotein (GI: 30173007). This protein contains 7 experimentally identified A*0201-restricted CD8 T-cell epitopes (Table 1) and we will use that knowledge to assess the predictive accuracy of various peptide-MHC binding motif-profiles.

3.1 Peptide Selection and Motif-Profile Building

We will build a motif-profile from all 9-mer peptides that are annotated in EPIMHC to bind with high affinity to the human MHC I molecule HLA-A*0201 (A*0201) (*see* Notes 1–6). To this end,



Rankpep: prediction of binding peptides to Class I and Class II MHC molecules

Description
 This server predicts peptide binders to MHC I and MHC II molecules from protein sequence/s or sequence alignments using Position Specific Scoring Matrices (PSSMs). In addition, it predicts those MHC I ligands whose C-terminal end is likely to be the result of proteasomal cleavage. A detailed explanation of the method can be found [here](#).

PSSM	SELECT PSSM (Check MHC I or MHC II)	
	<input checked="" type="radio"/> MHC I H2-Db (mouse) [8mer] H2-Db (mouse) [9mer] H2-Db (mouse) [10mer] H2-Db (mouse) [11mer] H2-Dd (mouse) [9mer]	<input type="radio"/> MHC II HLA-DP4 HLA-DP9(DPA1*0201xDPB1*0901) HLA-DPw4 HLA-DPw4(DPB1*0402) HLA-DQ1
OR, UPLOAD YOUR PSSM <input type="button" value="Choose File"/> no file selected		
INPUT	TYPE: <input checked="" type="radio"/> FASTA sequence/s <input type="radio"/> CLUSTALW multiple sequence alignment	
	Replace example with your query >A56881 PIR2 release 71.00 MWNLLHETDSAVATARRPRWLCAGALVLAGGFFLLGLFGLFGWFIKSSNEAT NITPKHNMKAFDELKAENIKKFLYNFTQIPHLAGTEQNFQLAKQIQSQW KEFGLDSELAHYDVLVLSYPNKTHPNYISIINEDGNEIFNTSLFEPPPG	
OR, UPLOAD SEQUENCES <input type="button" value="Choose File"/> no file selected		
BINDING THRESHOLD	<input type="radio"/> PERCENTAGE: 2%	<input checked="" type="radio"/> TOP NUMBER: 990
PROTEASOME CLEAVAGE	FILTER: OFF LMPC: One If Filter is ON only peptides predicted to be cleaved are shown	
ADVANCED OPTIONS		
RESTRICT RESULTS BY MW Lower Limit for Molecular Weight: 0.00 Upper Limit for Molecular Weight: 9999.00		VARIABILITY MASKING Select Variability Threshold: 1 Value must range between 0.0 and 4.3
<input type="button" value="Send"/> <input type="button" value="Clear Form"/>		

Fig. 3 RANKPEP Web server. The figure depicts a screenshot of the RANKPEP interface with the option of uploading custom-made profiles highlighted. RANKPEP is available for free public use at <http://imed.med.ucm.es/Tools/rankpep.html>

we first do a search in EPIMHC with the following selection criteria, leaving the remaining fields with default settings:

1. Select HLA-A*0201 in *MHC SELECTION*.
2. Select 9 in *LENGTH* (see Note 2).
3. Select high in *PEPTIDE BINDING LEVEL*.

Table 1
Known A*0201-restricted CD8 T-cell epitopes in SARS nucleoprotein

Epitope sequence	Location	References
ALNTPKDHI	139–147	[34]
LQLPQGTTL	160–168	[34]
LALLLLDRL	220–228	[34]
LLLDRLNQL	223–231	[34]
RLNQLESKV	227–235	[34]
GMSRIGMEV	317–325	[34]
ILLNKHIDA	352–360	[34]

