

# Bioinformatics Identification of Antigenic Peptide: Predicting the Specificity of Major MHC Class I and II Pathway Players.

**Lund, Ole; Karosiene, Edita; Lundegaard, Claus; Larsen, Mette Voldby; Nielsen, Morten**

*Published in:*
Antigen Processing

*Publication date:*
2013

*Document Version*
Publisher's PDF, also known as Version of record

# Chapter 19

# Bioinformatics Identification of Antigenic Peptide: Predicting the Specificity of Major MHC Class I and II Pathway Players

**Ole Lund, Edita Karosiene, Claus Lundegaard, Mette Voldby Larsen, and Morten Nielsen**

## Abstract

Bioinformatics methods for immunology have become increasingly used over the last decade and now form an integrated part of most epitope discovery projects. This wide usage has led to the confusion of defining which of the many methods to use for what problems. In this chapter, an overview is given focusing on the suite of tools developed at the Technical University of Denmark.

**Key words:** Immune, Epitope, MHC, HLA, Class I, Class II, Antigen processing, Proteasome, TAP, Visualization, Bioinformatics, Prediction, Web server.

## 1. Introduction

Experimental methods for analyzing antigenic peptide generation, transport, and binding to Major Histocompatibility Complex (MHC) class I molecules are expensive and time consuming. While bioinformatics methods can never replace experiments in the laboratory, they may in a highly cost-effective manner guide the experimental efforts in a direction that increases the likelihood of discovering immunologically important responses. At the Technical University of Denmark, we have over the last decade developed a number of methods for predicting which part of an antigen most likely is presented to the immune system. A complicating factor is that the MHC molecules associated with response to foreign antigens are encoded at several loci. Furthermore, these genes are the most polymorphic in the human genome and thousands of different alleles are known. Many of these alleles encode different variants of MHC molecules having different peptide binding specificities.

Peter van Endert (ed.), *Antigen Processing: Methods and Protocols*, Methods in Molecular Biology, vol. 960,
DOI 10.1007/978-1-62703-218-6_19, 

However, it is possible to cluster alleles with similar specificities into functional groups called supertypes, first described by Sette and Sidney (1). The pioneering methods for predicting binding to MHC class I molecules such as BIMAS (2) and SYPEITHI (3) helped initiate the field of immunological bioinformatics, but these methods have since been surpassed by newer methods like the ones described in this chapter, and we propose that experimental efforts may be minimized by basing the experiments on these newer methods.

## 2. Binding of Peptides to MHC

In recent years numerous methods for predicting binding to MHC molecules have been proposed. These methods can broadly be divided into two classes: one being the allele-specific and one being the pan-specific methods. Allele-specific methods are constructed for a given allele, and can interpolate between different ligands and give predictions for peptides for which no binding data are available. An obvious limitation by these methods is that predictions can only be made for alleles for which a number of binding data is already available. This requirement has been circumvented by the so-called pan-specific methods, which can also interpolate between different MHC alleles and thus make predictions for alleles for which no known binders are available. This strongly increases the number of alleles for which predictions can be obtained, from the few hundreds for which binding data is available to the more than 3,000 for which the protein sequence is known.

The accuracy of methods for MHC peptide binding prediction depends critically on the available data characterizing the binding specificity of the MHC molecules. This makes it very difficult for the nonexpert user to choose the most suitable method for predicting binding to a given MHC molecule. To complicate things even further, it has been demonstrated that consensus methods defined as combinations of two or more different methods led to improved prediction accuracy.

## 3. Prediction of MHC Class I Peptide Binding

To benefit from the consensus approach and to guide the nonexpert user on selecting the most appropriate binding prediction method for a given MHC class I molecule, we have recently developed the *NetMHCcons* method. The method is available at http://www.cbs.dtu.dk/services/NetMHCcons.

