



# Analysis Tools

[tools.iedb.org](https://tools.iedb.org)

Presented by: Alessandro Sette, IEDB Principal Investigator

# Analysis tools with broad applications

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

## IEDB Analysis Resource

[Overview](#) [T Cell Tools](#) [B Cell Tools](#) **[Analysis Tools](#)** [Tools-API](#) [Usage](#) [Download](#) [Datasets](#) [Contribute Tools](#) [References](#)

### Analysis Tools

#### Analysis Tools

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

##### [Population Coverage](#)

This tool calculates the fraction of individuals predicted to respond to a given set of epitopes with known MHC restrictions. This calculation is made on the basis of HLA genotypic frequencies assuming non-linkage disequilibrium between HLA loci.

##### [Epitope Conservancy Analysis](#)

This tool calculates the degree of conservancy of an epitope within a given protein sequence set at different degrees of sequence identity. The degree of conservation is defined as the fraction of protein sequences containing the epitope at a given identity level.

##### [Epitope Cluster Analysis](#)

This tool groups epitopes into clusters based on sequence identity. A cluster is defined as a group of sequences which have a sequence similarity greater than the minimum sequence identity threshold specified.

##### [Computational Methods for Mapping Mimotopes to Protein Antigens](#)

This page provides information on available methods for mimotope mapping, how to search the IEDB for mimotopes, and an example of a mimotope dataset and the results of its mapping, using the available web servers hosted outside the IEDB.

##### [RATE \(Restrictor Analysis Tool for Epitopes\)](#)

The RATE is an automated method that can infer HLA restriction for a set of given epitopes from large datasets of T cell responses in HLA typed subjects. The tool takes two data files, one containing the alleles expressed by the subjects and the other containing the response of the peptides in the subjects. The tool calculates the odds ratio and estimates its significance using Fisher's exact test. It also calculates a parameter called relative frequency similar to odds ratio. The tool was developed with a focus on class II alleles but can also be applied to class I alleles.


This tool groups epitopes into clusters based on sequence identity. A cluster is defined as a group of sequences which have a sequence similarity greater than the minimum sequence identity threshold specified. User can also select the minimum and maximum length of peptide and also one of the three approaches for clustering of peptides.

##### [ImmunomeBrowser](#)

The tool is helpful to aggregate and visualize immune reactivity from epitope data in different assays/donors in given reference proteins using user-defined identity thresholds. The tool also accepts predicted epitopes.

##### [PepSySco](#)

Given a set of peptide sequences, Peptide Synthesis Score (PepSySco) predicts the likelihood that they can be synthesized successfully.

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



# Population Coverage

<http://tools.iedb.org/population/>

- Calculates the fraction of individuals projected to bind and/or respond to a given set of epitopes with defined reactivity
- Based on
  - Epitope known HLA binding/restrictions
  - HLA genotypic frequencies
- HLA genotypic frequencies vary in different ethnicities
  - <http://allelefrequencies.net>

# Population Coverage – example

<http://tools.iedb.org/population/>

For a set of 11 MHC class II restricted epitopes with promiscuous HLA binding, what is the population coverage in different North African populations?

Home Help Example Reference Download Contact

Please Note: The genotypic frequency is currently not working as expected and the IEDB team is looking into this.

## Population Coverage

Number of epitope(s):

Query by:  \*

Select area(s) and/or population(s):	Select calculation option(s):
<ul style="list-style-type: none"><li>North Africa</li><li>Algeria</li><li>Algeria Arab</li><li>Ethiopia</li><li>Ethiopia Black</li><li>Sudan</li><li>Sudan Arab</li><li>Sudan Black</li><li>Sudan Mixed</li><li>Morocco</li><li>Morocco Arab</li><li>Morocco Caucasoid</li><li>Tunisia</li></ul>	<input type="checkbox"/> Class I separate <input checked="" type="checkbox"/> Class II separate <input type="checkbox"/> Class I and II combined

Add user populations(s)  No file selected.

Enter epitope / MHC restriction data in the form below or select a file  No file selected.

\* Population datasets generously provided by Derek Middleton at [The Allele Frequency Net Database](#)

