



MHC Binding Predictions

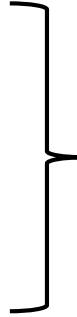
tools.iedb.org

Presented by:

Bjoern Peters, Professor

Raphael Trevizani, Bioinformatics Postdoc

Outline

- MHC class I binding prediction
 - MHC class II binding prediction
 - TepiTool
 - Datasets availability
 - Benchmarking of class I tools
 - Contributing tools
- 
- How the tool works
 - Recommendations
 - Interpreting results
 - Exercises

MHC binding predictions

- MHC molecules are **highly polymorphic** – thousands of different variants exist
- MHC-peptide binding is **promiscuous** in nature
- Experimental characterization of peptide–MHC interactions is highly **cost-intensive**
- Prediction methods facilitate selection of potential epitopes from a pool of peptides

Peptide binding data HLA-A*01:01

Peptide	IC ₅₀ (nM)
ASFCGSPY	51.4
LTDFGLSK	739.3
FTSFFYRY	1285.0
KSVFNSLY	1466.0
RDWAHNSL	1804.6
FSSCPVAY	1939.4
RNWAHSSL	2201.7
LSCAASGF	2830.1
LASIDLKY	3464.0



Machine learning algorithms



MHC I Binding Prediction

tools.iedb.org/main/tcell/

IEDB Analysis Resource

Overview | **T Cell Tools** | B Cell Tools | Analysis Tools | Tools-API | Usage | Download | Datasets | Contribute Tools | References

T Cell Epitope Prediction Tools

T Cell Epitopes - MHC Binding Prediction

These tools predict IC50 values for peptides binding to specific MHC molecules. Note that binding to MHC is necessary but not sufficient for recognition by T cells.

[Peptide binding to MHC class I molecules](#)

This tool will take in an amino acid sequence, or set of sequences and determine each subsequence's ability to bind to a specific MHC class I molecule.

[Peptide binding to MHC class II molecules](#)

This tool employs different methods to predict MHC Class II epitopes, including a consensus approach which combines NN-align, SMM-align and Combinatorial library methods.

[TepiTool:](#)

The Tepitool provides prediction of peptides binding to MHC class I and class II molecules. Tool is designed as a wizard with 6 steps as described below. Each field (except sequences and alleles) is filled with default recommended settings for prediction and selection of optimum peptides. The input parameters can be adjusted as per your specific needs. You can go back to previous steps to change your selection before submission of the job. Once you submit the job (at the end of step-6), you will not be able to make any more changes and will have to start the prediction all over again with updated input parameters.

T Cell Epitopes - Processing Prediction

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

[Proteasomal cleavage/TAP transport/MHC class I combined predictor](#)


This tool combines predictors of proteasomal processing, TAP transport, and MHC binding to produce an overall score for each peptide's intrinsic potential of being a T cell epitope.

[Neural network based prediction of proteasomal cleavage sites \(NetChop\) and T cell epitopes \(NetCTL and NetCTLpan\)](#)

NetChop is a predictor of proteasomal processing based upon a neural network. NetCTL and NetCTLpan are predictors of T cell epitopes along a protein sequence. It also employs a neural network architecture.

[MHC-NP: Prediction of peptides naturally processed by the MHC](#)

MHC-NP employs data obtained from MHC elution experiments in order to assess the probability that a given peptide is naturally processed and binds to a given MHC molecule. This tool was the winner of the [2nd Machine Learning Competition in Immunology](#).

 [MHCII-NP:](#)


This tool utilizes MHC II ligand elution data to predict naturally processed MHC II ligands by scanning the given peptide sequences.

T Cell Epitopes - Immunogenicity Prediction

This tool predicts the relative ability of a peptide/MHC complex to elicit an immune response.

[T cell class I pMHC immunogenicity predictor](#)

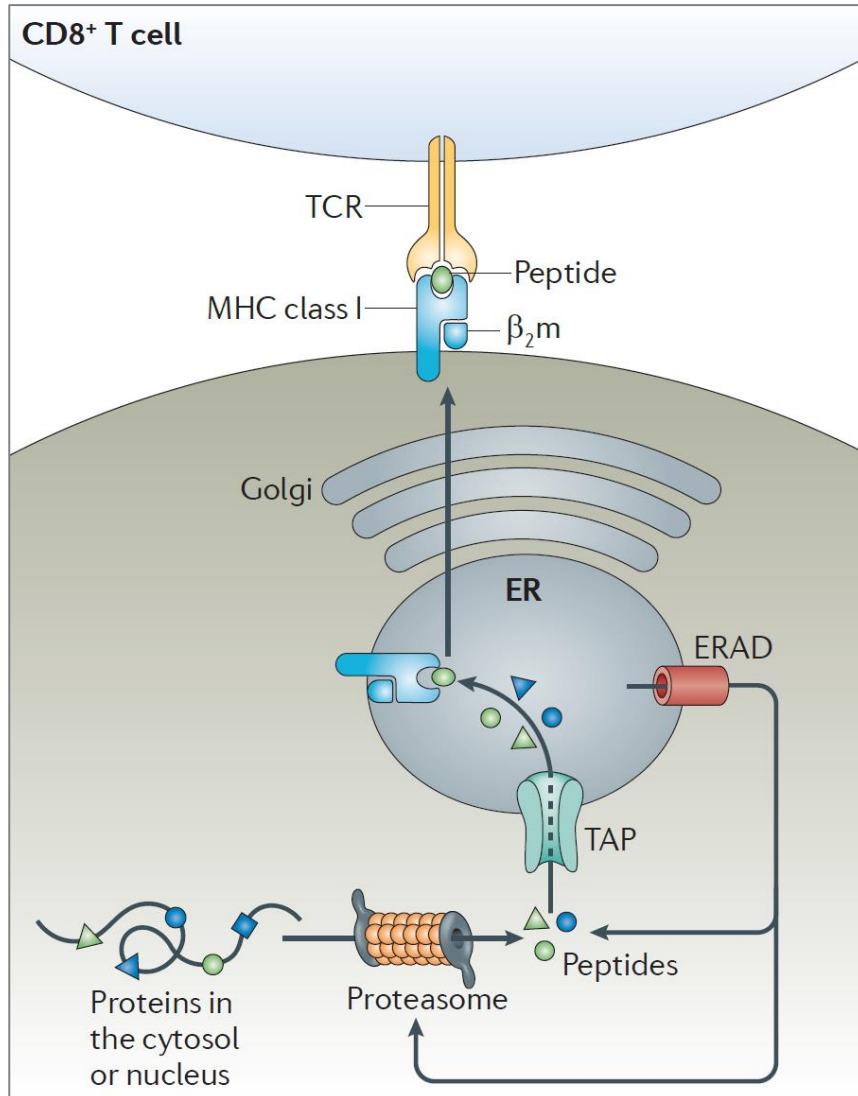
This tool uses amino acid properties as well as their position within the peptide to predict the immunogenicity of a class I peptide MHC (pMHC) complex.

 [Deimmunization:](#)

The deimmunization tool is attempt to identify immunodominant regions in a given therapeutically important protein, and suggest amino-acid substitutions that create non-immunogenic versions of the proteins. So we have opted a two steps process; 1) In the first step, the deimmunization tool will list all the immunogenic regions or peptides based on selected threshold. These peptides will be generated from the protein with 15mer window size and 10mer overlap. 2) In the second step, the user can select one or more



Endogenous antigen processing pathway (class I)



- Antigen generated within the cell
 - Viral particles
 - Self proteins
 - DRiPs (Defective Ribosomal Particles)
- Different factors influence peptide being “epitope”

Class I MHC molecule

- Expressed by almost all nucleated cells
- Presents antigen to **CD8+ T cells** (Cytotoxic T cells)
- One MHC encoded polymorphic chain (α) (2nd chain – β 2-microglobulin)
- The binding groove is **closed** at both ends and can accommodate peptides of **8-15 AA**
- Only **α chain** impacts binding

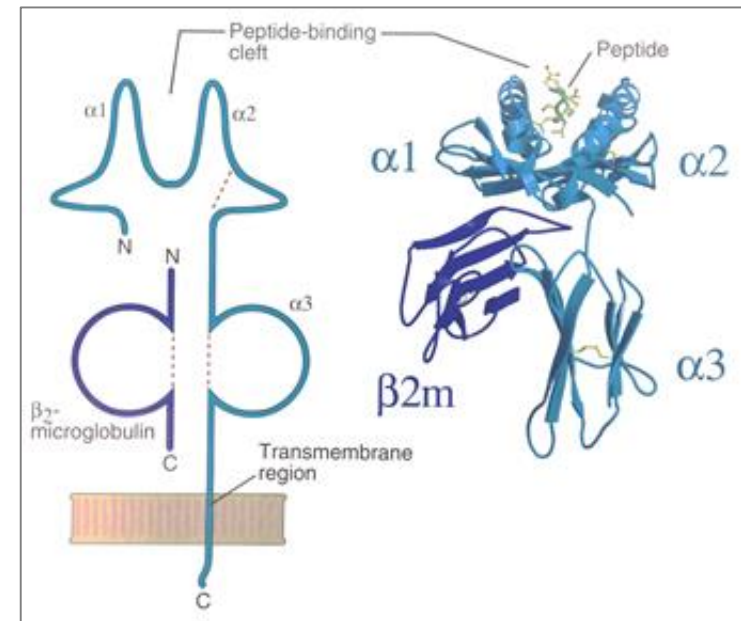


Figure Source
Cellular & Molecular Immunology, 5th Ed by Abbas and Lichtman

MHC-I binding prediction - example

tools.iedb.org/mhci/

Home Help Example Reference Download Contact

MHC-I Binding Predictions

Prediction Method Version v2.24 [\[Older versions\]](#)

Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences.

```
>LCMV Armstrong, Protein GP
MGQIVTMFEALPHIIDDEVINIVIIIVLIVITGIKAVYNFATCGIFALISFLLLAGRSCGM
YGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHYISMGTSGLELFTNDSII
SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA
QSAQSQCRTRFRGRVLDMFRTAFGGKYMMSGWGWGTGSDGKTTWCSQTSYQYLIQNRTWE
NHCTYAGPFGMSRILLSQEKTFFTRRLAGFTWTLSDSSGVENPGGYCLTKWMILAAE
LKCFGNTAVAKCNVNHDAEFCMDMLRLIDYNKAALSFKEDVESALHLFKTTVNSLISDQ
LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVSIFLHLVKIPTHRHIKGGSCPKP
HRLTNKGICSCGAFKVPGVKTVWKRR
```

Or select file containing sequence(s) No file chosen

**Epitope
sequence**
(copy or upload)

MHC-I binding prediction - example

tools.iedb.org/mhci/

Home Help Example Reference Download Contact

MHC-I Binding Predictions

Prediction Method Version: v2.24 [\[Older versions\]](#)

Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences.

```
>LCMV Armstrong, Protein GP
MGQIVTMFEALPHIDEVINIVIVLIVITGKAVYNFATCGIFALISFLLLAGRSCGM
YGLKGPDIYKGVYQFKSVFEDMSHLNLTMPNACSAANSHHYISMGTSGLELFTNDSII
SHNFCNLTSAFNKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA
QSAQSQCRTRFRGRVLDMFRTAFGGKYMRSWGWTGSDGKTTWCSQTSYQYLIQNRWTE
NHCTYAGPFGMSRILLSQEKTKFFRRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE
LKCFGNTAVAKCNVNHDAEFCMDMLRIDYNTKAALSFKKEDVESALHLFKTTVNSLISDQ
LLMRNHLRDLMGVPCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTAYLVSIPLHLVKIPTRHIKGGGSCPKP
HRLTNKGICSCGAFKVPGVKTVWKRR
```

Or select file containing sequence(s): No file chosen

Choose a Prediction Method

Prediction Method [?]
Show all the method versions:

MHC source species

Show only frequently occurring alleles: [?]
Select MHC allele(s)

Select HLA allele reference set: [?]
(Specify MHC allele sequence)

Sort peptides by

Show: All predictions

Output format: XHTML table

Email address (optional): [?]

Prediction Method Selections:

- IEDB recommended 2020.09 (NetMHCpan EL 4.1) [Help on prediction method selections](#)
- IEDB recommended 2020.09 (NetMHCpan EL 4.1)
- Consensus
- NetMHCpan BA 4.1
- IEDB recommended 2020.04 (NetMHCpan EL 4.0)
- NetMHCpan BA 4.0
- ANN 4.0
- SMMPMBEC
- SMM
- CombLib_Sidney2008
- PickPocket
- netMHCcons
- netMHCstabpan

Prediction method

MHC class I binding prediction methods available

Method	Reference	Performance Reported
NetMHCpan EL - 4.1	Reynisson et al., 2020	0.978 AUC (evaluated on EL data)
NetMHCpan EL - 4.0	Paul et al., 2020	0.977 AUC (average)
NetMHCpan BA - 4.1	Reynisson et al., 2020	0.893 AUC (evaluated on BA data)
NetMHCpan BA - 4.0	Paul et al., 2020	0.975 AUC (average)
Consensus	Moutaftsi et al., 2006	
ANN (NetMHC - 4.0)	Andreatta & Nielsen, 2016	0.887 AUC (average)
SMM with Peptide:MHC Binding Energy Covariance matrix (SMMPMBEC)	Kim et al., 2009	0.894 AUC (average)
Stabilized matrix method (SMM)	Peters & Sette, 2005	0.887 AUC (average) (Kim et. al., 2009)
Combinatorial library (CombLib)	Sidney et al., 2008	0.909 AUC (HLA-A*0201)
PickPocket - 1.1	Zhang et al., 2009	0.895 AUC (average)
NetMHCcons - 1.1	Karosiene et al., 2012	0.729 PCC (average)
NetMHCstabpan - 1.0	Rasmussen et al., 2016	0.980 AUC (average)

Guidelines: Choosing the prediction method

- Suggested method = “IEDB recommended”
 - Employs NetMHCpan EL 4.1 across all alleles
 - Provides binding affinity & percentile rank for each method separately as well
- Recommendation will change with the new benchmark studies

IEDB Tools Version	Recommended Method
2020.09 (current)	NetMHCPan 4.1 EL
2020.04	NetMHCPan 4.0 EL
2.22 and earlier	Consensus, if available; otherwise, NetMHCpan

MHC-I binding prediction – example

tools.iedb.org/mhci/

Home Help Example Reference Download Contact

MHC-I Binding Predictions

Prediction Method Version: v2.24 [\[Older versions\]](#)

Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences.

```
>LCMV Armstrong, Protein GP
MGQIVTMFEALPHIIDVINIVIVLIVITGIAVYNFATCGIFALISFLLLAGRSCGM
YGLKGPDIYKGVYQFQKSVFDMSHLNLMPNACSAANSHHYISMGTSGLELFTNDSII
SHNFCNLTSAFNKKTDFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA
QSAQSQCRTRFRGRVLDMFRTAFGGKYMRSWGWTGSDGKTTWCSQTSYQYLIQNRTWE
NHCTYAGPFGMSRILLSQEKTKFFTRRLAGFTFTWLSDSGVENPGGYCLTKWMILAAE
LKCFGNTAVAKCNVNHDAEFCMDLRLIDYNKAALSFKFEDVESALHLFKTTVNSLISDQ
LLMRNHLRDLMGVPCYNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSQIEQEA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTAYLVSIFLHLVKIPTRHIKGGSCPKP
HRLTNKGICSCGAFKVPGVKTVWKRR
```

Or select file containing sequence(s): No file chosen

Choose a Prediction Method

Prediction Method [?]: IEDB recommended 2020.09 (NetMHCpan EL 4.1) [Help on prediction method selections](#)

Show all the method versions:

Specify what to make binding predictions for

MHC source species: **human** (dropdown menu open showing: chimpanzee, cow, gorilla, human, macaque, mouse, pig, dog, horse)

Show only frequently occurring alleles: [?]

Select MHC allele(s): Length: [?]

Select HLA allele reference set: [?] [\(Specify MHC allele sequence\)](#)

Specify Output

Sort peptides by:

Show: All predictions

Output format: XHTML table

Email address (optional): [?]

Choose species

MHC-I binding prediction – example

tools.iedb.org/mhci/

The screenshot shows the MHC-I Binding Predictions web interface. Red boxes and arrows highlight key features: a 'Complete set' of protein sequences in the input field, 'Reference alleles' selected in the 'Choose a Prediction Method' dropdown, and 'Specify allele(s) & peptide length' in the 'Specify what to make binding predictions for' section. A red dashed box at the bottom right shows the 'Upload format' for the selected alleles and lengths.

Home Help Example Reference Download Contact

MHC-I Binding Predictions

Prediction Method Version: v2.24 [\[Older versions\]](#)

Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences.

```
>LCMV Armstrong, Protein GP
MGQIVTMFEALPHIIDEVINIVIVLITGKAVYNFATCGIFALISFLLAGRSCGM
YGLKGPDIYKGVYQFKSVEFDMSHLNLTPNACSANNSSHYYISMGTSGLELFTNDSII
SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQNLTFSDA
QSAQSQCRTRFRGRVLDMFRATFAGGKYMRSWGWGTGSDGKTTWCQSQTSYQYLIIQNRTWE
NHCTYAGPFGMSRILLSQEKTKFFRRLAGTFTWTLSDSSGVENPGGYCLTKWMLAAE
LKCFGNTAVAKCNVNHDAEFCMDLRLIDYNKAALSFKFEDVESALHLFKTTVNSLISDQ
LLMRNHLRDLMGVPCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSQIQEFA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTAYLVSIFLHLVKIPTRHRHKGGSCKPK
HRLTNKGICSCGAFKVPVGVKTVWKR
```

Or select file containing sequence(s) No file chosen

Choose a Prediction Method

Prediction Method Show all the method versions:

IEDB recommended 2020.09 (NetMHCpan EL 4.1) [Help on prediction method selections](#)

Specify what to make binding predictions for

MHC source species: human

Show only frequently occurring alleles:

Select MHC allele(s):

Allele	Length	
HLA-A*01:01	9	<input type="checkbox"/>
HLA-B*07:02	10	<input type="checkbox"/>

[Upload allele file](#) (?)

