



Analysis Resource Overview

tools.iedb.org

Presented by: Bjoern Peters, Professor

IEDB Analysis Resource

tools.iedb.org

T cell tools

MHC binding prediction

- MHC I binding prediction
- MHC II binding prediction
- TepiTool

MHC processing prediction

- NetChop/NetCTL/NetCTLpan
- MHC-NP, MHCII-NP

Immunogenicity prediction

- CD4 & CD8 T cell immunogenicity prediction tools

Remove epitopes from therapeutic proteins

- Deimmunization tool

B cell tools

Linear epitope prediction

- BepiPred
- Other methods

Discontinuous epitope prediction

- DiscoTope
- ElliPro

Antibody and TCR structure prediction

- LYRA
- SCEptRe

Analysis tools

Population coverage of epitope set

- Population coverage tool

Degree of conservation

- Conservancy analysis tool

Group peptides based on sequence identity

- Cluster 2.0

Infer restriction in HLA typed subjects

- RATE

Aggregate heterogeneous immune response

- Immunome-browser

Development of prediction tools

- Epitope prediction tools
 - Vaccine, cancer neoantigen studies
- Based on machine learning techniques
 - Experimentally derived data
 - E.g. MHC binding data, Eluted ligand data, T cell assay data

Training data

species	mhc	peptide_length	sequence	inequality	meas
human	HLA-A*01:01	8	ASFCGSPY	=	51.4
human	HLA-A*01:01	8	LDFGLSK	=	739.385479
human	HLA-A*01:01	8	FTSFFVRY	=	1285
human	HLA-A*01:01	8	KSVFNLSY	=	1466
human	HLA-A*01:01	8	RDWAHNSL	=	1804.675523
human	HLA-A*01:01	8	FSSCPVAY	=	1939.46663
human	HLA-A*01:01	8	RNWAHSSL	=	2201.794454
human	HLA-A*01:01	8	LSCAASGF	=	2830.055894
human	HLA-A*01:01	8	LASIDLKY	=	3464
human	HLA-A*01:01	8	RAKFKQLL	>	5000
human	HLA-A*01:01	8	LVESGGGL	=	5886.338262
human	HLA-A*01:01	8	NIILKANF	=	8920.343726
human	HLA-A*01:01	8	RGYVFQGL	=	15645.52954
human	HLA-A*01:01	8	HHINQNL	=	18923.65903
human	HLA-A*01:01	8	FVNRPLV	>	20000
human	HLA-A*01:01	8	RYSHWTKL	>	20000
human	HLA-A*01:01	8	STASSWSY	>	20000
human	HLA-A*01:01	8	YDPVVIKV	>	20000

Peptide data for prediction

peptide	allele
ALPHIIDE	HLA-A*01:01
TGIKAVYN	HLA-A*01:01
VLIVITGI	HLA-A*01:01
EVINIVII	HLA-A*01:01
IVTMFEAL	HLA-A*01:01
HIIDEVIN	HLA-A*01:01
TCGIFALI	HLA-A*01:01
KAVYNFAT	HLA-A*01:01
NIVIIVLI	HLA-A*01:01
YNFATCGI	HLA-A*01:01
MGQIVTMF	HLA-A*01:01
IFALISFL	HLA-A*01:01
LLLAGRSC	HLA-A*01:01
SFLLLAGR	HLA-A*01:01



Prediction method/model

Prediction results

peptide	allele	predicted score
ALPHIIDE	HLA-A*01:01	80
TGIKAVYN	HLA-A*01:01	87
VLIVITGI	HLA-A*01:01	64
EVINIVII	HLA-A*01:01	68
IVTMFEAL	HLA-A*01:01	33
HIIDEVIN	HLA-A*01:01	85
TCGIFALI	HLA-A*01:01	32
KAVYNFAT	HLA-A*01:01	78
NIVIIVLI	HLA-A*01:01	71
YNFATCGI	HLA-A*01:01	23
MGQIVTMF	HLA-A*01:01	27
IFALISFL	HLA-A*01:01	34
LLLAGRSC	HLA-A*01:01	81
SFLLLAGR	HLA-A*01:01	70

Accessing the Analysis Resource



iedb.org

IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE

Home Specialized Searches **Analysis Resource** Help More IEDB

Epitope Prediction and Analysis

Welcome to the Immune Epitope Database Analysis Resource. It provides a collection of tools for the prediction and analysis of epitopes. It serves as a companion site to the Immune Epitope Database (IEDB), a manually curated database of experimentally determined immune epitopes.

The tools contained fall into the following categories:

T Cell Epitope Prediction Tools
This set of tools includes MHC class I & II binding predictions, peptide processing predictions and immunogenicity predictions.

B Cell Epitope Prediction Tools
The tools here are intended to predict regions of an antigen that may be recognized as epitopes in the context of a B cell.

Analysis Tools
The epitope analysis tools are intended for the analysis of an epitope sequence or group of sequences.

Welcome

The Immune Epitope Database (IEDB) is a freely available resource funded by NIAID. It catalogs experimental data on antibody and T cell epitopes studied in humans, non-human primates, and other animal species in the context of infectious disease, allergy, autoimmunity and transplantation. The IEDB also hosts tools to assist in the prediction and analysis of epitopes.

START YOUR SEARCH HERE

Epitope

- Any Epitopes
- Linear Epitope
- Discontinuous Epitopes
- Non-peptidic Epitopes

Exact Mz: Ex: SIINFEKL

Assay

- Positive Assays Only
- T Cell Assays
- B Cell Assays
- MHC Ligand Assays

Ex: neutralization

Antigen

Organism: Ex: influenza, peanut

Antigen Name: Ex: core, capsid, myosin

MHC Restriction

- Any MHC Restriction
- MHC Class I
- MHC Class II
- MHC Nonclassical

Ex: HLA-A*02:01

Host

- Any Host
- Humans
- Mice
- Non-human Primates

Ex: dog, camel

Disease

- Any Disease
- Infectious Disease
- Allergic Disease
- Autoimmune Disease

Ex: asthma, diabetes

Epitope Analysis Resource

T Cell Epitope Prediction

Scan an antigen sequence for amino acid patterns indicative of:

- MHC I Binding
- MHC II Binding
- MHC I Processing (Proteasome, TAP)
- MHC I Immunogenicity

B Cell Epitope Prediction

Predict linear B cell epitopes using:

- Antigen Sequence Properties

Predict discontinuous B cell epitopes using antigen structure via:

- Discotope
- ElliPro

Epitope Analysis Tools

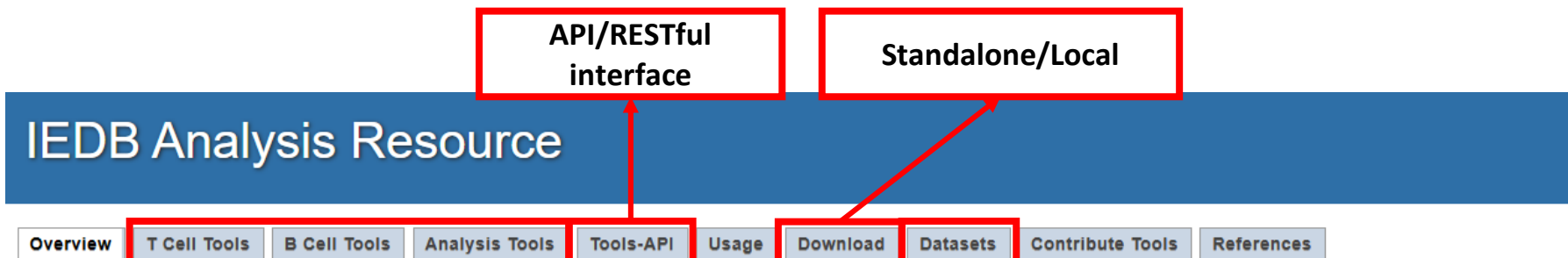
Analyze epitope sets of:

- Population Coverage
- Conservation Across Antigens
- Clusters with Similar Sequences

Summary Metrics

Peptidic Epitopes	592,423
Non-Peptidic Epitopes	2,810
T Cell Assays	364,879
B Cell Assays	484,136
MHC Ligand Assays	1,207,448
Epitope Source Organisms	3,761
Restricting MHC Alleles	785
References	20,589

Available Resources



Epitope Prediction and Analysis Tools

Welcome to the Immune Epitope Database Analysis Resource. This site provides a collection of tools for the prediction and analysis of immune epitopes. It serves as a companion site to the [Immune Epitope Database \(IEDB\)](#), a manually curated database of experimentally characterized immune epitopes.

The tools contained fall into the following categories:

T Cell Epitope Prediction Tools

This set of tools includes MHC class I & II binding predictions, as well as peptide processing predictions and immunogenicity predictions.

B Cell Epitope Prediction Tools

The tools here are intended to predict regions of proteins that are likely to be recognized as epitopes in the context of a B cell response.

Analysis Tools

The epitope analysis tools are intended for the detailed analysis of a known epitope sequence or group of sequences.

IEDB-AR News

• We're Hiring

[Bioinformatician for the Immune Epitope Analysis Resource](#)

• Python 3 availability

As of 2 July 2020, all code has been ported to Python 3, including the standalone tools. Thanks for your patience and be sure to [inform us](#) of any issues that arise.