None of these epitopes has been used for profile building in EPIMHC

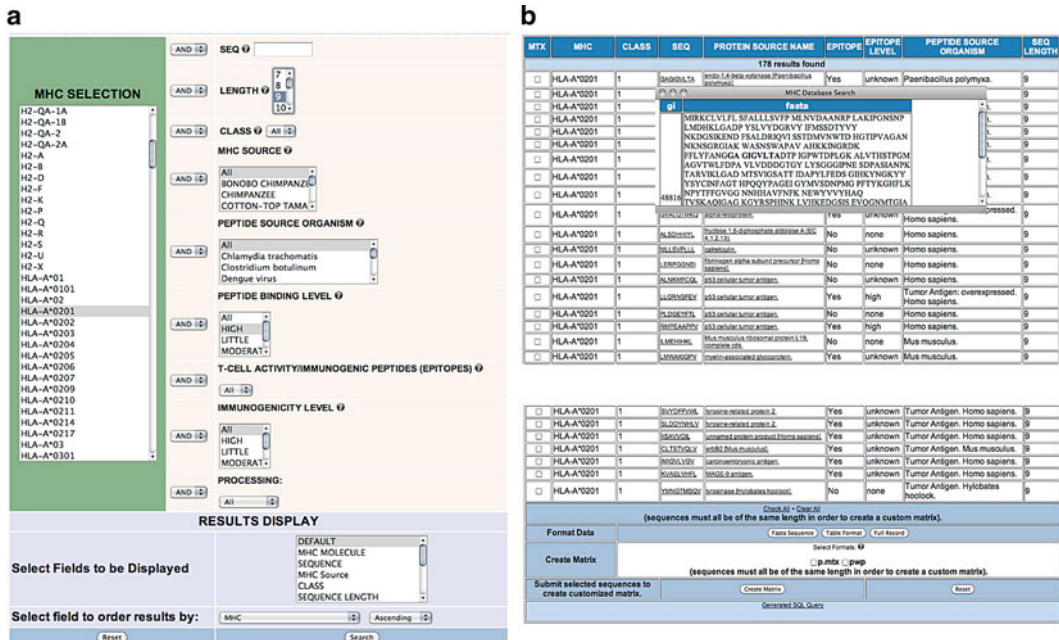


Fig. 4 EPIMHC search example and output. The figure illustrates a search example in the EPIMHC resource for peptides binding to HLA-A*0201 with high affinity (a) and the result of that specific search (b)

In Fig. 4a we show a screen capture of the selection described. Upon submitting the search, we get 178 peptides (Fig. 4b). EPIMHC search results consist of a tabulated list, rows, of records fitting the search criteria. The table columns provide the information fields selected by the users in the query form. The default fields are those shown in Fig. 4b and include the MHC restriction element (*MHC*), the MHC class (I or II) (*CLASS*), the sequence

A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
1.242	0.511	-3.95	-6.57	0.868	1.588	1.939	0.437	0.183	-0.04	-0.47	-2.39	-5.13	-0.35	0.65	1.619	-2.53	0.18	0.974	2.959
-2.85	-7.01	-4.94	-5.05	-5.89	-10.8	-9.46	1.65	-4.4	7.78	4.35	-10.3	-11.4	-3.34	-9.33	-9.83	-0.81	1.236	-9.48	-8.34
-0.32	0.362	1.139	-3.5	0.567	0.459	2.033	0.959	-3.15	0.618	2.244	0.982	-3.94	-2.05	-0.83	-1.42	-3.04	1.318	5.311	2.123
-1.43	-1.02	3.157	2.219	-6.8	2.367	-5.67	0.641	0.4	-2.54	0.544	-0.58	2.026	-1.14	-0.65	-0.42	-0.61	-3.5	0.286	-1.77
-1.23	-0.58	-1.03	-1.3	1.308	2.357	0.944	0.333	-0.16	-0.64	-1.86	-0.22	-0.12	-0.3	-1.7	-2.67	-0.09	0.13	3.087	3.421
-2.24	3.381	-0.87	-3.12	0.527	-1.23	-1.19	3.077	-1.71	1.318	1.057	-1.49	-0.98	-1.86	-4.7	-0.17	0.915	2.648	-1.18	0.249
1.233	0.416	-3.41	-3.08	1.554	-1.54	2.554	1.719	-3.57	1.286	0.828	-0.21	2.362	-1.35	-4.06	-3.31	-0.16	1.889	3.005	-1.94
-1.8	-1.51	-2.67	1.279	0.087	-0.63	1.255	-0.88	-0.42	0.574	-0.17	-2.09	-1.91	1.059	0.347	1.467	1.8	-0.91	1.317	1.002
-0.77	-7.92	-12.2	-11.6	-7.63	-11.9	-11.1	3.887	-10.7	5.368	-0.05	-11.8	-12.6	-1.09	-11	-4.99	-9.15	7.339	-11.1	-9.67