[Select HLA allele reference set: \(Specify MHC allele sequence\)](#) (?)

Specify Output

Sort peptides by: Predicted IC50

Show: All predictions

Output format: XHTML table

Email address (optional):

Complete set

Reference alleles

Specify allele(s) & peptide length (select or upload)

Upload format:
HLA-A*01:01,9
HLA-B*07:02,10

Natural length distribution in epitope prediction

- Alleles differ in their preference for lengths on binding and presentation of peptides

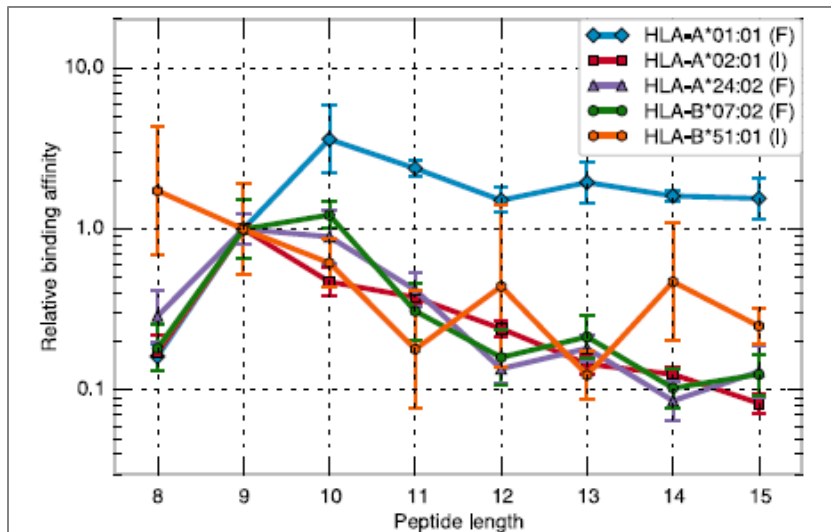


FIGURE 1. Peptide binding-length preference for five common HLA alleles. The length preference for each HLA was determined by measuring

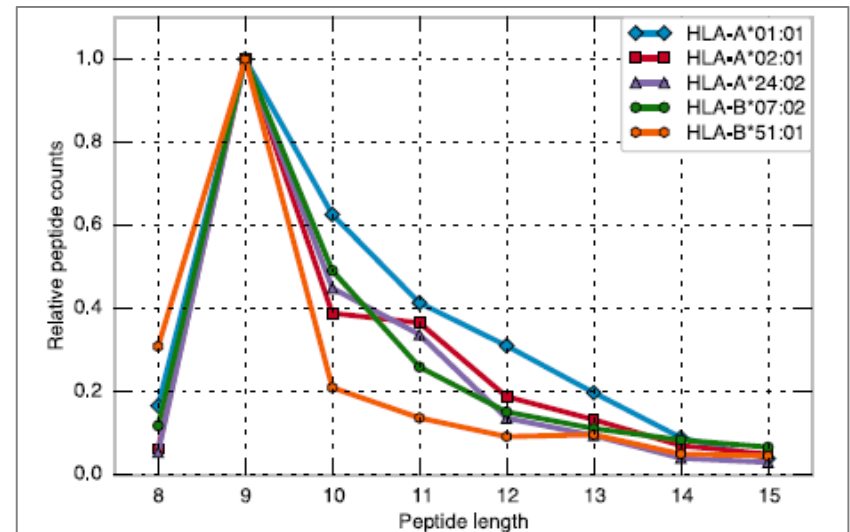


FIGURE 2. Length profiles of naturally presented peptides for five HLA molecules. Large datasets of HLA-I ligands were determined by the elu-

[J Immunol.](#) 2016 Feb 15;196(4):1480-7. doi: 10.4049/jimmunol.1501721. Epub 2016 Jan 18.

The Length Distribution of Class I-Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference.

Trolle T¹, McMurtrey CP², Sidney J³, Bardet W², Osborn SC², Kaever T³, Sette A³, Hildebrand WH², Nielsen M⁴, Peters B⁵.

PMID: 26783342 PMCID: PMC4744552 DOI: 10.4049/jimmunol.1501721

Allele selection – Reference set for global coverage

- Reference set of 27 alleles
- Covers > 97% of population

HLA-A	Frequency	HLA-B	Frequency
A*01:01	16.2	B*07:02	13.3
A*02:01	25.2	B*08:01	11.5
A*02:03	3.3	B*15:01	5.2
A*02:06	4.9	B*35:01	6.5
A*03:01	15.4	B*40:01	10.3
A*11:01	12.9	B*44:02	9.2
A*23:01	6.4	B*44:03	7.6
A*24:02	16.8	B*51:01	5.5
A*26:01	4.7	B*53:01	5.4
A*30:01	5.1	B*57:01	3.2
A*30:02	5.0	B*58:01	3.6
A*31:01	4.7		
A*32:01	5.7		
A*33:01	3.2		
A*68:01	4.6		
A*68:02	3.3		

<http://iedb.zendesk.com/entries/25054538-HLA-allele-frequencies>

Prediction method dependent allele selection

tools.iedb.org/mhci/

NetMHCpan prediction methods allow **FASTA sequence input**

Choose a Prediction Method

Prediction Method ?
Show all the method versions: IEDB recommended 2020.09 (NetMHCpan EL 4.1) ▼ [Help on prediction method selections](#)

Specify what to make binding predictions for

MHC source species: human ▼

Select MHC allele(s)
[Select HLA allele reference set:](#) ?
[Input FASTA sequence \(Select MHC allele\(s\)\)](#)

Paste a single full length MHC protein sequence in [FASTA](#) format:

Select "Specify MHC allele sequence"

MHC-I binding prediction – example

tools.iedb.org/mhci/

Home Help Example Reference Download Contact

MHC-I Binding Predictions

Prediction Method Version: v2.24 [\[Older versions\]](#)

Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences.

```
>LCMV Armstrong, Protein GP
MGQIVTMFEALPHIIDDEVINIVIVLIVITGKAVYNFATCGIFALISFLLLAGRSCGM
YGLKGPDIYKGVYQKFSVEFDMSHLNLTPNACANNSSHYYISMGTSGLELFTNDSII
SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQNLTFSDA
QSAQSQCRFRGRVLD MFRTAFGGKYMRSWGWGTGSDGKTTWCSQTSYQYLIQNRTWE
NHCTYAGPFGMSRILLSQEKTFFTRRLAGTFTWLS DSSGVENPGGYCLTKWMLAAE
LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSLSDQ
LLMRNHLRDLMGVPCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDDQIEQEA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVSIFLHLVKIPTRHRHKGSGCPKP
HRLTNKGICSGAFKVPGVKTVWKRR
```

Or select file containing sequence(s): No file chosen

Choose a Prediction Method

Prediction Method: Show all the method versions: IEDB recommended 2020.09 (NetMHCpan EL 4.1) [Help on prediction method selections](#)

Specify what to make binding predictions for

MHC source species: human

Allele	Length	
HLA-A*01:01	9	<input type="checkbox"/>
HLA-B*07:02	10	<input type="checkbox"/>

Show only frequently occurring alleles: [?](#)
Select MHC allele(s)
[Select HLA allele reference set:](#) [?](#)
[\(Specify MHC allele sequence\)](#) [?](#)
 [?](#)

Specify Output

Sort peptides by: Predicted IC50

Show: Position in sequence

Output format: All predictions

IC50 below [cutoff] nM

Percent rank below [cutoff]

Email address (optional): [?](#)

Input

Output



How the tool works

- Breaks the sequence into all possible peptides (of chosen length).
- Predicts the binding affinity for each peptide based on the method.
- Compares the predicted affinity to that of a large set of randomly selected peptides.
- Assigns a percentile rank depending on individual predicted affinity.
- Consensus picks the median rank of the methods used.

MHC-I binding prediction – example

tools.iedb.org/mhci/

Home Help Example Reference Download Contact

MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	LCMV Armstrong, Protein GP	MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFL LLAGRSCGMGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNHHY ISMGTSGLELTFITNDSIISHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIR GNSNYKAVSCDFNNGITIQYNLTFSDAQSAQSQCRTRFRGRVLDMFRFAFG GKYMRSWGWTGSDGKTTWCSQTSYQYLIIQNRTWENHCYAGPFGMSRI LLSQEKTFFFTRRLAGFTWTLSDDSSGVENPGGYCLTKWMILAAELKCFG NTAVAKCNVNHDAEFCMDLRLIDYNKAALSKFKEDVESALHLFKTTVNSL ISDQLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYL NETHFSQDIEQEADNMIETMLRKDYIKRQGSTPLALMDLLMFSTSAYLVS IFLHLVKIPTHRIKGGSCPKPHRLTNKGICSCGAFKVPGVKTVWKRR

Prediction method: **IEDB recommended 2.22** | **Low Percentile Rank = good binders**
[Download result](#)

Citations
 Check to expand the result:

Allele	#	Start	End	Length	Peptide	Method used	Percentile Rank
HLA-A*01:01	1	92	100	9	CSANNSHHY	Consensus (ann/smm)	0.2
HLA-A*01:01	1	405	413	9	FSDQIEQEA	Consensus (ann/smm)	0.34
HLA-B*07:02	1	471	480	10	KPHRLTNKGI	Consensus (ann/smm)	0.35
HLA-A*01:01	1	417	425	9	ITEMLRKDY	Consensus (ann/smm)	0.6
HLA-A*01:01	1	361	369	9	LRDLMGVPY	Consensus (ann/smm)	0.68
HLA-A*01:01	1	233	241	9	RTWENHCTY	Consensus (ann/smm)	0.69
HLA-A*01:01	1	217	225	9	TTWCSQTSY	Consensus (ann/smm)	0.71
HLA-A*01:01	1	439	447	9	LLMFSTSAY	Consensus (ann/smm)	0.75
HLA-A*01:01	1	147	155	9	LSIRGNSNY	Consensus (ann/smm)	1.25
HLA-B*07:02	1	425	434	10	YIKRQGSTPL	Consensus (ann/smm)	1.27
HLA-B*07:02	1	243	252	10	GPFGMSRILL	Consensus (ann/smm)	1.35
HLA-A*01:01	1	191	199	9	VLDMERTAF	Consensus (ann/smm)	1.6
HLA-A*01:01	1	174	182	9	SDPAGTGG	Consensus (ann/smm)	1.75

Input sequence

Output
 (sorted low-to-high by
 percentile rank)

A **percentile rank** for a peptide is the percentage of randomly sampled peptides scoring better than the peptide.

MHC-I binding prediction – example

tools.iedb.org/mhci/

Individual scores for different methods

Prediction method: IEDB recommended 2.22 | Low Percentile Rank = good binders

[Download result](#) 

Citations



Check to expand the result:





Allele	#	Start	End	Length	Peptide	Method used	Percentile Rank	ANN IC50(nM)	ANN rank	SMM IC50(nM)	SMM rank
HLA-A*01:01	1	92	100	9	CSANNSHHY	Consensus (ann/smm)	0.2	25.62	0.09	173.60	0.3
HLA-A*01:01	1	405	413	9	FSDQIEQEA	Consensus (ann/smm)	0.34	121.15	0.27	360.21	0.4
HLA-B*07:02	1	471	480	10	KPHRLTNKGI	Consensus (ann/smm)	0.35	46.84	0.2	112.67	0.5
HLA-A*01:01	1	417	425	9	ITEMLRKDY	Consensus (ann/smm)	0.6	591.06	0.71	426.14	0.5
HLA-A*01:01	1	361	369	9	LRDLMGVPY	Consensus (ann/smm)	0.68	799.14	0.85	421.26	0.5
HLA-A*01:01	1	233	241	9	RTWENHCTY	Consensus (ann/smm)	0.69	552.60	0.68	694.30	0.7
HLA-A*01:01	1	217	225	9	TTWCSQTSY	Consensus (ann/smm)	0.71	604.36	0.72	653.96	0.7
HLA-A*01:01	1	439	447	9	LLMFSTSAY	Consensus (ann/smm)	0.75	724.33	0.8	728.70	0.7
HLA-A*01:01	1	147	155	9	LSIRGNSNY	Consensus (ann/smm)	1.25	3116.42	2.0	448.28	0.5
HLA-B*07:02	1	425	434	10	YIKRQGSTPL	Consensus (ann/smm)	1.27	59.83	0.24	575.20	2.3
HLA-B*07:02	1	243	252	10	GPFGMSRILL	Consensus (ann/smm)	1.35	418.14	1.2	351.41	1.5
HLA-A*01:01	1	191	199	9	VLDMFRTAF	Consensus (ann/smm)	1.6	2586.86	1.8	1457.30	1.4
HLA-A*01:01	1	174	182	9	FSDAQAQS	Consensus (ann/smm)	1.75	2437.12	1.7	1934.42	1.8

Downloaded prediction results

	A	B	C	D	E	F	G	H	I	J	K	L
1	allele	seq_num	start	end	length	peptide	method	Percentile Rank	ann_ic50	ann_rank	smm_ic50	smm_rank
2	HLA-A*01:01	1	92	100	9	CSANNSHHY	Consensus (ann/smm)	0.2	25.62	0.09	173.6	0.3
3	HLA-A*01:01	1	405	413	9	FSDQIEQEA	Consensus (ann/smm)	0.34	121.15	0.27	360.21	0.4
4	HLA-B*07:02	1	471	480	10	KPHRLTNKGI	Consensus (ann/smm)	0.35	46.84	0.2	112.67	0.5
5	HLA-A*01:01	1	417	425	9	ITEMLRKDY	Consensus (ann/smm)	0.6	591.06	0.71	426.14	0.5
6	HLA-A*01:01	1	361	369	9	LRDLMGVPI	Consensus (ann/smm)	0.68	799.14	0.85	421.26	0.5
7	HLA-A*01:01	1	233	241	9	RTWENHCTY	Consensus (ann/smm)	0.69	552.6	0.68	694.3	0.7
8	HLA-A*01:01	1	217	225	9	TTWCSQTSY	Consensus (ann/smm)	0.71	604.36	0.72	653.96	0.7
9	HLA-A*01:01	1	439	447	9	LLMFSTSAY	Consensus (ann/smm)	0.75	724.33	0.8	728.7	0.7
10	HLA-A*01:01	1	147	155	9	LSIRGNSNY	Consensus (ann/smm)	1.25	3116.42	2	448.28	0.5
11	HLA-B*07:02	1	425	434	10	YIKRQGSTPL	Consensus (ann/smm)	1.27	59.83	0.24	575.2	2.3
12	HLA-B*07:02	1	243	252	10	GPFGMSTRILL	Consensus (ann/smm)	1.35	418.14	1.2	351.41	1.5
13	HLA-A*01:01	1	191	199	9	VLDMFRTAF	Consensus (ann/smm)	1.6	2586.86	1.8	1457.3	1.4
14	HLA-A*01:01	1	174	182	9	FSDAQSAQS	Consensus (ann/smm)	1.75	2437.12	1.7	1934.42	1.8
15	HLA-A*01:01	1	52	60	9	LAGRSCGMY	Consensus (ann/smm)	2.05	4721.07	2.5	1692.58	1.6
16	HLA-A*01:01	1	220	228	9	CSQTSYQYL	Consensus (ann/smm)	2.15	5007.72	2.6	1826.21	1.7
17	HLA-A*01:01	1	219	227	9	WCSQTSYQY	Consensus (ann/smm)	2.2	2051.4	1.6	3009.89	2.8
18	HLA-A*01:01	1	86	94	9	LTMPNACSA	Consensus (ann/smm)	2.25	4423.31	2.4	2215.9	2.1
19	HLA-B*07:02	1	320	329	10	RLIDYNKAAL	Consensus (ann/smm)	2.25	1113.26	2.2	595.42	2.3
20	HLA-B*07:02	1	190	199	10	RVLDMFRTAF	Consensus (ann/smm)	2.4	567.7	1.5	816.24	3.3
21	HLA-A*01:01	1	272	280	9	LSDSSGVEN	Consensus (ann/smm)	2.45	8300.79	3.9	913.17	1
22	HLA-A*01:01	1	369	377	9	YCNYSKFWY	Consensus (ann/smm)	2.45	5677.63	2.9	2145.61	2
23	HLA-A*01:01	1	436	444	9	LMDLLMFST	Consensus (ann/smm)	2.5	3758.17	2.2	3037.74	2.8
24	HLA-B*07:02	1	432	441	10	TPLALMDLLM	Consensus (ann/smm)	2.6	767.22	1.8	854.71	3.4
25	HLA-A*01:01	1	166	174	9	ITIQYNLTF	Consensus (ann/smm)	2.75	8692.54	4	1583.25	1.5
26	HLA-A*01:01	1	364	372	9	LMGVYPYCN	Consensus (ann/smm)	2.75	5142.58	2.7	3009.89	2.8
27	HLA-A*01:01	1	104	112	9	GTSGLELTF	Consensus (ann/smm)	2.8	7192.3	3.4	2374.38	2.2
28	HLA-A*01:01	1	222	230	9	QTSYQYLII	Consensus (ann/smm)	2.9	8442.18	4	1873.05	1.8
29	HLA-A*01:01	1	448	456	9	LVSIFLHLV	Consensus (ann/smm)	2.95	5023.73	2.7	3424.13	3.2
30	HLA-B*07:02	1	252	260	10	PLACTETMTL	Consensus (ann/smm)	2.95	1227.49	2.4	1062.7	4.1

Emailed prediction results

IEDB Tools MHC class I prediction result (2019-10-07 10:05:35) Inbox x  



IEDB Tools <Prediction-results-noreply@iedb.org>  10:05 AM (0 minutes ago)   

to me ▾



Your MHC class I prediction completed on the IEDB servers (<http://tools.iedb.org/mhci/>) and the result is attached in csv format.