IEDB-AR Release Notes

[IEDB Analysis Resource v2.24 release notes \(1 Oct 2020\)](#)

2020-1

NEW: I

web, A

BA)...

[IEDB Analysis Resource v2.23 release notes](#)

[IEDB Analysis Resource v2.22 release notes](#)

Solutions Center:
Tutorials, Q&A

help@iedb.org

IEDB

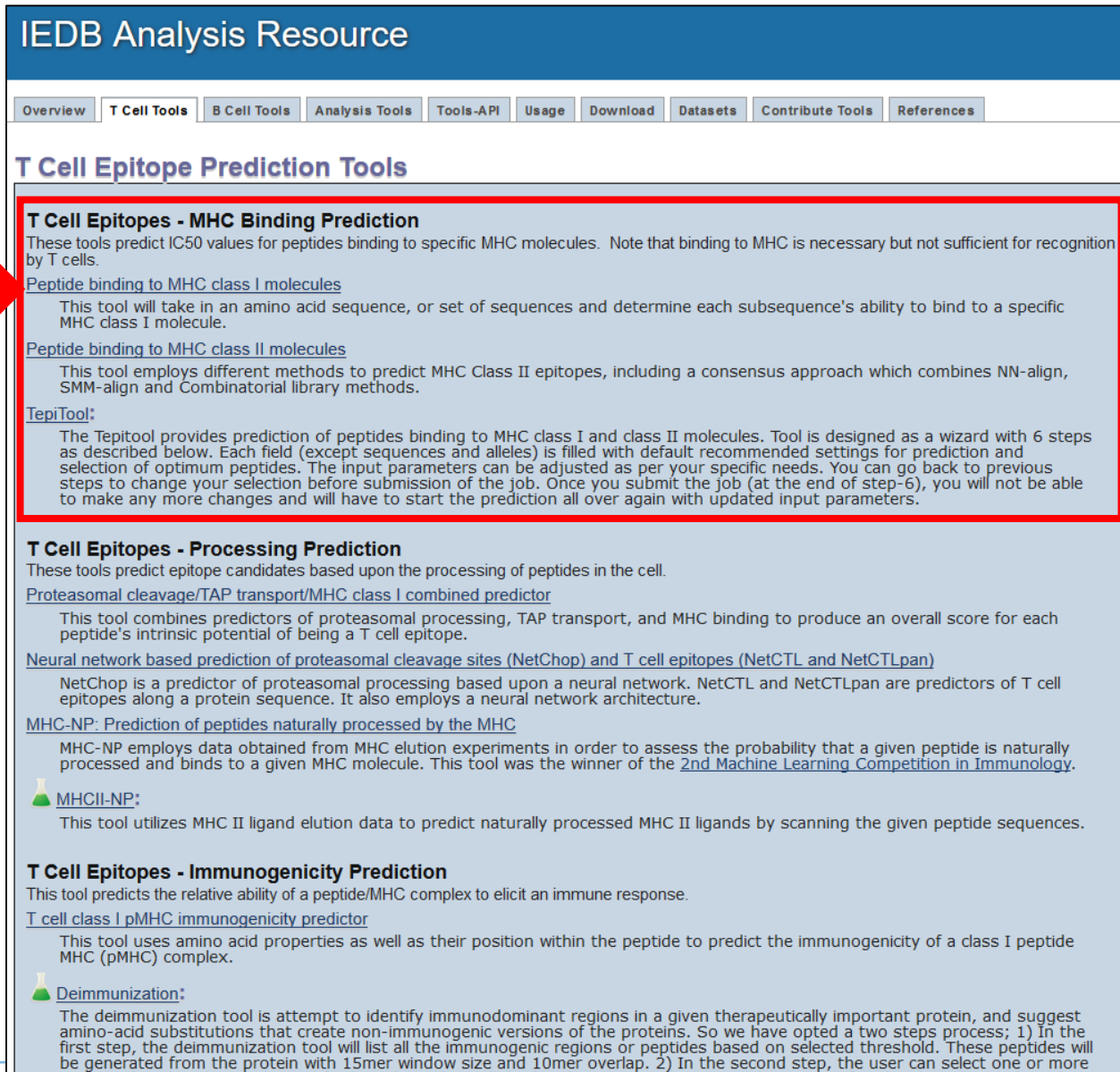
Help

Contact

This site is best viewed with current versions of [Mozilla Firefox](#) or [Google Chrome](#).

© 2005-2020 | [IEDB Home](#)

Supported by a contract from the [National Institute of Allergy and Infectious Diseases](#), a component of the National Institutes of Health in the Department of Health and Human Services.



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

MHC I binding prediction

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
MGQIVTMFEALPHIIDDEVINIVIIVLIVITGKAVYNFATCGIFALISFLLLAGRSCGM
YGLKGPDIYKGVYQKSVFEFDMSHLNLMPNACSAANSHHYISMGTSGLELFTNDNSII
SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA
QSAQSQCRTRFRGRVLDMFRFAFGGKYMRSGWGWWTGSDGKTTWCQSQTSYQYLIQNRWE
NHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSGIVENPGGYCLTKWMILAAE
LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSFKFEDVESALHLFKTTVNSLISDQ
LLMRNHLRLDMGVPCYNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTAYLVSIHLVLIPTHRHIKGGSCPKP
HRLTNKGICSCGAFKVPVGVKTVWKR
```

Or select file containing sequence(s) No file chosen

Choose a Prediction Method

Prediction Method [Help on prediction method selections](#)

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

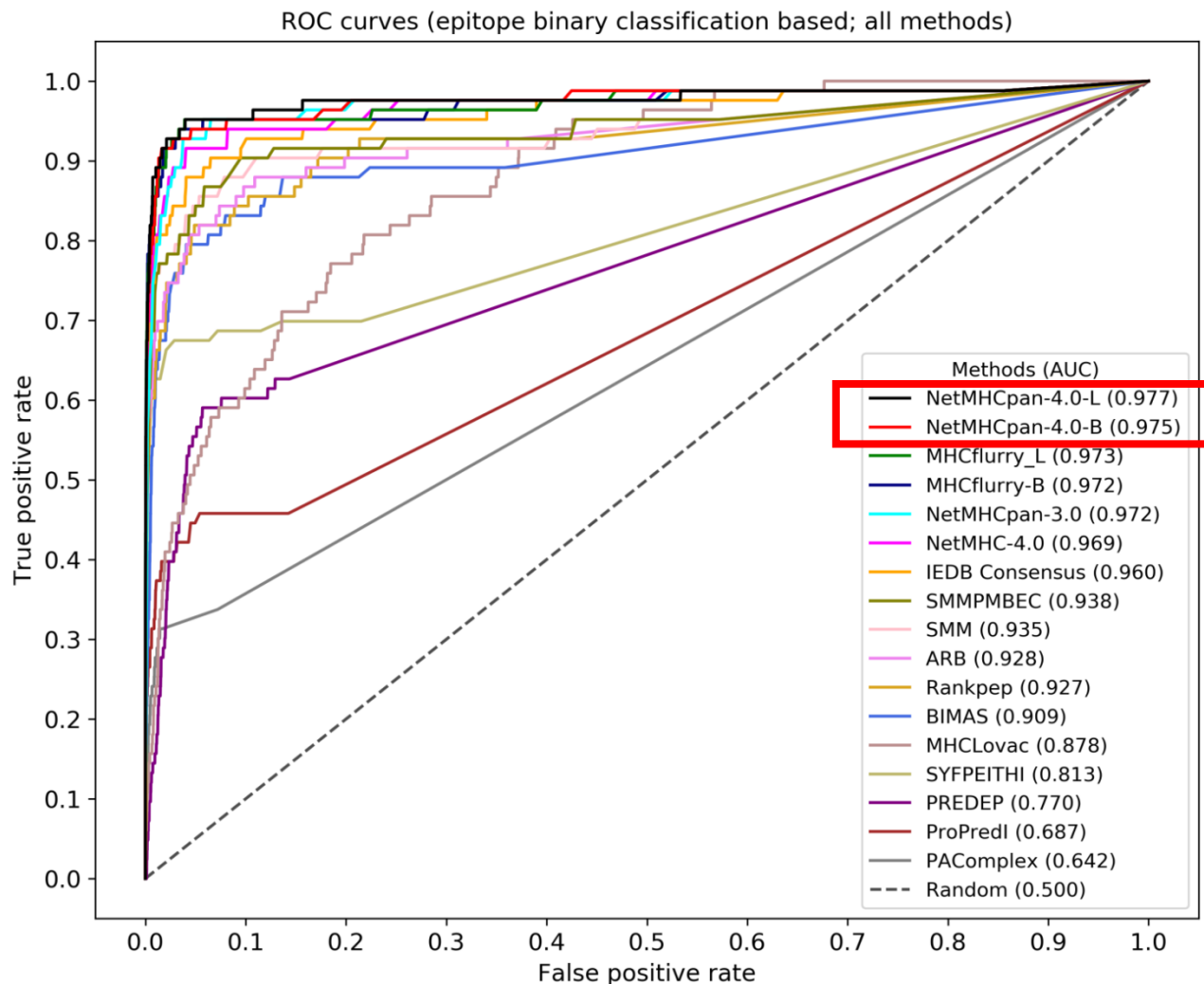
Output format

Email address (optional)

Sequence

Prediction method

MHC I binding prediction methods - benchmarking



MHC I binding prediction

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
MGQIVTMFEALPHIIDVINIVIVITGIKAVYNFATCGIFALISFLLLAGRSCGM
YGLKGPDIYKGVYQFKSVEFDMSHLNLTMFNACSAANSHHYISMGTSGLELFTNDSII
SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA
QSAQSQRCRTFRGRVLDMFRTAFGGKYMRSWGWTGSDGKTTWCQSQTSYQYLIQNRWTWE
NHCTYAGPFGMSRILLSQEKTFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMLAAE
LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSLISDQ
LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTAYLVSIFLHLVKIPITHRHIKGGSCPKP
HRLTNKGICSCGAFKVPGVKTVWKR
```