The method integrates predictions from three well-established prediction methods (*NetMHC* (4, 5), *NetMHCpan* (6, 7), and

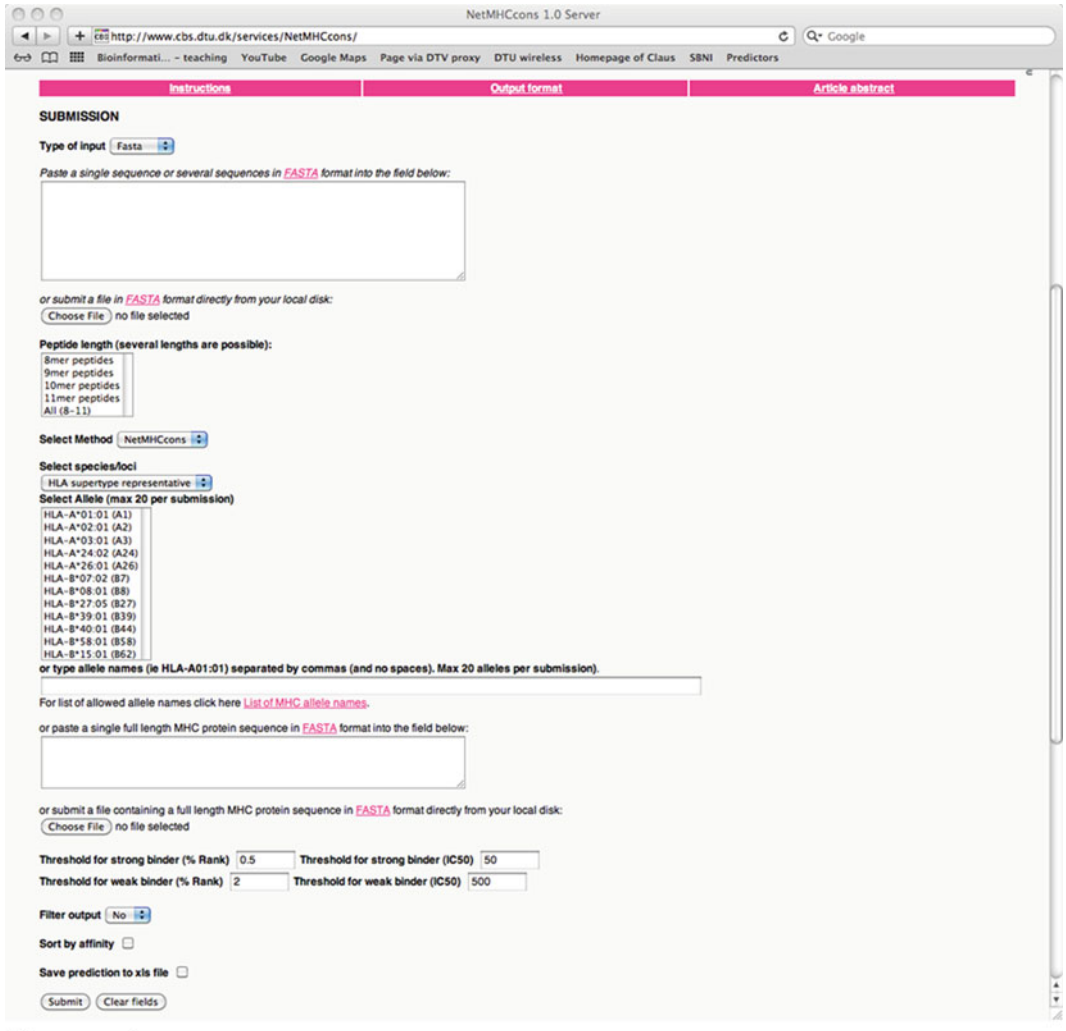


Fig. 1. Submission site of *NetMHCcons* server. Two submission types are handled—a list of peptides or protein sequence(s). The server provides a possibility for the user to choose MHC molecules in question from a list of alleles or alternatively upload a full-length MHC protein sequence of interest. The user has a choice of setting the threshold for defining strong and weak binders based on predicted affinity (IC50) or %Rank. The output can be sorted based on predicted binding affinity as well as filtered on the user-specified thresholds.

*PickPocket* (8)) and allows the user in an automatic manner to obtain the most accurate predictions for any given MHC class I molecule of known protein sequence. The three methods included in *NetMHCcons* are state of the art and have performed well in recent benchmarks (9–14). For MHC class I alleles with well-characterized binding specificity, the method is defined as a combination of the *NetMHC* and *NetMHCpan* methods, and for alleles with unknown binding specificity, the method is defined in terms of the *NetMHCpan* method combined with *PickPocket*. For details on the method and its benchmark performance refer to (15).

The submission site of the server can be seen in Fig. 1.