Fig. 6 EPIMHC profile with position-based weights generated from 178 9-mer peptides binding to HLA-A*0201 with high affinity

Upon hitting the *create matrix* button, EPIMHC opens a simplified version of the RANKPEP Web interface that incorporates the custom-made motif-profile from the selected peptides (Fig. 5b). The profile appears under the *File* field of the form and can be downloaded with a mouse right click (PC) or by a mouse click-and-hold (MAC). The profile thus generated, shown in Fig. 6, has the format required by the MAST-motif search algorithm [31] and can be uploaded to the original RANKPEP Web server to produce custom predictions of peptide-MHC binding and T-cell epitopes. However, the RANKPEP interface launched by EPIMHC allows a more direct and simple way to achieve such a task (Fig. 5b). Under *SET DISPLAY OPTIONS*, users can select between two options to set the number of peptides to be returned by the algorithm: one is as a fixed *number of top scoring peptides* and the other as a *percentage of top scoring peptides*. Users can also restrict the peptides sorted by RANKPEP by molecular weight (MW) so that only peptides within a MW window will be returned. By default, MW filtering is not applied. The RANKPEP interface also provides three models for proteasomal cleavage predictions [17, 22]. By default, model *one* is selected. These models will be applied regardless of the class of the MHC targeted for predictions but the predictions are only meaningful for MHC I-restricted peptides (*see Note 4*).

3.2 Prediction of Peptide-MHC Binding and T-Cell Epitopes With EPIMHC Custom-Made Profiles

To target SARS nucleoprotein for T-cell epitope predictions using the RANKPEP form launched by EPIMHC with the custom-made profile we carry on as follows:

1. Set peptides to display to 2 % of top scoring peptides (*see Note 5*).
2. Paste the SARS nucleoprotein, FASTA format, in the text box INPUT section.
3. Click on the matrix check box.
4. Click on the action button *Run Rankpep*.

The indicated steps are highlighted in Fig. 7a and we describe next the RANKPEP output (Fig. 7b)

The top part of the RANKPEP output shows the matrix (profile) used for the predictions, a consensus sequence that would reach the largest score, optimal (largest) score and a binding threshold (BT). The later is an important feature. Large scores lead

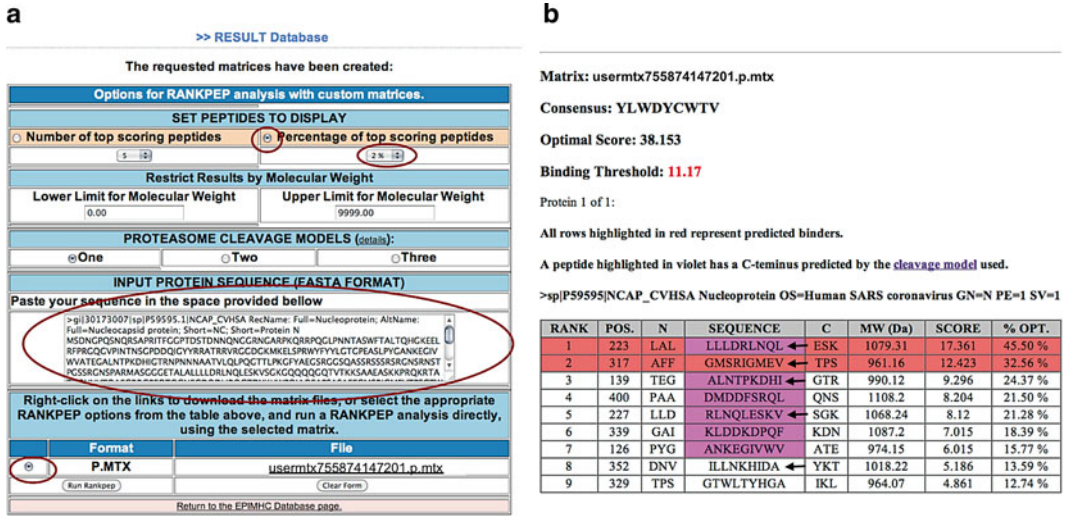


Fig. 7 Tailored prediction of peptide–MHC binding using the RANKPEP form launched by EPIMHC (a) The figure illustrate the steps to predict HLA-A*0201-restricted peptides from SARS nucleoprotein. (b) RANKEP output showing the prediction results