Input parameters
Method: recommended
Number of sequences: 1
Input sequences: attached
Alleles: HLA-A*01:01, HLA-B*07:02
Lengths: 9, 10

Job parameters
Submission date: 2019-10-07 10:05:35
Completion date: 2019-10-07 10:05:51
Total walltime since submission: 16 seconds

2 Attachments  

id	allele	seq_len	start	end	length	method	result	Percent
1	HLA-A*01:01	9	86	94	9	B*02010101	Combinatorial	
2	HLA-A*01:01	9	86	94	9	B*02010101	Combinatorial	
3	HLA-A*01:01	9	87	95	9	B*02010101	Combinatorial	
4	HLA-A*01:01	9	87	95	9	B*02010101	Combinatorial	
5	HLA-A*01:01	9	88	96	9	B*02010101	Combinatorial	
6	HLA-A*01:01	9	88	96	9	B*02010101	Combinatorial	
7	HLA-A*01:01	9	89	97	9	B*02010101	Combinatorial	
8	HLA-A*01:01	9	89	97	9	B*02010101	Combinatorial	
9	HLA-A*01:01	9	90	98	9	B*02010101	Combinatorial	
10	HLA-A*01:01	9	90	98	9	B*02010101	Combinatorial	
11	HLA-A*01:01	9	91	99	9	B*02010101	Combinatorial	
12	HLA-A*01:01	9	91	99	9	B*02010101	Combinatorial	
13	HLA-A*01:01	9	92	100	9	B*02010101	Combinatorial	
14	HLA-A*01:01	9	92	100	9	B*02010101	Combinatorial	
15	HLA-A*01:01	9	93	101	9	B*02010101	Combinatorial	
16	HLA-A*01:01	9	93	101	9	B*02010101	Combinatorial	
17	HLA-A*01:01	9	94	102	9	B*02010101	Combinatorial	
18	HLA-A*01:01	9	94	102	9	B*02010101	Combinatorial	
19	HLA-A*01:01	9	95	103	9	B*02010101	Combinatorial	
20	HLA-A*01:01	9	95	103	9	B*02010101	Combinatorial	
21	HLA-A*01:01	9	96	104	9	B*02010101	Combinatorial	
22	HLA-A*01:01	9	96	104	9	B*02010101	Combinatorial	
23	HLA-A*01:01	9	97	105	9	B*02010101	Combinatorial	
24	HLA-A*01:01	9	97	105	9	B*02010101	Combinatorial	
25	HLA-A*01:01	9	98	106	9	B*02010101	Combinatorial	
26	HLA-A*01:01	9	98	106	9	B*02010101	Combinatorial	
27	HLA-A*01:01	9	99	107	9	B*02010101	Combinatorial	
28	HLA-A*01:01	9	99	107	9	B*02010101	Combinatorial	
29	HLA-A*01:01	9	100	108	9	B*02010101	Combinatorial	
30	HLA-A*01:01	9	100	108	9	B*02010101	Combinatorial	

 **predict_result.csv**  **input_sequences.txt**

Selection of “binders”

- Pick peptides **below percentile rank 1.0**
- Pick peptides **below predicted binding affinity of 500 nM**
 - IC50 < 50 nM - high affinity
 - IC50 < 500 nM - intermediate affinity
 - IC50 < 5000 nM - low affinity
 - Sette et al. 1994, J. Immunology (PMID: 7527444)
 - Ensures that all peptides have reasonable affinity
- Pick **top 1% of peptides** for each allele/length combination to cover most of immune responses
 - Moutaftsi et al. 2006 (PMID: 16767078)
 - Kotturi et al. 2007 (PMID: 17329346)
 - Ensures equal number of peptides per allele
- Select based on **allele specific binding affinity** threshold

Different peptide-binding repertoires

- The size of the peptide repertoire binding at a given affinity varies between alleles

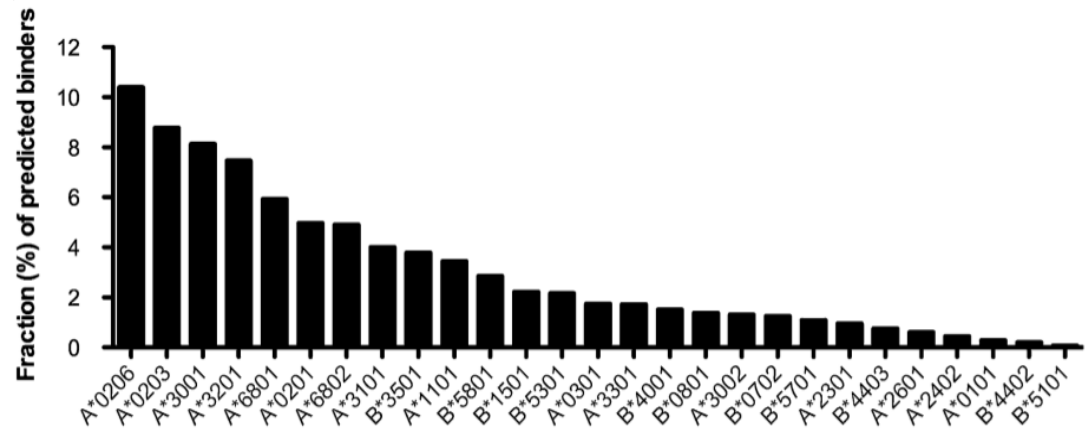
● All peptides

★ Binders

HLA-A*02:01



HLA-A*01:01



Allele-specific affinity cutoffs

J Immunol. 2013 Dec 15;191(12):5831-9. doi: 10.4049/jimmunol.1302101. Epub 2013 Nov 4.

HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity.

Paul S^{#1}, Weiskopf D^{#1}, Angelo MA¹, Sidney J¹, Peters B¹, Sette A¹.

PMID: 24190657 PMCID: PMC3872965 DOI: 10.4049/jimmunol.1302101

Allele-specific thresholds

Home **Help** Example Reference Download Contact

MHC-I binding predictions - Tutorial

Guidelines for selecting thresholds (cut-offs) for predictions can be found [here](#).

How to obtain predictions

This website provides access to predictions

Selecting thresholds (cut-offs) for MHC class I and II binding predictions



Ward Fleri
posted this on May 21, 2013 04:33 PM

MHC class I

For MHC class I T cell epitope predictions, selection of predicted binders can be done based on the percentile rank or MHC binding affinity. The IEDB currently recommends making selections based on a percentile rank of $\leq 1\%$ for each (MHC allele, length) combination to cover most of the immune responses.^{1,2} Alternatively, a binding affinity (IC50) threshold of 500 nM identifies peptide binders recognized by T cells and this threshold can be used to select peptides.³ Recently, a paper from our group showed that absolute binding affinity threshold correlates better with immunogenicity and also that, for even better correlation, MHC-specific thresholds should be used.⁴ The tables below show the allele-specific thresholds for the 38 most common HLA-A and HLA-B alleles, representative of the nine major supertypes. The tables can also be downloaded as an RTF file (see attached file).

Alleles sorted by population frequency			Alleles sorted by name		
Allele	Population frequency of allele	Allele specific affinity cutoff (IC50 nM)	Allele	Population frequency of allele	Allele specific affinity cutoff (IC50 nM)
A*0201	25.2	255	A*0101	16.2	884
A*2402	16.8	849	A*0201	25.2	255
A*0101	16.2	884	A*0203	3.3	92
A*0301	15.4	602	A*0206	4.9	60
B*0702	13.3	687	A*0301	15.4	602
A*1101	12.9	382	A*1101	12.9	382
B*0801	11.5	663	A*2301	6.4	740
B*4001	10.3	639	A*2402	16.8	849
B*4402	9.2	904	A*2501	2.5	795
B*4403	7.6	780	A*2601	4.7	815
B*3501	6.5	348	A*2902	2.9	641
A*2301	6.4	740	A*3001	5.1	109
A*3201	5.7	131	A*3002	5	674
B*5101	5.5	939	A*3101	4.7	329
B*5301	5.4	538	A*3201	5.7	131
B*1501	5.2	528	A*3301	3.2	606
A*3001	5.1	109	A*6801	4.6	197
A*3002	5	674	A*6802	3.3	259

Recommendations

- Both approaches (affinity and ranking) are reasonable, and have been applied in numerous studies
- Thresholds can be combined (peptides in top 1% and IC50 <500nM)
- Current studies suggest that allele specific thresholds can be derived

Alternate approaches for selecting binders

- Change threshold values depending on your need
 - e.g. in case you have too few or too many predicted binders.
- Set a desired percentage within your peptide set (irrespective of IEDB percentile rank) in case you want to study a fixed number of best possible peptides.

Class II MHC Binding Prediction

- Basic structure and principles same as class I binding prediction tool



IEDB Analysis Resource

Overview | **T Cell Tools** | B Cell Tools | Analysis Tools | Tools-API | Usage | Download | Datasets | Contribute Tools | References

T Cell Epitope Prediction Tools

T Cell Epitopes - MHC Binding Prediction

These tools predict IC50 values for peptides binding to specific MHC molecules. Note that binding to MHC is necessary but not sufficient for recognition by T cells.

[Peptide binding to MHC class I molecules](#)

This tool will take in an amino acid sequence, or set of sequences and determine each subsequence's ability to bind to a specific MHC class I molecule.

[Peptide binding to MHC class II molecules](#)

This tool employs different methods to predict MHC Class II epitopes, including a consensus approach which combines NN-align, SMM-align and Combinatorial library methods.

[TepiTool:](#)

The Tepitool provides prediction of peptides binding to MHC class I and class II molecules. Tool is designed as a wizard with 6 steps as described below. Each field (except sequences and alleles) is filled with default recommended settings for prediction and selection of optimum peptides. The input parameters can be adjusted as per your specific needs. You can go back to previous steps to change your selection before submission of the job. Once you submit the job (at the end of step-6), you will not be able to make any more changes and will have to start the prediction all over again with updated input parameters.

T Cell Epitopes - Processing Prediction

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

[Proteasomal cleavage/TAP transport/MHC class I combined predictor](#)


This tool combines predictors of proteasomal processing, TAP transport, and MHC binding to produce an overall score for each peptide's intrinsic potential of being a T cell epitope.

[Neural network based prediction of proteasomal cleavage sites \(NetChop\) and T cell epitopes \(NetCTL and NetCTLpan\)](#)

NetChop is a predictor of proteasomal processing based upon a neural network. NetCTL and NetCTLpan are predictors of T cell epitopes along a protein sequence. It also employs a neural network architecture.

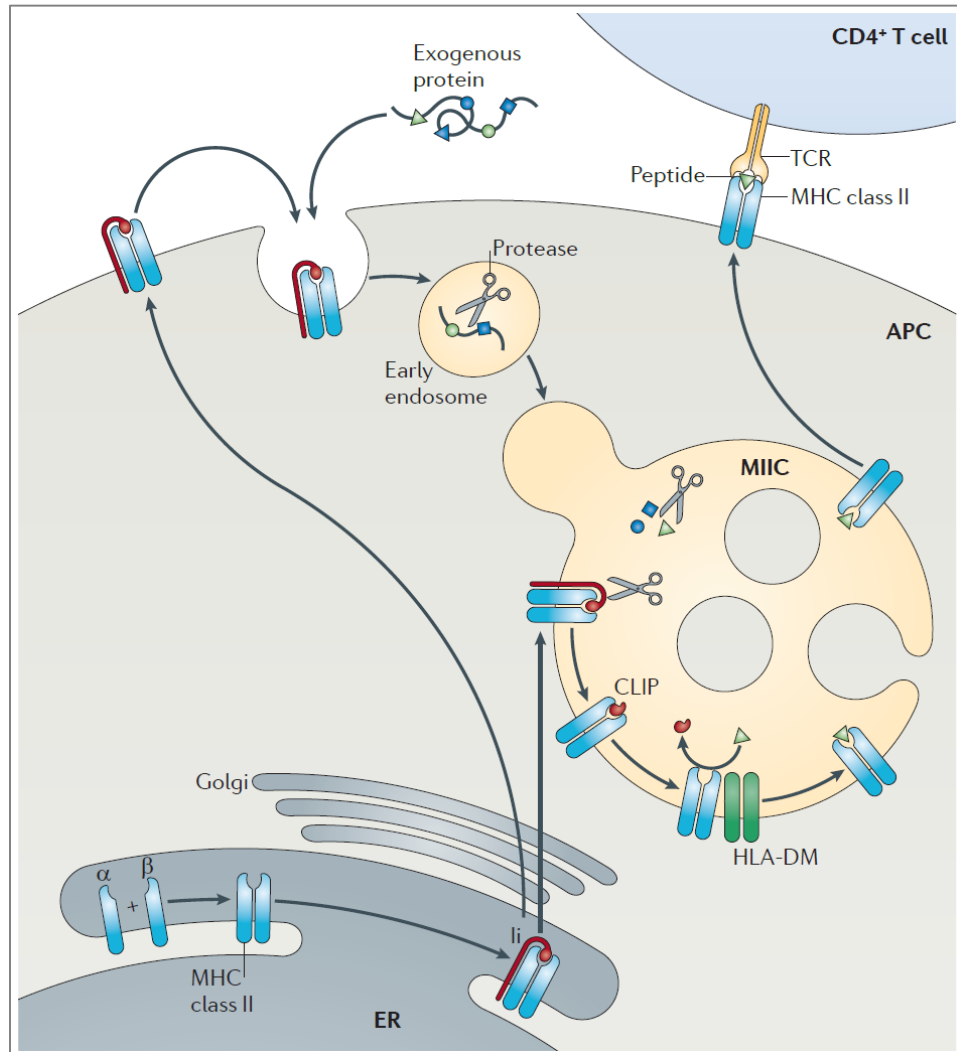
[MHC-NP: Prediction of peptides naturally processed by the MHC](#)

MHC-NP employs data obtained from MHC elution experiments in order to assess the probability that a given peptide is naturally processed and binds to a given MHC molecule. This tool was the winner of the [2nd Machine Learning Competition in Immunology](#).

 **MHCII-NP:**

This tool utilizes MHC II ligand elution data to predict naturally processed MHC II ligands by scanning the given peptide sequences.

Exogenous antigen processing pathway (class II)



- Antigenes generated outside the cell
 - Entered through inhalation, ingestion, injection
 - Bacteria, Allergens, Parasites etc.

Nat Rev Immunol. 2011 Nov 11;11(12):823-36. doi: 10.1038/nri3084.

Towards a systems understanding of MHC class I and MHC class II antigen presentation.

Neefjes J¹, Jongstra ML, Paul P, Bakke O.

PMID: 22076556 DOI: 10.1038/nri3084

Class II MHC molecule

- Only expressed by professional antigen presenting cells
- Two MHC encoded polymorphic chains (α , β)
- Both α and β chains impact binding
- Binding groove is open
- Can bind **longer peptides** (13-25 AA)
- Presents peptides to **CD4+ T cells**

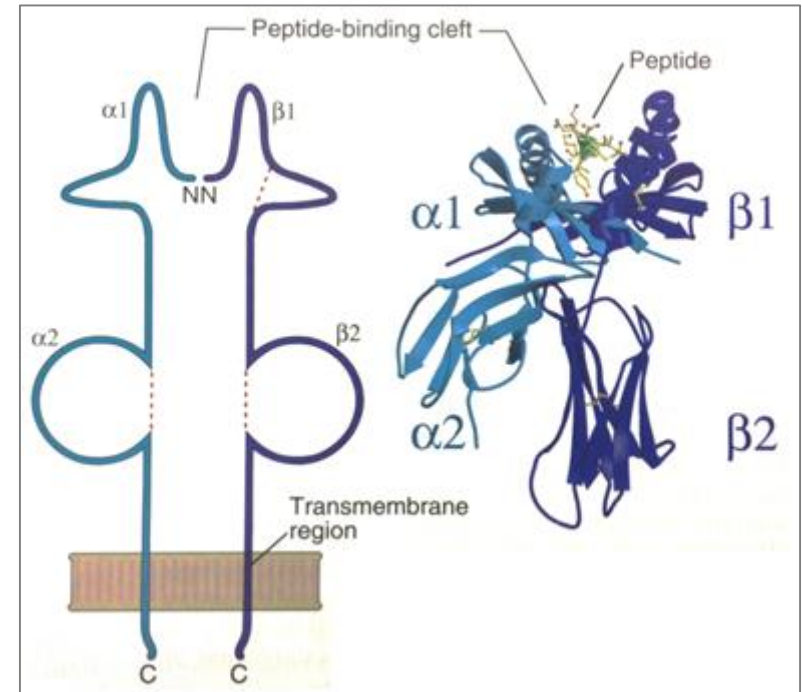


Figure Source
Cellular & Molecular Immunology, 5th Ed by Abbas and Lichtman

HLA Nomenclature

- Class I:
 - Only α chain is variable
 - HLA-B*07:02 (“B” is locus)
 - β 2-microglobulin
- Class II:
 - Both α and β chains are variable for DP & DQ loci
 - HLA-DPA1*01:03/DPB1*02:01
 - HLA-DQA1*01:01/DQB1*05:01
 - Only β chain is variable for DR locus
 - HLA-DRB1*01:01

Class II binding peptide “Binding core”

- 9 AA core within the peptide that interacts with the binding groove of MHC molecule

Binding Core

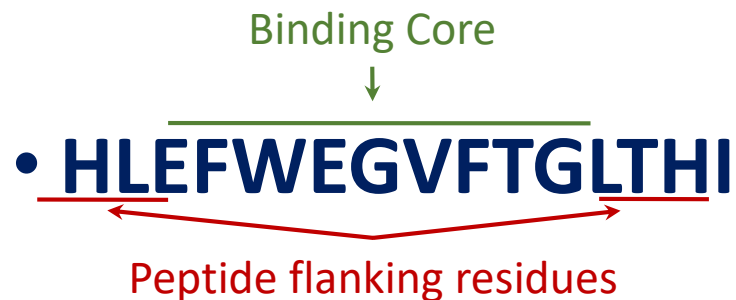


• **HLEFWEGVFTGLTHI**

- Challenge: Correct identification of the binding core
- Needs proper alignment of the binding core with the binding groove

“Peptide flanking residues” (PFR)

- Residues flanking the binding core - interacts with MHC molecule outside the groove.



- Challenge: PFR length & composition (to a lesser degree) influence binding.

