

T-Cell Epitope Prediction

George N. Konstantinou

Abstract

An epitope is a part of an immunogenic protein that can be recognized by the immune system. The peptides that can be recognized by the T-cell receptors after a particular antigen has been intracellularly processed, bound to at least one MHC molecule and expressed on the surface of the antigen presenting cell as a MHC-peptide complex, are called a T-cell epitope. Individuals who have at least one MHC molecule able to most avidly bind to allergenic amino acid sequences from an allergen, and at the same time have the appropriate T-cell clone that can recognize this MHC-peptide complex, are expected to be genetically prone to allergic reactions against that allergen. This possibility can be examined *in silico* by utilizing modern computational techniques that are based on sophisticated mathematics and statistics. The design principles of these techniques are different and therefore variations in their predictions are expected. The available software programs that have been developed on this basis are able to combine the increasing amount and complexity of the available experimental data that have been organized in immunoinformatics databases to predict potential allergen T-cell epitopes. All relevant T-cell epitope prediction methods can be accessed online as a freeware.

Key words T-Cell epitope, Prediction, *In silico*, Bioinformatics, Peptide, Allergy, Allergen, Antigen, Processing

1 Introduction

1.1 Epitope

Definition

All parts of an antigen that can be recognized and bind to either immune cells (e.g., T cells and B cells), free molecules (e.g., antibodies), or cell-surface proteins [e.g., major histocompatibility complex (MHC)] are characterized as epitopes. The epitopes are distinguished into two types according to the primary, secondary, or tertiary structure of the molecule they belong:

1. A continuous stretch of amino acid residues within a peptide is known as a *linear*, *sequential*, or *continuous* epitope.
2. Noncontinuous amino acid segments that are distantly separated within the same molecule but are brought into proximity by the folding of the molecule are called *conformational* or *discontinuous* epitopes.

B-cell epitopes can be either linear or conformational, but T-cell epitopes are always linear [1].

All antigens are displayed by antigen presenting cells (APC) after being internalized (most of the times by phagocytosis) and digested into smaller fragments/peptides. These peptides are presented in MHC molecules on APC surfaces. MHC molecules can bind only short peptides and consequently, all antigenic proteins have to be converted into peptides, a procedure that is known as *antigen processing*. There are two types of antigens that can be processed: intracellular or class I antigens and extracellular or class II antigens. The intracellular antigens are displayed on the MHC class I molecules that are found on the cell surface of all nucleated cells. The extracellular antigens are displayed on the MHC class II molecules found basically on dendritic cells that have phagocytosed exogenous pathogens (e.g., bacteria, parasites, toxins, allergens, etc.). In general, MHC class I molecules obtain peptides mainly from cytosolic proteins and MHC class II molecules from proteins in intracellular vesicles [2].

Each MHC molecule consists of an extracellular peptide-binding cleft or groove formed by appropriate folding of the amino termini of the MHC-encoded proteins. Each cleft has two walls and a floor that may vary on the amino acids composition among the different MHC alleles and is responsible for binding with the peptides that will be displayed to the antigen receptors of the T cells [2].

MHC molecules and peptides are bound with a noncovalent interaction mediated by amino acid residues on both the MHC clefts and the processed peptides. This interaction is class dependent. The association of a peptide with the MHC class I cleft depends mainly on the electrostatic interactions of the positively charged amino (N)-terminus and the negatively charged carboxyl (C)-terminus of the processed peptide to the MHC molecule. In most MHC class I molecules, the floor of the cleft contains hydrophobic pockets where hydrophobic amino acid residues of the C-terminal end of these peptides (e.g., valine, isoleucine, leucine, or methionine) can bind. Some MHC class I molecules have a preference for a basic amino acid residue at the C terminus (most of the times lysine or arginine). The hydrophobic bonds do not exclude complementary binding of other amino acid residues of a peptide (called anchor residues because they literally anchor the peptide in the cleft of the MHC molecule) with electrostatic, hydrogen, or van der Waals interactions that vary in strength according to avidity and affinity of this bond. In MHC class II molecules, such interactions may occur not only with the floor but also with the walls of the cleft contributing additionally to peptide binding and its strength [2].