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

Allele & length

Email ID

MHC I binding prediction - result

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

tools.iedb.org/mhci/

MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	LCMV Armstrong Protein GP	MGQIVTMFEALPHIIDDEVINIVIIVLIVITGIKAVYNFATCGIFALISFLLLAGRSCGMVGLKGPDIYKG VYQFKSVFEDMSHLNLTMPNACSAANNSHHYISMGTSGLELFTTNDISIHNFCNLTSFAFNKKT DHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDAQSAQSQCRFTFRGRVDMFRFA FGGKYMRSQGWGWTGSDGKTTWCSQTSYQYLIQNRWENHCTYAGPFGMSRILLSQEKT FTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAELKCFGNTAVAKCNVNHDAEFCDMLRLI DYNKAAALSFKEDVESALHLFKTTVNSLISDQLLMRNHLRDLMGVPCYCNYSKFWYLEHAKTGE TSVPCWLVLTNGSYLNETHFSDQIEQEADNMITMLRKDYIKRQGSTPLALMDLLMFSTAYLV SIFLHLVKIPTHRIKGGSCPKPHRLTNKGGICSCGAFKVPVGVKTVWVRR

NetMHCpan allele distance

Input Allele	Closest Allele	Distance
HLA-A*02:01	HLA-A*02:01	0.000

Prediction method: NetMHCpan EL 4.1 **High Score = good binder**

* The 'IEDB recommended' method was updated in September 2020 to NetMHCpan EL 4.1 More information is available on the [help](#) page.

[Download result](#)

Citations

Allele	#	Start	End	Length	Peptide	Core	Icore	Score	Percentile Rank
HLA-A*02:01	1	6	14	9	TMFEALPHI	TMFEALPHI	TMFEALPHI	0.942547	0.03
HLA-A*02:01	1	10	18	9	ALPHIIDEV	ALPHIIDEV	ALPHIIDEV	0.920331	0.03
HLA-A*02:01	1	137	145	9	TLMSIVSSL	TLMSIVSSL	TLMSIVSSL	0.882391	0.04
HLA-A*02:01	1	447	455	9	YLVSIPLHL	YLVSIPLHL	YLVSIPLHL	0.855633	0.06
HLA-A*02:01	1	14	22	9	IIDEVINIV	IIDEVINIV	IIDEVINIV	0.807736	0.08
HLA-A*02:01	1	13	21	9	HIIDEVINI	HIIDEVINI	HIIDEVINI	0.779939	0.09
HLA-A*02:01	1	339	347	9	ALHLFKTTV	ALHLFKTTV	ALHLFKTTV	0.565798	0.22
HLA-A*02:01	1	450	458	9	SIFLHLVKI	SIFLHLVKI	SIFLHLVKI	0.530163	0.24
HLA-A*02:01	1	45	53	9	ALISFLLLA	ALISFLLLA	ALISFLLLA	0.525738	0.25
HLA-A*02:01	1	349	357	9	SLISDQLLM	SLISDQLLM	SLISDQLLM	0.510305	0.26
HLA-A*02:01	1	440	448	9	LMFSTSAYL	LMFSTSAYL	LMFSTSAYL	0.483579	0.27
HLA-A*02:01	1	320	328	9	RLIDYNKAA	RLIDYNKAA	RLIDYNKAA	0.438929	0.32
HLA-A*02:01	1	435	443	9	ALMDLLMFS	ALMDLLMFS	ALMDLLMFS	0.386032	0.39
HLA-A*02:01	1	27	35	9	IVITGIKAV	IVITGIKAV	IVITGIKAV	0.36266	0.42
HLA-A*02:01	1	42	50	9	GIFALISFL	GIFALISFL	GIFALISFL	0.285315	0.57

MHC II binding prediction

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
FNCLGMSNRDFLEGVSGATWVDLVLEGDSCVTIMSKDKPTIDVKMMNMEANLAEVRSYCYLATVSDLS
T
KAAACPTMGEAHNDKRADPAFVCRQGVVDRGWNGCGLFGKGSIDTCAKFACTSKAIGRTILKENIKYEVA
IFVHGPTTVESHGNYSTQVGATQAGRFSITPAAPSYTLKLEGEVETVDCPRSGIDTNAVYVMTVGTGT
FLVHREWFMDLNLPSWSSAGSTVWRNRETLMEFEPEPHATKQSVIALGSQEGALHQAAGAIPEVFSNTVK
LTSGHKCRVKMEKQLKGTTYGVCSKAFKFLGTPADTGHGTVVLELQYTGTDGPKVPISSVASLNDLT
PVGRLLTVNPFVSVATANAKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTLTKGAQRLAA
LGDTAWDFGSGVGGVFTSVGKAVHQVFGGAFRSLFSGGMSWITQGLLGALLLWGMGINARDRSIALTFLAVG
GVLLFLSVNVHA
```

FASTA format detected.

Or select file containing sequence(s) No file chosen

Choose a Prediction Method

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

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: [?]

DRB1*01:01 [?]

[Upload allele file](#) [?]

Select length(s) [?]

<input type="checkbox"/> 11	<input type="checkbox"/> 12	<input type="checkbox"/> 13	<input type="checkbox"/> 14	<input checked="" type="checkbox"/> 15	<input type="checkbox"/> 16	<input type="checkbox"/> 17	<input type="checkbox"/> 18	<input type="checkbox"/> 19	<input type="checkbox"/> 20
<input type="checkbox"/> 21	<input type="checkbox"/> 22	<input type="checkbox"/> 23	<input type="checkbox"/> 24	<input type="checkbox"/> 25	<input type="checkbox"/> 26	<input type="checkbox"/> 27	<input type="checkbox"/> 28	<input type="checkbox"/> 29	<input type="checkbox"/> 30

Specify Output

Sort peptides by

Output format

Email address (optional) [?]

TepiTool – MHC I and II binding prediction

tools.iedb.org/tepitool/

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.

```
>LCMV Armstrong, Protein GP
MGQIVTMFEALPHIIDVINIVIIVLIVITGIKAVYNFATCGIFALISF
```

FASTA format detected.

Or upload file containing sequence(s) No file chosen

TepiTool – MHC I and II binding prediction

tools.iedb.org/tepitool/

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	1
Host species	Human
Allele class	Class II
Alleles	1.DRB1*01:01
Duplicate peptides	Removed
Peptide lengths selected	15mers (Only one length for class II)
Approx no. of peptides included	8
Peptide overlap	10 AA residues
Conservancy analysis	Peptides conserved in at least % sequences
Prediction method	NetMHCIIpan
Peptide selection criterion	Based on predicted consensus percentile rank (Cutoff selected = 10)

Job details:


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

(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.)

TepiTool – MHC I and II binding prediction



tools.iedb.org/tepitool/

TepiTool

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

Seq # ^v	Peptide start ^v	Peptide end ^v	Peptide sequence ^v	Consensus percentile rank ^v	Allele ^v
1	26	40	LIVITGIKAVYNFAT	4.64	HLA-DRB1*01:01

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

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 NetMHCIIpan method [2,3].

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
2. 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 HLA-DQ. *Immunogenetics.* 65(10): 711.
3. 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 class II allele. *Nat. Methods.* 4(7): e1000107.

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

Input sequences:

Seq #	Seq title	Sequence
1	LCMV Armstrong, Protein GP	MGQIVTMFEALPHIIDEVINIVIMLIVITGIKAVYNFATCGIFALISF

B Cell Tools

IEDB Analysis Resource

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

B Cell Epitope Prediction Tools

B Cell Epitope Prediction


[Prediction of linear epitopes from protein sequence](#)
A collection of methods to predict linear B cell epitopes based on sequence characteristics of the antigen using amino acid scales and HMMs.


[Discotope - Prediction of epitopes from protein structure](#)
This method incorporates solvent-accessible surface area calculations, as well as contact distances into its prediction of B cell epitope potential along the length of a protein sequence.


[ElliPro - Epitope prediction based upon structural protrusion](#)
This method predicts epitopes based upon solvent-accessibility and flexibility.

[Methods for modeling and docking of antibody and protein 3D structures](#)
This page provides information on available methods for modeling and docking of antibody and protein 3D structures.

Structure Tools

 [LYRA \(Lymphocyte Receptor Automated Modelling\):](#)
The LYRA server predicts structures for either T-Cell Receptors (TCR) or B-Cell Receptors (BCR) using homology modelling. Framework templates are selected based on BLOSUM score, and complementary determining regions (CDR) are then selected if needed based on a canonical structure model and grafted onto the framework templates.

 [SCEptRe: Structural Complexes of Epitope Receptor](#)
SCEptRe provides weekly updated, non-redundant, user customized benchmark datasets with information on the immune receptor features for receptor-specific epitope predictions. This tool extracts weekly updated 3D complexes of antibody-antigen, TCR-pMHC and MHC-ligand from the Immune Epitope Database (IEDB) and clusters them based on antigens, receptors and epitopes to generate benchmark datasets. Users can customize structural quality and clustering parameters (e.g. resolution, R free factors, antigen or epitope sequence identity) to generate these datasets based on their need.

 : Tools under AR Labs which are experimental and are not quite ready for production yet. They are intended for further research, updates and testing.



B Cell epitope prediction – sequence based

tools.iedb.org/bcell/

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

Antibody Epitope Prediction

Specify Input

Enter a Swiss-Prot ID (example: P02185)

Or enter a protein sequence in plain format (50000 residues maximum):

