Analysis Resource Overview

tools.iedb.org

Presented by: Sinu Paul, Bioinformatics Scientist

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
- MHCNP, MHCIINP

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

Analysis tools

Population coverage of epitope set
- Population coverage tool

Degree of conservation
- Conservancy analysis tool

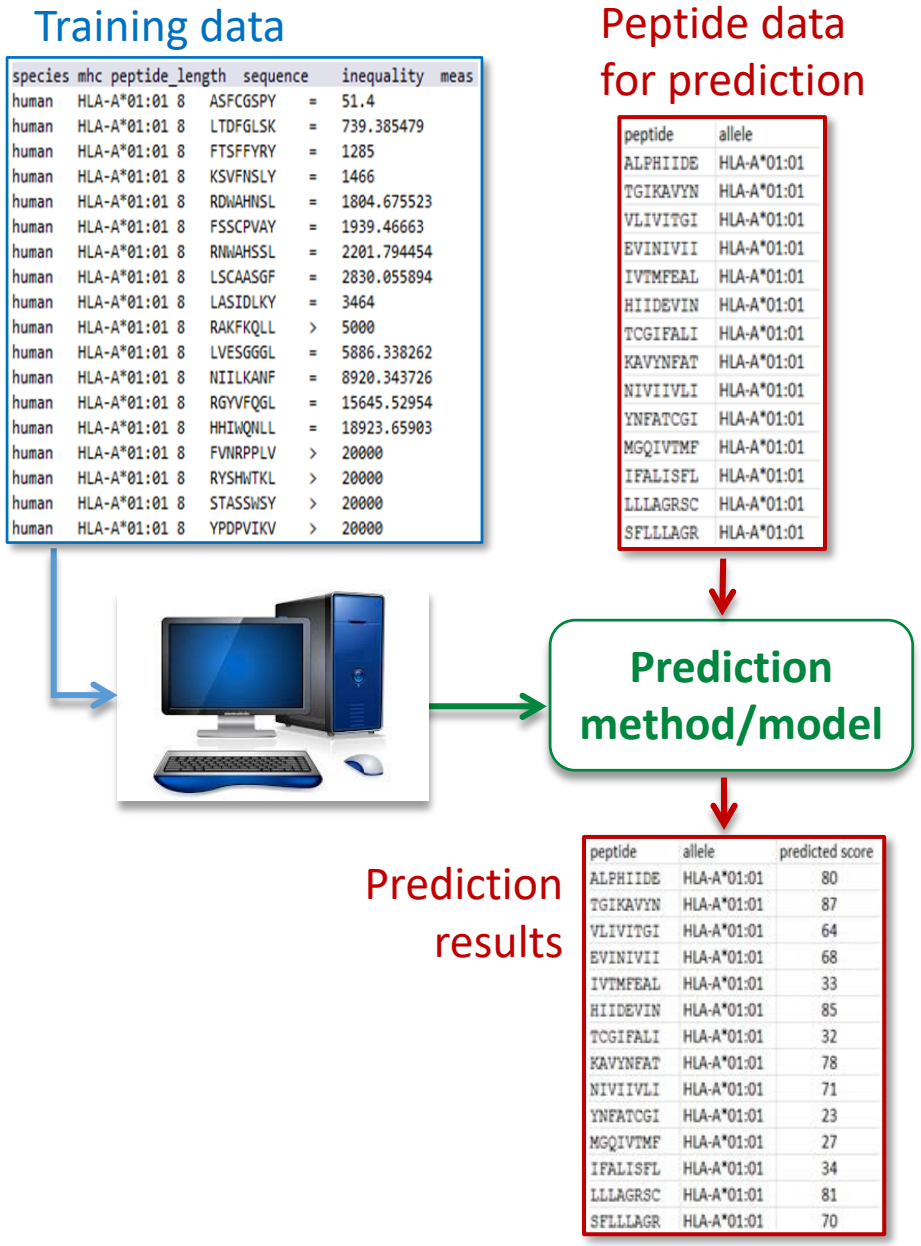
Group peptides based on sequence identity
- Cluster 2.0

Infer restriction in HLA typed subjects
- RATE

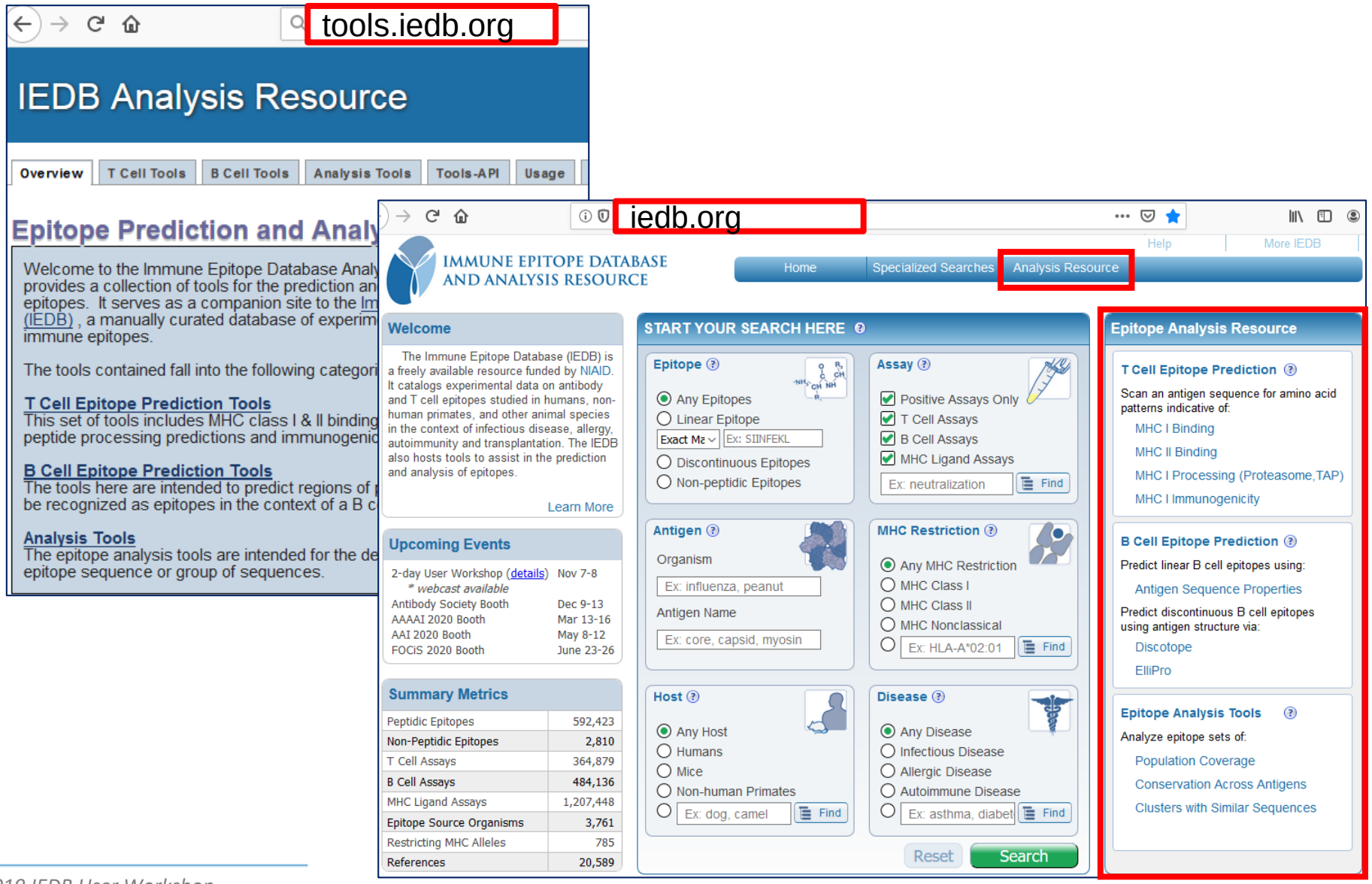
Aggregate heterogeneous immune response
- Immunome-browser