1. Select method. By default, the consensus method (*NetMHCcons*) is selected but each of the three individual prediction methods can be run separately.
2. Select Allele(s). To aid in navigation, the alleles listed by default are limited to the human supertype representatives, but all alleles from different human/animal loci can be selected under "Select species/loci" (the list of selectable alleles is limited to alleles with well-characterized binding specificity when using the *NetMHC* method). In the MHC allele selection field, multiple alleles can be selected but the selection is limited to 20 alleles per submission. Multiple alleles can also be inputted as a comma-separated list. For the pan-specific methods (*NetMHCcons*, *NetMHCpan*, and *PickPocket*) the user can upload a file containing the protein sequence of an MHC class I molecule that is not among the available, selectable alleles, and the method will perform peptide binding predictions for this molecule.
3. Provide input sequence. The input can either be in peptide raw text or protein FASTA format. In peptide format, each line is assumed to be a separate peptide. All peptides must be of equal length. In FASTA format, the sequence of each protein must be preceded by a line beginning with a ">." When FASTA input is used, multiple different epitope lengths from 8 to 11 residues can be selected.
4. Select output formatting. By default the output is sorted by the residue number, but the user can choose to sort the output by the predicted binding affinity. Predictions for all the input peptides given are by default but by setting "Filter output" to "Yes," only the peptides predicted to bind stronger than the defined thresholds are given in the output. The output can optionally be saved to a file readable by spreadsheet applications for further processing by the user.
5. Press submit.
6. Wait for the server to produce output. The output from the server consists of a list of peptides, each associated with three prediction values: 1-log50k(aff), Affinity, and %Rank. The 1-log50k value is the raw score provided by the prediction method, and is related to the predicted binding affinity value as 1-log(Aff)/log(50,000). The %Rank score gives % rank of the prediction score to a set of 200,000 random natural 9mer peptides. Thresholds can be selected for which peptides to report as strong binders (SB) and week binders (WB). The peptides are labeled as a strong binder if the %Rank score or the binding affinity is below the specified thresholds for the strong binders. Likewise, peptides are labeled as weak binders if the %Rank or the binding affinity is above the thresholds of strong binders, but below the specified threshold for the weak binders.

References to other well-performing methods for prediction of MHC class I binding can be found in one of the several reviews that have been written on the subject including a recent one from our group (14).

### 3.1. Prediction of MHC Class II Peptide Binding

For class I, alignment-free methods like the ones described earlier can readily be applied, since the binding motif is well characterized and most natural peptides that bind MHC class I are of the same length. For MHC class II, the situation is quite different due to the great variability in the length of natural MHC-binding peptides. This variation in ligand length makes alignment a crucial and integrated part of estimating the MHC-binding motif and predicting peptide binding. During the last decade, large efforts have been invested in developing data-driven prediction methods for MHC class II peptide binding. For an overview of these refer to one of the many reviews written on the theme including the one written by our group (16).

The binding of a peptide to a given MHC class II molecule is predominantly determined by the amino acids present in the peptide-binding core. However, peptide residues flanking the binding core (the so-called peptide flanking residues, PFR) do also to some degree affect the binding affinity of a peptide (17–19). Most published methods for MHC class II binding prediction focus on identifying the peptide-binding core only, ignoring the effects on the binding affinity of PFRs. In the work by (19) it was demonstrated that the additional information provided by the PFR leads to significantly improved predictions.

Two high-performing methods for MHC class II binding prediction developed by our group are *NetMHCII* (19) and *NetMHCIIpan* (20, 21). The *NetMHCII* method is allele-specific and allows for peptide–MHC binding predictions to a set of 14 HLA-DR, six HLA-DQ, six HLA-DP, and two mouse H2 class II alleles. *NetMHCIIpan* is HLA-DR pan specific, allowing for prediction of peptide binding to all HLA-DR molecules with a known protein sequence. Several benchmark studies have demonstrated these methods to be high performing and state of the art (22–25).