The peptide-binding cleft of MHC class I molecules is large enough to accommodate peptides no larger than 8–11 amino acids long because the ends of their cleft are closed (Fig. 1). In contrast,

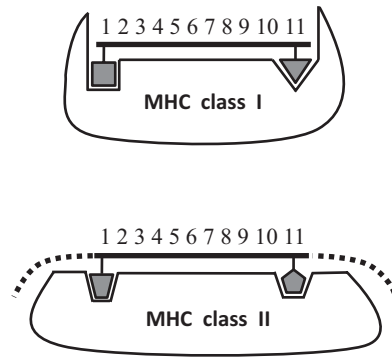


Fig. 1 Schematic representation of structure of MHC/peptide complexes

MHC class II molecules have open ends of the peptide-binding cleft so that they can bind peptides of 10–30 residues long or even longer, with the optimal length being 12–16 residues [2].

1.2 Antigen Processing T-Cell Epitope Dependence

The antigen processing procedure is dependent on the intracellular mechanisms of cleavage, transportation, and formation of the MHC binding.

In the MHC class I intracellular pathway, antigens in the cytosol are processed by the proteolytic activity of the proteasomes. Because MHC class I molecules are synthesized in the endoplasmic reticulum (ER), a mechanism is needed to deliver cytosolic cleaved peptides into the ER. This delivery is mediated by the transporter associated with antigen processing (TAP). TAP exhibits a broad range of specificities and therefore it is expected to have higher transport affinity for some peptides over others. Its optimal activity, however, needs peptides 8–16 amino acids long that contain basic or hydrophobic carboxyl termini. The TAP-peptide dimer is transported to the ER, where they bind to class I MHC molecules. Once MHC class I molecules are loaded with the peptide, they exit the ER and are translocated to the cell surface.

The MHC Class II processing pathway involves the endocytosis of extracellular antigens into vesicles and their enzymatic degradation that takes place in these vesicles. MHC Class II molecules cannot bind with any peptides in the ER so that all ER-generated peptides are always associated with the MHC class I molecules as aforementioned. The MHC class II molecules are transported into exocytic vesicles toward the cell surface where they meet with the endocytic vesicles containing the processed antigens. The generated peptides are then associated with the appropriated MHC class II molecules in this fusion vesicle that is transported to the cell surface where the peptide-MHC class II complex is expressed [2].

1.3 MHC-Peptide Complex T-Cell Epitope Recognition

The antigen receptors of the T cells recognize the complex of the MHC molecule bound to a specific peptide. The fine specificity of antigen recognition is attributed to the peptide while the MHC

residues are responsible for the MHC restriction of the T cells. Part of the bound peptide in the MHC cleft is exposed from the open top of the cleft, along with the adjacent amino acid residues of this MHC cleft. Only these exposed amino acid side chains can be recognized by a specific for this antigen complex T-cell receptor. It is self-explanatory that variations in either the peptide or the MHC cleft influence presentation and recognition of that peptide by the T cells (e.g., existence of a specific amino acid residue on either the MHC or the presenting peptide may enhance the immunogenicity of that peptide if this residue strengthens the binding between MHC-peptide complex and T-cell receptor). Thus, the structure of a processed peptide and the physicochemical properties of the component amino acids define the antigenic specificity of this epitope. Any change in the structure, including a single amino acid substitution or destruction, may influence antigenicity.

1.4 Immunogenicity of Protein Antigens

The immunogenicity of each antigen depends on the ability of the antigen-processing pathways to generate peptides that are able to bind with high avidity and affinity to the MHC molecules this organism has inherited. The MHC genes are the most highly polymorphic genes known. Because of this polymorphism a single MHC allele can present many different peptides to T cells and therefore the different MHC molecules are expected to present a great variety of different amino acid sequences and to cover all immunogenic amino acid sequences that exist in nature.

Each MHC molecule can accommodate only one peptide at a time. Therefore, only individuals that express MHC alleles that can bind this particular peptide may respond to that peptide. However, it has to be noted that a single T cell will be able to recognize only one of these huge amounts of MHC-peptide complexes. For instance, the individuals who have at least one MHC class II molecule able to most avidly bind to a peptide that is derived from grass pollen, and at the same time have the appropriate T-cell clone that can recognize this complex, are expected to be genetically prone to allergic reactions against that pollen.