Development of prediction tools

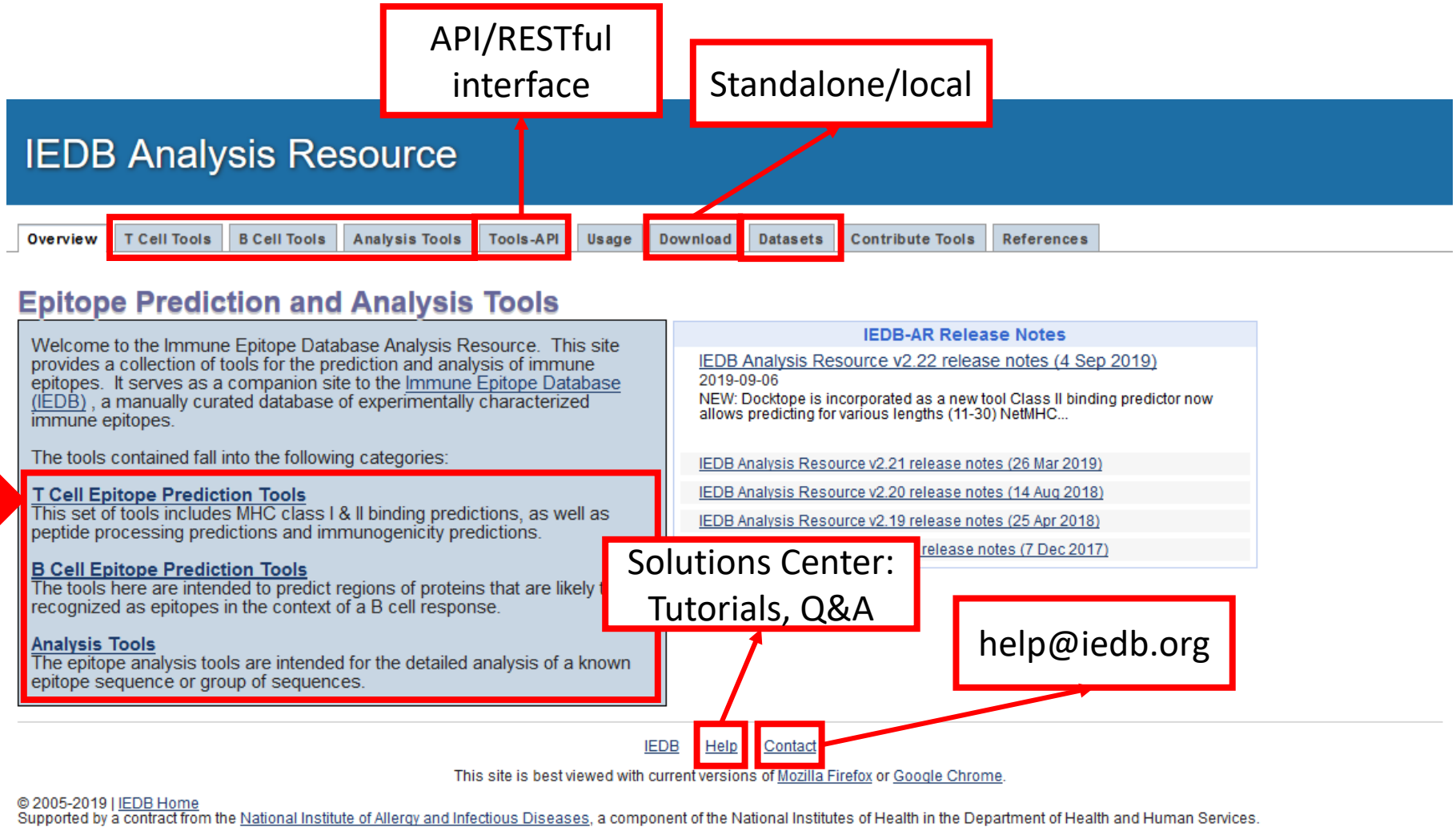
- Epitope prediction tools
  - Vaccine, cancer neoantigen studies
- Based on machine learning techniques
  - Experimentally derived data
  - E.g. MHC binding data, Eluted ligand data, T cell assay data