Other differences between class I & II tools

- Lesser accuracy compared to class-I tool

Class I		Class II	
Method	AUC*	Method	AUC*
NetMHCpan-4.0	0.960 ¹	NetMHCIIpan-3.2	0.781 ²
SMM	0.894 ³	SMM-align	0.763 ⁴

** The AUCs reported here are from different studies and obtained from different data sets*

- Less stringent threshold for selecting binders than class-I

1. Jurtz et al., 2017, J of Immunology
2. Jensen et al. 2018, Immunology
3. Kim et al. 2009, BMC Bioinformatics
4. Wang et al. 2010, BMC Bioinformatics

MHC-II binding prediction interface

- Tool entry point layout similar to class I

The screenshot shows the MHC-II Binding Predictions interface. At the top, a navigation bar contains links for Home, Help, Example, Reference, Download, and Contact. The main heading is "MHC-II Binding Predictions".

Specify Sequence(s)

Enter protein sequence(s) in FASTA format
(Browse for sequences in NCBI)

Or select file containing sequence(s) No file selected.

Choose a Prediction Method

Prediction Method [?] IEDB recommended 2.22 [Help on prediction method selections](#)

Show all the method versions:

Specify what to make binding predictions for

Select species/locus: Human, HLA-DR

Select MHC allele(s)
Select α & β chains separately if applicable: [?] Allele

Select full HLA reference set: [?]

Select 7-allele HLA reference set: [?]

Select length(s) [?]

default 12-18 as is

11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30

Specify Output

Sort peptides by: Adjusted Rank

Output format: XHTML table

Email address (optional): [?]

tools.iedb.org/mhcii/

Insert protein sequence(s)

Select prediction method

Select species

Select allele(s) & length

Output options

MHC-II binding prediction - example

tools.iedb.org/mhcii/

Home Help Example Reference Download Contact

MHC-II Binding Predictions

Specify Sequence(s)

Enter protein sequence(s) in FASTA format

```
>West Nile virus envelope glycoprotein
FNCLGMSNRDFLEGVSGATWVDLVLEGDSVITMSKDKPTIDVKMMNMEAANLAEVRSYCYLATVSDLST
KAACPTMGEAHNDKRAPAFVCRQGVVDRGWGNGCGLFGKGSIDTCAKFACTKAIGRTILKENIKYEVA
IFVHGPTTVESHGNYSTQVGATQAGRFSITPAAPSYTLKLGEGEVTVDCEPRSGIDTNAYVMTVGTKT
FLVHREWFMDLNLWPSSAGSTVWRNRETLMEFEEPHATKQSVIALGSQEGALHQALAGAIPVEFSNTVK
LTSGHLKCRVKMEKLQKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDGPCKVPISVASLNDLT
PVGRLVTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLKGAQRLAA
LGDTAWDFGSVGGVFTSVGKAVHQVFGGAFRSLFGGMSWITQGLLALLWGMGINARDRSIALTFLAVGG
VLLFLSVNVHA
```

Or select file containing sequence(s) No file chosen

Choose a Prediction Method

Prediction Method [?]
Show all the method versions:

Specify

Select species/locus

Select MHC allele(s)
Select α & β chains separately if applicable: [?]

Select full HLA reference set: [?]
Select 7-allele HLA reference set: [?]

Select length(s) [?]

default 12-18 as is

11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30

Specify Output

Sort peptides by: Adjusted Rank

Output format: XHTML table

Email address (optional): [?]

Prediction method

Guidelines: Choosing the prediction method

- Suggested method = “IEDB recommended”
 - Employs Consensus (Combination of NN-align, SMM-align & CombLib/Sturniolo) or NetMHCIIpan depending on the allele
 - Provides binding affinity & percentile rank for each method separately as well
- Recommendation will change with the new benchmark studies

MHC-II binding prediction – example

Home Help Example Reference Download Contact

tools.iedb.org/mhcii/

MHC-II Binding Predictions

Specify Sequence(s)

Enter protein sequence(s) in FASTA format

```
>West Nile virus envelope glycoprotein
FNCLGMSNRDFLEGVSGATWVDLVLEGDSCTIMSKDKPTIDVKMMNMEEANLAEVRSYCYLATVSDLST
KAACPTMGEAHNDRADPAFVCRQGVVDRGWGNGCGLFGKGSIDTCAKFACTSKAIGRTILKENIKYEVA
IFVHGPTTVESHGNYSTQVGATQAGRFSITPAAPSYTLKLGEGEVTVDCEPRSGIDTNAYYVMTVGTKT
FLVHREWFMDLNLWPSSAGSTVWRNRETLMEFEHPATKQSVIALGSQEGALHQAALAGAIPEVSSNTVK
LTSGHLKCRVKMEKLQKGGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTDGPKVPISSVASLNDLT
PVGRLVTVPFVSVATANAKVLIIEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTLKGAQRLAA
LGDTAWDFGSGVGGVTSVGKAVHQVFGGAFRSLFGGMSWITQGLLGALLLWMMGINARDRSIALTFLAVGG
VLLFLSVNVHA
```

Or select file containing sequence(s) No file chosen

Choose a Prediction Method

Prediction Method [?] IEDB recommended 2.22 [Help on prediction method selections](#)
Show all the method versions:

Specify what to make binding predictions for

Select species/locus **Human, HLA-DR** [?]

Select MHC allele(s)
Select α & β chains separately if applicable: [?]

Select full HLA reference set: [?]
Select 7-allele HLA reference set: [?]

[Upload allele file](#) [?]

Select length(s) [?]

default 12-18 as is

11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30

Specify Output

Sort peptides by Adjusted Rank [?]

Output format XHTML table [?]

Email address (optional) [?]

Choose species & locus

Select or upload alleles

Upload format:

H2-IAb

HLA-DPA1*01/DPB1*04:01

HLA-DRB1*01:01

Allele selection - α and β chains separately

tools.iedb.org/mhcii/

Specify what to make binding predictions for

Select species/locus: Human, HLA-DQ

Select MHC allele(s)
Select α & β chains separately if applicable: ?

Select full HLA reference set: ?
Select 7-allele HLA reference set: ?

Select length(s): ?

Allele

DQA1*01:01

Upload allele file ?

default	12-18				
11	12	13	14	19	20
21	22	23	24	29	30

Specify O

Sort peptides by: Adjusted Rank

Output format: XHTML table

Email address (optional):

Submit Reset

DQB1*02:01
DQB1*02:02
DQB1*02:03
DQB1*02:04
DQB1*02:05
DQB1*02:06
DQB1*03:01
DQB1*03:02
DQB1*03:03
DQB1*03:04
DQB1*03:05
DQB1*03:06
DQB1*03:07
DQB1*03:08
DQB1*03:09
DQB1*03:10
DQB1*03:11
DQB1*03:12
DQB1*03:13

Allele selection – 27 allele reference set

tools.iedb.org/mhcii/

Specify what to make binding predictions for

Select species/locus: Human, HLA-DQ ▾

Select MHC allele(s)

Select α & β chains separately if applicable: ?

Select full HLA reference set: ?

Select 7-allele HLA reference set: ?

Allele	
HLA-DRB1*01:01	<input type="radio"/>
HLA-DRB1*03:01	<input type="radio"/>
HLA-DRB1*04:01	<input type="radio"/>
HLA-DRB1*04:05	<input type="radio"/>
HLA-DRB1*07:01	<input type="radio"/>
HLA-DRB1*08:02	<input type="radio"/>
HLA-DRB1*09:01	<input type="radio"/>
HLA-DRB1*11:01	<input type="radio"/>
HLA-DRB1*12:01	<input type="radio"/>
HLA-DRB1*13:02	<input type="radio"/>
HLA-DRB1*15:01	<input type="radio"/>
HLA-DRB3*01:01	<input type="radio"/>
HLA-DRB3*02:02	<input type="radio"/>
HLA-DRB4*01:01	<input type="radio"/>
HLA-DRB5*01:01	<input type="radio"/>
HLA-DQA1*05:01/DQB1*02:01	<input type="radio"/>
HLA-DQA1*05:01/DQB1*03:01	<input type="radio"/>
HLA-DQA1*03:01/DQB1*03:02	<input type="radio"/>
HLA-DQA1*04:01/DQB1*04:02	<input type="radio"/>
HLA-DQA1*01:01/DQB1*05:01	<input type="radio"/>
HLA-DQA1*01:02/DQB1*06:02	<input type="radio"/>
HLA-DPA1*02:01/DPB1*01:01	<input type="radio"/>
HLA-DPA1*01:03/DPB1*02:01	<input type="radio"/>
HLA-DPA1*01/DPB1*04:01	<input type="radio"/>
HLA-DPA1*03:01/DPB1*04:02	<input type="radio"/>
HLA-DPA1*02:01/DPB1*05:01	<input type="radio"/>
HLA-DPA1*02:01/DPB1*14:01	<input type="radio"/>

DQA1*01:01 ▾ ▾ [Upload allele file](#) ?

Select length(s)

default 12-18 as is

11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30

Allele selection – 7 allele set

tools.iedb.org/mhcii/

Specify what to make binding predictions for

Select species/locus: Human, HLA-DR ▾

Select MHC allele(s)

Select α & β chains separately if applicable: ?

Select full HLA reference set: ?

Select 7-allele HLA reference set: ?

Allele	
HLA-DRB1*03:01	<input type="radio"/>
HLA-DRB1*07:01	<input type="radio"/>
HLA-DRB1*15:01	<input type="radio"/>
HLA-DRB3*01:01	<input type="radio"/>
HLA-DRB3*02:02	<input type="radio"/>
HLA-DRB4*01:01	<input type="radio"/>
HLA-DRB5*01:01	<input type="radio"/>

[Upload allele file](#) ?

[J Immunol Methods](#). 2015 Jul;422:28-34. doi: 10.1016/j.jim.2015.03.022. Epub 2015 Apr 7.

Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes.

Paul S¹, Lindestam Arlehamn CS², Scriba TJ³, Dillon MB², Oseroff C², Hinz D², McKinney DM², Carrasco Pro S⁴, Sidney J², Peters B², Sette A².

PMID: 25862607 PMCID: [PMC4458426](#) DOI: [10.1016/j.jim.2015.03.022](#)

MHC-II binding prediction – example

tools.iedb.org/mhcii/

Home Help Example Reference Download Contact

MHC-II Binding Predictions

Specify Sequence(s)

Enter protein sequence(s) in FASTA format
([Browse for sequences in NCBI](#))

```
>West Nile virus envelope glycoprotein
FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEAANLA
EVRSYCYLATVSDLST
KAACPTMGEAHNDKRAPAFVCRQGVVDRGWGNGCGLFGKGSIDTCAKFCACST
KAIGRTLKENIKYEVA
IFVHGPTTVESHGNYSTQVGATQAGRFSITPAAPSYTLKLEGEYGEVTVDCEPRSGI
DTNAYYVMTVGTKT
FLVHREWFMDLNLNPWSSAGSTVWRNRETLMFEFEPHATKQSVIALGSQEGALHQ
ALAGAIPVEFSSNTVK
LTSGLKCRVKMEKLQLKGTTYGVC SKAFKFLGTPADTGHGTVVLELQYTGTDGP
```

FASTA format detected.

Or select file containing sequence(s) No file chosen

Choose a Prediction Method

Prediction Method [?] IEDB recommended 2.22 [Help on prediction method selections](#)
Show all the method versions:

Specify what to make binding predictions for

Select species/locus: Human, HLA-DP

Select MHC allele(s)
Select α & β chains separately if applicable: [?]
Allele: DPA1*01/DPB1*04:01
Select full HLA reference set: [?]
Select 7-allele HLA reference set: [?]
 [Upload allele file](#) [?]

Select length(s) [?]
default 12-18 as is

11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30

Specify Output

Sort peptides by: Adjusted Rank

Output format: XHTML table

Email address (optional): spaul@lji.org [?]

Length selection

Output format



How the tool works

- Breaks sequence into all possible peptides of chosen lengths
- Predicts the binding affinity for each peptide based on the method
- Compares the predicted affinity to that of a large set of randomly selected peptides
- Assigns a percentile rank depending on individual predicted affinity
- Consensus picks median rank of the methods used – consensus percentile rank

MHC-II binding prediction – example

tools.iedb.org/mhcii/

Home Help Example Reference Download Contact

MHC-II Binding Prediction Results

Input Sequences

#	Name	Sequence
1	West Nile virus envelope glycoprotein	FNCLGMSNRDFLEGVSGATWDLVLEGDSCVTIMSKDKPTIDVKMMNHEA ANLAEVRSYCYLATVSDLSTKAACPTMGEAHNDKRADPAFVCRQGVVDRG WGNCGCLFGKGSIDTCAKACSTKAIGRTILKENIKYEVAIFVHGPTTVE SHGNYSIQVGATQAGRFSTIPAAPSYTLKLGVEYGEVTVDCPRSGIDTNA YYVMTVGTKTFLVHREWFMDLNLWSSAGSTVWRNRETLMFEFEPHATKQ SVIALGSQEGALHQALAGAIPEVFSSTVTKLTSGLHKCRVKMEKQLQKGT TYGVCSKAFKFLGTPADTGHGTVVLELQYGTGTGDPCKVPISSVASLNDLT PVGRVLTVPFVSVATANAKVLEIEPPFGDSYIVVGRGEQQINHHWHKS GSSIKAFTTTLKGAQRLAALGDTAWDFGSGVGGVFTSVGKAVHQVFGGAF RSLFGGMSWITQGLLQALLLWGINARDRSIALTF LAVGGVLLFLSVNVH A

Prediction method: IEDB Recommended | Low adjusted_rank = good binders

[Download result](#)

Citations

Check to expand the result:

Allele	#	Start	End	Length	Method used	Peptide	Percentile Rank	Adjusted rank
HLA-DPA1*01/DPB1*04:01	1	476	490	15	Consensus (comb.lib./smm)	ARDRSIALTF LAVGG	2.85	2.85
HLA-DPA1*01/DPB1*04:01	1	474	488	15	Consensus (comb.lib./smm)	INARDRSIALTF LAV	2.85	2.85
HLA-DPA1*01/DPB1*04:01	1	475	489	15	Consensus (comb.lib./smm)	NARDRSIALTF LAVG	2.85	2.85
HLA-DPA1*01/DPB1*04:01	1	477	491	15	Consensus (comb.lib./smm)	RDRSIALTF LAVGGV	2.85	2.85
HLA-DPA1*01/DPB1*04:01	1	478	492	15	Consensus (comb.lib./smm)	DRSIALTF LAVGGVLL	2.95	2.95
HLA-DPA1*01/DPB1*04:01	1	207	221	15	Consensus (comb.lib./smm)	GTKTF LVHREWFMDL	3.55	3.55
HLA-DPA1*01/DPB1*04:01	1	209	223	15	Consensus (comb.lib./smm)	KTF LVHREWFMDLNL	3.60	3.60
HLA-DPA1*01/DPB1*04:01	1	208	222	15	Consensus (comb.lib./smm)	TKTF LVHREWFMDLN	3.60	3.60
HLA-DPA1*01/DPB1*04:01	1	479	493	15	Consensus (comb.lib./smm)	RSIALTF LAVGGVLL	3.95	3.95
HLA-DPA1*01/DPB1*04:01	1	200	214	15	Consensus (comb.lib./smm)	AYYVMTVGTKTF LVH	4.05	4.05
HLA-DPA1*01/DPB1*04:01	1	202	216	15	Consensus (comb.lib./smm)	YVMTVGTKTF LVHRE	4.05	4.05
HLA-DPA1*01/DPB1*04:01	1	203	217	15	Consensus (comb.lib./smm)	VMTVGTKTF LVHREW	4.10	4.10
HLA-DPA1*01/DPB1*04:01	1	201	215	15	Consensus (comb.lib./smm)	YYVMTVGTKTF LVHR	4.10	4.10
HLA-DPA1*01/DPB1*04:01	1	483	497	15	Consensus (comb.lib./smm)	LTF LAVGGVLL FLSV	4.50	4.50
HLA-DPA1*01/DPB1*04:01	1	204	218	15	Consensus (comb.lib./smm)	MTVGTKTF LVHREWF	4.71	4.71
HLA-DPA1*01/DPB1*04:01	1	440	454	15	Consensus (comb.lib./smm)	KAVHQVFGGAF RSLF	4.95	4.95
HLA-DPA1*01/DPB1*04:01	1	441	455	15	Consensus (comb.lib./smm)	AVHQVFGGAF RSLFG	5.00	5.00
HLA-DPA1*01/DPB1*04:01	1	443	457	15	Consensus (comb.lib./smm)	HQVFGGAF RSLFGGH	5.00	5.00
HLA-DPA1*01/DPB1*04:01	1	442	456	15	Consensus (comb.lib./smm)	VHQVFGGAF RSLFGG	5.10	5.10
HLA-DPA1*01/DPB1*04:01	1	439	453	15	Consensus (comb.lib./smm)	GKAVHQVFGGAF RSL	5.20	5.20

Input sequence

Output
(sorted low-to-high by adjusted rank)

The **adjusted rank** is the percentile rank adjusted based on the frequency of peptide lengths.