```
VLSEGEWQLVLHVWAKVEADVAGHGGQDILIRLFKSHPETLEKDFRFKHLKTEAEMKASEDLKKHGVTVLTA  
LGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLSRHPGNFGADAGGAMNKALELFRKDIAAK  
YKELGYQG
```

Choose a method:

- [Bepipred Linear Epitope Prediction](#)
- [Bepipred Linear Epitope Prediction 2.0](#)
- [Chou & Fasman Beta-Turn Prediction](#)
- [Emini Surface Accessibility Prediction](#)
- [Karplus & Schulz Flexibility Prediction](#)
- [Kolaskar & Tongaonkar Antigenicity](#)
- [Parker Hydrophilicity Prediction](#)

B Cell epitope prediction – sequence based

tools.iedb.org/bcell/

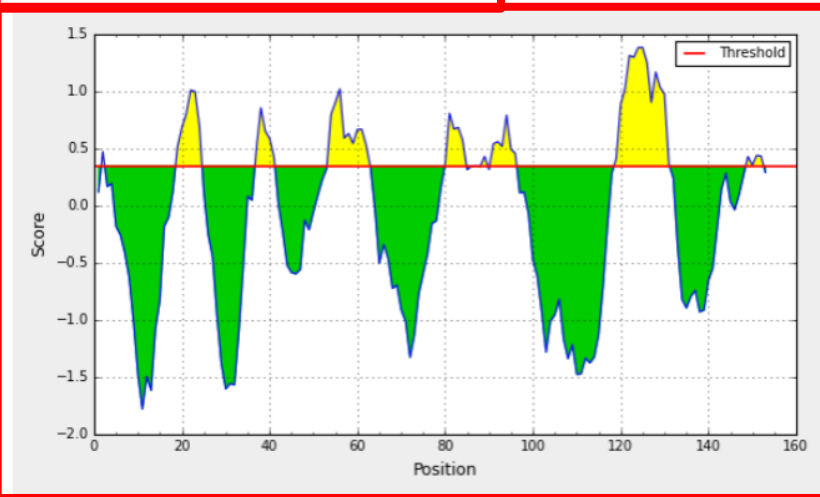
Home Help Example Reference Download Contact

Bepipred Linear Epitope Prediction Results

Input Sequences

1 VLSEGEWQLV LHWAKVEAD VAGHGQDILI RLFKSHPETL EKFDREKHLK TEEMKASED
 61 LKKHGVTVLT ALGAILKKKG HHEAELKPLA QSHATKHKIP IKYLEFISEA IIHVLHSRHP
 121 GNFGADAGGA MNKALELFRK DIAAKYKELG YQG

Center position: 4 Threshold: 0.350



Average: -0.105 Minimum: -0.028 Maximum: 1.390

Predicted peptides:

No. ↕	Start ↕	End ↕	Peptide ↕	Length ↕
1	2	2	L	1
2	19	24	ADVAGH	6
3	37	41	PETLE	5
4	54	62	EMKASEDLK	9
5	80	84	GHHEA	5
6	87	87	K	1
7	89	89	L	1
8	91	96	QSHATK	6
9	119	131	HPGNFGADAGGAM	13
10	149	152	LGYQ	4

Predicted residue scores:

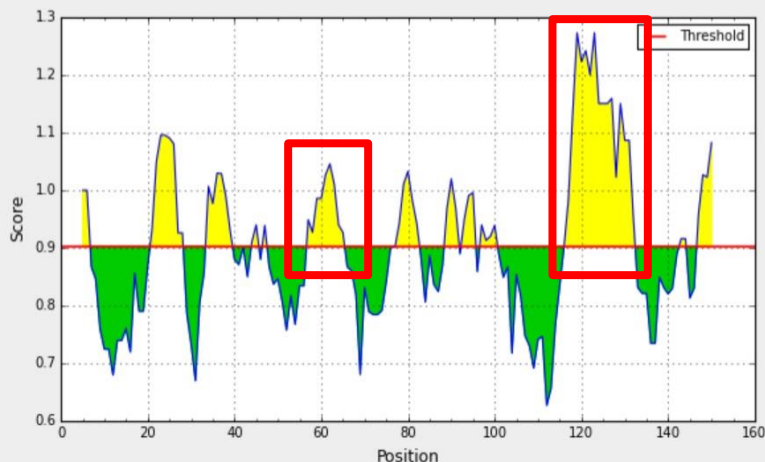
Position ↕	Residue ↕	Score ↕	Assignment ↕
1	V	0.121	.
2	L	0.476	E
3	S	0.168	.

Chou & Fasman Beta-Turn Prediction

Input Sequences

```
1 VLSEGEWQLV LHWAKVEAD VAGHGQDILI RLFKSHPETL EKFDKFKHLK TEAEMKASED
61 LKKHGVTVLT ALGAILKKKG HHEAELKPLA QSHATKHKIP IKYLEFISEA IIHVLHSRHP
121 GNFGADAGGA MNKALELFRK DIAAKYKELG YQG
```

Center position: 4 Window size: 7 Threshold: 0.903

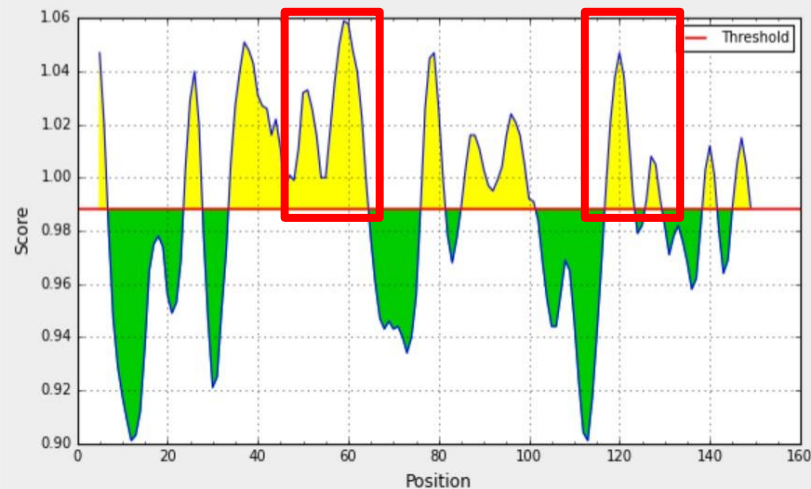


Karplus & Schulz Flexibility Prediction

Input Sequences

```
1 VLSEGEWQLV LHWAKVEAD VAGHGQDILI RLFKSHPETL EKFDKFKHLK TEAEMKASED
61 LKKHGVTVLT ALGAILKKKG HHEAELKPLA QSHATKHKIP IKYLEFISEA IIHVLHSRHP
121 GNFGADAGGA MNKALELFRK DIAAKYKELG YQG
```

Center position: 4 Window size: 7 Threshold: 0.988

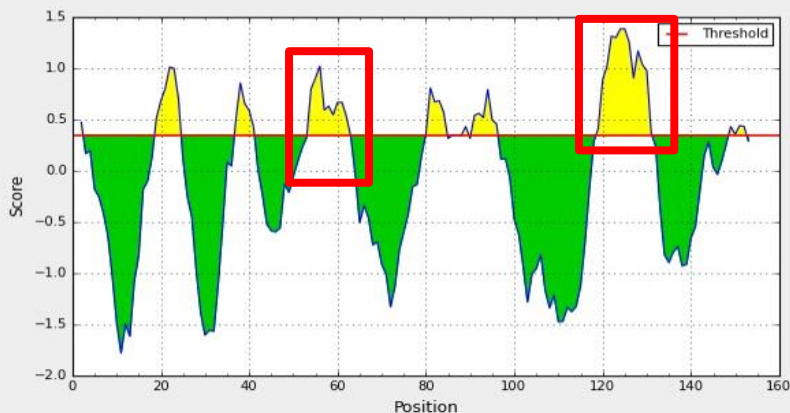


Bepipred Linear Epitope Prediction

Input Sequences