2 Materials

Because of the MHC polymorphism, every single individual is expected to be unique. However, many of the different MHC molecules express similar peptide specificities because they have structural similarities according to which they are grouped into MHC supertypes [3]. A particular supertype is expected to bind to the same peptides. There are experimental data defining peptide sequences as epitopes able to bind with known MHC molecules and peptides that do not activate the MHC processing pathway (non-epitopes). These data are organized in immunoinformatics

databases. To exploit the opportunities given by this huge amount of information, mathematical and statistical algorithms have been developed to analyze these peptide sequences so that T-cell epitope predictions can become an option. Based on the aforementioned antigen processing, this prediction cannot be simplified to directly utilizing peptide sequences and MHC or T-cell receptor structures. Proteosomal cleavage and TAP transportation should be taken into account since they are of equal importance, although unfortunately, there is scarcity of such experimental data.

There are online platforms that can combine the available information from databases that are constantly updated. This increasing amount and complexity of the available information complicates even more the prediction methods, highlighting the need to develop modern and sophisticated prediction algorithms based on statically learning theory. All these available tools, which can be accessed online and are available as freeware, will be presented in the following section.

The opportunity of *in silico* epitope prediction offers alternative options to minimize the cost of producing all the synthetic peptides needed to be experimentally tested and the enormous working time needed to perform all relevant experiments.

The accuracy of T-cell epitope predictions has been tested in Allergology, too. The potential of utilizing this analysis in Allergy field has implications not only in diagnostics but also for potential development of therapeutic options such as allergen immunotherapy. Utilizing the same online platforms and software, allergen T-cell epitope prediction has been proven to be of high precision when bioinformatics and experimental data are compared [4, 5]. This precision is very important for allergen epitope determination, for peptide-based allergen immunotherapy design, for allergenicity evaluation and modification, and for understanding immune tolerance mechanisms.

The constantly improved accuracy along with the cost and time minimization by following these approaches explain why *in silico* allergen T-cell epitope prediction is gaining ground in experimental settings.

3 Methods

There are three types of T-cell epitope prediction tools: MHC-binding prediction, Processing Prediction and Immunogenicity Prediction. The Immune Epitope Database (IEDB) and Analysis Resources platform incorporates all relevant available information of comparative studies having a state-of-the-art dynamic structure (*see Note 1*). At the same time, it follows continuous evaluation of new methods and datasets, and implements these data in a database that is constantly updated (*see Note 2*). Therefore, its platform is

recommended as one of the best options to perform T-cell epitope predictions utilizing almost all available tools or, even better, a combination of them (*see* **Notes 3** and **4**). The following steps are based on the IEDB interface (*see* **Note 5**).

3.1 MHC-Binding Prediction

The peptide-binding strength is estimated by using two different types of thresholds to specify strong and weak binders: the binding affinity and a percentile rank.

The binding affinity is expressed in nanomolar of the half maximal inhibitory concentration (IC_{50}) values. The peptides are distinguished, according to the binding strength of the core peptides that are bound in the MHC cleft, into strong binders (predicted $IC_{50} < 50$ nM) or weak binders ($IC_{50} \leq 500$ nM). Nonsignificant peptides have $IC_{50} > 500$ nM but not above 5000 nM. MHC-binding prediction algorithms predict IC_{50} values for the specific peptide of interest that is supposed to bind a specific MHC molecule. Although, binding to MHC is necessary, it is not sufficient by itself for T-cell recognition.

The percentile rank is generated by comparing the peptide's IC_{50} against those of a set of random peptides derived, e.g., from Universal Protein Resource (UniProt) database. The smaller the numbered percentile rank is, the higher the expected affinity. In some algorithms the median percentile rank of the methods used is reported as the representative percentile rank.

There are tools that predict peptide binding to MHC class I or MHC class II molecules (*see* **Note 6**).

3.1.1 MHC-I-Binding Predictions

1. Specify sequences: Specify the sequences that will be scanned for binding peptides. Sequences can be entered either directly into the appropriate field or can be taken from appropriate protein databases like RCSB Protein Data Bank, NCBI protein database, or UniProt (*see* **Note 7**). The FASTA format is almost always the standard for inputting protein/peptide sequences.
2. Choose a prediction method: There are several different prediction methods that can be used to MHC class I binding:
 - (a) Artificial neural network (ANN) [6].
 - (b) Stabilized matrix method (SMM) [7].
 - (c) SMM with a Peptide: MHC-Binding Energy Covariance matrix (SMMPMBEC) [8].
 - (d) Scoring Matrices derived from Combinatorial Peptide Libraries (Comlib_Sidney2008) [9].
 - (e) Consensus [10].
 - (f) NetMHCpan [11].
 - (g) NetMHCcons [12].
 - (h) PickPocket [13].