MHC-II binding prediction – example

tools.iedb.org/mhcii/

Individual scores for different methods

Prediction method: IEDB recommended | Low adjusted_rank = good binders
[Download result](#)

Citations
 Check to expand the result

Allele	#	Start	End	Length	Method used	Peptide	Percentile Rank	Adjusted rank	Comblib. core	Comblib. score	Comblib. percentile rank	Comblib. adjusted rank	SMM align core	SMM align IC50(nM)	SMM align percentile rank	SMM align adjusted rank
HLA-DPA1*01:DPB1*04:01	1	476	490	15	Consensus (comb.lib./simm)	ARDRSIALTFLAVGG	2.85	2.85	RSIALTFLA	0.01	2.80	2.80	SIALTFLAV	208.00	2.90	2.90
HLA-DPA1*01:DPB1*04:01	1	474	488	15	Consensus (comb.lib./simm)	INARDRSIALTFLAV	2.85	2.85	RSIALTFLA	0.01	2.80	2.80	RSIALTFLA	207.00	2.90	2.90
HLA-DPA1*01:DPB1*04:01	1	475	489	15	Consensus (comb.lib./simm)	NARDRSIALTFLAVG	2.85	2.85	RSIALTFLA	0.01	2.80	2.80	SIALTFLAV	203.00	2.90	2.90
HLA-DPA1*01:DPB1*04:01	1	477	491	15	Consensus (comb.lib./simm)	RDRSIALTFLAVGGV	2.85	2.85	RSIALTFLA	0.01	2.80	2.80	SIALTFLAV	205.00	2.90	2.90
HLA-DPA1*01:DPB1*04:01	1	478	492	15	Consensus (comb.lib./simm)	DRSIALTFLAVGGVL	2.95	2.95	RSIALTFLA	0.01	2.80	2.80	SIALTFLAV	221.00	3.10	3.10
HLA-DPA1*01:DPB1*04:01	1	207	221	15	Consensus (comb.lib./simm)	GKTFVLVHREWFMDL	3.55	3.55	KTFVLVHREW	0.03	3.90	3.90	FLVHREWFMDL	230.00	3.20	3.20
HLA-DPA1*01:DPB1*04:01	1	209	223	15	Consensus (comb.lib./simm)	KTFVLVHREWFMDLNL	3.60	3.60	KTFVLVHREW	0.03	3.90	3.90	VHREWFMDL	232.00	3.30	3.30
HLA-DPA1*01:DPB1*04:01	1	208	222	15	Consensus (comb.lib./simm)	TKTFVLVHREWFMDLN	3.60	3.60	KTFVLVHREW	0.03	3.90	3.90	VHREWFMDL	232.00	3.30	3.30
HLA-DPA1*01:DPB1*04:01	1	479	493	15	Consensus (comb.lib./simm)	RSIALTFLAVGGVLL	3.95	3.95	RSIALTFLA	0.01	2.80	2.80	SIALTFLAV	348.00	5.10	5.10
HLA-DPA1*01:DPB1*04:01	1	200	214	15	Consensus (comb.lib./simm)	AYYVMTVGTKTFLVH	4.05	4.05	TVGKTFLV	0.01	0.01	0.01	TVGKTFLV	579.00	8.10	8.10
HLA-DPA1*01:DPB1*04:01	1	202	216	15	Consensus (comb.lib./simm)	YVMTVGTKTFLVHRE	4.05	4.05	TVGKTFLV	0.01	0.01	0.01	VGTKTFLVH	583.00	8.10	8.10
HLA-DPA1*01:DPB1*04:01	1	203	217	15	Consensus (comb.lib./simm)	VMTVGTKTFLVHREW	4.10	4.10	TVGKTFLV	0.01	0.01	0.01	VGTKTFLVH	593.00	8.20	8.20
HLA-DPA1*01:DPB1*04:01	1	201	215	15	Consensus (comb.lib./simm)	YYVMTVGTKTFLVHR	4.10	4.10	TVGKTFLV	0.01	0.01	0.01	VGTKTFLVH	585.00	8.20	8.20
HLA-DPA1*01:DPB1*04:01	1	483	497	15	Consensus	LTF LAVGGVLLFLSV	4.50	4.50	AVGGVLLFL	0.03	3.90	3.90	FLAVGGVLL	346.00	5.10	5.10

Guidelines: Selecting binders

- Based on Percentile rank or MHC binding affinity?
Recommendation: **IEDB Percentile rank**
- Threshold guidelines:
 - Percentile rank \leq **10.0** (Percentile rank on linear scale (0-100), lower value = better binder)
 - MHC binding affinity $IC_{50} \leq$ **1000nM**
- Select all peptides with IEDB percentile rank \leq 10.0

Alternate approaches for selecting binders

- Change threshold values depending on your need
 - e.g. in case you have too few or too many predicted binders.
- Set a desired percentage within your peptide set (irrespective of IEDB percentile rank) in case you want to study a fixed number of best possible peptides.

Issue of overlapping peptides

- The tool breaks the sequence into all possible 15-mers - Peptides overlapping by 14 amino acid residues

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	allele	seq_n	start	end	peptide	method	percent	comblib_core	comblib	comblib	smm_align_core	smm_al	smm_al	nn_align_core	nn_a
2	HLA-DPA1*01/DPB1*0401	1	527	541	DHLEFWEGVFTGLTH	Consensus (com	2.52	FWEGVFTGL	6.9	19.59	FWEGVFTGL	310	0.62	EFWEGVFTG	3
3	HLA-DPA1*01/DPB1*0401	1	528	542	HLEFWEGVFTGLTHI	Consensus (com	2.57	FWEGVFTGL	6.9	19.59	FWEGVFTGL	302	0.58	FWEGVFTGL	4
4	HLA-DPA1*01/DPB1*0401	1	526	540	QDHLEFWEGVFTGLT	Consensus (com	2.62	FWEGVFTGL	6.9	19.59	FWEGVFTGL	310	0.62	EFWEGVFTG	4
5	HLA-DPA1*01/DPB1*0401	1	529	543	LEFWEGVFTGLTHID	Consensus (com	3.13	FWEGVFTGL	6.9	19.59	FWEGVFTGL	308	0.61	FWEGVFTGL	5
6	HLA-DPA1*01/DPB1*0401	1	525	539	CQDHLEFWEGVFTGL	Consensus (com	3.26	FWEGVFTGL	6.9	19.59	EFWEGVFTG	320	0.66	FWEGVFTGL	5
7	HLA-DPA1*01/DPB1*0401	1	39	53	AAQTFLATCINGVCW	Consensus (com	3.8	QTFLATCIN	728.23	45.84	FLATCINGV	742	3.36	FLATCINGV	6
8	HLA-DPA1*01/DPB1*0401	1	262	276	GSPITYSTYGKFLAD	Consensus (com	4.07	TYSTYGKFL	2.38	15.24	ITYSTYGKF	827	4.07	ITYSTYGKF	
9	HLA-DPA1*01/DPB1*0401	1	40	54	AQTFLATCINGVCWT	Consensus (com	4.08	QTFLATCIN	728.23	45.84	FLATCINGV	746	3.39	FLATCINGV	7
10	HLA-DPA1*01/DPB1*0401	1	263	277	SPITYSTYGKFLADG	Consensus (com	4.08	TYSTYGKFL	2.38	15.24	ITYSTYGKF	828	4.08	ITYSTYGKF	5
11	HLA-DPA1*01/DPB1*0401	1	38	52	TAAQTFLATCINGVC	Consensus (com	4.13	TAAQTFLAT	52.52	29.77	FLATCINGV	478	1.49	FLATCINGV	7
12	HLA-DPA1*01/DPB1*0401	1	37	51	STAAQTFLATCINGV	Consensus (com	4.56	TAAQTFLAT	52.52	29.77	TAAQTFLAT	464	1.41	FLATCINGV	8
13	HLA-DPA1*01/DPB1*0401	1	261	275	TGSPITYSTYGKFLA	Consensus (com	4.78	TYSTYGKFL	2.38	15.24	ITYSTYGKF	908	4.78	ITYSTYGKF	
14	HLA-DPA1*01/DPB1*0401	1	530	544	EFWEGVFTGLTHIDA	Consensus (com	5	FWEGVFTGL	6.9	19.59	FWEGVFTGL	664	2.75	FWEGVFTGL	9
15	HLA-DPA1*01/DPB1*0401	1	102	116	SDLYLVRHADVIPV	Consensus (com	7.45	LVTRHADVI	23.49	25.43	YLVRHADV	1194	7.45	YLVRHADV	14
16	HLA-DPA1*01/DPB1*0401	1	41	55	QTFLATCINGVCWTV	Consensus (com	7.57	QTFLATCIN	728.23	45.84	FLATCINGV	829	4.09	FLATCINGV	16
17	HLA-DPA1*01/DPB1*0401	1	101	115	SSDLYLVRHADVIP	Consensus (com	7.57	LVTRHADVI	23.49	25.43	YLVRHADV	1206	7.57	YLVRHADV	16
18	HLA-DPA1*01/DPB1*0401	1	260	274	TTGSPITYSTYGKFL	Consensus (com	7.71	TYSTYGKFL	2.38	15.24	ITYSTYGKF	1221	7.71	ITYSTYGKF	10
19	HLA-DPA1*01/DPB1*0401	1	100	114	GSSDLYLVRHADVI	Consensus (com	7.85	GSSDLYLVT	0.74	11.33	YLVRHADV	1183	7.34	YLVRHADV	17
20	HLA-DPA1*01/DPB1*0401	1	531	545	FWEGVFTGLTHIDAH	Consensus (com	7.97	FWEGVFTGL	6.9	19.59	FWEGVFTGL	728	3.24	FWEGVFTGL	17
21	HLA-DPA1*01/DPB1*0401	1	103	117	DLYLVRHADVIPVR	Consensus (com	8.57	LVTRHADVI	23.49	25.43	YLVRHADV	1307	8.57	YLVRHADV	16

Issue of overlapping peptides: Solution

- Pre-processing:
 - Generate 15mers overlapping by 10 AA residues and do the prediction

APITAYAQQTRGLLGCIITSLTGRD

APITAYAQQTRGLLG-----

-----YAQQTRGLLGCIITS-----

-----RLLGCIITSLTGRD

- 15 is mostly preferred length for class II
- 10 AA overlap captures minimal 15-mers with all possible 9-mer binding cores with at least 1 flanking residue
- Python/Perl script or Excel

Issue of overlapping peptides: Solution

- Post-processing:
 - Remove largely overlapping peptides after prediction (based on same binding core or position)

	A	B	C	D	E	G	H	I	J	K	L	M	N	O
1	allele	seq_n	start	end	peptide	percent	comblib_core	comblib	comblib	smm_align_core	smm_al	smm_al	nn_align_core	nn_a
2	HLA-DPA1*01/DPB1*0401	1	527	541	DHLEFWEGVFTGLTH	2.52	FWEGVFTGL	6.9	19.59	FWEGVFTGL	310	0.62	EFWEGVFTG	3
3	HLA-DPA1*01/DPB1*0401	1	528	542	HLEFWEGVFTGLTHI	2.57	FWEGVFTGL	6.9	19.59	FWEGVFTGL	302	0.58	FWEGVFTGL	4
4	HLA-DPA1*01/DPB1*0401	1	526	540	QDHLEFWEGVFTGLT	2.62	FWEGVFTGL	6.9	19.59	FWEGVFTGL	310	0.62	EFWEGVFTG	4
5	HLA-DPA1*01/DPB1*0401	1	529	543	LEFWEGVFTGLTHID	3.13	FWEGVFTGL	6.9	19.59	FWEGVFTGL	308	0.61	FWEGVFTGL	5
6	HLA-DPA1*01/DPB1*0401	1	525	539	CQDHLEFWEGVFTGL	3.26	FWEGVFTGL	6.9	19.59	EFWEGVFTG	320	0.66	FWEGVFTGL	5
7	HLA-DPA1*01/DPB1*0401	1	39	53	AAQTFLATCINGVCW	3.8	QTFLATCIN	728.23	45.84	FLATCINGV	742	3.36	FLATCINGV	6
8	HLA-DPA1*01/DPB1*0401	1	262	276	GSPITYSTYGKFLAD	4.07	TYSTYGKFL	2.38	15.24	ITYSTYGKF	827	4.07	ITYSTYGKF	
9	HLA-DPA1*01/DPB1*0401	1	40	54	AQTFLATCINGVCWT	4.08	QTFLATCIN	728.23	45.84	FLATCINGV	746	3.39	FLATCINGV	7
10	HLA-DPA1*01/DPB1*0401	1	263	277	SPITYSTYGKFLADG	4.08	TYSTYGKFL	2.38	15.24	ITYSTYGKF	828	4.08	ITYSTYGKF	5

Promiscuous binders

- Peptides that bind to more than one MHC molecule.
- Significance:
 - Associated with stronger antigenicity & larger population coverage
 - Important in reducing immunogenicity of therapeutic proteins
 - Can be predicted based on binding affinity
- Consensus percentile rank threshold \leq **20.0**

[J Immunol](#). 2010 Jul 15;185(2):943-55. doi: [10.4049/jimmunol.1000405](#). Epub 2010 Jun 16.

Molecular determinants of T cell epitope recognition to the common Timothy grass allergen.

[Oseroff C](#)¹, [Sidney J](#), [Kotturi MF](#), [Kolla R](#), [Alam R](#), [Broide DH](#), [Wasserman SI](#), [Weiskopf D](#), [McKinney DM](#), [Chung JL](#), [Petersen A](#), [Grey H](#), [Peters B](#), [Sette A](#).

PMID: 20554959 PMCID: [PMC3310373](#) DOI: [10.4049/jimmunol.1000405](#)

Promiscuous binders - Multiple alleles

MHC-II Binding Predictions

Specify Sequence(s)

Enter protein sequence(s) in FASTA format
[\(Browse for sequences in NCBI\)](#)

```
>HCV_NS3
APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCINGVCWTVYHGAGTRTIASPKGP
VIQMYTNVDQDLVGWPAAPQGSRLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPISYLKGS
GPLLCPAGHAVGLFRAAVCTRGVAKAVDFIPVENLETTMRSPVFTDNSSPPAVQSFQVAHLHAPTGSGK
STKVPAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGVDPNIRTGVRTITGSPITYSTYGKFLADGGCS
GGAYDIIICDECHSTDATSI LGIGTVLDQAETAGARLVV LATATPPG SVTVSHPNIEEVALSTTGEIPFY
GKAIPLEVIKGRHLIFCHSKKKCDELAAKLVALGINAVAYYRGLDVSVIPTSGDVVVVSTDALMTGFTG
DFDSVIDCNTCVTQTVDFSLDPTFTIETTLPQDAVSRTRRRGRTGRGKPGIYRFVAPGERPSGMFDSSV
LCECYDAGCAWYELTPAETT VRLRAYMNT PGLPVCQDHLEFWEGVFTGLTHIDAHFLSQT KQSGENFPYL
VAYQATVCARAQAPPPSWDQMWKCLIRL KPTLHGPTPLLYRLGAVQNEVTLTHPITKYIMTCMSADLEVV
T
```

FASTA format detected.

Or select file containing sequence(s)

Browse...

No file selected.

Choose a Prediction Method

Prediction Method ?

Show all the method versions:

IEDB recommended 2.22
▼
[Help on prediction method selections](#)

Specify what to make binding predictions for

Select species/locus

Select MHC allele(s)

Select α & β chains separately if applicable: ?

Select full HLA reference set: ?

Select 7-allele HLA reference set: ?

Human, HLA-DR ▼

Allele

DPA1*01/DPB1*04:01
⊗

DPA1*03:01/DPB1*04:02
⊗

DPA1*02:01/DPB1*05:01
⊗

DRB1*01:01
⊗

▼
[Upload allele file](#) ?

Select length(s) ?

default

12-18

as is

11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30

Multiple alleles - result

Allele	#	Start	End	Length	Method used	Peptide	Percentile Rank	Adjusted rank
HLA-DRB1*01:01	1	222	236	15	Consensus (comb.lib./smm/nn)	GYKVLVLNPSVAATL	0.14	0.14
HLA-DRB1*01:01	1	221	235	15	Consensus (comb.lib./smm/nn)	QGYKVLVLNPSVAAT	0.14	0.14
HLA-DRB1*01:01	1	220	234	15	Consensus (comb.lib./smm/nn)	AQGYKVLVLNPSVAA	0.39	0.39
HLA-DRB1*01:01	1	223	237	15	Consensus (comb.lib./smm/nn)	YKVLVLNPSVAATLG	0.39	0.39
HLA-DRB1*01:01	1	224	238	15	Consensus (comb.lib./smm/nn)	KVLVLNPSVAATLGF	1.30	1.30
HLA-DRB1*01:01	1	219	233	15	Consensus (comb.lib./smm/nn)	AAQGYKVLVLNPSVA	1.80	1.80
HLA-DRB1*01:01	1	378	392	15	Consensus (comb.lib./smm/nn)	AAKLVALGINAVAYY	3.60	3.60
HLA-DRB1*01:01	1	379	393	15	Consensus (comb.lib./smm/nn)	AKLVALGINAVAYYR	3.60	3.60
HLA-DRB1*01:01	1	375	389	15	Consensus (comb.lib./smm/nn)	DELAAKLVALGINAV	3.60	3.60
HLA-DRB1*01:01	1	376	390	15	Consensus (comb.lib./smm/nn)	ELAAKLVALGINAVA	3.60	3.60
HLA-DRB1*01:01	1	377	391	15	Consensus (comb.lib./smm/nn)	LAAKLVALGINAVAY	3.60	3.60
HLA-DRB1*01:01	1	225	239	15	Consensus (comb.lib./smm/nn)	VLVLNPSVAATLGF	3.90	3.90
HLA-DRB1*01:01	1	386	400	15	Consensus (comb.lib./smm/nn)	INAVAYYRGLDVSVI	6.00	6.00
HLA-DRB1*01:01	1	512	526	15	Consensus (comb.lib./smm/nn)	RLRAYMNTPLPVCQ	6.00	6.00
HLA-DRB1*01:01	1	388	402	15	Consensus (comb.lib./smm/nn)	AVAYYRGLDVSVIPT	6.30	6.30
HLA-DRB1*01:01	1	389	403	15	Consensus (comb.lib./smm/nn)	VAYYRGLDVSVIPTS	6.30	6.30
HLA-DRB1*01:01	1	511	525	15	Consensus (comb.lib./smm/nn)	VRLRAYMNTPLPVC	6.40	6.40
HLA-DRB1*01:01	1	555	569	15	Consensus (comb.lib./smm/nn)	ENFPYLVAIQATVCA	6.50	6.50
HLA-DRB1*01:01	1	557	571	15	Consensus (comb.lib./smm/nn)	FPYLVAIQATVCARA	6.50	6.50
HLA-DRB1*01:01	1	554	568	15	Consensus (comb.lib./smm/nn)	GENFPYLVAIQATVC	6.50	6.50
HLA-DRB1*01:01	1	387	401	15	Consensus (comb.lib./smm/nn)	NAVAYYRGLDVSVIP	6.50	6.50
HLA-DRB1*01:01	1	556	570	15	Consensus (comb.lib./smm/nn)	NFPYLVAIQATVCAR	6.50	6.50
HLA-DRB1*01:01	1	553	567	15	Consensus (comb.lib./smm/nn)	SGENFPYLVAIQATV	6.50	6.50
HLA-DRB1*01:01	1	380	394	15	Consensus (comb.lib./smm/nn)	KLVALGINAVAYYRG	7.30	7.30
HLA-DRB1*01:01	1	513	527	15	Consensus (comb.lib./smm/nn)	LRAYMNTPLPVCQD	7.50	7.50
HLA-DRB1*01:01	1	558	572	15	Consensus (comb.lib./smm/nn)	PYLVAIQATVCARAQ	7.60	7.60
HLA-DRB1*01:01	1	559	573	15	Consensus (comb.lib./smm/nn)	YLVAIQATVCARAQA	7.60	7.60
HLA-DRB1*01:01	1	372	386	15	Consensus (comb.lib./smm/nn)	KKCELAAKLVALGI	8.00	8.00
HLA-DRB1*01:01	1	514	528	15	Consensus (comb.lib./smm/nn)	RAYMNTPLPVCQDH	8.00	8.00
HLA-DRB1*01:01	1	373	387	15	Consensus (comb.lib./smm/nn)	KCELAAKLVALGIN	8.20	8.20
HLA-DPA1*03:01/DPB1*04:02	1	164	178	15	Consensus (comb.lib./smm)	AKAVDFIPVENLETT	8.30	8.30
HLA-DRB1*01:01	1	226	240	15	Consensus (comb.lib./smm/nn)	LVLNPSVAATLGF	8.60	8.60
HLA-DPA1*03:01/DPB1*04:02	1	37	51	15	Consensus (comb.lib./smm)	STATQTFLATCINGV	8.80	8.80
HLA-DPA1*03:01/DPB1*04:02	1	163	177	15	Consensus (comb.lib./smm)	VAKAVDFIPVENLET	8.85	8.85