1. Select input sequences. Both methods accept input either as individual peptides in raw text format or as protein sequence(s) uploaded in FASTA format (see earlier). If protein sequences are uploaded, the user can specify the peptide length and predictions are made for each overlapping peptide of the specified length. Multiple MHC alleles can be specified.
2. Customize search. The input to (and output from) the *NetMHCIIpan* method is very similar to that of *NetMHCII*. Only does the *NetMHCIIpan* method (as was the case for MHC class I methods described earlier) allow the user to upload a file containing the protein sequence of an HLA-DR

molecule that is not among the available, selectable alleles, and the method will perform binding predictions for this molecule. Likewise the user can define the prediction score threshold values used to classify prediction as strong and weak binders. Also can the output from the *NetMHCIIpan* server be saved to a file readable by most spreadsheet applications for further processing by the user.

3. Select output formatting. By default the output is sorted by the residue number but the output can also be sorted by affinity. Predictions for all peptides are by default given but by setting a "Threshold," only the peptides predicted to bind stronger than the defined threshold (in 1-log50k units) are given in the output.
4. Press Submit.
5. Wait for output. As for the MHC class I prediction server described earlier, the output from the MHC class II prediction servers consists of a list of peptides, each associated with the predicted binding core and three prediction values: 1-log50k(aff), Affinity, and %Rank. The 1-log50k value is the raw score provided by the prediction method, and is related to the predicted binding affinity value as 1-log(Aff)/log(50,000). The %Rank score gives % rank of the prediction score to a set of 200,000 random natural peptides. Peptides are labeled as a strong binder if the binding affinity is below 50 nM. Likewise, peptides are labeled as a weak binder if the binding affinity is below 500 nM.

## 4. MHCMotif-Viewer: Browsing and Visualization of MHC Class I and Class II Binding Motifs

The number and binding specificity diversity of MHC molecules can be overwhelming for most users. To help get an overview, we have developed the *MHCMotifViewer* server (http://www.cbs.dtu.dk/biotools/MHCMotifViewer/). The homepage is shown in Fig. 2.

1. Select species/loci. By clicking on "Human alleles," different loci can be selected. For other species the user is taken directly to a list of alleles.
2. Select allele. Clicking on one of the thumbnail pictures will create a larger logo for that allele. This is shown for HLA-A*01:03 in the right panel of Fig. 2. On the x-axis the nine positions in the binding motif are given. The height of the columns of letters at each position corresponds to the predicted contribution to binding on that position calculated according to the formula developed by Kullback–Leibler (26). The amino acids for which their frequency differs the most

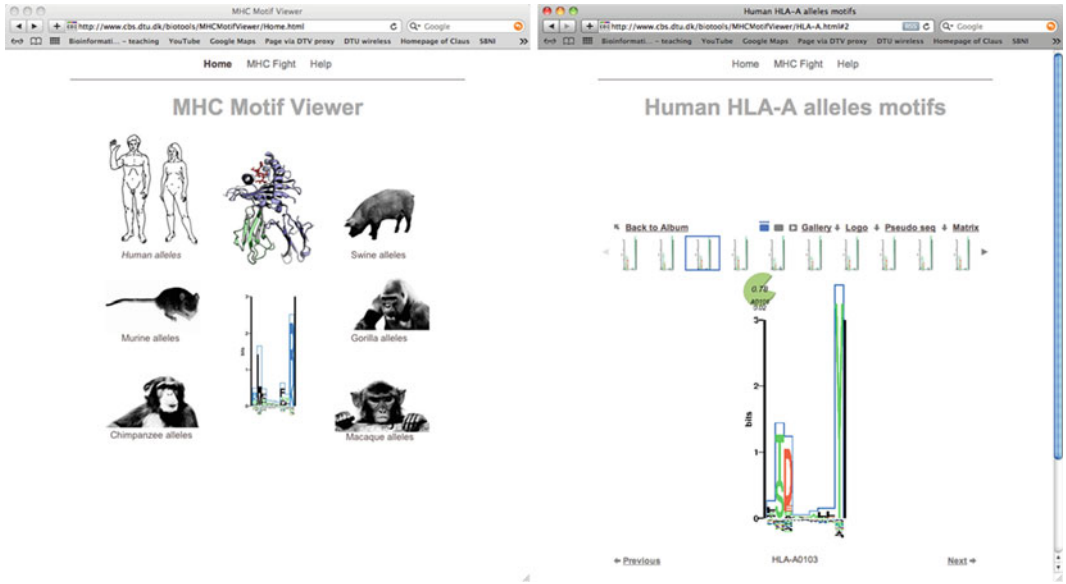


Fig. 2. The *MHCMotifviewer* server. Left panel shows the homepage of the *MHCMotifViewer* server where the organism can be selected. Human, murine, chimpanzee, swine, gorilla, and macaque alleles can be browsed. In the right panel an allele from the Human HLA-A loci (HLA-A*01:03) is selected and its motif is displayed as the sequence logo representation.

from the background frequency for that amino acid in proteins in general are shown with the highest letters. The overrepresented amino acids are shown above the x-axis, and the underrepresented ones below.