```
1 VLSEGEWQLV LHWAKVEAD VAGHGQDILI RLFKSHPETL EKFDKFKHLK TEAEMKASED
61 LKKHGVTVLT ALGAILKKKG HHEAELKPLA QSHATKHKIP IKYLEFISEA IIHVLHSRHP
121 GNFGADAGGA MNKALELFRK DIAAKYKELG YQG
```

Center position: 4 Window size: 7 Threshold: 0.35



**Using a consensus of
different methods may
be better.**

B cell epitope prediction – structure based

tools.iedb.org/discotope/

Home Help Example Reference Download Contact

DiscoTope: Structure-based Antibody Prediction

Step 1: Please enter the 4-letter PDB ID
Or upload a PDB file

(example: 1z40)

No file selected.

Step 2: Please enter PDB Chain ID

Step 3: Select version

RCSB PDB 155687 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or ligands

Advanced Search | Browse by Annotations

PDB-101 PDB EMDataResource Wellcome Open Access WorldWide Protein Data Bank Foundation

Structure Summary 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment

Biological Assembly 1



1Z40
AMA1 from Plasmodium falciparum
DOI: 10.2210/pdb1Z40/pdb
Classification: [UNKNOWN FUNCTION](#)
Organism(s): [Plasmodium falciparum \(isolate 3D7\)](#)
Expression System: [Escherichia coli BL21\(DE3\)](#)
Deposited: 2005-03-14 Released: 2005-08-16
Deposition Author(s): [Bai, T.](#), [Becker, M.](#), [Gupta, A.](#), [Strike, P.](#), [Murphy, V.J.](#), [Anders, R.F.](#), [Batchelor, A.H.](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.901 Å
R-Value Free: 0.236
R-Value Work: 0.192

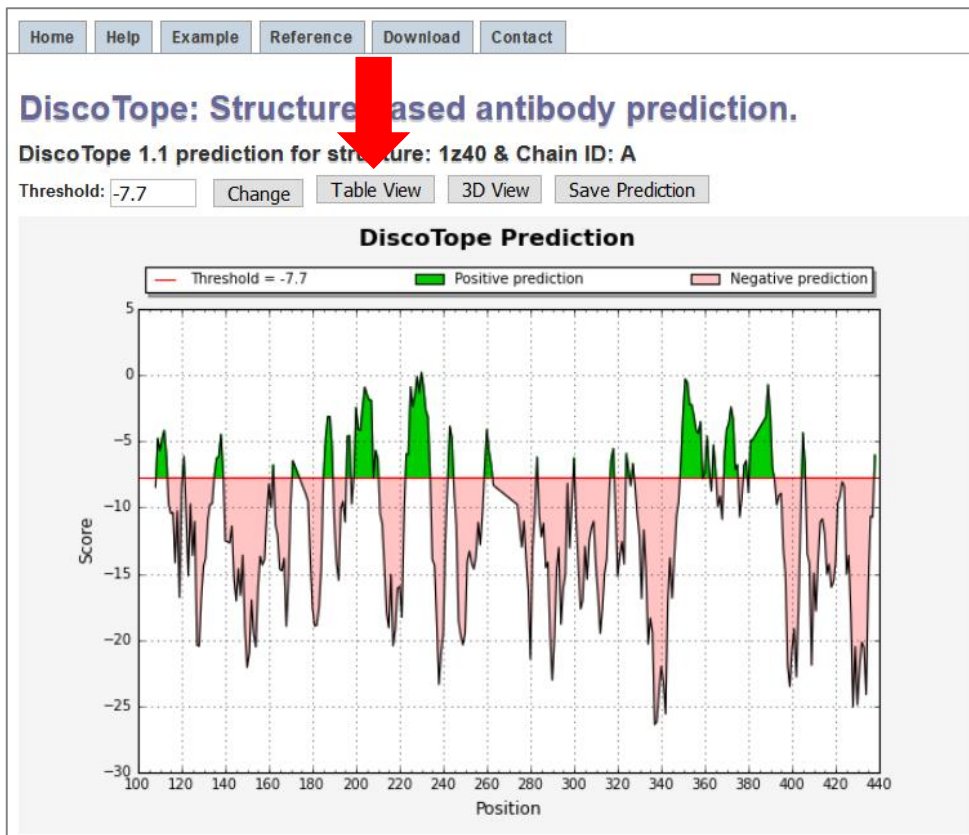
wwPDB Validation

Metric	Percentile Ranks	Value
Rfree		0.234
Clashscore		5
Ramachandran outliers		1.0%
Sidechain outliers		3.2%
RSRZ outliers		16.2%

3D View: [Structure](#) | [Electron Density](#) | [Ligand Interaction](#)

Standalone Viewers
[Oryx3](#) | [MolView](#) | [Jmol](#) | [Pymol](#) | [Ribbon](#) | [Surface](#) | [Cartoon](#) | [Stick](#) | [Ball](#) | [Wireframe](#) | [Schematic](#) | [Schematic](#) | [Schematic](#)

B cell epitope prediction – structure based



Home Help Example Reference Download Contact

DiscoTope - Result

DiscoTope 1.1 prediction for structure: 1z40 & Chain ID: A
The positive predictions are displayed in green.

Chain ID	Residue ID	Residue Name	Contact Number	Propensity Score	DiscoTope Score
A	108	ASN	14	-1.459	-8.459
A	109	PRO	11	0.724	-4.776
A	110	TRP	13	0.804	-5.696
A	111	THR	12	1.211	-4.789
A	112	GLU	11	1.331	-4.169
A	113	TYR	14	0.929	-6.071
A	114	MET	18	-0.779	-9.779
A	115	ALA	20	-0.444	-10.444
A	116	LYS	21	0.122	-10.378
A	117	TYR	24	-2.172	-14.172
A	118	ASP	21	0.257	-10.243
A	119	ILE	32	-0.783	-16.783
A	120	GLU	21	1.954	-8.546
A	121	GLU	15	1.366	-6.134
A	122	VAL	20	-0.374	-10.374
A	123	HIS	28	-1.144	-15.144
A	124	GLY	22	1.274	-9.726
A	125	SER	29	0.887	-13.613
A	126	GLY	28	2.951	-11.049
A	127	ILE	35	-2.881	-20.381
A	128	ARG	29	-5.973	-20.473
A	129	VAL	30	-1.817	-16.817
A	130	ASP	31	1.048	-14.452
A	131	LEU	31	1.727	-13.773
A	132	GLY	25	1.617	-10.883
A	133	GLU	19	-0.26	-9.76

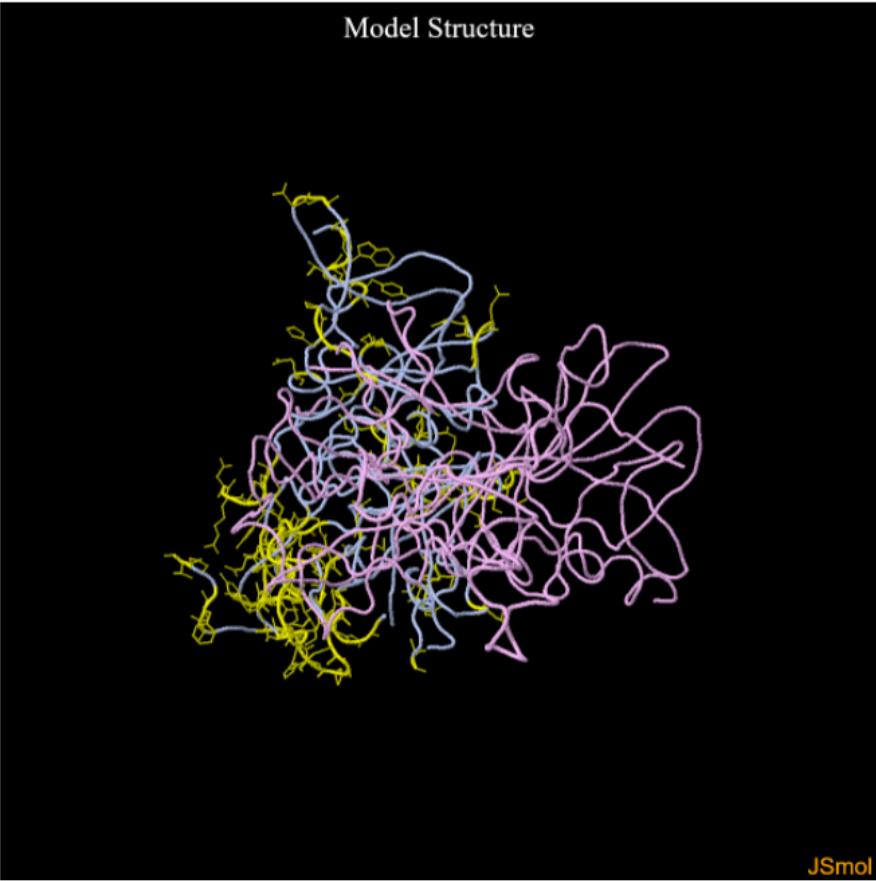
B cell epitope prediction – structure based

Home Help Example Reference Download Contact

JSmol-Rendered PDB Structure

Chart View Table View Save Prediction

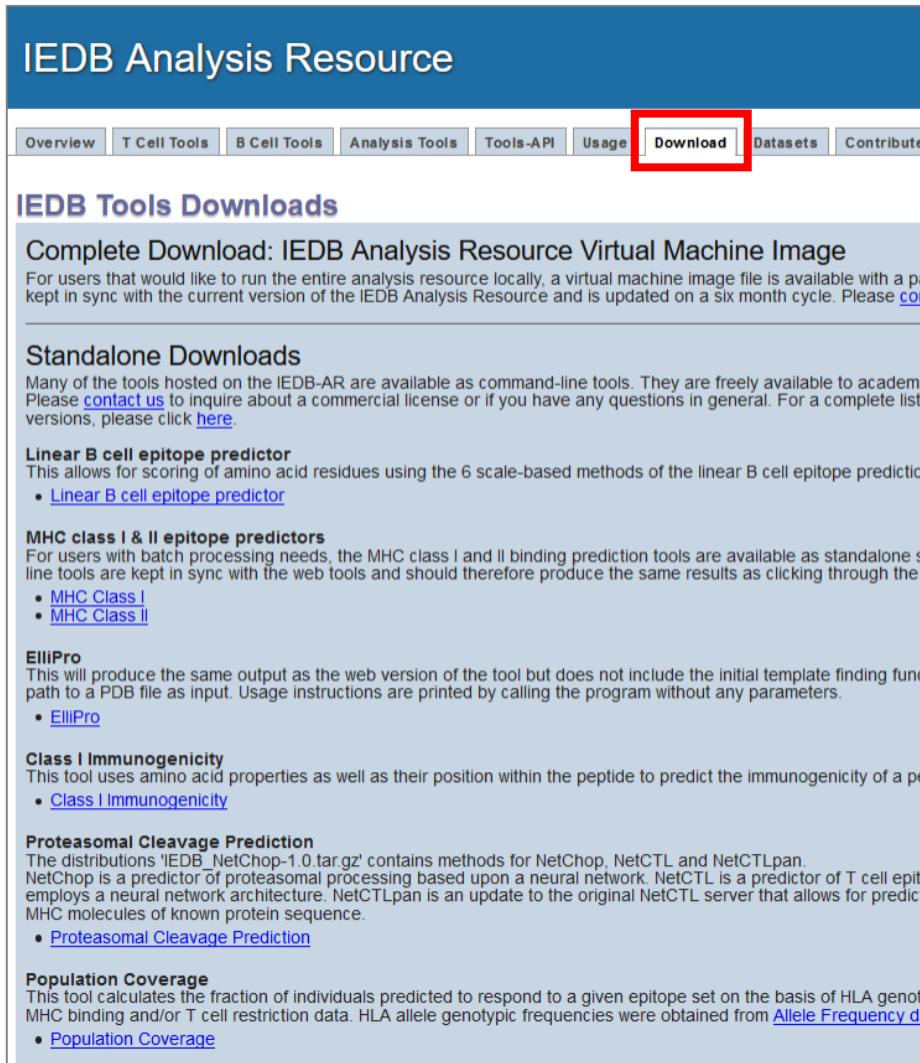
Model Structure



Chain ID	Residue ID	Residue Name	Contact Number	Propensity Score	Discoptope Score	View
A	109	PRO	11	0.724	-4.776	CPK
A	110	TRP	13	0.804	-5.696	CPK
A	111	THR	12	1.211	-4.789	CPK
A	112	GLU	11	1.331	-4.169	CPK
A	113	TYR	14	0.929	-6.071	CPK
A	121	GLU	15	1.366	-6.134	CPK
A	135	ALA	15	-0.017	-7.517	CPK
A	136	GLU	15	1.225	-6.275	CPK