The IEDB has a default prediction method for each MHC I allele which is based on availability of predictors and previously observed predictive performances. According to current (May 2016) performances the expected predictive performance of the aforementioned methods in decreasing order is Consensus, ANN, SMM, NetMHCpan, and CombLib.

3. Specify predictions details: Predictions are based on species-specific available data. The length and the type of the MHC molecule influence the predictions, too. Multiple allele/length pairs can be submitted at a time. For some allele/peptide length combinations, no prediction tools exist because there is too little experimental data available to generate them. Therefore, the selection of the allele/peptide length may influence the prediction outcome. Some platforms allow selection of distinct set of MHC alleles and peptide lengths according to existing data. If not a particular MHC allele of interest exist, there is an option to choose among frequently occurring alleles. Frequently occurring alleles are those that occur in at least 1% of the human population. A small set of such alleles can be compiled to cover most of the population of interest.
4. Specify the output: If the prediction will be based on IC_{50} there is an option to limit the outcomes displayed by defining an upper cutoff value. This option may influence the time needed for the calculations to be generated and the outcome to be displayed.
5. Submit the prediction: Click the submit button and wait for the result screen to appear.
6. Interpreting predicted affinities and percentile ranks: The predicted output can be given in units of $IC_{50}nM$. In that case, a lower number indicates higher affinity. Most known epitopes have high or intermediate affinity. Some epitopes have low affinity, but until now there is no known T-cell epitope with an IC_{50} value greater than 5000 nM. It has to be noted that although the predicted values seem to be quantitative, deviations are anticipated from the expected experimental IC_{50} values. This expected variability can be explained mainly by the mathematical/biostatistical way the prediction algorithms function and the differences in training and evaluation datasets that have been used in the developing process of these algorithms.

3.1.2 MHC-II-Binding Predictions

1. Specify sequences: Same as **step 1** in Subheading 3.1.1. MHC-I-binding predictions.
2. Choose a prediction method: The mostly recommended prediction methods for MHC class II binding are:
 - (a) Consensus method [14].
 - (b) Combinatorial library.

- (c) NN-align (netMHCII-2.2) [15].
- (d) SMM-align (netMHCII-1.1) [16].
- (e) Sturniolo [17].
- (f) NetMHCIIpan [18].

The IEDB has a default prediction method for each MHC class II molecule, which is based on availability of predictors and previously observed predictive performance that are based on large-scale evaluations of the performance of the different MHC class II binding algorithms [14, 19, 20].

3. Specify predictions details: Same as **step 3** in Subheading 3.1.1 MHC-I-binding predictions (specify species and select MHC alleles of interest). As in the MHC class I prediction method, the available predictions options are limited to alleles in which each of the prediction method has been trained and has available data.

Steps 4 and 5 same as **steps 4 and 5** in Subheading 3.1.1. MHC-I-binding predictions.

6. Interpreting predicted results: The predicted output can be given in units of IC₅₀nM for CombLib and SMM_align that are interpreted as mentioned in Subheading 3.1.1, **step 6**. The prediction result for Sturniolo is given as raw score in which higher score indicates higher affinity.

3.2 Processing Prediction

These tools predict epitope candidates based upon the processing of peptides in the cell.

3.2.1 MHC-I Processing Predictions

With this prediction method predictors of proteasomal processing, TAP transport, and MHC binding are combined to calculate an overall score for each antigen that is processing through the MHC class I antigen presentation pathway to become a potential MHC class I ligand.

1. Specify sequences: Same as **step 1** in Subheading 3.1.1 MHC-I-binding predictions.
2. Choose a prediction method: Same as **step 2** in Subheading 3.1.1 MHC-I-binding predictions [according to current (May 2016) performances and comparisons the best prediction method seems to be the NetMHCpan because it supplies IC₅₀ values and is available for all MHC alleles).
3. Specify predictions details: Same as **step 3** in Subheading 3.1.1 MHC-I-binding predictions (*see Note 8*).
4. Choose a proteasomal cleavage prediction type: There are two types of proteasomes: the constitutively expressed and the inducible immune-proteasomes regulated by IFN- γ secretion. Inducible proteasomes seem to enhance antigen presentation

and should be the preferable option to make a prediction when no more information is available.