Accessing the Analysis Resource

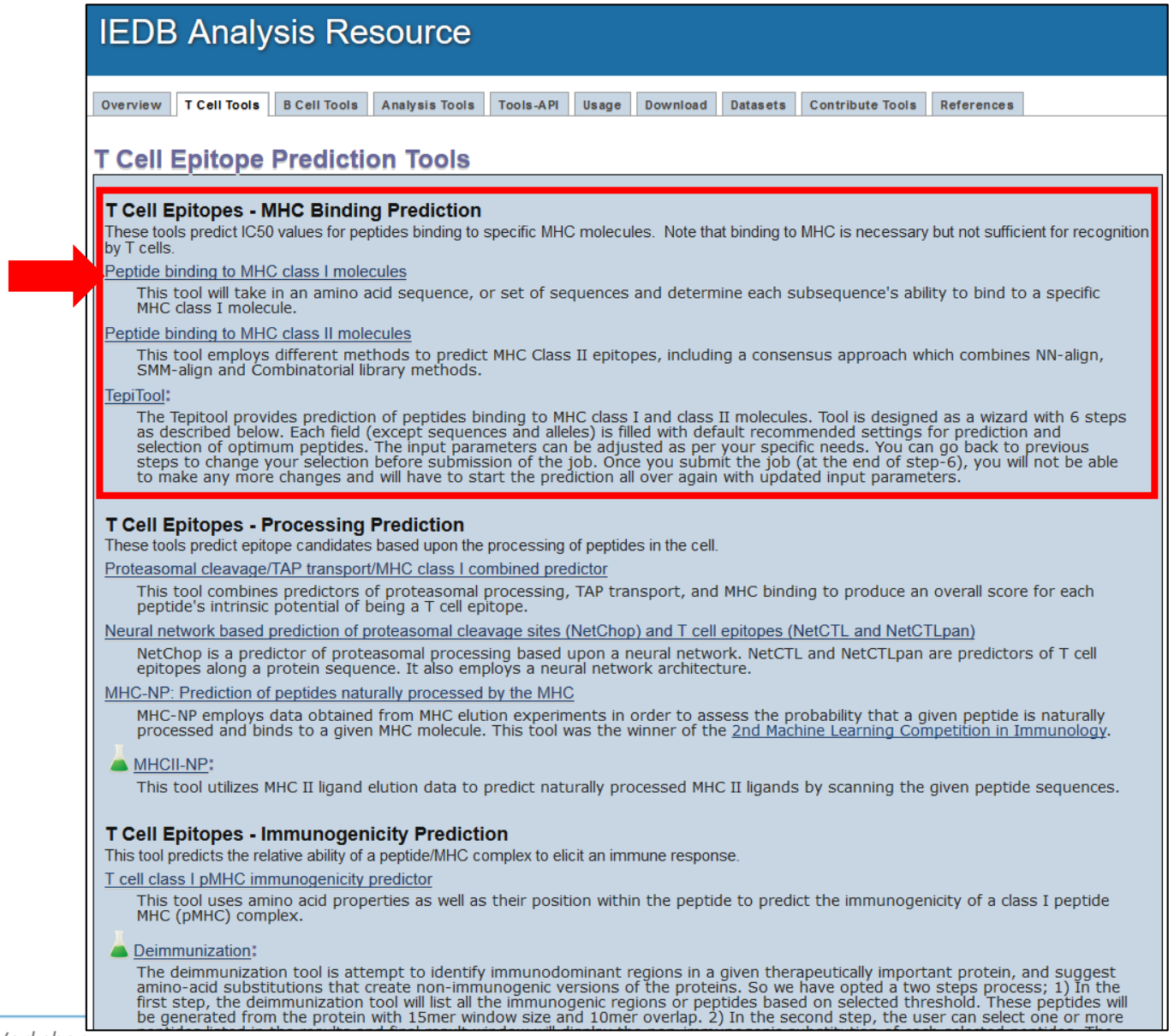


Available Resources



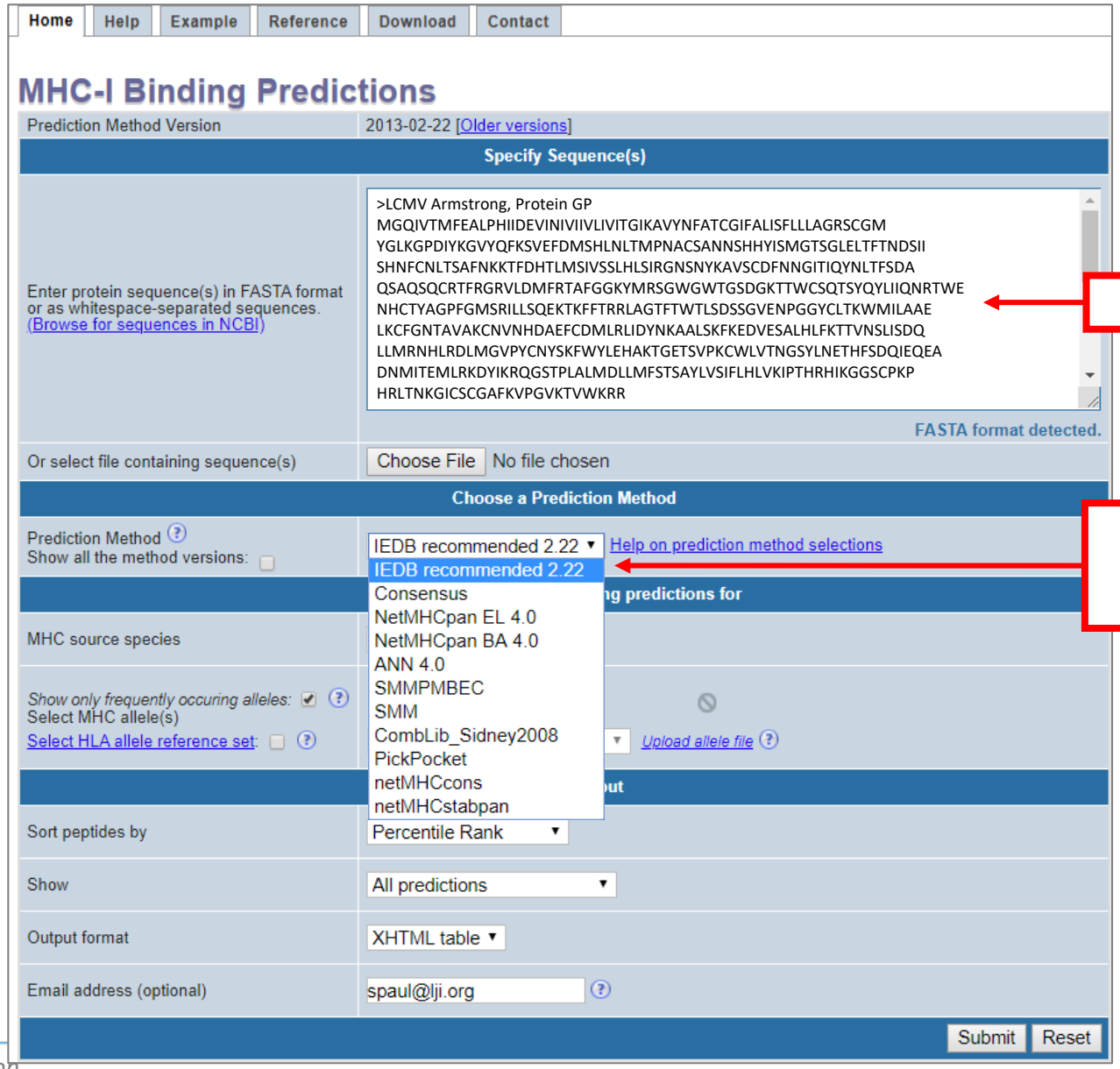
T Cell Tools

tools.iedb.org/main/tcell/



MHC I binding prediction

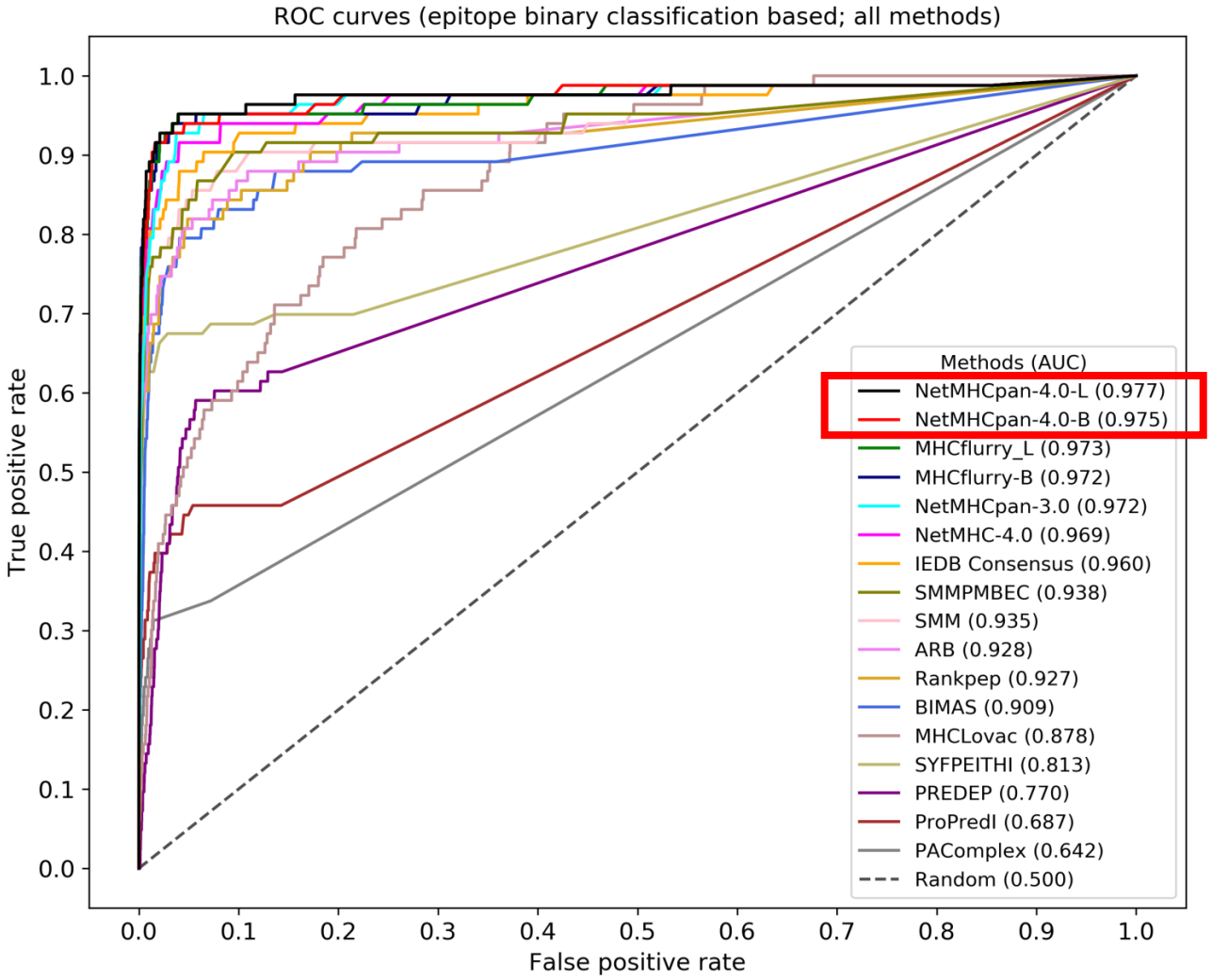
tools.iedb.org/mhci/



Sequence

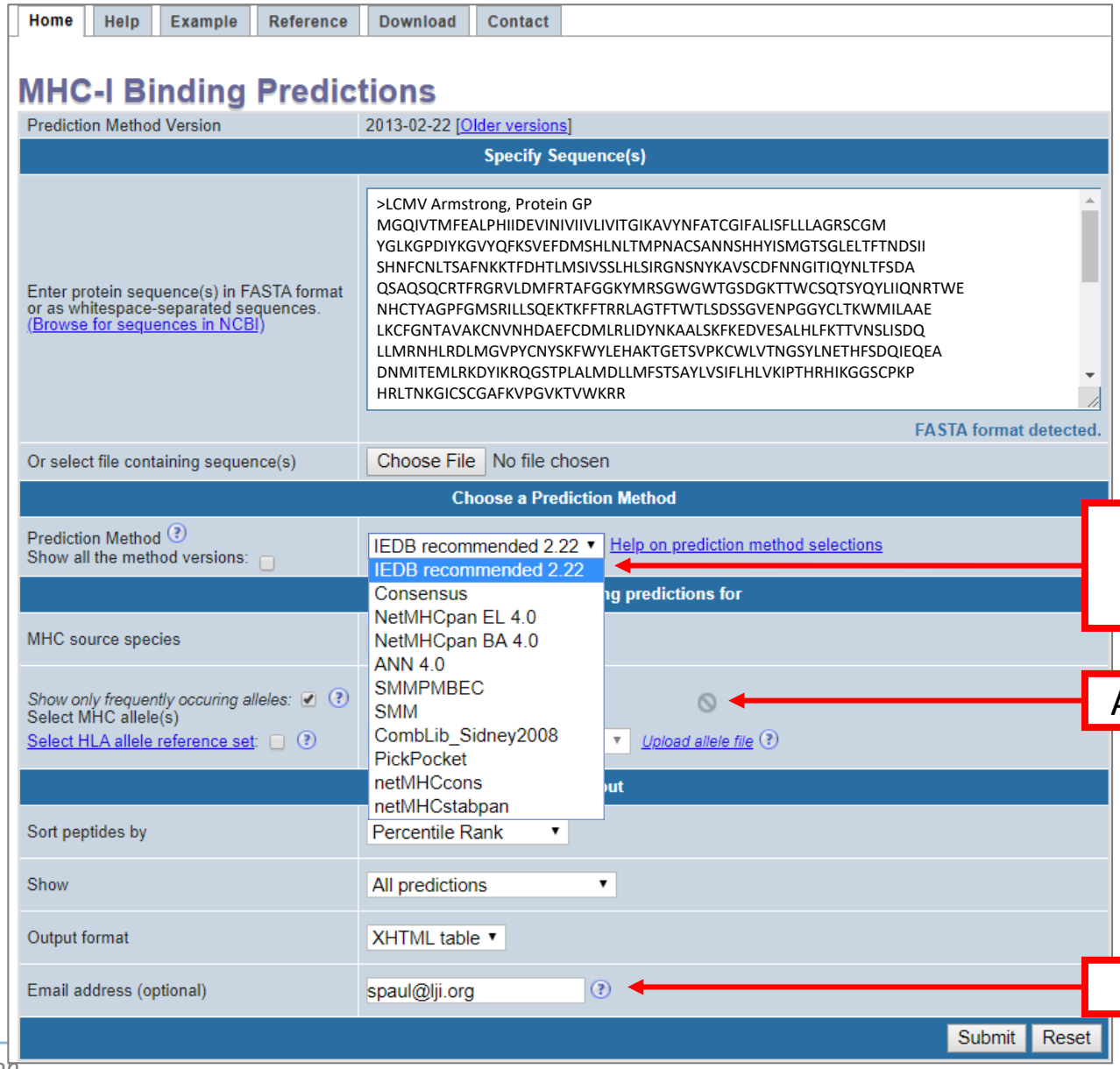
Prediction method

MHC I binding prediction methods - benchmarking



MHC I binding prediction

tools.iedb.org/mhci/



Prediction method

Allele & length

Email ID

MHC I binding prediction - result

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 LLAGRSCGMYGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHY ISMGTSGLELTFTNDSIISHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIR GNSNYKAVSCDFNNGITIQYNLTFSDAQSAQSQCRTFRGRVLDMFRTAFG GKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWENHCTYAGPFGMSRI LLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAELKCFG NTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSL ISDQLLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYL NETHFSDQIEQEADNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVS IFLHLVKIPTHRHIKGGSCPKPHRLTNKGICSCGAFKVPGVKTVWKRR
2	LCMV Armstrong, Protein NP	MSLSKEVKSFQWTQALRRELQSFTSDVKAAVIKDATNLLNGLDFSEVSNV QRIMRKEKRDDKDLQRLRSLNQTVHSLVDLKSTSKKNVLKVGRLSAEELM SLAADLEKLKAKIMRSERPQASGVYMGNLTTQQLDQRSQILQIVGMRKPQ QGASGVVRVWDVKDSSLLNNQFGTMPSLTMACMAKQSQTPLNDVVQALTD LGLLYTVKYPNLNDLERLKDKHPVLGVITEQQSSINISGYNFSLGAAVKA GAALLDGGNMLESILIKPSNSEDLLKAVLGAKRKLNMFVSDQVGDRNPYE NILYKVCLSGEGWPYIACRTSIVGRAWENTTIDLTSEKPAVNSPRPAPGA AGPPQVGLSYSQTMLLKDLMGGIDPNAPTWIDIEGRFNDPVEIAIFQPQN GQFIHFYREPVDQKQFKQDSKYSHGMDLADLFNAQPGLTSSVIGALPQGM VLSCQGSDDIRKLLDSQNRKDIKLIDVEMTREASREYEDKVWDKYGWLCK MHTGIVRDKKKKEITPHCALMDCIIFESASKARLPDLKTVHNILPHDLIF RGPNVVTL