A	137	VAL	22	4.846	-6.154	CPK
A	138	ALA	15	3.024	-4.476	CPK
A	139	GLY	12	-1.166	-7.166	CPK
A	162	ASN	11	-1.279	-6.779	CPK
A	171	THR	9	-1.963	-6.463	CPK
A	186	THR	10	0.194	-4.806	CPK
A	187	GLU	7	0.361	-3.139	CPK
A	188	PRO	7	0.361	-3.139	CPK
A	189	LEU	11	0.049	-5.451	CPK
A	196	ASP	12	1.384	-4.616	CPK
A	197	GLU	15	2.938	-4.562	CPK
A	199	ARG	20	2.492	-7.508	CPK
A	200	HIS	13	4.057	-2.443	CPK

Standalone (local) version

tools.iedb.org/main/download/



The screenshot shows the IEDB Analysis Resource website. The header is blue with the text 'IEDB Analysis Resource'. Below the header is a navigation bar with tabs: Overview, T Cell Tools, B Cell Tools, Analysis Tools, Tools-API, Usage, Download (highlighted with a red box), Datasets, and Contribute. The main content area is titled 'IEDB Tools Downloads' and contains several sections:

- Complete Download: IEDB Analysis Resource Virtual Machine Image**
For users that would like to run the entire analysis resource locally, a virtual machine image file is available with a patch kept in sync with the current version of the IEDB Analysis Resource and is updated on a six month cycle. Please [contact us](#)
- Standalone Downloads**
Many of the tools hosted on the IEDB-AR are available as command-line tools. They are freely available to academic users. Please [contact us](#) to inquire about a commercial license or if you have any questions in general. For a complete list of versions, please click [here](#).
- Linear B cell epitope predictor**
This allows for scoring of amino acid residues using the 6 scale-based methods of the linear B cell epitope predictor.
 - [Linear B cell epitope predictor](#)
- MHC class I & II epitope predictors**
For users with batch processing needs, the MHC class I and II binding prediction tools are available as standalone command-line tools. These tools are kept in sync with the web tools and should therefore produce the same results as clicking through the web interface.
 - [MHC Class I](#)
 - [MHC Class II](#)
- EIIIPro**
This will produce the same output as the web version of the tool but does not include the initial template finding function. It takes a path to a PDB file as input. Usage instructions are printed by calling the program without any parameters.
 - [EIIIPro](#)
- Class I Immunogenicity**
This tool uses amino acid properties as well as their position within the peptide to predict the immunogenicity of a peptide.
 - [Class I Immunogenicity](#)
- Proteasomal Cleavage Prediction**
The distributions 'IEDB_NetChop-1.0.tar.gz' contains methods for NetChop, NetCTL and NetCTLpan. NetChop is a predictor of proteasomal processing based upon a neural network. NetCTL is a predictor of T cell epitopes. NetCTLpan employs a neural network architecture. NetCTLpan is an update to the original NetCTL server that allows for prediction of MHC molecules of known protein sequence.
 - [Proteasomal Cleavage Prediction](#)
- Population Coverage**
This tool calculates the fraction of individuals predicted to respond to a given epitope set on the basis of HLA genotype, MHC binding and/or T cell restriction data. HLA allele genotypic frequencies were obtained from [Allele Frequency Database](#).
 - [Population Coverage](#)

- Run programs on your local machine
- Advantages:
 - No internet needed
 - Very helpful in case of large data sets
 - Free for non-profit & academia
 - Available for industry at a nominal fee
- license@iedb.org

Standalone (local) version - example