5. Specify TAP prediction parameters: High affinity peptides exhibit high transport rates. The TAP score estimates an effective $-\log(\text{IC}_{50})$ values for the binding to TAP of a peptide or its N-terminal prolonged precursors.
6. Specify the output: Results can be sorted by proteasome, TAP, MHC, or processing score.
7. Submit the prediction: Click the submit button and wait for the result screen to appear.

3.2.2 Neural Network-Based Prediction of Proteasomal Cleavage Sites and T-Cell Epitopes

There are three methods to approach proteasomal processing and cleavage that are all based on neural network.

- (a) NetChop predicts cleavage sites of the human proteasome.
 - (b) NetCTL predicts cytotoxic T-cell (CTL) epitopes in protein sequences.
 - (c) NetCTLpan (NetCTL update) allows for prediction of CTL epitope restricted to any MHC molecules of known protein sequence.
1. Choose a prediction method: Select either NetCHOP, NetCTL, or NetCTLpan method.
 2. Specify sequences: Same as **step 1** in Subheading 3.1.1.
 3. Method specific option: Depending on the selected method, enter the appropriate parameters or use the default options.
 4. Submit the prediction: Click the submit button and wait for the result screen to appear.
 5. Interpreting predicted results: A chart and a table outcome provides the prediction scores. NetCTL and NetCTLpan algorithms produce proteasome, TAP, MHC, or processing scores along with a global score for each peptide sequence.

3.2.3 Prediction of Peptides Naturally Processed by the MHC

This approach utilizes data obtained from MHC elution experiments with which the probability that a given peptide is naturally processed and binds to a given MHC molecule is assessed. The algorithm is named MHC-NP and is still unpublished, although it is one of the T-cell epitopes prediction option in IEDB. This method combines two predictors: one that has been trained to distinguish binding peptides from nonbinding peptides for a given allele, and one that has been trained to distinguish eluted peptides from other binding peptides. The generated score is based on the estimated probability that a processed peptide is eluted and a good binder. The greater the score, the more likely the peptide is naturally processed by the MHC allele.

Step 1 same as **step 1** in Subheading 3.1.1. MHC-I-binding predictions.

Step 2 same as **step 3** in Subheading 3.1.1 MHC-I-binding predictions (specify species and select MHC alleles of interest).

3. Specify the output: For each peptide, a score is predicted. This score is between 0 and 1 and the greater the value, the more likely a peptide is naturally processed by a specific allele.
4. Submit the prediction: click the submit button and wait for the result screen to appear.

3.3 Immunogenicity Prediction

This tool predicts the relative ability of a peptide/MHC class I complex to elicit an immune response and is based on the fact that some peptides are more immunogenic than others and therefore more likely to be T-cell epitopes [21]. This method has been validated for 9-mer peptides, but can produce prediction for any length.

Step 1 same as **step 1** in Subheading 3.1.1 MHC-I-binding predictions.

2. Choose a masking option: The masked positions are dependent on the MHC class I molecule on which the peptide is presented. The first, second, and the C-terminal amino acid are masked by default, but additional amino acids may be masked depending on the expected anchor positions each MHC molecule exhibits.
3. Submit the prediction: Click the submit button and wait for the result screen to appear.

4 Notes

1. Although the IEDB Analysis Resource is a good option for performing T-cell epitope predictions, most of the previously described methods that are incorporated into IEDB website have their own websites for independent analysis.
2. The existing databases are updated almost weekly. Therefore, results from analyses performed different time points may differ. The same applies for the version of the same computational technique utilized.
3. The results among the different available online software may differ not only because of the mathematical approach but also because of the different datasets that have been used for the development and training of these algorithms.
4. The recommended method as the best performing at a certain time is usually based on comparative studies performed up to this time point. Since new techniques are developed and published, and new comparative studies are performed, the method of choice may change accordingly.

5. The software used for T-cell epitope prediction is available online as a freeware and in some cases they are available as standalone downloads.
6. For the allergen T-cell epitope prediction, the MHC Class II processing pathway is the one of interest. However, the autoallergy theory [22, 23] obliges the description of methods for MCH Class I processing, too.
7. Predictions can be made for molecules with unknown epitopes, provided that the datasets used to develop and train the computational technique do not include known data for this molecule of interest.
8. Select MHC alleles of interest and specify species although it must be highlighted that both the proteasome and TAP predictions have been developed from human experimental data that seem to be species-specific. Therefore, interpreting predictions for epitope processing in other species should be done with extra caution.