Panel of 27 class II alleles to allow for global coverage

Locus	Molecule	Phenotype frequency		Locus	Molecule	Phenotype frequency
DRB1	DRB1*01:01	5.4		DQA1/DQB1	DQA1*05:01/DQB1*02:01	11.3
	DRB1*03:01	13.7			DQA1*05:01/DQB1*03:01	35.1
	DRB1*04:01	4.6			DQA1*03:01/DQB1*03:02	19.0
	DRB1*04:05	6.2			DQA1*04:01/DQB1*04:02	12.8
	DRB1*07:01	13.5			DQA1*01:01/DQB1*05:01	14.6
	DRB1*08:02	4.9			DQA1*01:02/DQB1*06:02	14.6
	DRB1*09:01	6.2			Combined	81.6
	DRB1*11:01	11.8		DPA1/DPB1	DPA1*02:01/DPB1*01:01	16.0
	DRB1*12:01	3.9			DPA1*01:03/DPB1*02:01	17.5
	DRB1*13:02	7.7			DPA1*01/DPB1*04:01	36.2
	DRB1*15:01	12.2			DPA1*03:01/DPB1*04:02	41.6
	Combined	71.1			DPA1*02:01/DPB1*05:01	21.7

[Immunogenetics](#), 2011 Jun;63(6):325-35. doi: 10.1007/s00251-011-0513-0. Epub 2011 Feb 9.

Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes.

[Greenbaum J¹](#), [Sidney J](#), [Chung J](#), [Brander C](#), [Peters B](#), [Sette A](#).

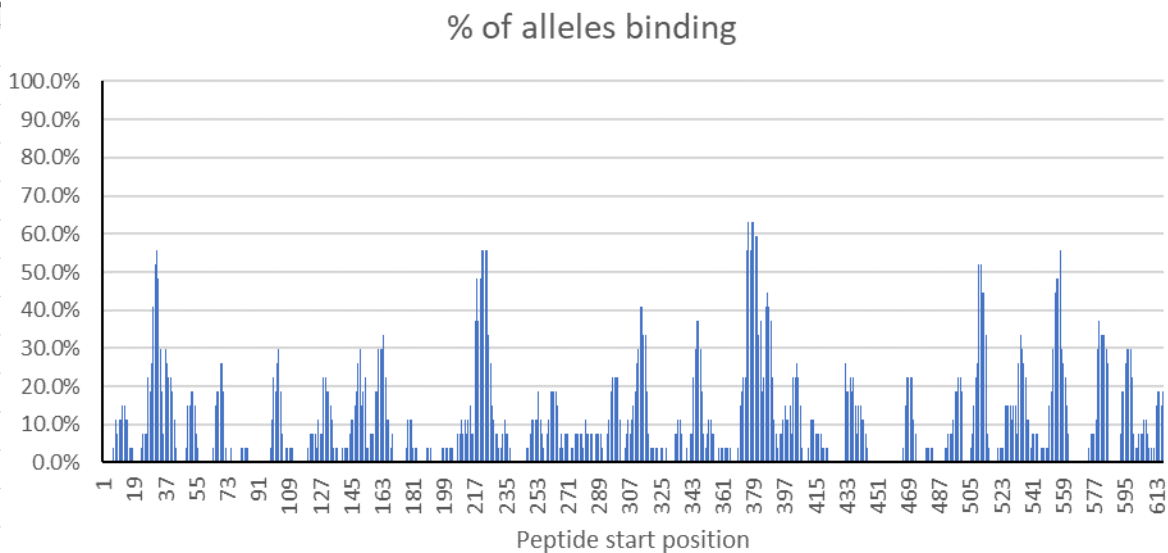
PMID: 21305276 PMID: [PMC3626422](#) DOI: [10.1007/s00251-011-0513-0](#)

DRB3/4/5	DRB3*01:01	26.1				
	DRB3*02:02	34.3				
	DRB4*01:01	41.8				
	DRB5*01:01	16.0				

Promiscuous binders

- Binders with $\geq 50\%$ alleles binding (consensus percentile ≤ 20.0) considered promiscuous binders

star	enc	peptide	Coun	%
31	45	GEVQIVSTATQTFLA	14	51.9%
32	46	EVQIVSTATQTFLAT	15	55.6%
221	235	QGYKVLVLNPSVAAT	15	55.6%
222	236	GYKVLVLNPSVAATL	15	55.6%
223	237	YKVLVLNPSVAATLG	15	55.6%
224	238	KVLVLNPSVAATLGF	15	55.6%
375	389	DELAAKLVALGINAV	15	55.6%
376	390	ELAAKLVALGINAVA	17	63.0%
377	391	LAAKLVALGINAVAY	15	55.6%
378	392	AAKLVALGINAVAYY	17	63.0%
379	393	AKLVALGINAVAYYR	17	63.0%
380	394	KLVALGINAVAYYRG	16	59.3%
381	395	LVALGINAVAYYRGL	16	59.3%
510	524	TVRLRAYMNTPGLPV	14	51.9%
511	525	VRLRAYMNTPGLPVC	14	51.9%
557	571	FPYLVAYQATVCARA	15	55.6%



“7-allele” method

- Aim was to capture maximum immune response with minimum no. of peptides
- 6 peptide datasets with measured immune responses (SFCs/106 PBMCs)
- 15 or 16-mer peptide sets with 10 AA residues overlapping

Dataset	Purpose	No. of Antigens	Total peptides	No. of donors	Reference
Der p/f (House dust mite)	Training data	4	156	20	Hinz et al., 2015, CEA
Phl p (Timothy grass)	Training data	10	425	25	Oseroff et al., 2010, JI
TB-1	Training data	4	71	18	Arlehamn et al., 2012, JI
TB-2	Training data	11	499	32	Arlehamn et al., 2016, PLoS Path
Cockroach	Validation data	6	463	19	Dillon et al., 2015, CEA
Pertussis	Validation data	9	785	23	Bancroft et al., 2016, CEA
TOTAL		44	2399	137	

“7-allele” method

- Optimal results obtained with a set of 7 alleles:
 - 3 DRB1 alleles with frequency $\geq 12\%$ (DRB1*03:01, DRB1*07:01, DRB1*15:01) and 4 DRB3/4/5 alleles (DRB3*01:01, DRB3*02:02, DRB4*01:01, DRB5*01:01)
- Top 21.41% peptides $\approx 50\%$ response
- **The median consensus percentile rank of the 7 alleles ≈ 20.0 - Universal prediction threshold**

[J Immunol Methods](#). 2015 Jul;422:28-34. doi: [10.1016/j.jim.2015.03.022](#). Epub 2015 Apr 7.

Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes.

[Paul S](#)¹, [Lindestam Arlehamn CS](#)², [Scriba TJ](#)³, [Dillon MB](#)², [Oseroff C](#)², [Hinz D](#)², [McKinney DM](#)², [Carrasco Pro S](#)⁴, [Sidney J](#)², [Peters B](#)², [Sette A](#)².

PMID: 25862607 PMCID: [PMC4458426](#) DOI: [10.1016/j.jim.2015.03.022](#)

“7-allele” method

- Generate 15-mers overlapping by 10 AA residues
- Do binding prediction for the **7 selected alleles**
- Estimate the **median consensus percentile rank** (of the 7 alleles)
- Select all peptides with median consensus percentile rank \leq **20.0**
- This set of peptides can capture \approx 50% of the response
- These 7 alleles can be selected as a set
- This is implemented in **Tepitool**

TepiTool

- New interface to prediction of class I and class II epitope candidates
- Motivation:
 - Make tools more user friendly
 - Provide recommendations as default
 - Provide a set of top peptides as concise results
- In the form of a step-by-step wizard (6 steps)
- Input parameters can be adjusted as desired
- New methods incorporated

Step 1: Sequence data

tools.iedb.org/tepitool/

IEDB Analysis Resource

[Home](#) [Help](#) [Reference](#) [Download](#) [Contact](#)

TepiTool

Steps **1** 2 3 4 5 6

SEQUENCE - Provide sequence data:

Enter sequence(s) in FASTA or PLAIN format.

```
>Seq_1
MKALIVLGLVLLSVTVQGKVFCEARTLKRLGMDGYRGISLANWMCLAKW
>Seq_2
MLLALVCLLSCLANSDF
>Seq_3
MKALIVLGLVLLSVTVQGKVFERCELAR
```

FASTA format detected.

Or upload file containing sequence(s) No file chosen

Step 2: Species & allele class

tools.iedb.org/tepitool/

IEDB Analysis Resource

Home Help Reference Download Contact

TepiTool

Steps 1 2 3 4 5 6

SPECIES & ALLELE CLASS - Select the host species and MHC allele class:

Host species	Human
Allele class	Class I

Start Over Back

- Chimpanzee
- Cow
- Gorilla
- Human
- Macaque
- Mouse
- Pig

Current selections:
No. of sequences 3

Step 3: Alleles - Class I

tools.iedb.org/tepitool/

IEDB Analysis Resource

Home Help Reference Download Contact

TepiTool

Steps 1 2 3 4 5 6

ALLELES - Specify alleles:

Human - Class I

- Select from list of frequently occurring alleles (Frequency > 1%)
- Select from list of all available alleles
- Select from list of representative alleles from different HLA supertypes
- Use panel of 27 most frequent A & B alleles
- Upload allele file

Alleles

- A*01:01
- A*02:01
- A*02:06
- A*03:01
- A*11:01
- A*23:01
- A*24:02
- A*25:01
- A*26:01
- A*29:02
- A*30:01
- A*30:02

Start Over Back Next

Current selections:	
No. of sequences	3
Host species	Human
Allele class	Class I
Selected alleles	1.A*02:01 2.A*02:06 3.A*03:01
	Reset alleles

Step 4: Peptides - Class I

tools.iedb.org/tepitool/

IEDB Analysis Resource

Home Help Reference Download Contact

TepiTool

Steps 1 2 3 **4** 5 6

PEPTIDES - Select peptides to be included in prediction:

Peptides to be included in prediction	<input checked="" type="radio"/> Apply default settings for low number of peptides <input type="radio"/> Apply default settings for moderate number of peptides <input type="radio"/> Apply default settings for high number of peptides <input type="radio"/> Custom selection - Select your own settings
	Handling of duplicate peptides: - Duplicate peptides will be removed.
	Peptide lengths to be considered in prediction: - Only peptide length 9 will be included 9mers = 58
Conservancy analysis (Uses only peptides conserved in specified % of sequences)	<input checked="" type="radio"/> No <input type="radio"/> Yes

Start Over Back Next

Current selections:	
No. of sequences	3
Host species	Human
Allele class	Class I
Selected alleles	1.A*02:01 2.A*02:06 3.A*03:01

Step 4: Peptides - Class I

tools.iedb.org/tepitool/

IEDB Analysis Resource

[Home](#) [Help](#) [Reference](#) [Download](#) [Contact](#)

TepiTool

Steps 1 2 3 **4** 5 6

PEPTIDES - Select peptides to be included in prediction:

Peptides to be included in prediction

Apply default settings for low number of peptides
 Apply default settings for moderate number of peptides
 Apply default settings for high number of peptides
 Custom selection - Select your own settings

Handling of duplicate peptides:
- Duplicate peptides will be removed.

Peptide lengths to be considered in prediction:
- Only peptide lengths 8-11 will be included
8mers = 60
9mers = 58
10mers = 56
11mers = 54

Conservancy analysis
(Uses only peptides conserved in specified % of sequences)
 No
 Yes

Current selections:	
No. of sequences	3
Host species	Human
Allele class	Class I
Selected alleles	1.A*02:01 2.A*02:06 3.A*03:01

[Start Over](#) [Back](#) [Next](#)

Step 4: Peptides - Class I

tools.iedb.org/tepitool/

IEDB Analysis Resource

Home Help Reference Download Contact

TepiTool

Steps 1 2 3 **4** 5 6

PEPTIDES - Select peptides to be included in prediction:

Peptides to be included in prediction	<input type="radio"/> Apply default settings for low number of peptides
	<input type="radio"/> Apply default settings for moderate number of peptides
	<input checked="" type="radio"/> Apply default settings for high number of peptides
	<input type="radio"/> Custom selection - Select your own settings

Handling of duplicate peptides:
- Duplicate peptides will not be removed.

Peptide lengths to be considered in prediction:
- All peptide lengths (8-14) will be included
8mers = 74
9mers = 71
10mers = 68
11mers = 65
12mers = 62
13mers = 59
14mers = 56

Conservancy analysis
(Uses only peptides conserved in specified % of sequences)

No
 Yes

Start Over Back Next

Current selections:

No. of sequences	3
Host species	Human
Allele class	Class I
Selected alleles	1.A*02:01 2.A*02:06 3.A*03:01

Step 4: Peptides - Class I

tools.iedb.org/tepitool/

IEDB Analysis Resource

Home Help Reference Download Contact

TepiTool

Steps 1 2 3 **4** 5 6

PEPTIDES - Select peptides to be included in prediction:

Peptides to be included in prediction

- Apply default settings for low number of peptides
- Apply default settings for moderate number of peptides
- Apply default settings for high number of peptides
- Custom selection - Select your own settings

Handling of duplicate peptides:

- Remove duplicate peptides
- Keep duplicate peptides

Peptide lengths to be considered in prediction:

- 8mers = 60
- 9mers = 58
- 10mers = 56
- 11mers = 54
- 12mers = 52
- 13mers = 50
- 14mers = 48

Conservancy analysis
(Uses only peptides conserved in specified % of sequences)

- No
- Yes

Use peptides conserved in

Start Over Back Next

Current selections:	
No. of sequences	3
Host species	Human
Allele class	Class I
Selected alleles	1.A*02:01 2.A*02:06 3.A*03:01

1 sequence

10% sequences

20% sequences

30% sequences

40% sequences

50% sequences

60% sequences

70% sequences

80% sequences

90% sequences

100% sequences

Step 5: Method - Class I

tools.iedb.org/tepitool/

IEDB Analysis Resource

[Home](#) [Help](#) [Reference](#) [Download](#) [Contact](#)

TepiTool

Steps 1 2 3 4 5 6

METHOD - Select prediction & peptide selection methods and cutoff values:

Prediction method to use	IEDB recommended
Selection of predicted peptides	Select peptides based on predicted percentile rank Select peptides with predicted consensus percentile rank \leq 1

[Start Over](#) [Back](#) [Next](#)

- IEDB recommended
- Consensus
- NetMHCpan
- ANN
- SMMPMBEC
- SMM
- CombLib_Sidney_2008

Current selections:

No. of sequences	3
Host species	Human
Allele class	Class I
Selected alleles	1.A*02:01 2.A*02:06 3.A*03:01
Duplicate peptides	Removed
Peptide lengths selected	9mers 10mers
No. of peptides included (Not considering conservancy analysis)	114
Conservancy analysis	Peptides conserved in at least 50% sequences

- Select peptides based on predicted percentile rank
- Select peptides based on predicted IC50
- Select peptides based on MHC specific predicted binding threshold*
- Select top x% of predicted peptides**
- Select top x number of predicted peptides**

Step 5: Method - Class I

tools.iedb.org/tepitool/

Steps 1 2 3 4 **5** 6

METHOD - Select prediction & peptide selection methods and cutoff values:

Prediction method to use: IEDB recommended

Selection of predicted peptides: **Select peptides based on predicted IC50**

Select peptides with predicted IC50 \leq 500 nM

Start Over Back Next

Steps 1 2 3 4 **5** 6

METHOD - Select prediction & peptide selection methods and cutoff values:

Prediction method to use: IEDB recommended

Selection of predicted peptides: **Select top x% of predicted peptides****

Select top 2% of 114 peptides = 2 peptide(s) per allele x 3 allele(s) = 6 peptides

(**Fi

Start Over Back Next

Steps 1 2 3 4 **5** 6

METHOD - Select prediction & peptide selection methods and cutoff values:

Prediction method to use: IEDB recommended

Selection of predicted peptides: **Select top x number of predicted peptides****

Select top 5 peptides per allele (Maximum possible = 114)

(*Peptide selection done based on percentile rank)

Start Over Back Next

Step 5: Method - Class I

tools.iedb.org/tepitool/

Steps 1 2 3 4 **5** 6

METHOD - Select prediction & peptide selection methods and cutoff values:

Prediction method to use: IEDB recommended ▼

Select peptides based on MHC specific predicted binding threshold ▼

(*Each MHC allele has its own IC50 threshold.
Predicted peptides will correspond to 75% of immune response.
Prediction method is **SMM**)

As of now, only the following alleles are covered by this method:

- A*01:01
- A*02:01
- A*02:03
- A*02:06
- A*03:01
- A*11:01
- A*23:01
- A*24:02
- A*25:01
- A*26:01
- A*29:02
- A*30:01
- A*30:02
- A*31:01
- A*32:01
- A*33:01
- A*68:01
- A*68:02
- B*07:02
- B*08:01
- B*14:02
- B*15:01
- B*18:01
- B*27:05
- B*35:01
- B*35:03
- B*38:01
- B*39:01
- B*40:01
- B*40:02
- B*44:02
- B*44:03
- B*46:01
- B*48:01
- B*51:01
- B*53:01
- B*57:01
- B*58:01

Selection of predicted peptides

Please refer this paper for more details: [Paul et al., \(2013\)](#) J of Immunol. 191(12): 5831-5839.