The binding motif of up to four different alleles can be shown side by side by clicking on "MHC Fight." By default, all four alleles are the same, but by clicking on the blinking curser, the allele name can be changed by deleting (part of) the name using the backspace key and typing the new name. By holding the curser over the "K" button, the display will shift between showing a Kullback–Leibler (K), and a Sequence frequency (S)-based logo. In a sequence frequency-based logo the relative height of each letter within a column is proportional to the frequency of the corresponding amino acid at that position. A more detailed explanation can be found in (27).

## 5. HLArestrictor: Patient-Specific HLA Restriction Elements and Optimal Epitopes Within Peptides

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. To reduce this complexity, one could conveniently exploit a commonly used approach of T cell epitope discovery: testing overlapping peptides (OLP) with a length of 15–18 amino acids in IFNγ release, ELISpot, or flow cytometric intracellular staining assays. Given a positive peptide it is, however, not a simple task to find the actual stimulatory peptide (minimal epitope) and the presenting HLA restriction element. By way of example, a 15mer peptide tested positive in a patient with six different HLA class I molecules could potentially be explained by any one of the possible 22 * 6 = 132 8–11mer HLA combinations. To lower this experimental burden, we have developed an immunoinformatics method, *HLArestrictor* (www.cbs.dtu.dk/services/HLArestroctor) (28), which has been tailored to support CTL epitope discovery in individual subjects. As inputs, the method requires the amino acid sequence of the positive peptide(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 peptides(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. The vehicle behind the *HLArestrictor* is the *NetMHCpan* method, and the Webpage interface bears a high resemblance to the interfaces for *NetMHCpan, NetMHCIIpan,* and *NetMHCcons.*

1. Select input sequences. Multiple peptide sequences can be uploaded in FASTA format.
2. Select HLA alleles. The host HLA allele names can be selected or typed in.
3. Select lengths of epitopes. The lengths of the predicted minimal epitopes can be specified.
4. Select prediction threshold. Threshold values defining how the prediction scores are interpreted can be specified in terms of threshold values for strong and weak binding peptides.

With default settings, the server will scan all possible 8, 9, 10, and 11mer peptides from the target peptides(s) for binding to all HLA alleles of the host and report peptides with %Rank score less than or equal to 0.5 or affinity stronger than 50 nM as strong binders, and peptides with %Rank score less than or equal to 2 or affinity stronger than 500 nM as weak binders.

## 6. Interpreting the Output from the Prediction Servers

All the prediction servers described here provide three prediction scores for each peptide, as well as a label classifying the peptides into groups of strong and weak binders. For the end user, these prediction values are meant to serve as a guide to make rational peptide selections for epitope discovery and/or interpretation of immune responses. This opens for questions on how to define relevant thresholds relating prediction values to likelihoods of a peptide being a T cell epitope. It is becoming apparent that not all MHC molecules present peptides at the same binding threshold (29, 30). The two distinct prediction values (affinity and %Rank) are included to capture these intrinsic differences between MHC molecules in terms of binding threshold for presentation of peptides. Large benchmark studies have demonstrated that the vast majority of known CTL epitopes are characterized by having a %Rank score less than or equal to 2 or an affinity stronger than 500 nM (28, 31, 32). These numbers are hence used as default values for the definition of weak binding peptides for all MHC class I prediction methods. For MHC class II the situation is less clear. While it is clear that the prediction values correlate strongly with the measured binding affinity, few studies have investigated the direct correlation between %Rank score, predicted affinity values, and the likelihood of a peptide being immunogenic. The default values for the classification of peptides as weak and strong binders are hence poorly justified for MHC class II, and the relationship to the likelihood of being immunogenic is at the best poorly investigated. However, for both MHC class I and class II it is clear that using the prediction score to rank peptides provides a highly cost-effective tool to guide the experimental efforts in a direction that increases the likelihood of discovering immunologically important responses.