References

1. Barlow DJ, Edwards MS, Thornton JM (1986) Continuous and discontinuous protein antigenic determinants. *Nature* 322(6081):747–748. doi:[10.1038/322747a0](https://doi.org/10.1038/322747a0)
2. Abbas AK, Lichtman AH, Pillai S (2015) Cellular and molecular immunology, 8th edn. Saunders/Elsevier, Philadelphia, PA
3. Reche PA, Reinherz EL (2007) Definition of MHC supertypes through clustering of MHC peptide-binding repertoires. *Methods Mol Biol* 409:163–173. doi:[10.1007/978-1-60327-118-9_11](https://doi.org/10.1007/978-1-60327-118-9_11)
4. Pascal M, Konstantinou GN, Masilamani M, Lieberman J, Sampson HA (2013) In silico prediction of Ara h 2 T cell epitopes in peanut-allergic children. *Clin Exp Allergy* 43(1):116–127. doi:[10.1111/cea.12014](https://doi.org/10.1111/cea.12014)
5. Ramesh M, Yuenyongviwat A, Konstantinou GN, Lieberman J, Pascal M, Masilamani M, Sampson HA (2016) Peanut T-cell epitope discovery: Ara h 1. *J Allergy Clin Immunol*. doi:[10.1016/j.jaci.2015.12.1327](https://doi.org/10.1016/j.jaci.2015.12.1327)
6. 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](https://doi.org/10.1110/ps.0239403)
7. Peters B, Sette A (2005) Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 6:132. doi:[10.1186/1471-2105-6-132](https://doi.org/10.1186/1471-2105-6-132)
8. Kim Y, Sidney J, Pinilla C, Sette A, Peters B (2009) Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. *BMC Bioinformatics* 10:394. doi:[10.1186/1471-2105-10-394](https://doi.org/10.1186/1471-2105-10-394)
9. Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, Peters B (2008) Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunol Res* 4:2. doi:[10.1186/1745-7580-4-2](https://doi.org/10.1186/1745-7580-4-2)
10. Moutafsi M, Peters B, Pasquetto V, Tschärke 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(7):817–819. doi:[10.1038/nbt1215](https://doi.org/10.1038/nbt1215)
11. 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](https://doi.org/10.1007/s00251-008-0341-z)
12. Karosiene E, Lundegaard C, Lund O, Nielsen M (2012) NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* 64(3):177–186. doi:[10.1007/s00251-011-0579-8](https://doi.org/10.1007/s00251-011-0579-8)
13. 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](https://doi.org/10.1093/bioinformatics/btp137)
14. 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](https://doi.org/10.1371/journal.pcbi.1000048)

15. 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](https://doi.org/10.1186/1471-2105-10-296)
16. Nielsen M, Lundegaard C, Lund O (2007) Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* 8:238. doi:[10.1186/1471-2105-8-238](https://doi.org/10.1186/1471-2105-8-238)
17. Sturniolo T, Bono E, Ding J, Radrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J (1999) Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* 17(6):555–561. doi:[10.1038/9858](https://doi.org/10.1038/9858)
18. 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](https://doi.org/10.1371/journal.pcbi.1000107)
19. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B (2010) Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics* 11:568. doi:[10.1186/1471-2105-11-568](https://doi.org/10.1186/1471-2105-11-568)
20. Zhang L, Udaka K, Mamitsuka H, Zhu S (2012) Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools. *Brief Bioinform* 13(3):350–364. doi:[10.1093/bib/bbr060](https://doi.org/10.1093/bib/bbr060)
21. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Kesmir C, Peters B (2013) Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol* 9(10): e1003266. doi:[10.1371/journal.pcbi.1003266](https://doi.org/10.1371/journal.pcbi.1003266)
22. Altrichter S, Peter HJ, Pisarevskaja D, Metz M, Martus P, Maurer M (2011) IgE mediated autoallergy against thyroid peroxidase--a novel pathomechanism of chronic spontaneous urticaria? *PLoS One* 6(4): e14794. doi:[10.1371/journal.pone.0014794](https://doi.org/10.1371/journal.pone.0014794)
23. Valenta R, Seiberler S, Natter S, Mahler V, Mossabeh R, Ring J, Stingl G (2000) Autoallergy: a pathogenetic factor in atopic dermatitis? *J Allergy Clin Immunol* 105(3):432–437. doi:[10.1067/mai.2000.104783](https://doi.org/10.1067/mai.2000.104783)