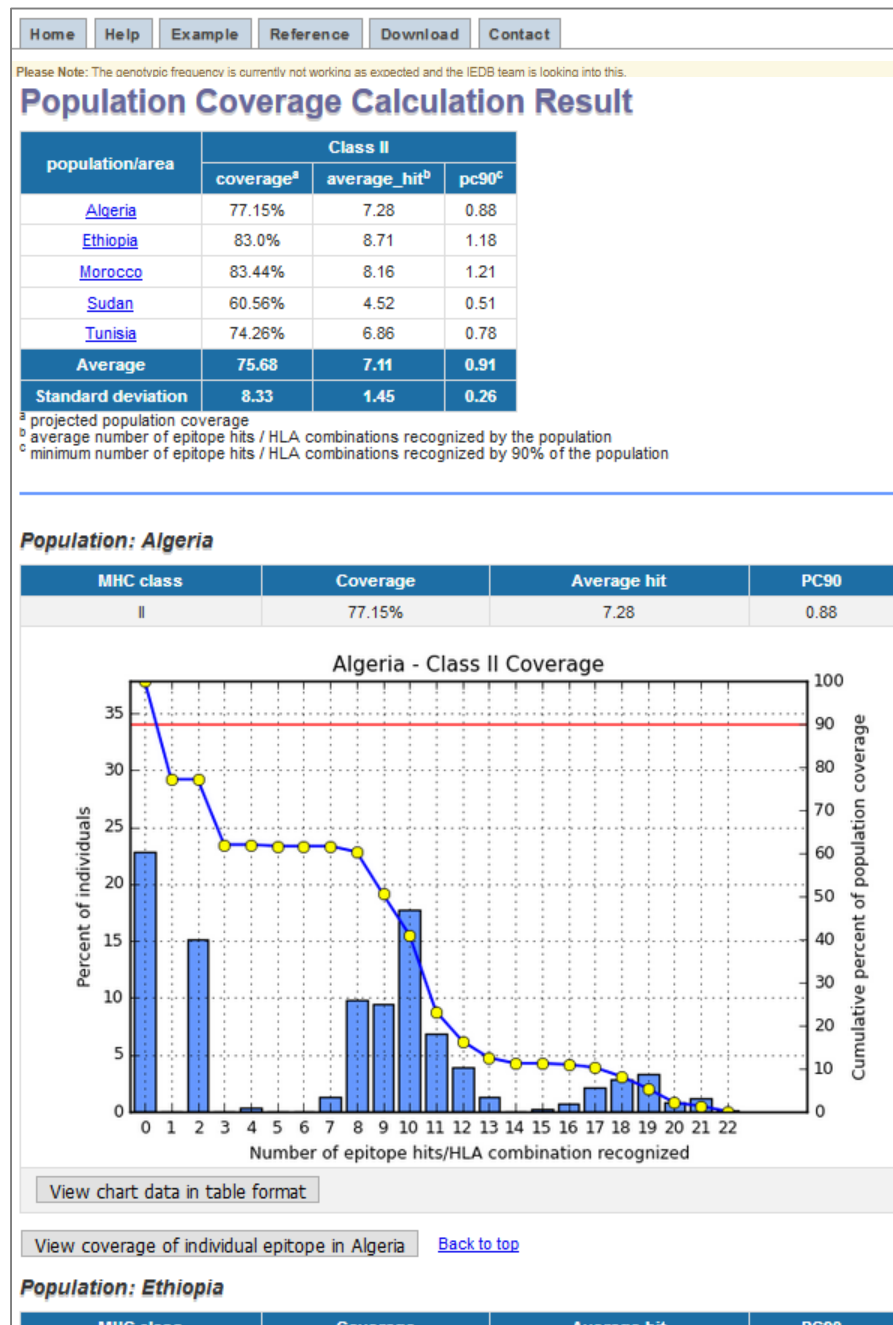
Epitope	MHC Restricted Allele(s)
<input type="text" value="Gag 171"/>	HLA-DRB1*01:01,HLA-DRB1*15:01,HLA-DRB1*04:01,HLA-DRB1*0 <input type="button" value="Browse..."/>
<input type="text" value="Gag 294"/>	HLA-DRB1*01:01,HLA-DRB1*15:01,HLA-DRB1*04:05,HLA-DRB1*1 <input type="button" value="Browse..."/>
<input type="text" value="Gag 298"/>	HLA-DRB1*01:01,HLA-DRB1*15:01,HLA-DRB1*03:01,HLA-DRB1*0 <input type="button" value="Browse..."/>
<input type="text" value="Pol 303"/>	HLA-DRB1*01:01,HLA-DRB1*15:01,HLA-DRB1*03:01,HLA-DRB1*0 <input type="button" value="Browse..."/>
<input type="text" value="Pol 335"/>	HLA-DRB1*01:01,HLA-DRB1*15:01,HLA-DRB1*04:05,HLA-DRB1*1 <input type="button" value="Browse..."/>
<input type="text" value="Pol 596"/>	HLA-DRB1*01:01,HLA-DRB1*15:01,HLA-DRB1*04:01,HLA-DRB1*0 <input type="button" value="Browse..."/>

# Population Coverage – example

<http://tools.iedb.org/population/>

## Results:

- Summarized in table format
- Plotted per population



# Epitope Conservancy Analysis

- Calculates the degrees of conservancy of one or more epitopes, within a given set of protein sequences
- Adjustable sequence identity threshold
- “Degree of conservation” = the fraction of protein sequences containing the epitope at a given identity level

# Epitope Conservancy – example

<http://tools.iedb.org/conservancy/>

Home Help Example Reference Download Contact

## Epitope Conservancy Analysis

**Step 1. Epitope Sequence(s)**

Enter epitope sequence(s) in PLAIN or FASTA format: ?

```
>NP 1
MSASKEVKSFLWTQS
>NP 21
SGYCSNIKLVVVKDA
>NP 41
GLDFSEVSNVQRLMR
>NP 61
DGDLLKRLRDLNQAQVN
>NP 81
KSTQQKSVLRVGTLS
>NP 101
MSASKEVKSFLWTQS
```

Or upload epitope sequence(s) from a file:  No file selected.