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

Download result

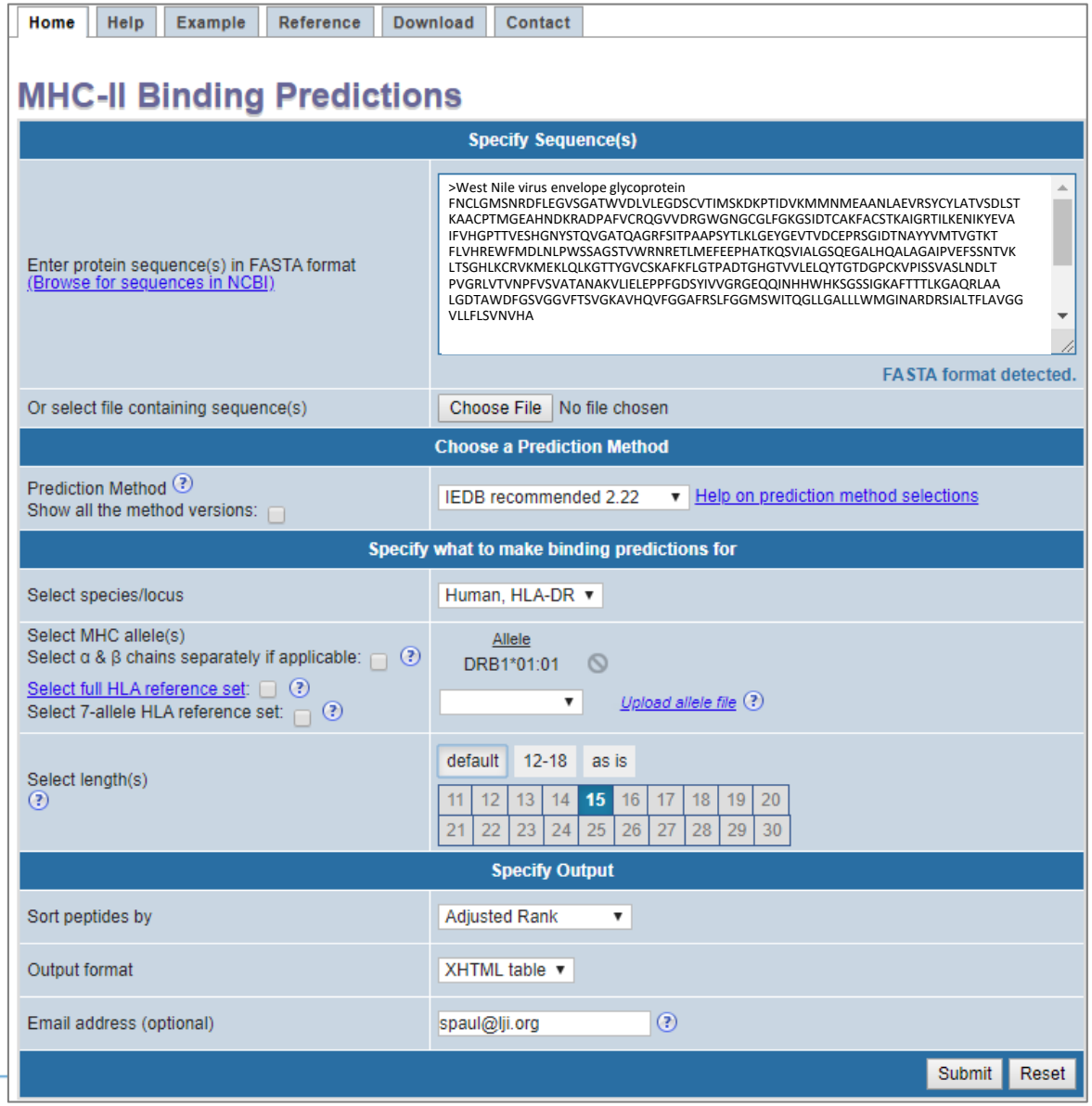
Citations

Check to expand the result:

Allele	#	Start	End	Length	Peptide	Method used	Percentile Rank
HLA-A*02:01	1	137	145	9	TLMSIVSSL	Consensus (ann/comblib_sidney2008/smm)	0.4
HLA-A*02:01	1	447	455	9	YLVSIFLHL	Consensus (ann/comblib_sidney2008/smm)	0.4
HLA-A*02:01	1	6	14	9	TMFEALPHI	Consensus (ann/comblib_sidney2008/smm)	0.5
HLA-A*02:01	1	45	53	9	ALISFLLLA	Consensus (ann/comblib_sidney2008/smm)	0.5
HLA-A*02:01	1	440	448	9	LMFSTSAYL	Consensus (ann/comblib_sidney2008/smm)	0.5
HLA-A*02:01	1	435	443	9	ALMDLLMFS	Consensus (ann/comblib_sidney2008/smm)	0.7
HLA-A*02:01	1	10	18	9	ALPHIIDEV	Consensus (ann/comblib_sidney2008/smm)	0.8
HLA-A*02:01	1	14	22	9	IIDEVINIV	Consensus (ann/comblib_sidney2008/smm)	1.4
HLA-A*02:01	2	403	411	9	FIHFYREPV	Consensus (ann/comblib_sidney2008/smm)	1.5
HLA-A*02:01	1	448	456	9	LVSIFLHLV	Consensus (ann/comblib_sidney2008/smm)	1.6
HLA-A*02:01	2	437	445	9	GLTSSVIGA	Consensus (ann/comblib_sidney2008/smm)	1.6
HLA-A*02:01	2	520	528	9	LMDCIIFES	Consensus (ann/comblib_sidney2008/smm)	1.6
HLA-A*02:01	1	42	50	9	GIFALISFL	Consensus (ann/comblib_sidney2008/smm)	1.8
HLA-A*02:01	2	253	261	9	ALLDGGNML	Consensus (ann/comblib_sidney2008/smm)	1.8

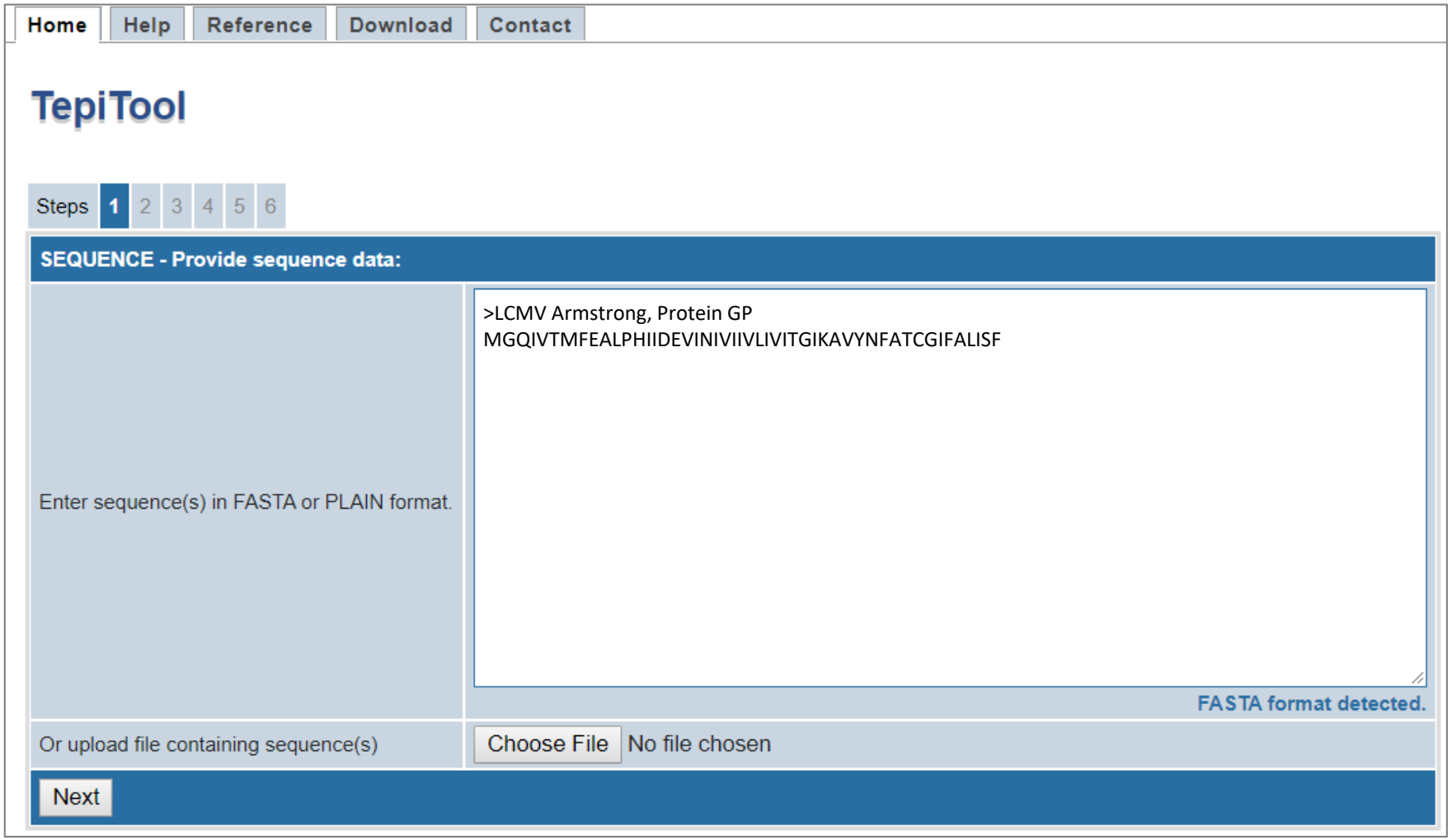
MHC II binding prediction

tools.iedb.org/mhcii/



TepiTool – MHC I and II binding prediction

tools.iedb.org/tepitool/



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)	sample
Email (optional - will notify when job is finished)	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.)