to top ranking peptides and are indicative of peptide–MHC binding. However, rank per se is insufficient to know whether a given peptide will bind to a particular MHC molecule, e.g., scoring a single peptide. Therefore, EPIMHC provides a profile-specific BT that serves to identify the most confident peptide–MHC binders and T-cell epitopes as those with a score \geq BT. The profile-specific BT provided by EPIMC is obtained by scoring all the peptides used to make the relevant profile matching the 90 percentile value of all peptide scores [18]. The next part in the RANKPEP output consists of a list of peptides from the input protein ranked by the scores obtained with the relevant profile. In our case, RANKPEP shows only 9 peptides from SARS nucleoprotein because we selected to display only the 2 % of top scoring peptides. For every peptide, RANKPEP shows its rank (*RANK*), location in the protein sequence (*POS*), sequence (*SEQUENCE*), three N-terminal (*N*) and C-terminal (*C*) flanking residues, score (*SCORE*), and relative score, in percentage, with regard to the optimum score (*%OPT*). Peptides whose scores are equal or greater than the BT score are highlighted in red, and those containing a C-terminal end predicted to be the result of proteasomal cleavage are shown in violet.

As we made a profile from peptides binding with high affinity to A*0201, we are predicting peptides from SARS nucleoprotein that bind to A*0201 and hence potential A*0201-restricted CD8 T-cell epitopes. In fact, in the results shown in Fig. 7b, it is possible to identify 5 out of the 7 known A*0201-restricted CD8 T-cell epitopes from SARS nucleoprotein (Table 2).

Table 2
Description of custom-made profiles used in this study

Profile name	EPIMHC search selection					EPIMHC search result and profile building	
	MHC molecule	PEP. binding level	PEP. source organism	T-cell activity	Length	Peptides	Method
Profile #1	HLA-A*0201	High	All	All	9	178	p.mtx ^b
Profile #2	HLA-A*0201	High	All	YES	9	95	p.mtx ^b
Profile #3	HLA-A*0201	High	Viruses ^a	All	9	48	p.mtx ^b
Profile #4	HLA-A*0201	High	Viruses ^a	YES	9	32	p.mtx ^b

^aThe viruses selected in EPIMHC were the following: Dengue virus, Epstein-Barr virus, Hepatitis B virus, Hepatitis C virus, Human immunodeficiency virus 1, Human immunodeficiency virus 1 OPT, Human immunodeficiency virus 2, Influenza A virus, Measles virus, and West Nile virus. None of the CD8 T-cell epitopes from SARS nucleoprotein are included in any of the peptide selections used for making these profiles

^bPSSMs generated using position-based weights [30]

3.3 Comparison of CD8 T-Cell Epitope Predictions Using Various Custom-Made Profiles

The goodness of peptide–MHC binding and T-cell epitope predictions provided by any predictive model, including motif-profiles, is tied to the data used for model building [32, 33]. To demonstrate such influence, here we will compare the predictions of A*0201-restricted CD8 T-cell epitopes from SARS nucleoprotein that are obtained with 4 distinct motif-profiles, including that described in the previous section (hereafter profile #1). The specific peptide selections that give rise to the different profiles used in this section are detailed in Table 2. Briefly, all profile-motifs are generated from peptides binding with high affinity to A*0201. In addition, profile #3 and #4 only include peptides from viruses and profile #2 and #4 only include peptides known to display T-cell activity (they are epitopes). To evaluate the predictive performance of these profiles, we scored and ranked all peptides from SARS nucleoprotein and compared the ranking achieved by the known SARS nucleoprotein A*0201-restricted CD8 T-cell epitopes shown in Table 1. These results are summarized in Table 3. All four motif-profiles produce related results, ranking the known CD8 T-cell epitopes among the top scoring peptides of SARS nucleoprotein. This is expected as A*0201-restricted CD8 T-cell epitopes and the peptides used for model building have in common the ability to bind to A*0201. However, there are also differences in the results. Thus, only the profiles derived from viral peptides are capable of predicting the known A*0201-restricted CD8 T-cell epitopes from SARS nucleoprotein among the top 11 scoring peptides (Table 3). Judging from the dispersion of the ranks (Table 3), the best overall epitope predictions are obtained