**Step 2. Protein Sequence(s)**

Enter protein sequence(s) in PLAIN or FASTA format: ?

```
>58643 Lassa NP
MSASKEVRSFLWTQSLRRELSGYCSNIKLVVVKDAQALLHGLDFSEVSNVQRLMRKQKRDGDLKRLRDL
NQAVNNLVELKSTQQKSVLRVGTLSDDLLILAADLEKLSKVTTRTERPLSSGVYMGNLSSQQLDQRRAL
LNMIGMTGVSGGGKASDGIIVRVNDVKNAELLNQFGTTPSLTLACLTKQGQVLDNDAVQALTDLGLIYT
AKYPNSSDLDRLSQSHPILNMIDTKKSSLNISGYNFSLGAAVKAGACMLDGGNMLETIKVSPQTMGILK
SILKVKKSLGMFVSDTPGERNPYENILYKICLSGGGWPIASRTSIVGRAWENTVVDLEQDNKPKIGNG
GSNKSLSAGFAAGLTYSQLMTLKDFFKFNLI PNAKTWMDIEGRPEDPVEIALYQPSGCVVHFFREPTD
LKQFKQDAKYSHGIDVTDLFAAQFGLTSAVIEALPRNMVITCQGSSEDIRKLESQGRRDIKLIDITLSKA
DSRKFENAVWDQFKDLCHMHTGVVVEKKRGGKEEITPHCALMDCIMFDAVSSGGLDAKVLRVVLPDMV
FRISTPKVVV
>9294837 Lassa NP
MSASKEVRSFLWTQSLRRELSGYCSNIKLVVVKDAQALLHGLDFSEVSNVQRLMRKQKRDGDLKRLRDL
```

Or upload protein sequence(s) from a file:  No file selected.

**Step 3. Calculation option(s)**

Analysis type: ?

Epitope linear sequence conservancy  
 Epitope discontinuous sequence conservancy

Sequence identity threshold: ?

>=  %


Remove duplicated protein sequences? ?

Known epitopes

Protein sequences

# Epitope Conservancy – example

<http://tools.iedb.org/conservancy/>

Home	Help	Example	Reference	Download	Contact		
<b>Epitope Conservancy Analysis Result</b>							
Download result 							
Epitope #	Epitope name	Epitope sequence	Epitope length	Percent of protein sequence matches at identity <= 100%	Minimum identity	Maximum identity	View details
22	NP 421	LKQFKQDAKYSHGID	15	98.44% (63/64)	93.33%	100.00%	<a href="#">Go</a>
20	NP 381	LDPNAKTWMDIEGRP	15	68.75% (44/64)	80.00%	100.00%	<a href="#">Go</a>
17	NP 321	ASRTSITGRAWENTV	15	67.19% (43/64)	86.67%	100.00%	<a href="#">Go</a>
21	NP 401	IALLYQSSGCYIHFF	15	65.62% (42/64)	86.67%	100.00%	<a href="#">Go</a>
25	NP 481	KLIDIALSKTDSRKY	15	54.69% (35/64)	66.67%	100.00%	<a href="#">Go</a>
26	NP 501	DQYKDLCHMHTGVVV	15	54.69% (35/64)	86.67%	100.00%	<a href="#">Go</a>
19	NP 361	FTAGLTYSQLMTLKD	15	50.00% (32/64)	80.00%	100.00%	<a href="#">Go</a>
24	NP 461	TCQGSDDIRKLLSQ	15	50.00% (32/64)	80.00%	100.00%	<a href="#">Go</a>
23	NP 441	AAQPGLTSAVIEALP	15	34.38% (22/64)	86.67%	100.00%	<a href="#">Go</a>
18	NP 341	DGKPKQAGSNNSNKS	15	28.12% (18/64)	60.00%	100.00%	<a href="#">Go</a>
2	NP 21	SGYCSNIKLQVVKDA	15	17.19% (11/64)	26.67%	100.00%	<a href="#">Go</a>
10	NP 181	SLTLACLTKQGQVDL	15	17.19% (11/64)	26.67%	100.00%	<a href="#">Go</a>
11	NP 201	ALTDLGLIYTAKYPN	15	17.19% (11/64)	26.67%	100.00%	<a href="#">Go</a>
3	NP 41	GLDFSEVSNVQRLMR	15	15.62% (10/64)	26.67%	100.00%	<a href="#">Go</a>
13	NP 241	ISGYNFSLGAAVKAG	15	15.62% (10/64)	26.67%	100.00%	<a href="#">Go</a>
16	NP 301	NPYENILYKICLSGD	15	14.06% (9/64)	20.00%	100.00%	<a href="#">Go</a>





# Epitope Conservancy – example

<http://tools.iedb.org/conservancy/>

Protein #	Protein name	Positions	Protein sub-sequence(s)	Identity
1	58643 Lassa NP	401-415	IALYQPSSGCVHFF	93.33%
2	9294837 Lassa NP	400-414	IALFQPSSGCVIHFF	93.33%
3	9294840 Lassa NP	400-414	IALYQPMMSGCVIHFF	93.33%
4	9294844 Lassa NP	82-96	IALYQPMMSGCVIHFF	93.33%
5	9294846 Lassa NP	82-96	IALYQPMMSGCVIHFF	93.33%
6	9294848 Lassa NP	82-96	IALYQPMMSGCVIHFF	93.33%
7	9294850 Lassa NP	82-96	IALYQPISGCVIHFF	93.33%
8	9294852 Lassa NP	82-96	IALYQPISGCVIHFF	93.33%
9	9294854 Lassa NP	82-96	IAIYQPMMSGCVIHFF	86.67%
10	9294856 Lassa NP	82-96	IALYQPMMSGCVIHFF	93.33%
11	9294858 Lassa NP	82-96	IALYQPMMSGCVIHFF	93.33%
12	9294860 Lassa NP	82-96	IALYQPSSGCVIHFF	100.00%
13	9294862 Lassa NP	82-96	IALYQPSSGCVIHFF	100.00%