TepiTool – MHC I and II binding prediction

tools.iedb.org/tepitool/

TepiTool

Prediction results - concise (Download table):

Seq #	Peptide start	Peptide end	Peptide sequence	Consensus percentile rank	Allele
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:

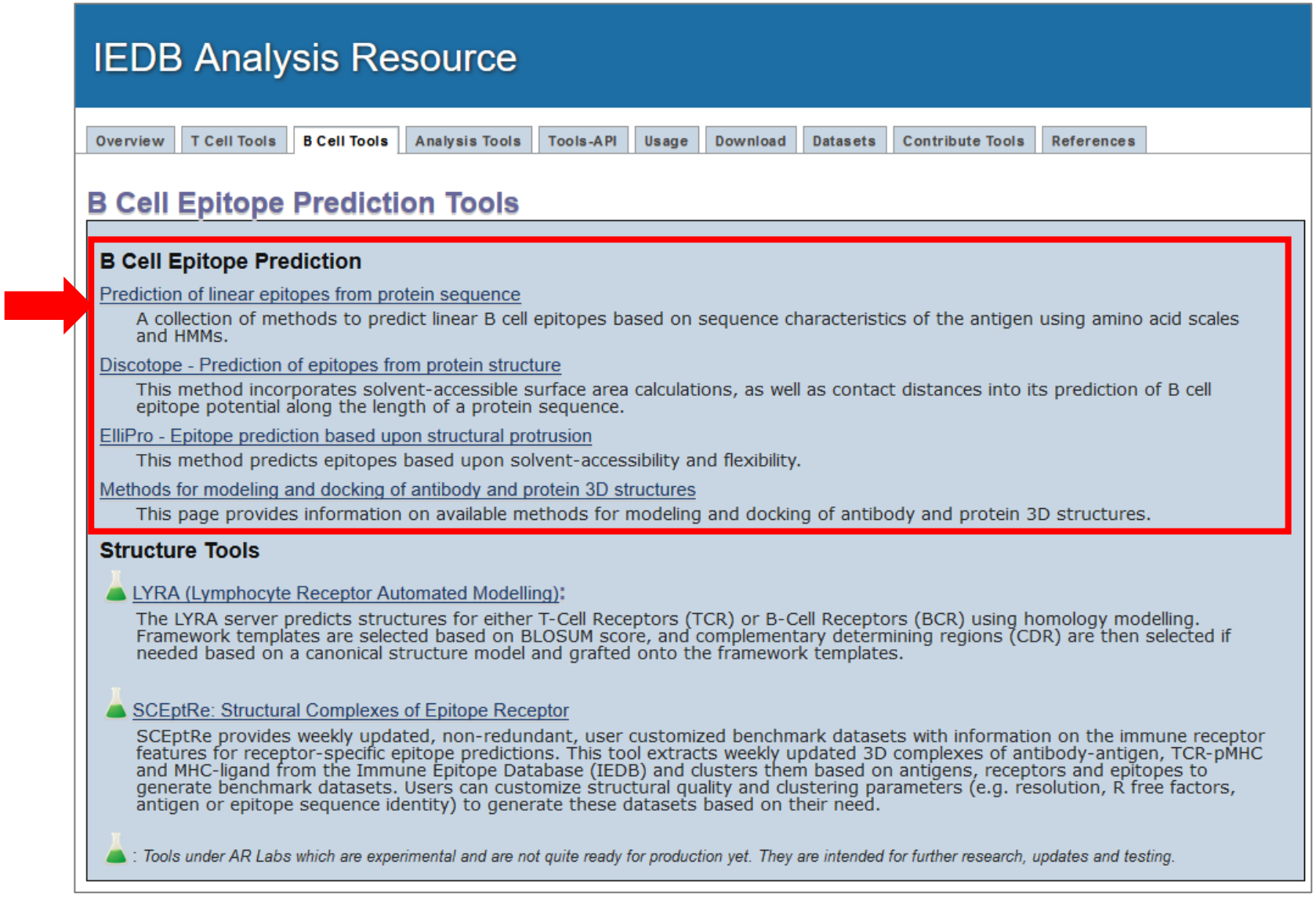
The MHC II binding predictions were done with IEDB analysis resource (TepiTool [1]) using NetMHCIIpan method [2,3].
1. Paul, S., Sidney, J., Sette, A., and Peters, B. 2016. TepiTool: A pipeline for computational prediction of T cell epitope candidates.*Curr. Protoc. Immunol.* 114
2. Karosiene E1, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. 2013. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method inclu and 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 4(7): e1000107.

For complete list of references please click here: References

Input sequences:

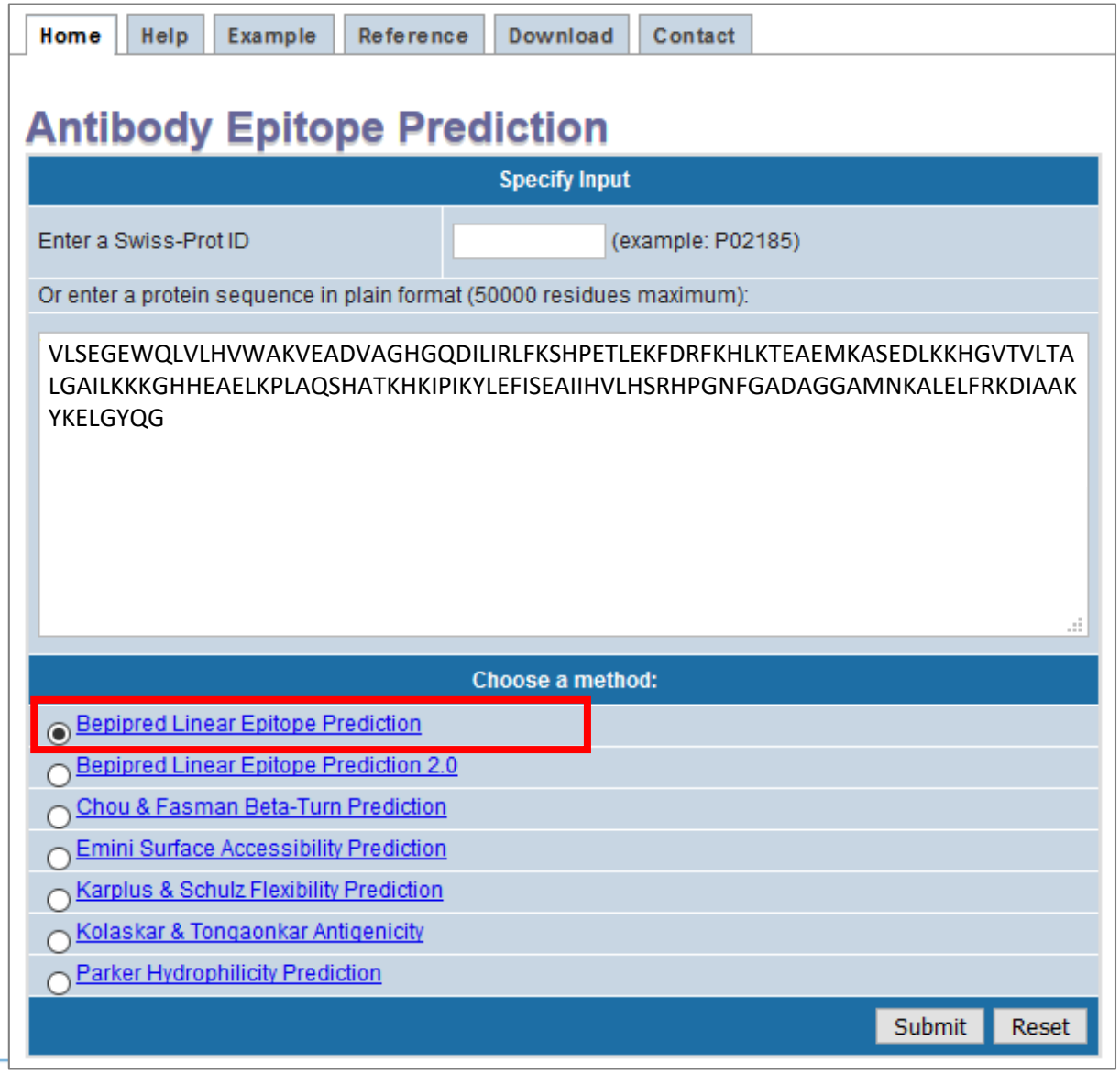
Seq #	Seq title	Sequence
1	LCMV Armstrong, Protein GP	MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISF

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/



B Cell epitope prediction – sequence based

tools.iedb.org/bcell/

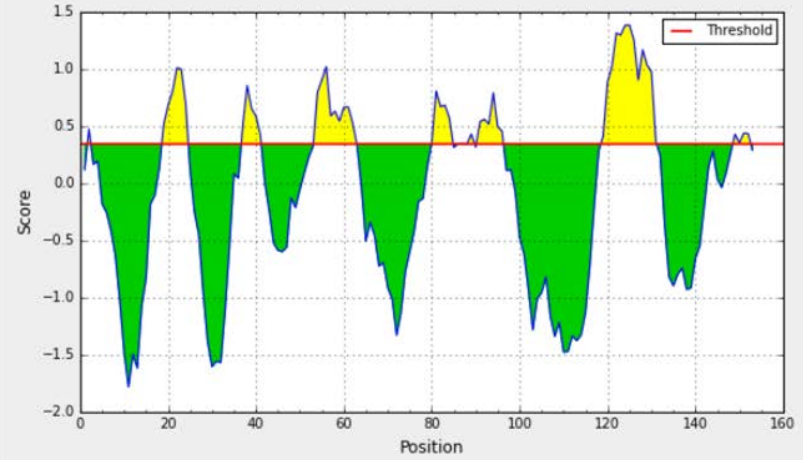
Home | Help | Example | Reference | Download | Contact

Bepipred Linear Epitope Prediction Results

Input Sequences

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

Center position: 4 Threshold: 0.350 Recalculate



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:

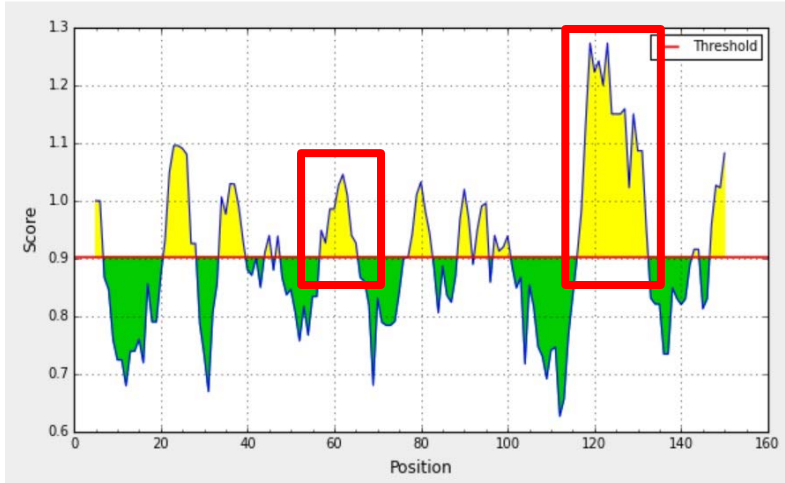
Position	Residue	Score	Assignment
1	V	0.121	.
2	L	0.476	E
3	S	0.168	.

Chou & Fasman Beta-Turn Prediction

Input Sequences

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

Center position: 4 Window size: 7 Threshold: 0.903 Recalculate

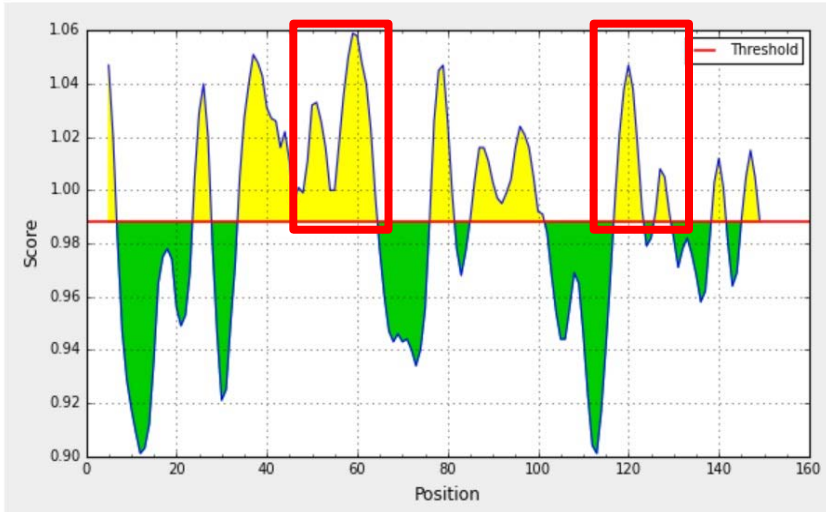


Karplus & Schulz Flexibility Prediction

Input Sequences

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

Center position: 4 Window size: 7 Threshold: 0.988 Recalculate

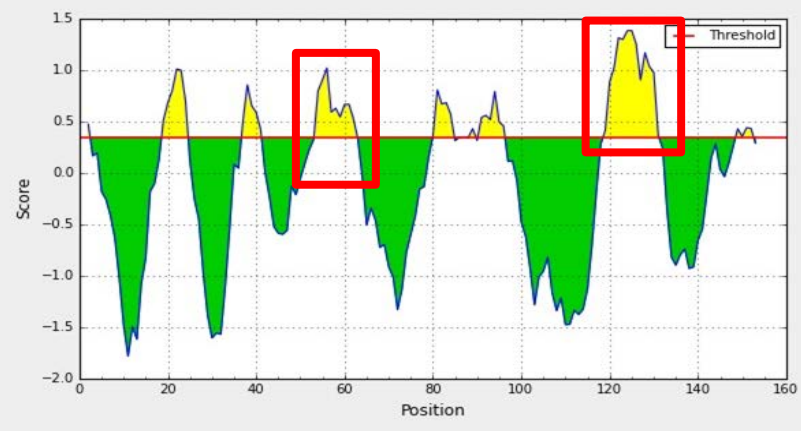


Bepipred Linear Epitope Prediction

Input Sequences

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

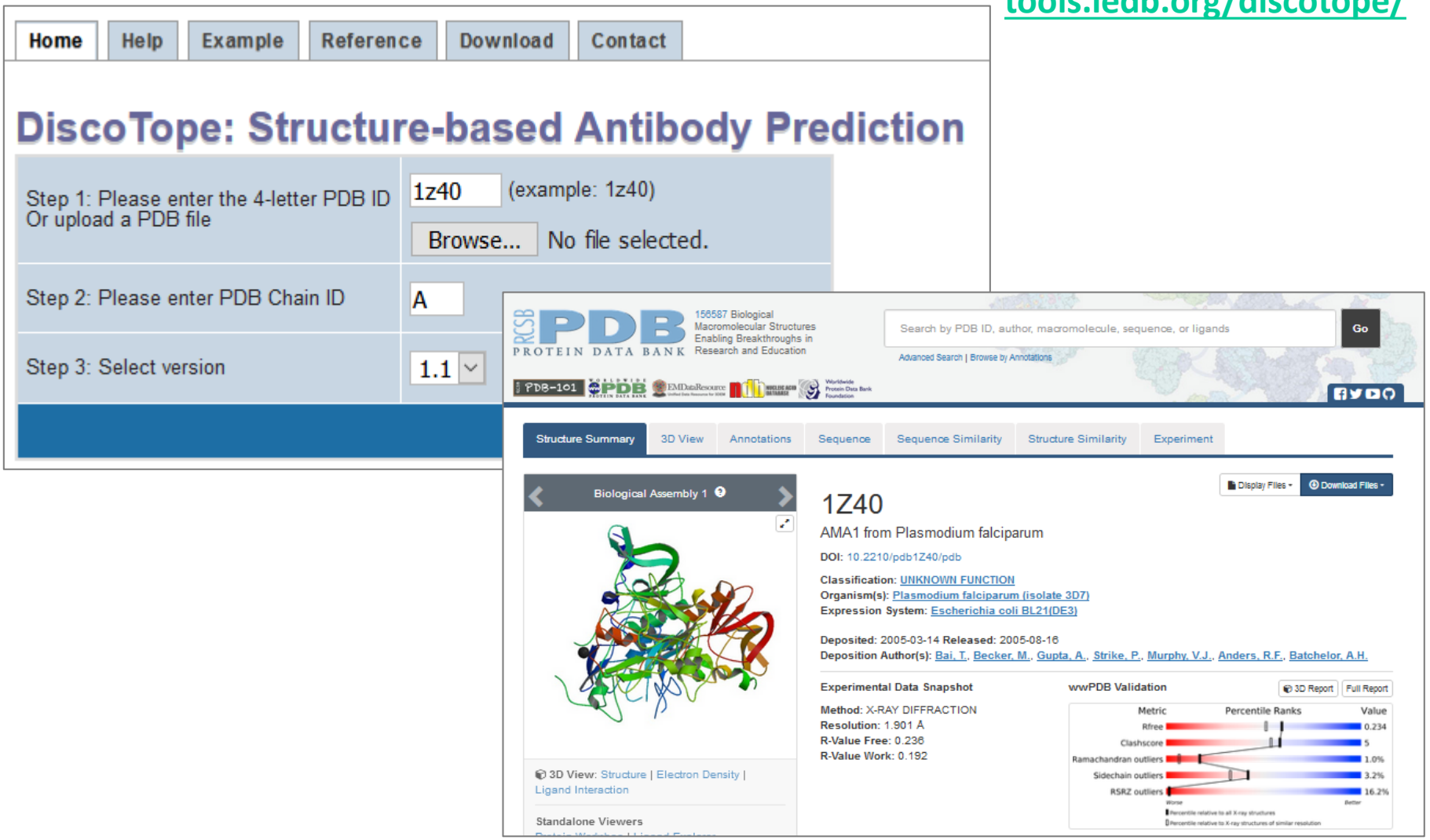
Center position: 4 Window size: 7 Threshold: 0.35 Recalculate