Table 3
Ranking and statistics of known SARS nucleoprotein A*0201-restricted CD8 T-cell epitopes using four custom-made motif-profiles

Epitope	Profile #1	Profile # 2	Profile #3	Profile # 4
ALNTPKDHI	3	3	4	4
GMSRIGMEV	2	1	1	3
ILLNKHIDA	8	7	5	5
LALLLLDRL	34	11	30	11
LLLDRLNQL	1	2	9	2
LQLPQGTTL	10	13	7	10
RLNQLESKV	5	4	3	1
		Rank	Statistics	
Median	5	4	5	4
Mean	9	5.8	5.4	5.1
Stdev	11.3	4.5	9.8	3.9
Range	1–34	1–13	1–30	1–11

Peptide ranks are obtained after scoring all peptides in SARS nucleoprotein with the relevant motif-profiles and depict the peptide score relative to that of all remaining peptides. Rank 1 means that the peptide has the largest score of all peptides

with a motif-profile derived from viral peptides with known T-cell activity (T-cell epitopes). It remains to be explored whether discarding peptide–MHC binders with no reported T-cell activity always improves the resulting T-cell epitope prediction models. It is important to note that none of the known epitopes used in these analyses have been used to derive any of the profiles. In fact, the current version of EPIMHC does not contain any SARS peptides at all.

In conclusion, profiles are very powerful at capturing nontrivial motifs and the results shown here support that epitope predictions can be improved using customized peptide–MHC binding profiles. EPIMHC is the only available resource readily suitable for that task.

4 Notes

1. In EPIMHC, users can make profiles from any peptide selection but the peptides must be related to some extent (e.g., binding to the same MHC) to produce profiles yielding meaningful predictions (*see* Subheading 3.1).
2. EPIMHC is better suited for making MHC I-specific profiles. Moreover, profiles can only be generated from peptides with the same length; otherwise EPIMHC returns an error

(see Subheading 3.1). There are practical and structure-based reasons for this limitation as discussed by Reche et al. [18].

3. EPIMHC can also produce profiles from peptides that have been selected to bind to MHC II molecules, provided that they have the same length (see Subheading 3.1). However, as data availability is limited, we recommend using a motif discovery program such as MEME [31] for making peptide-MHC II binding profiles from peptides of any length as described in previous reports [17, 18].
4. The proteasomal cleavage predictions should not be taken in consideration when predicting peptide binding to MHC II molecules: the proteasome is not involved in class II antigen processing. We are working in correcting this inconsistency.
5. For RANKPEP to return all peptides in a given protein sorted by score, users need to make the following selections in the RANKPEP input form: first set peptides to display by number and then select 990 from the pull-down menu.
6. To generate profiles that are capable of capturing the relevant peptide-MHC binding feature, we suggest using a minimum of five peptides.

Acknowledgements

This work was supported by grant SAF2009-08103 from Ministerio de Ciencia e Innovación to PAR.

References

1. Zinkernagel M, Doherty PC (1974) Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* 248:701–702
2. Margulies DH (1997) Interactions of TCRs with MHC-peptide complexes: a quantitative basis for mechanistic models. *Curr Opin Immunol* 9:390–395
3. Garcia KC, Teyton L, Wilson IA (1999) Structural basis of T cell recognition. *Annu Rev Immunol* 17:369–397
4. Wang J-H, Reinherz E (2001) Structural basis of T cell recognition of peptides bound to MHC molecules. *Mol Immunol* 38:1039–1049
5. Reche PA, Reinherz EL (2003) Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *J Mol Biol* 331:623–641
6. Tchernev G, Orfanos CE (2006) Antigen mimicry, epitope spreading and the pathogenesis of pemphigus. *Tissue Antigens* 68:280–286
7. Reche PA, Keskin DB, Hussey RE, Ancuta P, Gabuzda D, Reinherz EL (2006) Elicitation from virus-naive individuals of cytotoxic T lymphocytes directed against conserved HIV-1 epitopes. *Med Immunol* 5:1
8. Akdis CA, Akdis M, Blesken T, Wymann D, Alkan SS, Muller U, Blaser K (1996) Epitope-specific T cell tolerance to phospholipase A2 in bee venom immunotherapy and recovery by IL-2 and IL-15 *in vitro*. *J Clin Invest* 98:1676–1683
9. Stienekemeier M, Falk K, Rotzschke O, Weishaupt A, Schneider C, Toyka KV, Gold R, Strominger JL (2001) Vaccination, prevention, and treatment of experimental autoimmune neuritis (EAN) by an oligomerized T cell epitope. *Proc Natl Acad Sci U S A* 98:13872–13877
10. Lazoura E, Apostolopoulos V (2005) Insights into peptide-based vaccine design for cancer immunotherapy. *Curr Med Chem* 12:1481–1494