# Epitope Cluster Analysis Tool

- Analyzes how many epitopes in a set have significant sequence homology
- Groups epitopes into clusters based on having sequence identity greater than a specified threshold
- Three different clustering approaches are implemented
- Enables diverse applications such as generating epitope pools, and understanding cross-reactivity

# Clustering approaches

## 1. All connected peptides in a cluster

- All peptides homologous to specified threshold are clustered together (for example, 70%)
- Drawback: members of the cluster might be related by levels of homology lower than threshold (for example, 70%)

## 2. Fully interconnected clusters (cliques)

- All peptides in a cluster share homology higher than the given threshold
- Drawback: One peptide can be a part of multiple cliques

## 3. Cluster-break method (recommended method)

- An extension of first approach
- A cluster is broken down into subclusters based on consensus sequence computations

# Here is a set of sequences...as an example

TRAPPER  
CLAPPER  
SNAPPER  
RAPPERTIME  
DAYTIME  
ANYTIME  
PERTIME  
TIMELESS  
TIMER  
TWICE  
NICEDAY  
ICE  
CEDAR  
HAPPYDAYS  
HIPPY  
PAYDAY  
CALAMARI  
AMA  
ISQAVHAAHAEINEAGR

# Epitope Clustering – example

<http://tools.iedb.org/cluster/>

## Step 1:

Enter epitope sequences

step 1/3 Specify input peptides

Specify Sequence(s)

Enter epitope sequence(s) in PLAIN or FASTA format

TRAPPER  
CLAPPER  
SNAPPER  
RAPPETIME  
DAYTIME  
ANYTIME  
PERTIME  
TIMELESS  
TIMER  
TWICE

Or upload epitope sequence(s) from a file  No file chosen

## Step 2:

Specify parameters

Select Sequence Identity Threshold

Select minimum sequence identity threshold: 80% ▼

Select canonical peptides

Select minimum Peptide length: No minimum length ▼

Select maximum Peptide length: No Maximum Length ▼

## Step 3:

Select algorithm

step 3/3 Specify clustering algorithm

Select Clustering approach

Choose clustering method:

Cluster-break for clear representative sequence  
Fully interconnected clusters (cliques)

# Epitope Clustering – example

Show Table		Graphical Visualization			
Cluster Number	Peptide Number	Alignment	Position	Description	Peptide
1	Consensus	TRAPPERTIMECESS	-	-	-
1	1	TRAPPER-----	1	seq1	TRAPPER
1	2	-RAPPERTIME----	2	seq4	RAPPERTIME
1	3	----DAYTIME----	5	seq5	DAYTIME
1	4	----ANYTIME----	5	seq6	ANYTIME
1	5	----PERTIME----	5	seq7	PERTIME
1	6	-----TIMELESS	8	seq8	TIMELESS
1	7	-----TIMER---	8	seq9	TIMER
2	Consensus	TXICEDAX	-	-	-
2	1	TWICE---	1	seq10	TWICE
2	2	-NICEDAY	2	seq11	NICEDAY
2	3	--ICE---	3	seq12	ICE
2	4	---CEDAR	4	seq13	CEDAR
3	Consensus	HXPPYDAYS	-	-	-
3	1	HAPPYDAYS	1	seq14	HAPPYDAYS
3	2	HIPPY----	1	seq15	HIPPY
3	3	--PAYDAY-	3	seq16	PAYDAY
4	Consensus	CALAMARI	-	-	-
4	1	CALAMARI	1	seq17	CALAMARI
4	2	---AMA--	4	seq18	AMA
5	Singleton	CLAPPER	-	seq2	CLAPPER
6	Singleton	SNAPPER	-	seq3	SNAPPER
7	Singleton	ISQAVHAAHAEINEAGR	-	seq19	ISQAVHAAHAEINEAGR

<http://tools.iedb.org/cluster/>

TRAPPER and TIMELESS do not have 80% homology but are in the same cluster because each have 80% homology to other members of the cluster

# Epitope Clustering – example

<http://tools.iedb.org/cluster/>

**Step 1:**  
Enter epitope sequences

step 1/3 Specify input peptides

Specify Sequence(s)

Enter epitope sequence(s) in PLAIN or FASTA format

TRAPPER  
CLAPPER  
SNAPPER  
RAPPERTIME  
DAYTIME  
ANYTIME  
PERTIME  
TIMELESS  
TIMER  
TWICE

Or upload epitope sequence(s) from a file  No file chosen

**Step 2:**  
Specify parameters

Select Sequence Identity Threshold

Select minimum sequence identity threshold: 80% ▼

Select canonical peptides

Select minimum Peptide length: No minimum length ▼

Select maximum Peptide length: No Maximum Length ▼

**Step 3:**  
Select algorithm

Select Clustering approach

Choose clustering method: ▼

All the connected peptides in a cluster  
Cluster-break for clear representative sequence  
Fully interconnected clusters (cliques)