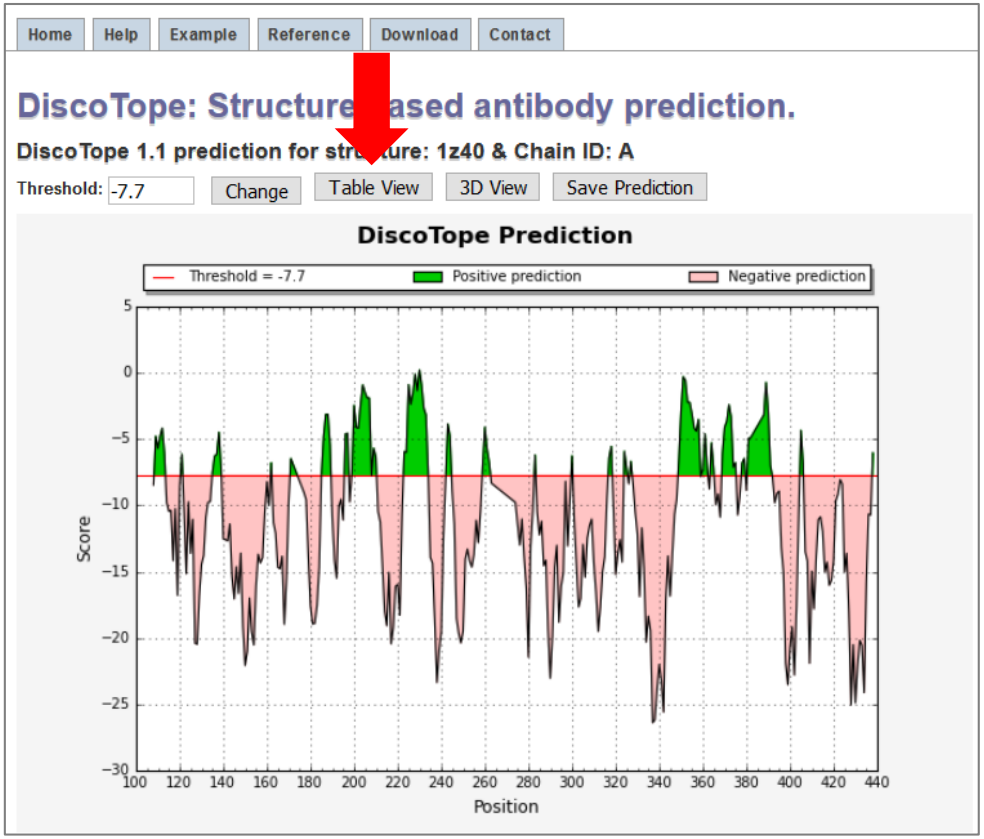
Using a consensus of different methods may be better.

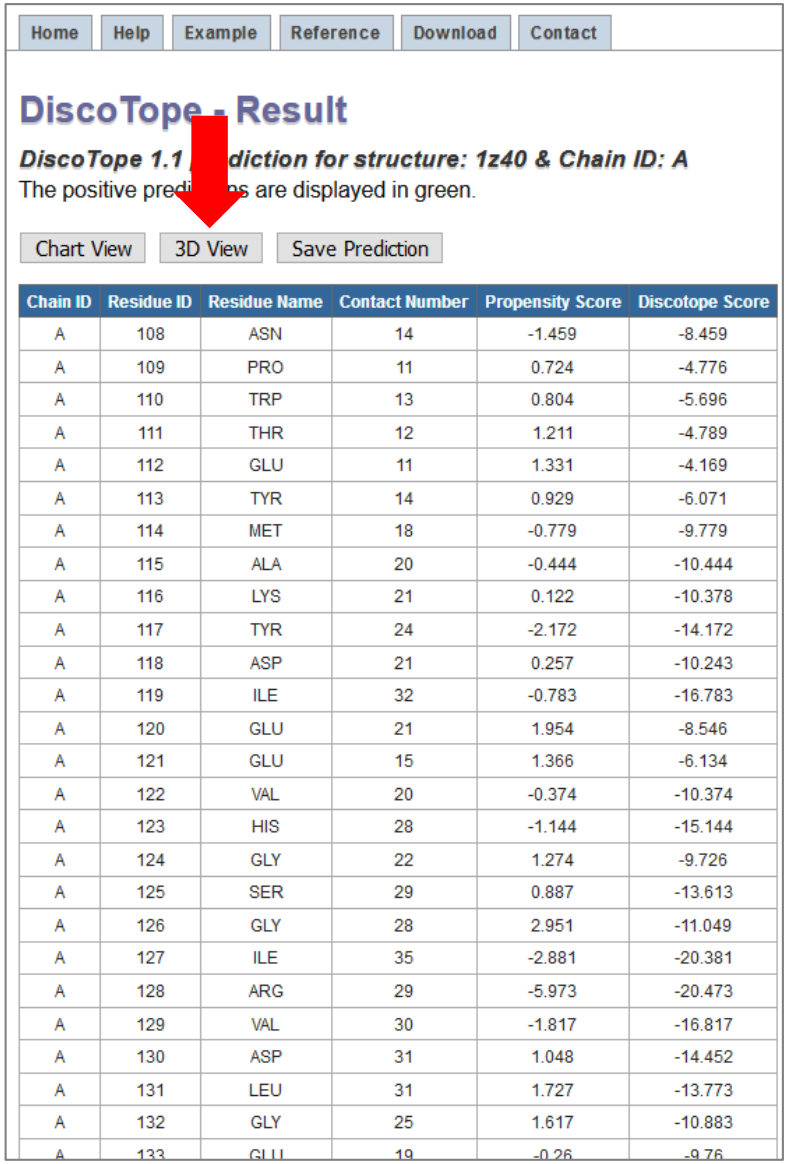
B cell epitope prediction – structure based

tools.iedb.org/discotope/



B cell epitope prediction – structure based





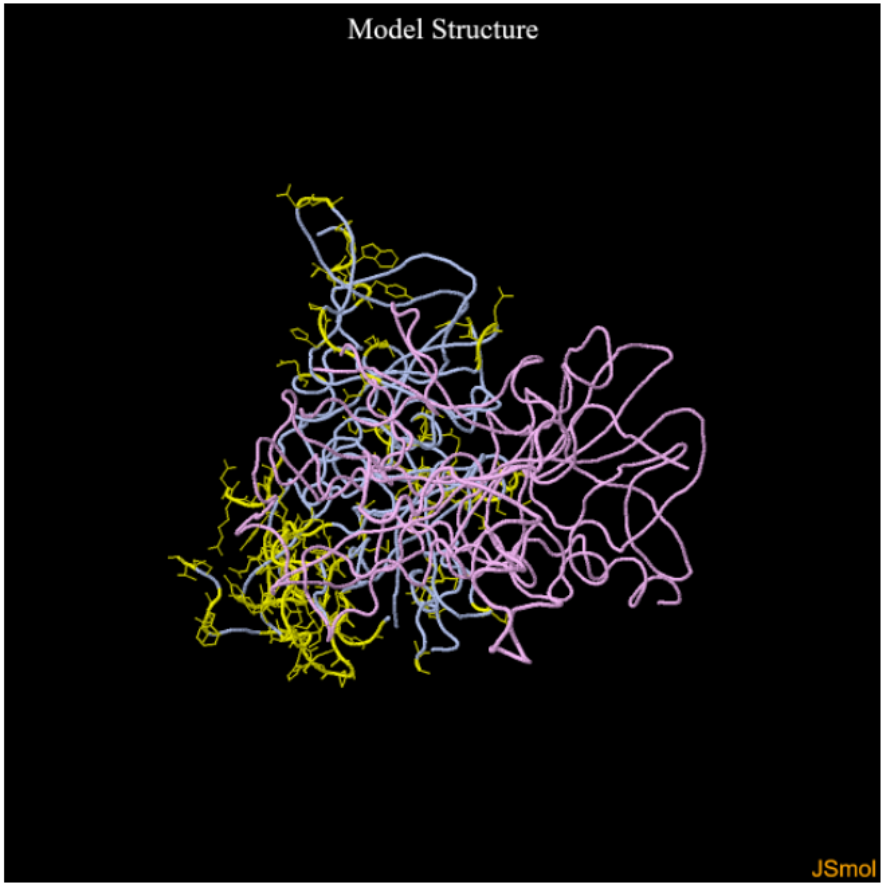
Chain ID	Residue ID	Residue Name	Contact Number	Propensity Score	Discotope Score
A	108	ASN	14	-1.459	-8.459
A	109	PRO	11	0.724	-4.776
A	110	TRP	13	0.804	-5.696
A	111	THR	12	1.211	-4.789
A	112	GLU	11	1.331	-4.169
A	113	TYR	14	0.929	-6.071
A	114	MET	18	-0.779	-9.779
A	115	ALA	20	-0.444	-10.444
A	116	LYS	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