## 7. The MHC Class I Antigen Presentation Pathway

As part of the protein recycling machinery, proteins in our cells are cut into shorter peptides by the proteasome. These peptides may bind to the Transporter associated with Antigen Processing (TAP) and be transferred to the Endoplasmic Reticulum (ER). Inside the ER, peptides may be further trimmed, bind the MHC class I molecules, and be transported along with it to the cell surface. If the peptide is of nonself origin, the peptide–MHC complex may bind to a T Cell Receptor (TCR) on a cytotoxic T cell, which will then initiate an immune response. More detailed descriptions of and

references to these processes can be found in other chapters of this book. The three most essential of the above steps (cleavage by the proteasome, transport by TAP, and binding to MHC class I) have been modeled by bioinformatics methods that can predict which peptides from a given protein/organism are most likely to be presented to the immune system.

## 8. NetChop: Proteasomal Cleavages (MHC Class I Ligands)

A method has been developed, which predicts proteasomal cleavage sites. The method is called *NetChop* (33), and a server is available at http://www.cbs.dtu.dk/services/NetChop/.

1. Select prediction method. Two different versions of the method exist: "C term 3.0" and "20S 3.0." They differ by the sets of data they have been trained on. While *NetChop 20S 3.0* has been trained on in vitro constitutive proteasome protein digests, *NetChop C term 3.0* has been trained on natural MHC class I ligands. The rationale for the latter is that the proteasome most likely has generated the ligand's C-terminal ends. *NetChop C term 3.0* predicts the C-terminal end of CTL epitopes with a higher specificity than *NetChop 20S 3.0* (has fewer false positives). The main reason for this is that since it is trained on natural ligands, it predicts a combination of MHC class I binding, TAP transport efficiency, and proteasomal cleavage.
2. Select input sequence. The input to the server is proteins or peptide fragments in FASTA format (see earlier). The method assigns a score in the range 0–1 to each residue in the input sequence. The higher the score, the more likely it is that the proteasome cleaves after this residue. Note that the score refers to cleavage of the peptide bond on the C-terminal side of the residue to which the score is assigned.
3. Select prediction threshold.

   By default, 0.5 is used as the threshold for predicted proteasomal cleavage. In the output, scores above the threshold are assigned an "S" in the C (cleavage) column, while lower scores are assigned a ".".

## 9. NetCTL and NetCTLpan: Integrated Class I Antigen Presentation

Two methods that integrate predictions of proteasomal C-terminal cleavage, TAP transport efficiency, and MHC class I binding for the overall prediction of MHC class I presentation called *NetCTL*

and *NetCTLpan* have been developed by our group. The *NetCTL* method (34) is available at http://www.cbs.dtu.dk/services/NetCTL/. For prediction of proteasomal cleavage, it uses *NetChop C term 3.0* (see above). Predictions of TAP transport efficiency are based on the weight matrix-based method described by Peters et al. (35). For predictions of MHC class I binding, *NetMHC* (see above) is used.

1. Select input sequence. The input to the server is proteins or peptide fragments in FASTA format (see earlier).
2. Select Allele/supertype. The user must specify for which of the 12 MHC class I supertypes the predictions should be performed (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, or B62; for a definition of supertypes see (1)). *NetCTL* integrates the individual scores from *NetChop*, the TAP matrix, and *NetMHC* into one, overall score. To allow for comparison between different MHC class I supertypes, the rescaled affinity is used (see (34) for details on how the rescaled affinity is calculated).
3. Select weighting of processing steps. As default, the relative weight of C-terminal cleavage is 0.15, while it is 0.05 for TAP transport efficiency. The default weights have been found to result in optimal performance, but can be changed by the user.
4. Select prediction threshold. The user can also specify which threshold to use for defining a CTL epitope. By default it is 0.75.
5. Select sorting of output. Lastly, the user can specify how the 9mers of the input sequence should be sorted in the output. In the default "no sort" option, the 9mers are listed according to the order in which they appear in the input sequence. Alternatively, they can be sorted according to the combined score, MHC binding, proteasomal cleavage, or TAP. For each 9mer sub-peptide in the input sequence, the output will list the predicted affinity and the prediction scores of proteasomal cleavage, TAP binding, and finally a combined score. If the combined score is above the selected threshold for defining an epitope, it is marked by an "E."