# Epitope Clustering – example

Show Table		Graphical Visualization			
Clique Number	Peptide Number	Alignment	Position	Description	Peptide
1	Consensus	PERTIMERAPPERTIME	-	-	-
1	1	PERTIME-----	1	seq7	PERTIME
1	2	---TIMER-----	4	seq9	TIMER
1	3	-----RAPPERTIME	8	seq4	RAPPERTIME
2	Consensus	TWICE	-	-	-
2	1	TWICE	1	seq10	TWICE
2	2	--ICE	3	seq12	ICE
3	Consensus	NICEDAY	-	-	-
3	1	NICEDAY	1	seq11	NICEDAY
3	2	-ICE---	2	seq12	ICE
4	Consensus	NICEDAX	-	-	-
4	1	NICEDAY	1	seq11	NICEDAY
4	2	--CEDAR	3	seq13	CEDAR
5	Consensus	CALAMARI	-	-	-
5	1	CALAMARI	1	seq17	CALAMARI
5	2	---AMA--	4	seq18	AMA
6	Consensus	ANYTIMER	-	-	-
6	1	ANYTIME-	1	seq6	ANYTIME
6	2	---TIMER	4	seq9	TIMER
7	Consensus	TIMEXESS	-	-	-
7	1	TIMELESS	1	seq8	TIMELESS
7	2	TIMER---	1	seq9	TIMER
8	Consensus	DAYTIMER	-	-	-
8	1	DAYTIME-	1	seq5	DAYTIME
8	2	---TIMER	4	seq9	TIMER
9	Consensus	HXPPYDAYS	-	-	-
9	1	HAPPYDAYS	1	seq14	HAPPYDAYS

<http://tools.iedb.org/cluster/>

TIMER (and several others) appear in multiple clusters because they have the specified level of homology to the other members of the clique





# Cluster-break method is the recommended method

<http://tools.iedb.org/cluster/>

**Step 1:**  
Enter epitope sequences

step 1/3 Specify input peptides

Specify Sequence(s)

Enter epitope sequence(s) in PLAIN or FASTA format

Or upload epitope sequence(s) from a file

Choose File No file chosen

Go to Step 2 Reset

Example sequences listed:

- TWICE
- NICEDAY
- ICE
- CEDAR
- HAPPYDAYS
- HIPPY
- PAYDAY
- CALAMARI
- AMA
- ISQAVHAAHAEINEAGR

**Step 2:**  
Specify parameters

Select Sequence Identity Threshold

Select minimum sequence identity threshold: 80% ▼

Select canonical peptides

Select minimum Peptide length: No minimum length ▼

Select maximum Peptide length: No Maximum Length ▼

Start Over Go to Step 3

**Step 3:**  
Select algorithm

Select Clustering approach

Choose clustering method: ▼

Go back to Step 2

All the connected peptides in a cluster

Cluster-break for clear representative sequence

Fully interconnected clusters (cliques)

# Epitope Clustering – example

Show Table **Graphical Visualization**

Cluster.Sub-Cluster Number	Peptide Number	Alignment	Position	Description	Peptide
1.1	Consensus	RAPXXXTIMEXESS	-	-	-
1.1	1	RAPPERTIME----	1	seq4	RAPPERTIME
1.1	2	---ANYTIME----	4	seq6	ANYTIME
1.1	3	---DAYTIME----	4	seq5	DAYTIME
1.1	4	---PERTIME----	4	seq7	PERTIME
1.1	5	-----TIMELESS	7	seq8	TIMELESS
1.1	6	-----TIMER---	7	seq9	TIMER
1.2	Singleton	TRAPPER	-	seq1	TRAPPER
2.1	Consensus	NICEDAX	-	-	-
2.1	1	NICEDAY	1	seq11	NICEDAY
2.1	2	-ICE---	2	seq12	ICE
2.1	3	--CEDAR	3	seq13	CEDAR
2.2	Singleton	TWICE	-	seq10	TWICE
3.1	Consensus	HXPPYDAYS	-	-	-
3.1	1	HAPPYDAYS	1	seq14	HAPPYDAYS
3.1	2	HIPPY----	1	seq15	HIPPY
3.1	3	--PAYDAY-	3	seq16	PAYDAY
4.1	Consensus	CALAMARI	-	-	-
4.1	1	CALAMARI	1	seq17	CALAMARI
4.1	2	---AMA--	4	seq18	AMA
5.1	Singleton	CLAPPER	-	seq2	CLAPPER
6.1	Singleton	SNAPPER	-	seq3	SNAPPER
7.1	Singleton	ISQAVHAAHAEINEAGR	-	seq19	ISQAVHAAHAEINEAGR

<http://tools.iedb.org/cluster/>

Show Table **Graphical Visualization**

Visualize a selected Sub-cluster  ▾


# Restrictor Analysis Tool for Epitopes (RATE)

- Automated method to infer HLA restriction of a given epitope set, from immune response data of HLA typed subjects
- The method is based on computing the frequency of all alleles expressed in the population studied in donors who had an immune response to each given epitope
- Compare those frequencies in donors that did not have a response to the same epitope