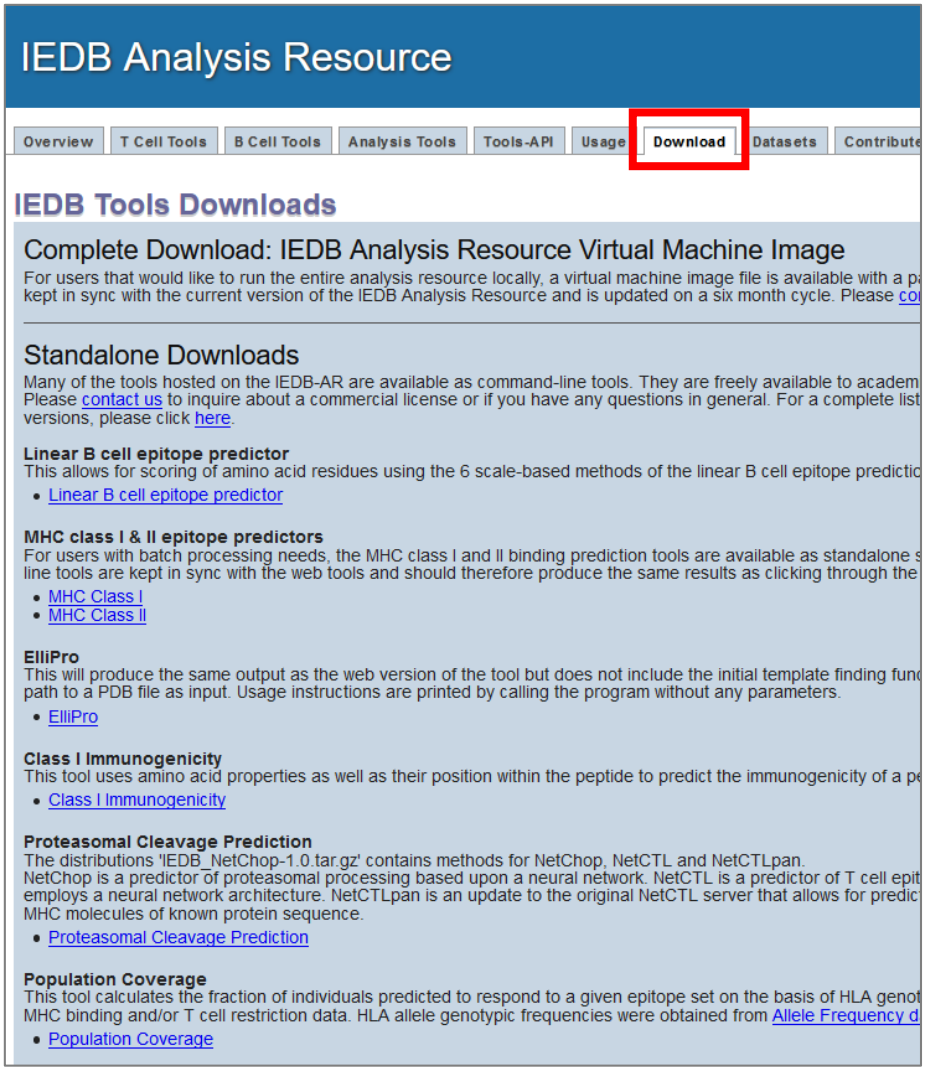
Chart View | Table View | Save Prediction



Chain ID	Residue ID	Residue Name	Contact Number	Propensity Score	Discotope 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/



- 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

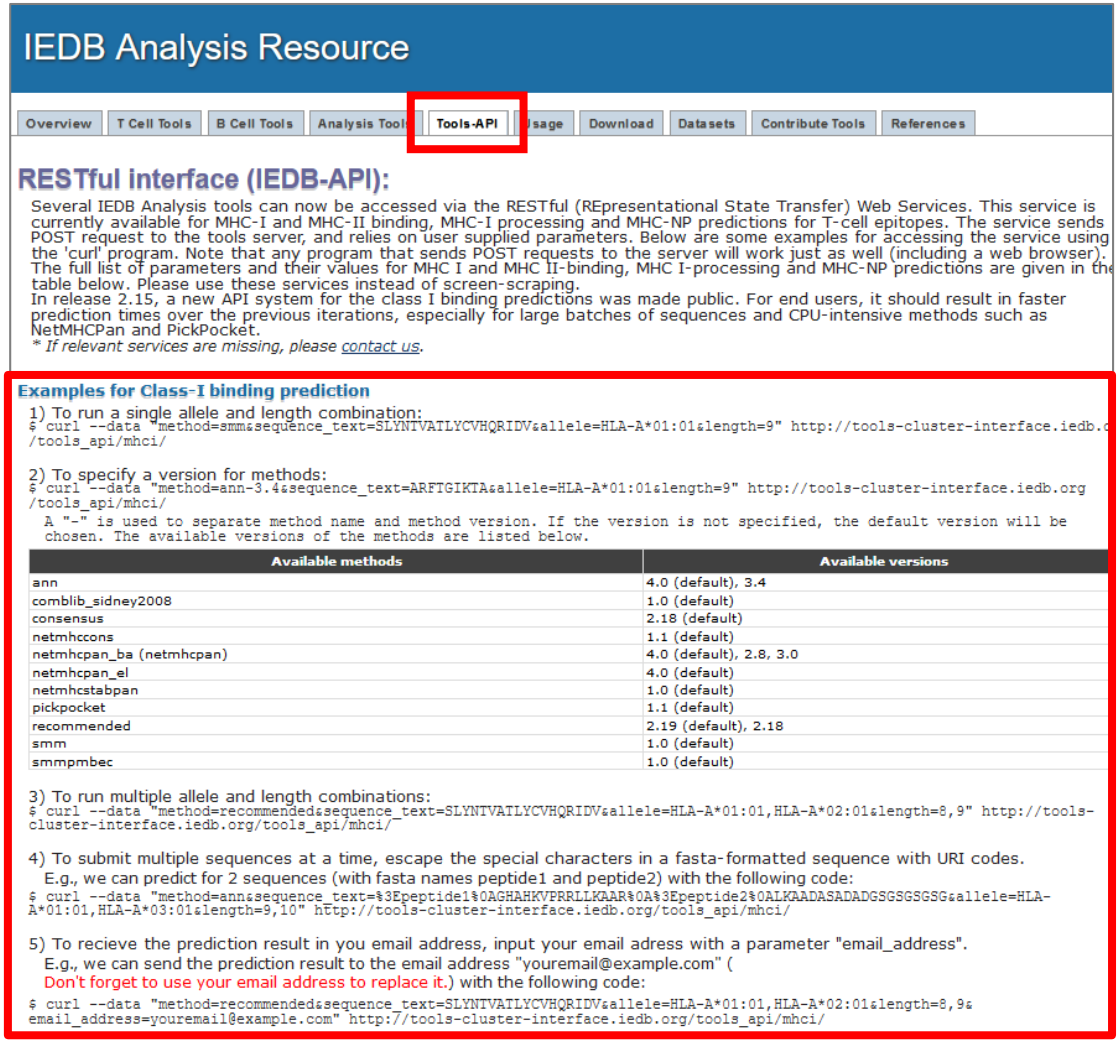
```
spaul@ubuntu:~/tools/mhc_i$ ./src/predict_binding.py netmhcpan "HLA-A*02:01" 9 test_sequence.fasta
allele  seq_num start   end     length  peptide ic50    rank
HLA-A*02:01     1       6       14      9       TMFEALPHI       4.3     0.03
HLA-A*02:01     1       10      18      9       ALPHIIDEV       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       EVINIVIIV       3263.2  7.2
HLA-A*02:01     1       18      26      9       VINIVIIVL       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       NIVIIVLIV       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       INIVIIVLI       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_i$ ./src/predict_binding.py netmhcpan "HLA-A*02:01" 9 test_s
spaul@ubuntu:~/tools/mhc_i$
```

prediction_results.txt

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

API version (RESTful interface)

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



IEDB Analysis Resource

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

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=SLYNTVATLYCVHQRIDV&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)
netmhcpan_ba (netmhcpan)	4.0 (default), 2.8, 3.0
netmhcpan_el	4.0 (default)
netmhcstabpan	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=SLYNTVATLYCVHQRIDV&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%0AGHAHKVPRRLLKAAR%0A%3Epeptide2%0ALKAADASADADGSGSGSGSG&allele=HLA-A*01:01,HLA-A*03:01&length=9,10" http://tools-cluster-interface.iedb.org/tools_api/mhci/`

5) To recieve the prediction result in you email address, input your email adress 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=SLYNTVATLYCVHQRIDV&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	QCRTFRGRVLDMF	HLA-B*53:01
11	GTSGLELTFTND	HLA-A*11:01
12	NLTSAFNKK	HLA-A*23:01
13	CDMLRLIDYNKAA	HLA-B*53:01
14	YIKRQGSTPL	HLA-A*26:01
15	YMRSGWGWTG	HLA-A*23:01
16	LVTNGSYLNETHF	HLA-B*58:01
17	TKFFTRRL	HLA-B*57:01
18	NVNHDAEFCDMLRL	HLA-B*08:01
19	HIKGGSCPKPH	HLA-A*30:01
20	DGKTTWCSQTS	HLA-A*32:01
21	HFSDQIEQEADNM	HLA-A*32:01
22	FSDQIEQEADNMI	HLA-B*57:01
23	CNYSKFWY	HLA-B*58:01
24	MSHLNLTMPNAC	HLA-A*02:01
25	SGVENPGGYC	HLA-B*44:03
26	VIIVLIVITGIK	H-2-Kb
27	FRGRVLDMFR	HLA-B*51:01
28	HIKGGSCPKPHR	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	AFGGKYMRSGWGWT	HLA-B*07:02
35	SRILLSQEKTKFFT	HLA-A*31:01
36	SALHLFKTTVNSLI	H-2-Kb
37	RKDYIKRQGSTP	HLA-B*58:01
38	FKSVEFDMSHLNLT	HLA-B*58:01

Sheet1 Sheet2 Sheet3

api_predictor.py

```
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,
        write_line = peptide + '\t' + allele + '\t' + str(
        file_to_write.write(write_line)
output_file.close()
```

prediction_results.txt

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