```

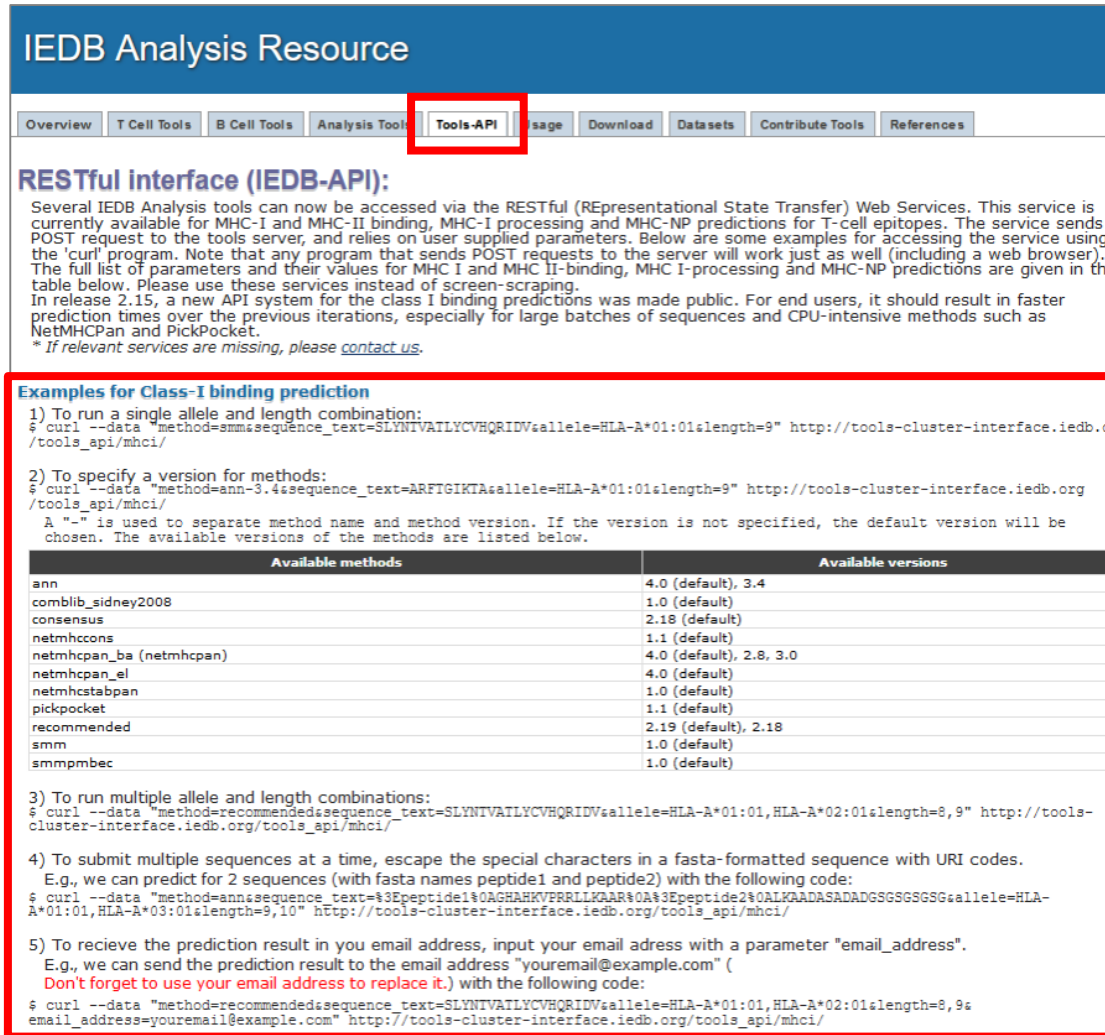
spaul@ubuntu:~/tools/mhc_1$ ./src/predict_binding.py netmhcpan "HLA-A*02:01" 9 test_sequence.fasta
allele seq_num start end length peptide ic50 rank
HLA-A*02:01 1 6 14 9 TMFEALPHI 4.3 0.03
HLA-A*02:01 1 10 18 9 ALPHIIDEV 12.6 0.12
HLA-A*02:01 1 14 22 9 IIDEVINIV 43.3 0.47
HLA-A*02:01 1 38 46 9 FATCGIFAL 65.2 0.64
HLA-A*02:01 1 13 21 9 HIIDEVINI 97.3 0.87
HLA-A*02:01 1 2 10 9 GQIVTMFEA 245.5 1.8
HLA-A*02:01 1 27 35 9 IVITGIKAV 324.4 2.1
HLA-A*02:01 1 24 32 9 IVLIVITGI 651.0 3.0
HLA-A*02:01 1 17 25 9 EVINIVIIIV 3263.2 7.2
HLA-A*02:01 1 18 26 9 VINIVIIIVL 3491.7 7.5
HLA-A*02:01 1 3 11 9 QIVTMFEAL 3523.2 7.5
HLA-A*02:01 1 20 28 9 NIVIIIVLIV 4861.5 9.0
HLA-A*02:01 1 21 29 9 IVIIVLIVI 5959.9 12
HLA-A*02:01 1 31 39 9 GIKAVYNFA 6530.5 12
HLA-A*02:01 1 7 15 9 MFEALPHII 11337.5 17
HLA-A*02:01 1 37 45 9 NFATCGIFA 11579.5 17
HLA-A*02:01 1 26 34 9 LIVITGIKA 12185.2 18
HLA-A*02:01 1 33 41 9 KAVYNFATC 12713.3 18
HLA-A*02:01 1 22 30 9 VIIVLIVIT 13298.3 19
HLA-A*02:01 1 19 27 9 INIVIIIVLI 14913.1 21
HLA-A*02:01 1 4 12 9 IVTMFEALP 15618.8 22
HLA-A*02:01 1 34 42 9 AVYNFATCG 18955.1 25
HLA-A*02:01 1 35 43 9 VYNFATCGI 19064.3 26
HLA-A*02:01 1 25 33 9 VLIVITGIK 19910.8 27
HLA-A*02:01 1 36 44 9 YNFATCGIF 21579.6 29
HLA-A*02:01 1 23 31 9 IIVLIVITG 22420.2 30
HLA-A*02:01 1 28 36 9 VITGIKAVY 28040.2 39
HLA-A*02:01 1 1 9 9 MGQIVTMFE 28618.9 40
HLA-A*02:01 1 32 40 9 IKAVYNFAT 28756.4 41
HLA-A*02:01 1 5 13 9 VTMFEALPH 28919.3 41
HLA-A*02:01 1 30 38 9 TGIKAVYNF 29272.8 42
HLA-A*02:01 1 11 19 9 LPHIIDEVI 29396.6 42
HLA-A*02:01 1 15 23 9 IDEVINIVI 32112.8 48
HLA-A*02:01 1 16 24 9 DEVINIVII 32893.9 50
HLA-A*02:01 1 9 17 9 EALPHIIDE 37902.4 66
HLA-A*02:01 1 29 37 9 ITGIKAVYN 38010.0 66
HLA-A*02:01 1 8 16 9 FEALPHIID 40600.3 76
HLA-A*02:01 1 12 20 9 PHIIDEVIN 48177.9 100
spaul@ubuntu:~/tools/mhc_1$ ./src/predict_binding.py netmhcpan "HLA-A*02:01" 9 test_s
spaul@ubuntu:~/tools/mhc_1$

```

prediction_results.txt x							
	allele	seq_num	start	end	length	peptide	ic50 rank
1	HLA-A*02:01	1	6	14	9	TMFEALPHI	4.3 0.03
2	HLA-A*02:01	1	10	18	9	ALPHIIDEV	12.6 0.12
3	HLA-A*02:01	1	14	22	9	IIDEVINIV	43.3 0.47
4	HLA-A*02:01	1	38	46	9	FATCGIFAL	65.2 0.64
5	HLA-A*02:01	1	13	21	9	HIIDEVINI	97.3 0.87
6	HLA-A*02:01	1	2	10	9	GQIVTMFEA	245.5 1.8
7	HLA-A*02:01	1	27	35	9	IVITGIKAV	324.4 2.1
8	HLA-A*02:01	1	24	32	9	IVLIVITGI	651.0 3.0
9	HLA-A*02:01	1	17	25	9	EVINIVIIIV	3263.2 7.2
10	HLA-A*02:01	1	18	26	9	VINIVIIIVL	3491.7 7.5
11	HLA-A*02:01	1	3	11	9	QIVTMFEAL	3523.2 7.5
12	HLA-A*02:01	1	20	28	9	NIVIIIVLIV	4861.5 9.0
13	HLA-A*02:01	1	21	29	9	IVIIVLIVI	5959.9 12
14	HLA-A*02:01	1	31	39	9	GIKAVYNFA	6530.5 12
15	HLA-A*02:01	1	7	15	9	MFEALPHII	11337.5 17
16	HLA-A*02:01	1	37	45	9	NFATCGIFA	11579.5 17
17	HLA-A*02:01	1	26	34	9	LIVITGIKA	12185.2 18
18	HLA-A*02:01	1	33	41	9	KAVYNFATC	12713.3 18
19	HLA-A*02:01	1	22	30	9	VIIVLIVIT	13298.3 19
20	HLA-A*02:01	1	19	27	9	INIVIIIVLI	14913.1 21
21	HLA-A*02:01	1	4	12	9	IVTMFEALP	15618.8 22
22	HLA-A*02:01	1	34	42	9	AVYNFATCG	18955.1 25
23	HLA-A*02:01	1	35	43	9	VYNFATCGI	19064.3 26
24	HLA-A*02:01	1	25	33	9	VLIVITGIK	19910.8 27
25	HLA-A*02:01	1	36	44	9	YNFATCGIF	21579.6 29
26	HLA-A*02:01	1	23	31	9	IIVLIVITG	22420.2 30
27	HLA-A*02:01	1	28	36	9	VITGIKAVY	28040.2 39
28	HLA-A*02:01	1	1	9	9	MGQIVTMFE	28618.9 40
29	HLA-A*02:01	1	32	40	9	IKAVYNFAT	28756.4 41
30	HLA-A*02:01	1	5	13	9	VTMFEALPH	28919.3 41
31	HLA-A*02:01	1	30	38	9	TGIKAVYNF	29272.8 42
32	HLA-A*02:01	1	11	19	9	LPHIIDEVI	29396.6 42
33	HLA-A*02:01	1	15	23	9	IDEVINIVI	32112.8 48
34	HLA-A*02:01	1	16	24	9	DEVINIVII	32893.9 50
35	HLA-A*02:01	1	9	17	9	EALPHIIDE	37902.4 66
36	HLA-A*02:01	1	29	37	9	ITGIKAVYN	38010.0 66
37	HLA-A*02:01	1	8	16	9	FEALPHIID	40600.3 76
38	HLA-A*02:01	1	12	20	9	PHIIDEVIN	48177.9 100
39	HLA-A*02:01	1	12	20	9	PHIIDEVIN	48177.9 100

API version (RESTful interface)

tools.iedb.org/main/tools-api/



The screenshot shows the IEDB Analysis Resource website. The 'Tools-API' tab is highlighted with a red box. Below the navigation bar, the 'RESTful interface (IEDB-API):' section is visible. A red box highlights the 'Examples for Class-I binding prediction' section, which contains five numbered examples and a table of available methods and versions.

RESTful interface (IEDB-API):

Several IEDB Analysis tools can now be accessed via the RESTful (Representational State Transfer) Web Services. This service is currently available for MHC-I and MHC-II binding, MHC-I processing and MHC-NP predictions for T-cell epitopes. The service sends POST request to the tools server, and relies on user supplied parameters. Below are some examples for accessing the service using the 'curl' program. Note that any program that sends POST requests to the server will work just as well (including a web browser). The full list of parameters and their values for MHC I and MHC II-binding, MHC I-processing and MHC-NP predictions are given in the table below. Please use these services instead of screen-scraping.
In release 2.15, a new API system for the class I binding predictions was made public. For end users, it should result in faster prediction times over the previous iterations, especially for large batches of sequences and CPU-intensive methods such as NetMHCpan and PickPocket.
** If relevant services are missing, please [contact us](#).*

Examples for Class-I binding prediction

- 1) To run a single allele and length combination:

```
$ curl --data "method=smm;sequence_text=SLYNTIVATLYCVHQRIQV;allele=HLA-A*01:01;length=9" http://tools-cluster-interface.iedb.org/tools_api/mhci/
```
- 2) To specify a version for methods:

```
$ curl --data "method=ann-3.4;sequence_text=ARFTGIKTA;allele=HLA-A*01:01;length=9" http://tools-cluster-interface.iedb.org/tools_api/mhci/
```

A "-" is used to separate method name and method version. If the version is not specified, the default version will be chosen. The available versions of the methods are listed below.

Available methods	Available versions
ann	4.0 (default), 3.4
comblib_sidney2008	1.0 (default)
consensus	2.18 (default)
netmhccons	1.1 (default)
netmhcpn_ba (netmhcpn)	4.0 (default), 2.8, 3.0
netmhcpn_el	4.0 (default)
netmhctabpan	1.0 (default)
pickpocket	1.1 (default)
recommended	2.19 (default), 2.18
smm	1.0 (default)
smmpmbec	1.0 (default)
- 3) To run multiple allele and length combinations:

```
$ curl --data "method=recommended;sequence_text=SLYNTIVATLYCVHQRIQV;allele=HLA-A*01:01,HLA-A*02:01;length=8,9" http://tools-cluster-interface.iedb.org/tools_api/mhci/
```
- 4) To submit multiple sequences at a time, escape the special characters in a fasta-formatted sequence with URI codes.
E.g., we can predict for 2 sequences (with fasta names peptide1 and peptide2) with the following code:

```
$ curl --data "method=ann;sequence_text=%3Epeptide1%0AGHAKVPRRLKAR%0A%3Epeptide2%0ALKAADASADAGSGSGSG;allele=HLA-A*01:01,HLA-A*03:01;length=9,10" http://tools-cluster-interface.iedb.org/tools_api/mhci/
```
- 5) To receive the prediction result in you email address, input your email address with a parameter "email_address".
E.g., we can send the prediction result to the email address "youremail@example.com" (**Don't forget to use your email address to replace it.**) with the following code:

```
$ curl --data "method=recommended;sequence_text=SLYNTIVATLYCVHQRIQV;allele=HLA-A*01:01,HLA-A*02:01;length=8,9;email_address=youremail@example.com" http://tools-cluster-interface.iedb.org/tools_api/mhci/
```

- Sends prediction request to the tools server at LJI
- No need to install tools on your machine
- Freely available to all users
- Can be incorporated in prediction pipelines
- Automatic update without reinstalling

API version (RESTful interface) - example

	A	B
1	peptide	allele
2	EALPHIIDEVINI	HLA-B*58:01
3	AVAKCNVNHDAEFC	HLA-A*68:01
4	SKFKEDVESA	HLA-A*68:02
5	SHLNLTMPNA	HLA-A*01:01
6	LMRNHLRDLMGV	HLA-A*32:01
7	NPGGYCLTKWMILA	HLA-A*26:01
8	AQSAQSQCRT	HLA-A*01:01
9	LSIRGNSNYKAVSC	HLA-A*03:01
10	QCRTFRGRVDMF	HLA-B*53:01
11	GTSGLELFTND	HLA-A*11:01
12	NLTSAFNKK	HLA-A*23:01
13	CDMLRLIDYNKAA	HLA-B*53:01
14	YIKRQGSTPL	HLA-A*26:01
15	YMRSGWGWG	HLA-A*23:01
16	LVTNGSYLNETHF	HLA-B*58:01
17	TKFFTRRL	HLA-B*57:01
18	NVNHDAEFCMLRL	HLA-B*08:01
19	HIKGGCCKPH	HLA-A*30:01
20	DGKTTWCSQTS	HLA-A*32:01
21	HFSQIEQADNM	HLA-A*32:01
22	FSDQIEQADNMI	HLA-B*57:01
23	CNYSKFWY	HLA-B*58:01
24	MSHLNLMPNAC	HLA-A*02:01
25	SGVENPGGYC	HLA-B*44:03
26	VIIVLIVITGIK	H-2-Kb
27	FRGRVDMFR	HLA-B*51:01
28	HIKGGCCKPHR	HLA-B*44:02
29	SIRGNSNYKAVS	H-2-Kb
30	IQYNLTFSDA	HLA-A*02:06
31	RTFRGRVL	HLA-B*15:01
32	DAQSAQSQCRTFRG	HLA-B*44:02
33	QNRTWENHCTYAGP	HLA-B*15:01
34	AFGGKYMRSWGWWT	HLA-B*07:02
35	SRILLSQEKTKFFT	HLA-A*31:01
36	SALHLFKTTVNSLI	H-2-Kb
37	RKDYIKRQGSTP	HLA-B*58:01
38	FKSVEFDMSHLNL	HLA-B*58:01

```

api_predictor.py x
import pandas as pd
import shlex, subprocess

peptide_data = pd.read_csv('peptides_to_predict.txt', sep='\t')
alleles = peptide_data['allele'].tolist()
peptides = peptide_data['peptide'].tolist()

def prediction(peptide, allele, length):
    command = 'curl --data "method=recommended&sequence_text=' + peptide + \
              '&allele=' + allele + \
              '&length=' + str(length) + \
              '" http://tools-cluster-interface.iedb.org/tools_api/mhci/'
    args = shlex.split(command)
    process = subprocess.Popen(args, stdout=subprocess.PIPE, stderr=subprocess.PIPE)
    output = process.communicate()
    consensus_percentile = output[0].decode('utf8').split('\t')[22]
    return consensus_percentile

output_file_name = 'prediction_results.txt'
output_file = open(output_file_name, 'w')
with open(output_file_name, 'a') as file to write:
    file to write.write('peptide\tallele\tlength\tpercentile_rank\n')
    for i in range(len(alleles)):
        peptide = peptides[i]
        allele = alleles[i]
        length = len(peptide)
        consensus_percentile = prediction(peptide, allele, length)
        write_line = peptide + '\t' + allele + '\t' + str(
            consensus_percentile)
        file to write.write(write_line)
output_file.close()
    
```

	peptide	allele	length	percentile_rank
1	EALPHIIDEVINI	HLA-B*58:01	13	53.0
2	AVAKCNVNHDAEFC	HLA-A*68:01	14	56.0
3	SKFKEDVESA	HLA-A*68:02	10	36.5
4	SHLNLTMPNA	HLA-A*01:01	10	46.5
5	LMRNHLRDLMGV	HLA-A*32:01	12	45.0
6	NPGGYCLTKWMILA	HLA-A*26:01	14	57.0
7	AQSAQSQCRT	HLA-A*01:01	10	28.5
8	LSIRGNSNYKAVSC	HLA-A*03:01	14	13.0
9	QCRTFRGRVDMF	HLA-B*53:01	13	41.0
10	GTSGLELFTND	HLA-A*11:01	12	46.0
11	NLTSAFNKK	HLA-A*23:01	9	28.5
12	CDMLRLIDYNKAA	HLA-B*53:01	13	53.0
13	YIKRQGSTPL	HLA-A*26:01	10	10.25
14	YMRSGWGWG	HLA-A*23:01	10	13.55

Versions of IEDB Analysis Resource tools

Web

<http://tools.iedb.org>

- Client uses browsers to submit data
- Predictions run on IEDB tools server
- Can be run on Windows/Mac/Linux
- Internet is needed
- May not be suitable for very very large data sets
- Automatically updated

Standalone

<http://tools.iedb.org/main/download>

- Uses command line interface
- Downloaded from IEDB website
- Installed and run on local machine
- Runs on Linux only (can use virtual machines to run Linux on other OS)
- Internet not needed once installed
- Better for very large data sets
- Need to update for every release
- Free to academia/non-profit; License fee for industry

API

<http://tools.iedb.org/main/tools-api>

- Uses command line interface
- Predictions run on IEDB server
- Clients send parameters to IEDB server using commands or scripts
- Internet is needed
- Can be used to make custom scripts for use with large data sets
- Automatically updated
- Free to all

Points to remember

- First stop is IEDB database
- Epitope prediction tools extrapolate from existing data to identify new candidate epitopes
 - ‘Machine learning’ approaches identify patterns
 - ROC curves / AUC values as preferred performance metrics
 - Prediction is a screening step, not confirmatory
 - Predicted peptides should be experimentally verified
- Analysis tools help to examine existing sets of epitopes and gain new knowledge
 - No single metric of performance
 - Broad array of applications