# RATE

<http://tools.iedb.org/rate/>

## IEDB Analysis Resource - Labs



**Labs designation**

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

### RATE (Restrictor Analysis Tool for Epitopes)

**Enter data**

Allele data	<input type="button" value="Choose File"/> No file chosen
Response data	<input type="button" value="Choose File"/> No file chosen
Cutoff for response to be considered positive	<input type="text"/>
Email id (optional)	<input type="text"/>
Job name (optional)	<input type="text"/>

**Two data input files need to be provided**

# RATE – Input: Allele data

- Data format = tab separated in plain text file
- Sample data shown here – 12 class II alleles of 6 loci for each donor are listed in separate columns

Donor-1	Donor-2	Donor-3	Donor-4	Donor-5	Donor-6	Donor-7	Donor-8	Donor-9	Donor-10	Donor-11	Donor-12	Donor-13
DRB1*03:01	DRB1*11:01	DRB1*01:01	DRB1*07:01	DRB1*07:01	DRB1*15:03	DRB1*12:01	DRB1*11:02	DRB1*08:04	DRB1*11:01	DRB1*		
DRB1*07:01	DRB1*15:03	DRB1*07:01	DRB1*15:03	DRB1*13:02	DRB1*15:03	DRB1*13:02	DRB1*15:02	DRB1*13:02	DRB1*	DRB1*		
DRB3*02:02	DRB3*02:02	DRB3*02:02	DRB3*02:02	DRB4*01:03	DRB3*01:01	DRB3*02:02	DRB5*01:02	DRB3*02:02	DRB3*	DRB3*		
DRB4*01:03	DRB5*01:01	DRB4*01:03	DRB5*01:01	DRB5*01:01	DRB3*03:01	n/a	DRB3*03:01	DRB3*02:02	DRB3*	DRB3*		
DQA1*02:01	DQA1*01:02	DQA1*02:01	DQA1*01:02	DQA1*01:01	DQA1*01:02	DQA1*01:03	DQA1*01:02	DQA1*01:02	DQA1*	DQA1*		
DQA1*05:01	DQA1*05:05	DQA1*05:01	DQA1*02:01	DQA1*01:02	DQA1*05:05	DQA1*04:01	DQA1*05:05	DQA1*05:05	DQA1*	DQA1*		
DQB1*02:01	DQB1*06:02	DQB1*03:02	DQB1*02:02	DQB1*05:01	DQB1*03:19	DQB1*03:19	DQB1*06:09	DQB1*06:09	DQB1*	DQB1*		
DQB1*05:03	DQB1*06:02	DQB1*03:02	DQB1*06:02	DQB1*06:02	DQB1*06:09	DQB1*06:01	DQB1*06:09	DQB1*06:09	DQB1*	DQB1*		
DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*01:03	DPA1*02:01	DPA1*01:03	DPA1*01:03	DPA1*	DPA1*		
DPA1*02:01	DPA1*03:01	DPA1*02:01	DPA1*02:01	DPA1*03:01	DPA1*01:03	DPA1*02:02	DPA1*01:03	DPA1*01:03	DPA1*	DPA1*		
DPB1*13:01	DPB1*03:01	DPB1*17:01	DPB1*01:01	DPB1*18:01	DPB1*34:01	DPB1*01:01	DPB1*02:01	DPB1*02:01	DPB1*	DPB1*		
DPB1*17:01	DPB1*105:01	DPB1*01:01	DPB1*02:01	DPB1*105:01	DPB1*02:01	DPB1*13:01	DPB1*105:01	DPB1*	DPB1*	DPB1*		

# RATE – Input: Response data

- Data format = tab separated in plain text file
- The response data (here SFC values) for each epitope in each of the donors are provided

Peptide #	Peptide_ID	Peptide_Seq	Donor-1	Donor-2	Donor-3	Donor-4	Donor-5	Donor-6	Donor-7	Donor-8
1	3531.0365	GINTIPIAINEAEYV 98	0	0	0	0	0	0	40	0
2	3531.0367	AAFQAAHARFVAAAA 68	0	50	0	0	0	0	0	0
3	3531.0514	AAVRFQEAANKQKQ 0	55	0	0	0	0	0	0	0
4	3531.0494	ELFVAAYVPYVAWL 0	0	0	0	n/a	n/a	0	0	0
5	3531.037	AAGTYVAADAAAAS 0	0	83	0	0	0	0	0	0

# RATE – how it works

- Restrictions are determined based on
  - “Odds Ratio” and significance estimated using Fisher's exact test.

$$OR = \frac{(A^+R^+) \times (A^-R^-)}{(A^-R^+) \times (A^+R^-)}$$


- “Relative Frequency”, a parameter estimated by the tool based on the frequency of alleles and donors responded

$$RF = \frac{A^+R^+ / (A^+R^+ + A^+R^-)}{(A^+R^+ + A^-R^+) / \text{Total donors}}$$