Start Over Back Next

Step 5: Method - Class I

tools.iedb.org/tepitool/

IEDB Analysis Resource

[Home](#) [Help](#) [Reference](#) [Download](#) [Contact](#)

TepiTool

Steps [1](#) [2](#) [3](#) [4](#) [5](#) **[6](#)**

REVIEW: Review selections, enter job details & submit data:

Summary:

No. of sequences	3
Host species	Human
Allele class	Class I
Alleles	1.A*02:01 2.A*02:06 3.A*03:01
Duplicate peptides	Removed
Peptide lengths selected	9mers 10mers
Approx no. of peptides included	114
Peptide overlap	N/A (all possible nmers are included in class I)
Conservancy analysis	Peptides conserved in at least 50% sequences
Prediction method	IEDB recommended
Peptide selection criterion	Based on predicted consensus percentile rank (Cutoff selected = 1)

Job details:

Job name (optional)	<input type="text"/>
Email (optional - will notify when job is finished)	<input type="text" value="spaul@lji.org"/>

[Start Over](#) [Back](#) [Submit](#)

(Please note that you will not be able to make any more changes once submitted. You will have to start again if you want to do so.)

Results: Class I

tools.iedb.org/tepitool/



[Home](#) [Help](#) [Reference](#) [Download](#) [Contact](#)

TepiTool

Prediction results - concise ([Download table](#) ):

Seq #	Peptide start	Peptide end	Peptide	Percentile rank	Allele	Conservancy
1	5	14	IVLGLVLLSV	0.3	HLA-A*02:06	67%
1	10	19	VLLSVTVQGK	0.36	HLA-A*03:01	67%
1	5	14	IVLGLVLLSV	0.77	HLA-A*02:01	67%
1	6	14	VLGLVLLSV	0.84	HLA-A*02:01	67%
1	11	19	LLSVTVQGK	0.89	HLA-A*03:01	67%

Download results details:

Complete results 	Prediction results of all peptides
Conservancy of peptides 	Conservancy of peptides in the sequences

Citation information:

If you use these predictions in a manuscript, please include the following in the method section:

The MHC I binding predictions were done with IEDB analysis resource (TepiTool [1]) using Consensus method [2] which employs SMM, ANN and Combinatorial library methods.

1. Paul, S., Sidney, J., Sette, A., and Peters, B. 2016. TepiTool: A pipeline for computational prediction of T cell epitope candidates. *Curr. Protoc. Immunol.* 114:18.19.1-18.19.24.
2. 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.
3. Wang P, Sidney J, Dow C, Mothé 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.

For complete list of references please click here: [References](#)

Differences in TepiTool workflow if Class II?

Step 3: Alleles - Class II

IEDB Analysis Resource

Home Help Reference Download Contact

TepiTool

Steps 1 2 **3** 4 5 6

ALLELES - Specify alleles:

Human - Class II

Predict for custom allele set
 Predict for pre-selected panel of alleles
 Predict using pre-selected allele sets & methods

Options:

Select from list of alleles
 Upload allele file

Alleles

Select α and β chains separately when applicable

DQ

- DQA1*01:01/DQB1*05:01
- DQA1*01:02/DQB1*06:02
- DQA1*03:01/DQB1*03:02
- DQA1*04:01/DQB1*04:02
- DQA1*05:01/DQB1*02:01
- DQA1*05:01/DQB1*03:01

Current selections:

No. of sequences	3
Host species	Human
Allele class	Class II
Selected alleles	
Reset alleles	

Start Over Back Next

Step 4: Peptides - Class II

tools.iedb.org/tepitool/

IEDB Analysis Resource

Home | Help | Reference | Download | Contact

TepiTool

Steps 1 2 3 **4** 5 6

PEPTIDES - Select peptides to be included in prediction:

Peptides to be included in prediction

- Apply default settings for low number of peptides
- Apply default settings for moderate number of peptides
- Apply default settings for high number of peptides
- Custom selection - Select your own settings

Handling of duplicate peptides
- Duplicate peptides will be removed.

Desired no. of overlapping residues for 15mers
- No. of overlapping residues fixed at 10.

Approximate no. of peptides to be considered for prediction = 12

Conservancy analysis
(Uses only peptides conserved in specified % of sequences)

No
 Yes

Use peptides conserved in

Start Over | Back | Next

*settings
summary* →

	Low	Moderate	High	Custom
Duplicates	removed	removed	not removed	user selects
Overlapping residues	8	10	10	user selects

Current selections:

No. of sequences	3
Host species	Human
Allele class	Class II
Selected alleles	1.DPA1*01:03/DPB1*02:01 2.DQA1*03:01/DQB1*03:02 3.DRB1*01:01 4.DRB1*01:02 5.DRB1*01:03

Step 5: Method - Class II

tools.iedb.org/tepitool/

IEDB Analysis Resource

Home Help Reference Download Contact

TepiTool

Steps 1 2 3 4 5 6

METHOD - Select prediction & peptide selection methods and cutoff values:

Prediction method to use: IEDB recommended

Selection of predicted peptides: Select peptides based on predicted percentile rank

Select peptides with predicted consensus percentile rank \leq 10

Start Over Back Next

IEDB recommended
Consensus
NetMHCIIpan
NN_align
SMM_align
Combinatorial library
Sturniolo

Current selections:

No. of sequences	3
Host species	Human
Allele class	Class II
Alleles selected	1.DPA1*01:03/DPB1*02:01 2.DQA1*03:01/DQB1*03:02 3.DRB1*01:01 4.DRB1*01:02 5.DRB1*01:03
Duplicate peptides	Removed
Peptide overlap	10 AA residues
Approx. no. of peptides included (Not considering conservancy analysis)	12
Conservancy analysis	Peptides conserved in at least 50% sequences

Select peptides based on predicted percentile rank
Select peptides based on predicted IC50
Select peptides based on predicted no. of alleles binding
Select top x% of peptides*
Select top x number of predicted peptides*

exclusive to class II

Select peptides based on predicted no. of alleles binding

Select peptides that bind to at least 50% alleles

(binding determined by IEDB consensus percentile rank \leq 20.0)

Results – Class II

IEDB Analysis Resource




[Home](#) [Help](#) [Reference](#) [Download](#) [Contact](#)

TepiTool

Prediction results - concise ([Download table](#) ):

Seq #	Peptide start	Peptide end	Peptide sequence	Consensus percentile rank	Allele	Conservancy
1	2	16	KALIVLGLVLLSVIV	2.30	HLA-DRB1*01:01	67.0%
1	7	21	LGLVLLSVTVQGKVF	8.70	HLA-DRB1*01:01	67.0%
1	3	17	ALIVLGLVLLSVIVQ	7.60	HLA-DRB1*01:02	67.0%

Download results details:

Non-redundant results 	Prediction results with redundant peptides within each sequence removed - Includes positives and negatives
Complete results 	Prediction results of all peptides
Conservancy of peptides 	Conservancy of peptides in the sequences

Citation information:

If you use these predictions in a manuscript, please include the following in the method section:

The MHC II binding predictions were done with IEDB analysis resource (TepiTool [1]) using Consensus method [2,3] which employs SMM_align, NN_align, Combinatorial library, Sturniolo methods and NetMHCIIpan [4,5].

1. Paul, S., Sidney, J., Sette, A., and Peters, B. 2016. TepiTool: A pipeline for computational prediction of T cell epitope candidates. *Curr. Protoc. Immunol.* 114:18.19.1-18.19.24.
2. 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.
3. Wang P, Sidney J, Dow C, Mothé 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.
4. Karosiene E1, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. 2013. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics.* 65(10): 711.
5. Nielsen M, Lundegaard C, Blicher T, Peters B, Sette A, Justesen S, Buus S, and Lund O. 2008. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput Biol.* 4(7): e1000107.

For complete list of references please click here: [References](#)

Differences if 7 allele method or promiscuous?

Step 3-5: Class II -7 allele method

tools.iedb.org/tepitool/

Steps 1 2 3 4 5 6

ALLELES - Specify alleles:

Human - Class II

Predict for custom allele set

Predict for pre-selected panel of alleles

Predict using pre-selected allele sets & methods

Alleles

Options:

Use the "7-allele method"

Use panel of 26 most frequent alleles for promiscuous binding

- Selection criterion is median of percentile ranks from the 7 alleles involved.

Current selections:	
No. of sequences	3
Host species	Human
Allele class	Class II
Selected alleles	1. DRB1*03:01 2. DRB1*07:01 3. DRB1*15:01 4. DRB3*01:01 5. DRB3*02:02 6. DRB4*01:01 7. DRB5*01:01

[Reset alleles](#)

Steps 1 2 3 4 5 6

PEPTIDES - Select peptides to be included in prediction:

Handling of duplicate peptides	Duplicate peptides will be removed
No. of overlapping residues for 15mer peptides to be generated (Peptide length is fixed at 15 for class II)	10
Approximate no. of peptides to be considered for prediction	12

[Start Over](#) [Back](#) [Next](#)

Steps 1 2 3 4 5 6

METHOD - Select prediction & peptide selection methods and cutoff values:

Prediction method to use	IEDB recommended
Selection of predicted peptides	Promiscuity based on "7-allele method" - Peptides considered as binders if median consensus percentile \leq <input type="text" value="20"/>

[Start Over](#) [Back](#) [Next](#)

Step 3-5: Class II – promiscuous

tools.iedb.org/tepitool/

Steps 1 2 **3** 4 5 6

ALLELES - Specify alleles:

Human - Class II

Predict for custom allele set

Predict for pre-selected panel of alleles

Predict using pre-selected allele sets & methods

Options:

Use the "7-allele method"

Use panel of 26 most frequent alleles for promiscuous binding

Current selections:

No. of sequences	3
Host species	Human
Allele class	Class II

Selected alleles
[Reset alleles](#)

1. HLA-DPA1*01/DPB1*04:01
2. HLA-DPA1*01:03/DPB1*02:01
3. HLA-DPA1*02:01/DPB1*01:01
4. HLA-DPA1*02:01/DPB1*05:01
5. HLA-DPA1*03:01/DPB1*04:02
6. HLA-DQA1*01:01/DQB1*05:01
7. HLA-DQA1*01:02/DQB1*06:02
8. HLA-DQA1*03:01/DQB1*03:02
9. HLA-DQA1*04:01/DQB1*04:02
10. HLA-DQA1*05:01/DQB1*02:01
11. HLA-DQA1*05:01/DQB1*03:01
12. HLA-DRB1*01:01
13. HLA-DRB1*03:01
14. HLA-DRB1*04:01
15. HLA-DRB1*04:05
16. HLA-DRB1*07:01
17. HLA-DRB1*08:02
18. HLA-DRB1*09:01
19. HLA-DRB1*11:01
20. HLA-DRB1*12:01
21. HLA-DRB1*13:02
22. HLA-DRB1*15:01
23. HLA-DRB3*01:01
24. HLA-DRB3*02:02
25. HLA-DRB4*01:01
26. HLA-DRB5*01:01

Steps 1 2 3 **4** 5 6

PEPTIDES - Select peptides to be included in prediction:

Handling of duplicate peptides	Duplicate peptides will be removed
No. of overlapping residues for 15mer peptides to be generated (Peptide length is fixed at 15 for class II)	10
Approximate no. of peptides to be considered for prediction	

[Start Over](#) [Back](#) [Next](#)

Steps 1 2 3 4 **5** 6

METHOD - Select prediction & peptide selection methods and cutoff values:

Prediction method to use	IEDB recommended
Selection of predicted peptides	Promiscuity based on no. of alleles binding (Peptide considered as binder if it binds to at least 50% of the 26 most frequent alleles)

[Start Over](#) [Back](#) [Next](#)

Datasets

the paper, different cross-validation strategies (i.e. `cv_rnd`, `cv_sr`, and `cv_gs`) were tested. Please see the Methods section for details of the cross-validation strategies.

- **Data format:** Text file format.
- **Dataset availability:** [benchmark_reliability.tar.gz](#)
- Dataset used for retraining the IEDB class I binding prediction tools.
 - **Description of the dataset:** The dataset is largely identical to that of Kim et al (2014), described above, but includes additional data that was not publicly available at the time.
 - **Date of the dataset generation:** 2013
 - **Details on the dataset generation:** The dataset was compiled from three sources: the IEDB, the Sette lab, and the Buus lab. If a peptide/allele combination had more than 1 measurement among the three sources, its geometric mean was taken.
 - **Data format:** Compressed text file containing binding data.
 - **Dataset availability:** [binding_data_2013.zip](#)
- Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior.
Kim Y, Sidney J, Pinilla C, Sette A, Peters B.
BMC Bioinformatics, 2009.
 - **Description of the dataset:** Cross-validated predictive performances for SMMPMBEC using the same binding data set as in [Peters et al. PLOS Comput Biol 2006].
 - **Date of the dataset generation:** 2009
 - **Details on the dataset generation:** Using the same cross-validation data partitions as was done for ANN and ARB in 2006, cross-validated predictions using SMMPMBEC were made.
 - **Data format:** A table in Excel file format.
 - **Dataset availability:** <http://www.biomedcentral.com/1471-2105/10/394/additional>
- A Community Resource Benchmarking Predictions of Peptide Binding to MHC-I Molecules.
Peters B, Bui HH, Frankild S, Nielsen M, Lundegaard C, Kostem E, Basch D, Lamberth K, Harndahl M, Fleri W, Wilson SS, Sidney J, Lund O, Buus S, Sette A.
PLOS Computational Biology, 2006.
 - **Description of the dataset:** Experimentally measured peptide binding affinities for MHC class I molecules from two sources: the Alessandro Sette lab at the La Jolla Institute and the Soren Buus lab at the University of Copenhagen. The dataset contains 48,828 affinities and covers a total of 48 mouse, human, macaque and chimpanzee MHC class I alleles.
 - **Date of the dataset generation:** 2006
 - **Details on the dataset generation:** Used two different assays to generate the binding data.
 - **Data format:** Compressed text files containing experimental binding data as well as cross-validated predicted affinities.
 - **Dataset availability:** [ANN](#), [ARB](#), [SMM](#)