*NetCTLpan* is an extended and improved version of *NetCTL*, which is available at http://www.cbs.dtu.dk/services/NetCTLpan/ and described in detail in (30). The C-terminal proteasomal cleavage and TAP transport efficiency are predicted as for the *NetCTL* method, while MHC class I binding is based on the *NetMHCpan* method. While *NetCTL* only allows for predictions of peptides restricted by one of the 12 MHC class I supertypes, *NetCTLpan* allows for predictions of CTL epitopes binding any MHC class I molecule for which the protein sequence is known. As for the above-described pan prediction methods, it is additionally possible

to paste in or upload a file containing the protein sequence of an MHC class I molecule that is not among the available, selectable alleles, and the method will perform CTL epitope predictions for this molecule. *NetCTLpan* furthermore performs predictions for 8–11mers. The Webpage interface of *NetCTLpan* bears a high resemblance to the interfaces of *NetCTL*. One difference is that it is possible to select a threshold that the combined score must exceed for the predictions to be displayed in the output page. By default, this threshold is –99.9, which results in all predictions being displayed. In the output page, the same values are listed as in the *NetCTL* output. Additionally, the %Rank value is given (see above for definition of the %Rank value).

## References


1. Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics 50(3–4):201–212
2. Parker KC, Bednarek MA, Coligan JE (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol 152(1):163–175
3. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50(3–4):213–219
4. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. Nucleic Acids Res 36(Web Server issue):W509–W512. doi:gkn202 (pii) 10.1093/nar/gkn202
5. Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, Brunak S, Lund O (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci 12(5):1007–1017. doi:10.1110/ps.0239403
6. 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):1–13. doi:10.1007/s00251-008-0341-z
7. 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(8):e796. doi:10.1371/journal.pone.0000796
8. Zhang H, Lund O, Nielsen M (2009) The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. Bioinformatics 25(10):1293–1299. doi:10.1093/bioinformatics/btp137
9. Peters B, Bui HH, Frankild S, Nielson M, Lundegaard C, Kostem E, Basch D, Lamberth K, Harndahl M, Fleri W, Wilson SS, Sidney J, Lund O, Buus S, Sette A (2006) A community resource benchmarking predictions of peptide binding to MHC-I molecules. PLoS Comput Biol 2(6):e65. doi:06-PLCB-RA-0058R2 (pii) 10.1371/journal.pcbi.0020065
10. 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. doi:1471-2172-9-8 (pii) 10.1186/1471-2172-9-8
11. Zhang GL, Ansari HR, Bradley P, Cawley GC, Hertz T, Hu X, Jojic N, Kim Y, Kohlbacher O, Lund O, Lundegaard C, Magaret CA, Nielsen M, Papadopoulos H, Raghava GP, Tal VS, Xue LC, Yanover C, Zhu S, Rock MT, Crowe JE Jr, Panayiotou C, Polycarpou MM, Duch W, Brusic V (2011) Machine learning competition in immunology—Prediction of HLA class I binding peptides. J Immunol Methods. doi:S0022-1759(11)00255-9 (pii) 10.1016/j.jim.2011.09.010
12. 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(1):83–89. doi:10.1093/bioinformatics/btn579