# RATE – example

**RATE (Restrictor Analysis Tool for Epitopes)**

Enter data	
Allele data	<input type="button" value="Choose File"/> iedb-rate_sa...ele_data.txt
Response data	<input type="button" value="Choose File"/> iedb-rate_sa...nse_data.txt
Cutoff for response to be considered positive	<input type="text" value="20.0"/>
Email id (optional)	<input type="text"/>
Job name (optional)	<input type="text"/>
<input type="button" value="Submit"/> <input type="button" value="Reset"/>	

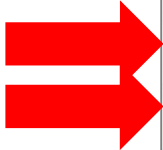




# RATE – example

RATE results	
<b>Input data summary</b>	
Allele file	iedb-rate_sample_input_allele_data.txt
Response file	iedb-rate_sample_input_response_data.txt
No. of peptides	5
No. of subjects	50
Cutoff for response to be considered positive	20.0
No. of alleles expressed by subjects	90
Total peptide-allele restriction examinations done	5 x 90 = 450
Re-formatted allele data	<input type="button" value="Allele data"/>
Re-formatted response data	<input type="button" value="Response data"/>
<b>HLA restriction results</b>	
No. of unique peptides for which restriction is defined (in selected HLA restrictions)	4
Total no. of peptide-allele restrictions (in selected HLA restrictions)	5
RATE results (selected HLA restrictions)	<input type="button" value="RATE results"/>
Complete report (analysis results for all peptide-allele combinations)	<input type="button" value="Complete report"/>
Job id	1570383651