11. Rammensee HG, Bachmann J, Emmerich NPN, Bacho OA, Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50:213–219
12. Blythe MJ, Doytchinova IA, Flower D (2002) JenPep: a database of quantitative functional peptide data for immunology. *Bioinformatics* 18:434–439
13. Bhasin M, Singh H, Raghava GP (2003) MHCBN: a comprehensive database of MHC binding and non-binding peptides. *Bioinformatics* 19:665–666
14. Diez-Rivero CM, Garcia-Boronat M, Reche PA (2008) Integrating T-cell epitope annotations with sequence and structural information using DAS. *Bioinformatics* 3:156–158
15. Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, Bui HH, Buus S, Frankild S, Greenbaum J et al (2008) Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res* 36:W513–W518
16. Reche PA, Zhang H, Glutting JP, Reinherz EL (2005) EPIMHC: a curated database of MHC-binding peptides for customized computational vaccinology. *Bioinformatics* 21: 2140–2141
17. Reche PA, Glutting J-P, Reinherz EL (2004) Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics* 56: 405–419
18. Reche PA, Glutting JP, Reinherz EL (2002) Prediction of MHC class I binding peptides using profile motifs. *Hum Immunol* 63:701–709
19. Reche PA, Reinherz EL (2007) Prediction of peptide-MHC binding using profiles. *Methods Mol Biol* 409:185–200
20. Jensen PE (2007) Recent advances in antigen processing and presentation. *Nat Immunol* 8: 1041–1048
21. Diez-Rivero CM, Chenlo B, Zuluaga P, Reche PA (2010) Quantitative modeling of peptide binding to TAP using support vector machine. *Proteins* 78:63–72
22. Diez-Rivero CM, Lafuente EM, Reche PA (2010) Computational analysis and modeling of cleavage by the immunoproteasome and the constitutive proteasome. *BMC Bioinformatics* 11:479
23. Lafuente EM, Reche PA (2009) Prediction of MHC-peptide binding: a systematic and comprehensive overview. *Curr Pharm Des* 15: 3209–3220
24. Gowthaman U, Agrewala JN (2008) In silico tools for predicting peptides binding to HLA-class II molecules: more confusion than conclusion. *J Proteome Res* 7:154–163
25. Gribskov M, McLachlan AD, Eisenberg D (1987) Profile analysis: detection of distantly related proteins. *Proc Natl Acad Sci U S A* 84: 4355–4358
26. Brusic V, Rudy G, Harrison LC (1994) MHCPEP: a database of MHC-binding peptides. *Nucleic Acids Res* 22:3663–3665
27. Korber BTM, Moor JP, Brander C, Walker BD, Haynes BF, Koup R (1998) HIV molecular immunology compendium. Los Alamos National Laboratory, Los Alamos, NM
28. Peters B, Tong W, Sidney J, Sette A, Weng Z (2003) Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules. *Bioinformatics* 19: 1765–1772
29. Thompson JD, Higgins DG, Gibson TJ (1994) Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Comput Appl Biosci* 10:19–29
30. Henikoff S, Henikoff JG (1994) Position-based sequence weights. *J Mol Biol* 243: 574–578
31. Bailey TL, Gribskov M (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 14: 48–54
32. Martinez-Naves E, Lafuente EM, Reche PA (2011) Recognition of the ligand-type specificity of classical and non-classical MHC I proteins. *FEBS Lett* 585:3478–3484
33. Martínez-Naves E, Lafuente EM, Reche PA (2011) In: Liò P, Nicosia G, Stibor T (eds.), 10th International conference on artificial immune systems. Springer-Verlag Cambridge, England, Vol. 6825, 55–65
34. Cheung YK, Cheng SC, Sin FW, Chan KT, Xie Y (2007) Induction of T-cell response by a DNA vaccine encoding a novel HLA-A*0201 severe acute respiratory syndrome coronavirus epitope. *Vaccine* 25:6070–6077