Benchmarking of class I epitope prediction methods

- List of 44 methods

- Freely available

- Selected 15 methods

- Trained models for H-2D^b & H-2K^b

- 2 variants

- NetMHCpan-4.0-B & L
- MHCflurry-B & L
- Total 17 methods

#	Method	Year first published*	URLs to method	Reference	URL to manuscript**	Included/Excluded (reason)
1	BIMAS	1994	https://www.bimas.cit.nih.gov/	Parker et al., 1994	http://www.jimmunol.org/	Included
2	PREDEP	1995	http://margalit.huji.ac.il/T/	Altuvia et al., 1995	https://www.sciencedirect.com/	Included
3	SYFPEITHI	1997	http://www.syfpeithi.de/	Rammensee et al., 1997	https://www.springer.com/	Included
4	Rankpep	2002	http://imed.med.ucm.es/	Reche et al., 2002	https://www.sciencedirect.com/	Included
5	ProPred1	2003	http://crdd.osdd.net/raghava/	Singh and Raghava, 2003	https://academic.oup.com/	Included
6	SMM	2005	http://tools.iedb.org/mhci/	Peters and Sette, 2005	https://bmcbioinformatics.org/	Included
7	ARB	2005	http://tools.iedb.org/mhci/	Bui et al., 2005	https://link.springer.com/	Included
8	IEDB Consensus	2006	http://tools.iedb.org/mhci/	Moutaftsi et al., 2006	https://www.nature.com/	Included
9	SMPMBEC	2009	http://tools.iedb.org/mhci/	Kim et al., 2009	https://bmcbioinformatics.org/	Included
10	PACComplex	2011	http://pacocomplex.life.nctu.edu.tw/	Liu et al., 2011	https://academic.oup.com/	Included
11	NetMHCpan-3.0	2016	http://www.cbs.dtu.dk/services/	Nielsen and Andreatta, 2016	https://genomedirect.com/	Included
12	NetMHC-4.0	2016	http://www.cbs.dtu.dk/services/	Andreatta and Nielsen, 2016	https://academic.oup.com/	Included
13	NetMHCpan-4.0***	2017	http://www.cbs.dtu.dk/services/	Jurtz et al., 2014	http://www.jimmunol.org/	Included
14	MHCflurry***	2018	https://openvax.github.io/	O'Donnell et al., 2018	https://www.sciencedirect.com/	Included
15	MHCLovac	Not published yet	https://github.com/stef3333/	Stojanovic, S.	-	Included
16	CTLPred	2004	http://crdd.osdd.net/raghava/	Bhasin and Raghava, 2004	https://www.sciencedirect.com/	Excluded
17	Epiln	2006	http://www.ddg-pharmfac.net/	Doytchinova et al., 2006	https://bmcbioinformatics.org/	Excluded (Trained model)
18	nHLAPred (ANNPred, ComPred)	2006	http://crdd.osdd.net/raghava/	Bhasin and Raghava, 2006	https://www.ias.ac.in/	Excluded
19	SVMHC	2006	http://abi.inf.uni-tuebingen.de/	Donnes and Kohlbach, 2006	https://academic.oup.com/	Excluded (Method not available)
20	SVRMHC	2006	http://c1.accurascience.com/	Wan et al., 2006	https://bmcbioinformatics.org/	Excluded (Not working)
21	KISS	2007	http://cbio.enscm.fr/kiss/	Jacob and Vert, 2008	https://academic.oup.com/	Excluded (Not working)
22	PickPocket	2009	http://www.cbs.dtu.dk/services/	Zhang et al., 2009	https://academic.oup.com/	Excluded (Author suggestion)
23	Multipred2	2011	http://cvc.dfci.harvard.edu/	Zhang et al., 2011	https://www.sciencedirect.com/	Excluded (Trained model)
24	NetMHCcons	2011	http://www.cbs.dtu.dk/services/	Karosiene et al., 2011	https://link.springer.com/	Excluded (Author suggestion)
25	KernelIRLspan	2014	https://github.com/guoxin/	Shen et al., 2014	https://www.sciencedirect.com/	Excluded (Trained model)
26	NIELuter	2015	http://immunet.cn/nie/cgi/	Tang et al., 2015	https://www.sciencedirect.com/	Excluded (Trained model)
27	HONN	2015	Method not available	Kuksa et al., 2015	https://academic.oup.com/	Excluded (Trained model)
28	ESMACS	2015	Method not available	Wan et al., 2015	https://pubs.acs.org/	Excluded (Too resource intensive)
29	sNebula	2016	Method not available	Luo et al., 2016	https://www.nature.com/	Excluded (Trained model)
30	HLaffy	2016	http://proline.biochem.iisc.ernet.in/	Mukherjee et al., 2016	https://academic.oup.com/	Excluded (Trained model)
31	ConvMHC	2017	http://jumong.kaist.ac.kr/	Han and Kim, 2017	https://bmcbioinformatics.org/	Excluded (Trained model)
32	PSSMHCpan	2017	https://github.com/BGI2017/	Liu et al., 2017	https://academic.oup.com/	Excluded (Trained model)
33	MixMHCpred	2017	https://github.com/Gfeller/	Bassani-Sternberg et al., 2017	https://journals.plos.org/	Excluded (Trained model)
34	HLA-CNN (HLA-bind)	2017	https://github.com/uci-cb/	Vang and Xie, 2017	https://academic.oup.com/	Excluded (Trained model)
35	EDGE	2018	Model provided as part of	Bulik-Sullivan et al.	https://www.nature.com/	Excluded (Not working)
36	MAM	2018	http://mhc.deepomics.org/	Xiao et al., 2018	https://bmcbioinformatics.org/	Excluded (Trained model)
37	DeepSeqPan	2019	https://github.com/pcpliu/	Liu et al., 2019	https://www.nature.com/	Excluded (Trained model)
38	ForestMHC	2019	https://github.com/kmboe/	Boehm et al., 2019	https://link.springer.com/	Excluded (Trained model)
39	DeepMHC	Not published yet	http://mieg.cse.sc.edu/deepmhc/	Hu and Liu	https://www.biorxiv.org/	Excluded (Trained model)
40	AI-MHC	Not published yet	https://baras.pathology.jhu.edu/	Sidhom et al.	https://www.biorxiv.org/	Excluded (Trained model)
41	MHCSeqNet	Not published yet	https://github.com/cmbcu/	Phloyphisut et al.	https://www.biorxiv.org/	Excluded (Trained model)
42	ACME	Not published yet	https://github.com/HYsxe/	Hu et al.	https://www.biorxiv.org/	Excluded (Trained model)
43	Deep-Learning-MHCI	Not published yet	https://github.com/altayg/	Altay, G.	https://www.biorxiv.org/	Excluded (Trained model)
44	MHCnuggets	Not published yet	https://karchinlab.org/app/	Bhattacharya et al.	https://www.biorxiv.org/	Excluded (Author suggestion)

RESEARCH ARTICLE

Benchmarking predictions of MHC class I restricted T cell epitopes in a comprehensively studied model system

Sinu Paul¹, Nathan P. Croft^{2,3}, Anthony W. Purcell^{2,3}, David C. Tscharke⁴, Alessandro Sette^{1,5}, Morten Nielsen^{6,7}, Bjoern Peters^{1,5*}

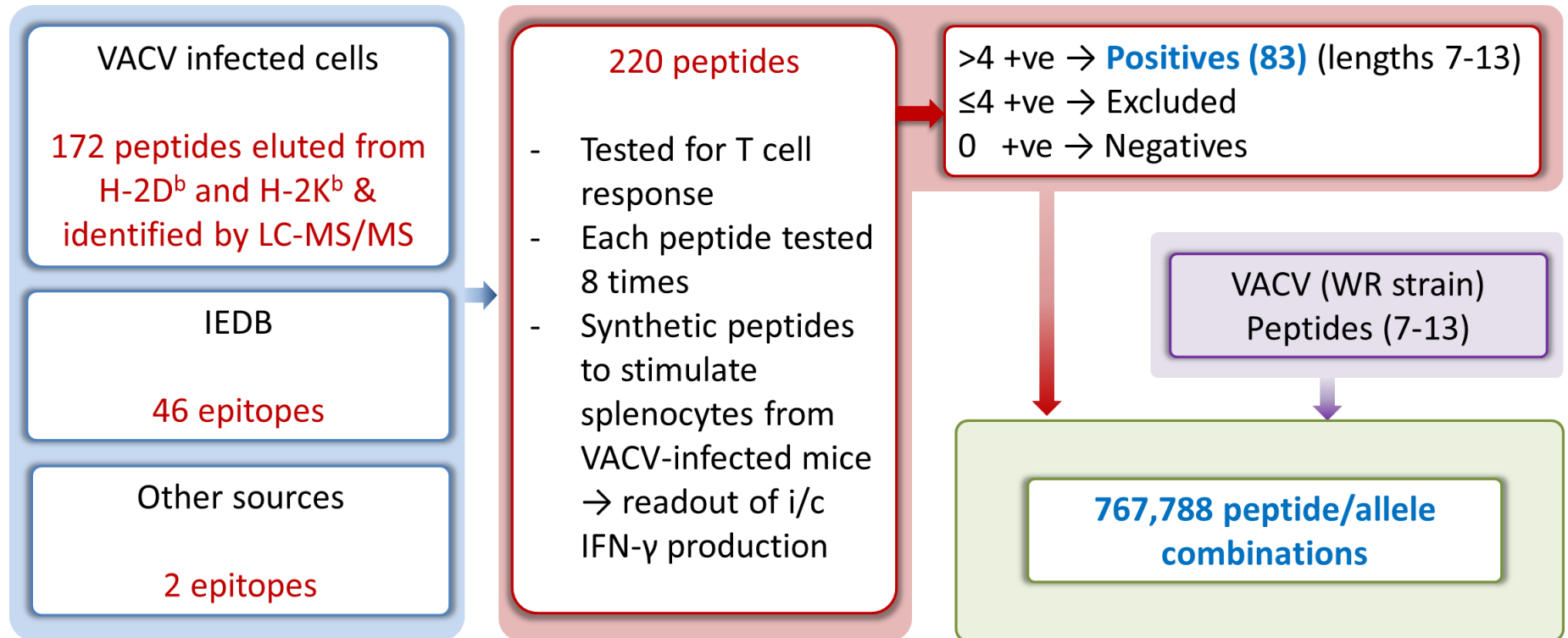
Dataset

- Comprehensive epitope dataset from Vaccinia virus

Most viral peptides displayed by class I MHC on infected cells are immunogenic

Nathan P. Croft^{a,b,1}, Stewart A. Smith^c, Jana Pickering^c, John Sidney^d, Bjoern Peters^{d,e}, Pouya Faridi^{a,b}, Matthew J. Witney^c, Prince Sebastian^c, Inge E. A. Flesch^c, Sally L. Heading^c, Alessandro Sette^{d,e}, Nicole L. La Gruta^{a,b,f}, Anthonv W. Purcell^{a,b,1,2}, and David C. Tschärke^{c,1,2}

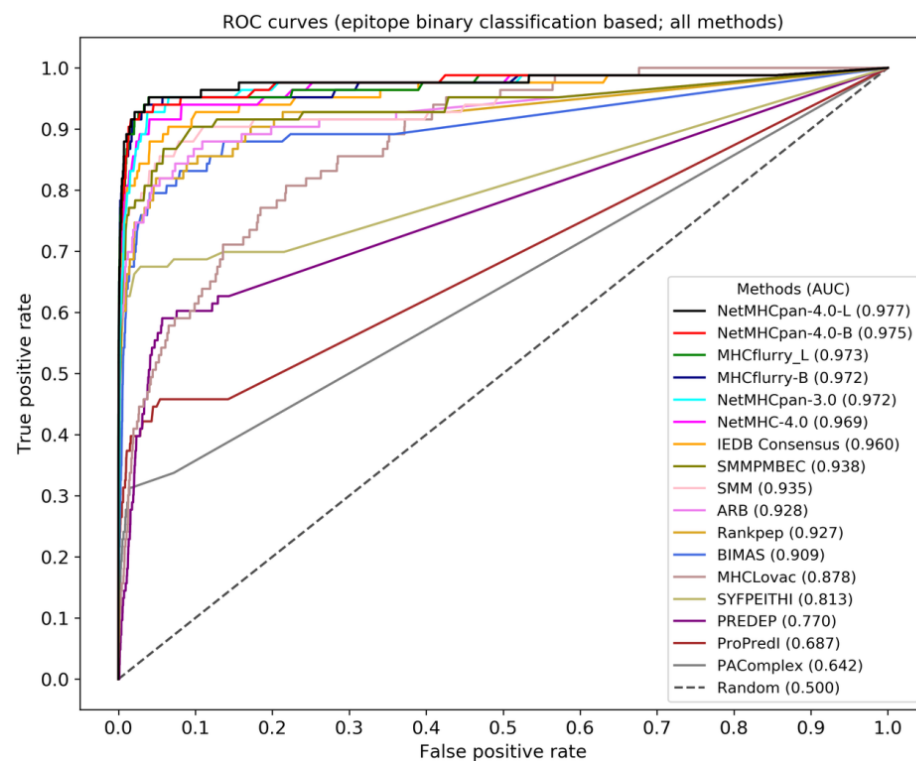
3112–3117 | PNAS | February 19, 2019 | vol. 116 | no. 8



Results

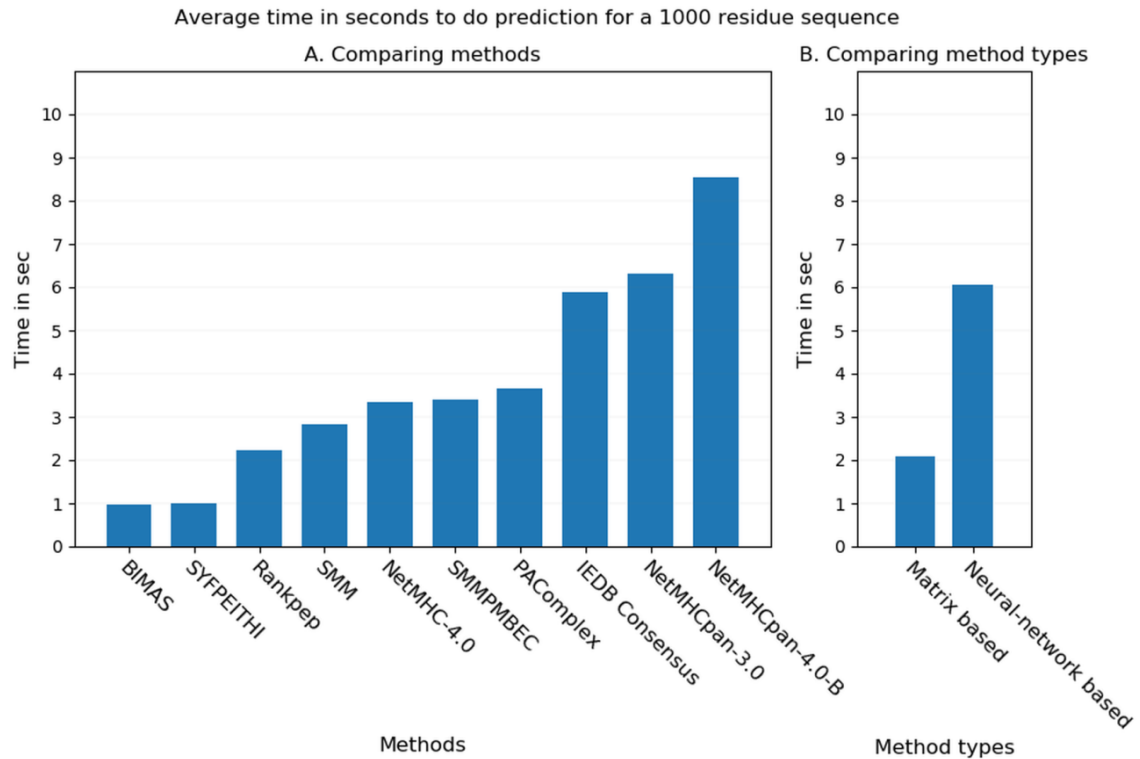
- Binary classification (epitope/non-epitope) & T cell response based
- In terms of AUC (Area under curve of ROC curves)

Method	Binary classification (epitope/non-epitope) based		T cell response based	
	AUC	Rank	AUC	Rank
NetMHCpan-4.0-L	0.977	1	0.979	1
NetMHCpan-4.0-B	0.975	2	0.978	2
MHCflurry-L	0.973	3	0.977	3
MHCflurry-B	0.972	4	0.976	4
NetMHCpan-3.0	0.972	5	0.975	5
NetMHC-4.0	0.969	6	0.974	6
IEDB Consensus	0.960	7	0.961	7
SMMPMBEC	0.938	8	0.940	8
SMM	0.935	9	0.938	10
ARB	0.928	10	0.939	9
Rankpep	0.927	11	0.894	12
BIMAS	0.909	12	0.918	11
MHCLovac	0.878	13	0.863	13
SYFPEITHI	0.813	14	0.778	14
PREDEP	0.770	15	0.737	15
ProPred-I	0.687	16	0.651	17
PACComplex	0.642	17	0.652	16



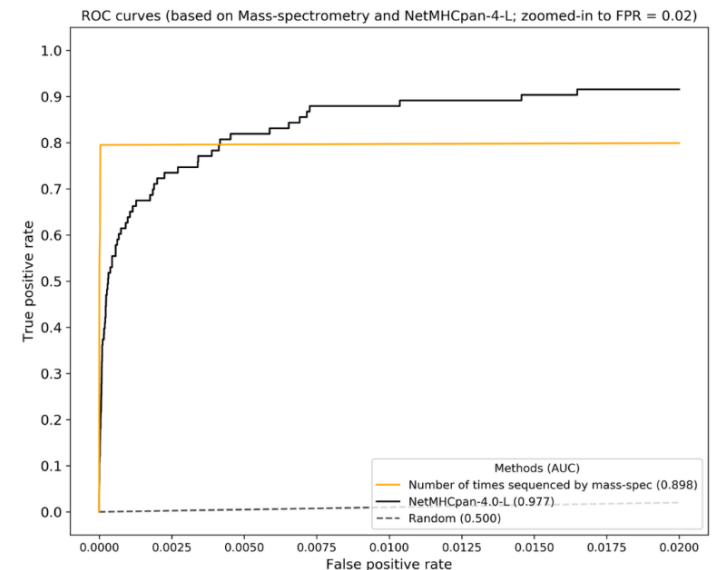
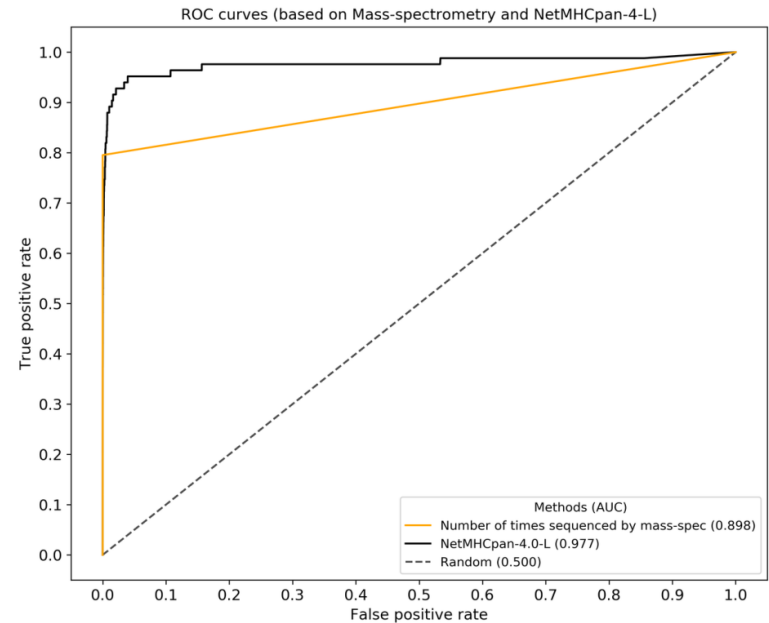
Prediction speed comparison

- 5 random sequences of 1000 residues each
- Matrix-based methods faster compared to Neural network-based methods



Results – peptide selection by MS vs. prediction

- Number of times peptides were identified by MS used as “MS score”
- Compared with NetMHCpan-4.0-L
- MS needs much less peptides to capture 50% epitopes (0.01%, N=48 vs. 0.04%, N = 277 for NetMHCpan-4.0-L)
- NetMHCpan-4.0-L better when considering all epitopes



Summary

- MHC class I binding prediction
- MHC class II binding prediction
- TepiTool
- Datasets
- Benchmarking of class I epitope prediction tools