13. Zhang L, Udaka K, Mamitsuka H, Zhu S (2011) Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools. Brief Bioinform. doi:bbr060 (pii) 10.1093/bib/bbr060
14. Lundegaard C, Hoof I, Lund O, Nielsen M, Lundegaard C, Hoof I, Lund O, Nielsen M (2010) State of the art and challenges in sequence based T-cell epitope prediction. Immunome Res 6(Suppl 2):S3. doi:1745-7580-6-S2-S3 (pii) 10.1186/1745-7580-6-S2-S3
15. Karosiene E, Lundegaard C, Lund O, Nielsen M (2011) NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. Immunogenetics. doi:10.1007/s00251-011-0579-8
16. Nielsen M, Lund O, Buus S, Lundegaard C (2010) MHC class II epitope predictive algorithms. Immunology 130(3):319–328. doi:10.1111/j.1365-2567.2010.03268.x
17. Godkin AJ, Smith KJ, Willis A, Tejada-Simon MV, Zhang J, Elliott T, Hill AV (2001) Naturally processed HLA class II peptides reveal highly conserved immunogenic flanking region sequence preferences that reflect antigen processing rather than peptide-MHC interactions. J Immunol 166(11): 6720–6727
18. Lovitch SB, Pu Z, Unanue ER (2006) Amino-terminal flanking residues determine the conformation of a peptide-class II MHC complex. J Immunol 176(5):2958–2968. doi:176/5/2958 (pii)
19. Nielsen M, Lund O (2009) NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinformatics 10:296. doi:10.1186/1471-2105-10-296
20. Nielsen M, Lundegaard C, Blicher T, Peters B, Sette A, Justesen S, Buus S, Lund O (2008) Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. PLoS Comput Biol 4(7):e1000107. doi:10.1371/journal.pcbi.1000107
21. Nielsen M, Justesen S, Lund O, Lundegaard C, Buus S (2010) NetMHCIIpan-2.0—Improved pan-specific HLA-DR predictions using a novel concurrent alignment and weight optimization training procedure. Immunome Res 6:9. doi:1745-7580-6-9 (pii) 10.1186/1745-7580-6-9
22. Lin HH, Zhang GL, Tongchusak S, Reinherz EL, Brusic V (2008) Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. BMC Bioinformatics 9(Suppl 12):S22. doi:10.1186/1471-2105-9-S12-S22
23. Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B (2008) A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 4(4):e1000048. doi:10.1371/journal.pcbi.1000048
24. Bordner AJ, Mittelmann HD (2010) MultiRTA: a simple yet reliable method for predicting peptide binding affinities for multiple class II MHC allotypes. BMC Bioinformatics 11:482. doi:10.1186/1471-2105-11-482
25. Zhang GL, Deluca DS, Keskin DB, Chitkushev L, Zlateva T, Lund O, Reinherz EL, Brusic V (2010) MULTIPRED2: a computational system for large-scale identification of peptides predicted to bind to HLA supertypes and alleles. J Immunol Methods. doi:S0022-1759(10)00345-5 (pii) 10.1016/j.jim.2010.11.009
26. Kullback S, Leibler RA (1951) On Information and Sufficiency. Ann Math Statist 22: 76–86
27. Rapin N, Hoof I, Lund O, Nielsen M. The MHC motif viewer: a visualization tool for MHC binding motifs. Curr Protoc Immunol. 2010; Chapter 18: Unit 18 17. doi:10.1002/0471142735.im1817s88.
28. Erup Larsen M, Kloverpris H, Stryhn A, Koofhethile CK, Sims S, Ndung'u T, Goulder P, Buus S, Nielsen M (2011) HLArestrictor–a tool for patient-specific predictions of HLA restriction elements and optimal epitopes within peptides. Immunogenetics 63(1):43–55. doi:10.1007/s00251-010-0493-5
29. 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(3):1526–1532. doi:182/3/1526 (pii)
30. Stranzl T, Larsen MV, Lundegaard C, Nielsen M (2010) NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics 62(6):357–368. doi:10.1007/s00251-010-0441-4
31. Hoof I, Perez CL, Buggert M, Gustafsson RK, Nielsen M, Lund O, Karlsson AC (2010) Interdisciplinary analysis of HIV-specific CD8+ T cell responses against variant epitopes reveals restricted TCR promiscuity. J Immunol 184(9):5383–5391. doi:jimmunol.0903516 (pii) 10.4049/jimmunol.0903516
32. Larsen MV, Lelic A, Parsons R, Nielsen M, Hoof I, Lamberth K, Loeb MB, Buus S, Bramson J, Lund O (2010) Identification of CD8+ T cell epitopes in the West Nile virus polyprotein by reverse-immunology using

NetCTL. PLoS One 5(9):e12697. doi:10.1371/journal.pone.0012697

33. Nielsen M, Lundegaard C, Lund O, Kesmir C (2005) The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. Immunogenetics 57(1–2):33–41. doi:10.1007/s00251-005-0781-7
34. Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, Nielsen M (2005) An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. Eur J Immunol 35(8):2295–2303. doi:10.1002/eji.200425811
35. Peters B, Bulik S, Tampe R, Van Endert PM, Holzhutter HG (2003) Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. J Immunol 171(4):1741–1749