All buttons  
download as  
.csv files



# RATE – results (complete report)

- All epitopes

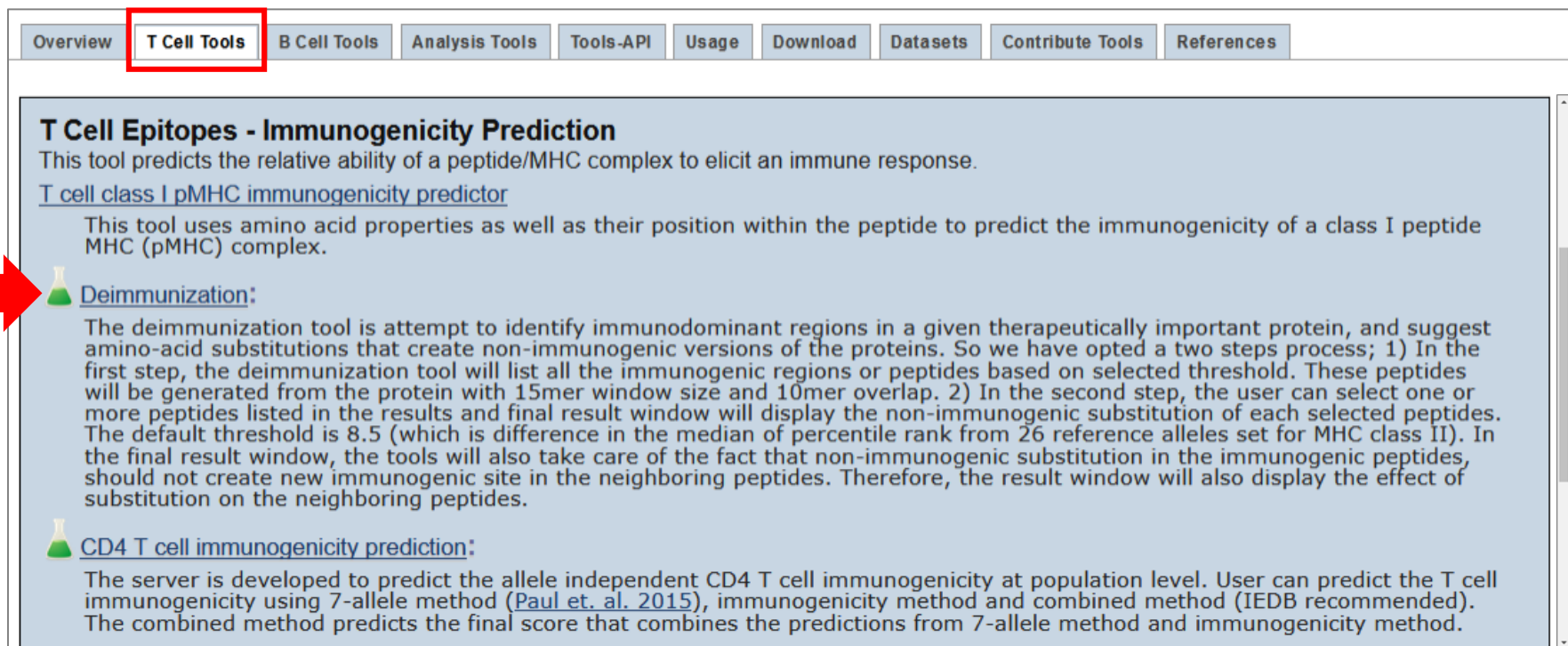
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Peptide#	Peptide_id	peptide_seq	Allele#	Allele	A+R+	A-R+	A+R-	A-R-	No_of_Donors	Response_n/a	Relative_freq	Odds_ratio	P-value
2	1	3531.037	GINTIPIAINEAEYV	1	DPA1*01:03	5	1	34	10	50	0	1.06838	1.47059	1
3	2	3531.037	AAFQAAHARFVAAAA	1	DPA1*01:03	10	2	29	9	50	0	1.06838	1.55172	1
4	3	3531.051	AAVVRFQEAAANKQKQ	1	DPA1*01:03	8	1	28	10	47	3	1.16049	2.85714	0.6631
5	4	3531.049	ELFVAAYVPYVAVLV	1	DPA1*01:03	1	0	36	11	48	2	1.2973	inf	1
6	5	3531.037	AAGTYVAADAAAAST	1	DPA1*01:03	5	1	34	10	50	0	1.06838	1.47059	1
7	1	3531.037	GINTIPIAINEAEYV	2	DPA1*01:04	0	6	1	43	50	0	0	0	1
8	2	3531.037	AAFQAAHARFVAAAA	2	DPA1*01:04	0	12	1	37	50	0	0	0	1
9	3	3531.051	AAVVRFQEAAANKQKQ	2	DPA1*01:04	0	9	1	37	47	3	0	0	1
10	4	3531.049	ELFVAAYVPYVAVLV	2	DPA1*01:04	0	1	1	46	48	2	0	0	1
11	5	3531.037	AAGTYVAADAAAAST	2	DPA1*01:04	0	6	1	43	50	0	0	0	1
12	1	3531.037	GINTIPIAINEAEYV	3	DPA1*02:01	3	3	20	24	50	0	1.08696	1.2	1
13	2	3531.037	AAFQAAHARFVAAAA	3	DPA1*02:01	5	7	18	20	50	0	0.9058	0.79365	1
14	3	3531.051	AAVVRFQEAAANKQKQ	3	DPA1*02:01	3	6	18	20	47	3	0.74603	0.55556	0.7111
15	4	3531.049	ELFVAAYVPYVAVLV	3	DPA1*02:01	0	1	23	24	48	2	0	0	1
16	5	3531.037	AAGTYVAADAAAAST	3	DPA1*02:01	2	4	21	23	50	0	0.72464	0.54762	0.674
17	1	3531.037	GINTIPIAINEAEYV	4	DPA1*02:02	1	5	10	34	50	0	0.75758	0.68	1
18	2	3531.037	AAFQAAHARFVAAAA	4	DPA1*02:02	2	10	9	29	50	0	0.75758	0.64444	1
19	3	3531.051	AAVVRFQEAAANKQKQ	4	DPA1*02:02	1	8	10	28	47	3	0.47475	0.35	0.6631
20	4	3531.049	ELFVAAYVPYVAVLV	4	DPA1*02:02	0	1	11	36	48	2	0	0	1
21	5	3531.037	AAGTYVAADAAAAST	4	DPA1*02:02	1	5	10	34	50	0	0.75758	0.68	1
22	1	3531.037	GINTIPIAINEAEYV	5	DPA1*03:01	0	6	9	35	50	0	0	0	0.5756
23	2	3531.037	AAFQAAHARFVAAAA	5	DPA1*03:01	2	10	7	31	50	0	0.92593	0.88571	1
24	3	3531.051	AAVVRFQEAAANKQKQ	5	DPA1*03:01	2	7	6	32	47	3	1.30556	1.52381	0.6388
25	4	3531.049	ELFVAAYVPYVAVLV	5	DPA1*03:01	0	1	8	39	48	2	0	0	1
26	5	3531.037	AAGTYVAADAAAAST	5	DPA1*03:01	1	5	8	36	50	0	0.92593	0.9	1
27	1	3531.037	GINTIPIAINEAEYV	6	DPA1*04:01	1	5	0	44	50	0	8.33333	inf	0.12
28	2	3531.037	AAFQAAHARFVAAAA	6	DPA1*04:01	0	12	1	37	50	0	0	0	1
29	3	3531.051	AAVVRFQEAAANKQKQ	6	DPA1*04:01	0	9	1	37	47	3	0	0	1
30	4	3531.049	ELFVAAYVPYVAVLV	6	DPA1*04:01	0	1	1	46	48	2	0	0	1

# RATE – results (selected HLA restrictions)

- Epitope-allele combinations with RF  $\geq 1.3$  and p-value  $< 0.01$  are reported (not Bonferroni corrected)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Peptide#	Peptide_id	peptide_seq	Allele#	Allele	A+R+	A-R+	A+R-	A-R-	No_of_Donors	Response_n/a	Relative_freq	Odds_ratio	P-value
2	1	3531.0365	GINTIPIAINEAEYV	84	DRB3*02:02	6	0	18	26	50	0	2.08333	inf	0.00847
3	2	3531.0367	AAFQAAHARFVAAAA	54	DRB1*01:01	4	8	0	38	50	0	4.16667	inf	0.00215
4	2	3531.0367	AAFQAAHARFVAAAA	65	DRB1*07:01	5	7	2	36	50	0	2.97619	12.85714	0.00593
5	3	3531.0514	AAVVRFQEAANKQKQ	89	DRB5*01:01	7	2	8	30	47	3	2.43704	13.125	0.0025
6	5	3531.037	AAGTYVAADAAAAS	54	DRB1*01:01	4	2	0	44	50	0	8.33333	inf	7.00E-05


# Deimmunization




The screenshot shows the IEDB website's navigation menu with 'T Cell Tools' highlighted in a red box. A red arrow points to the 'Deimmunization' section in the main content area.

**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 peptides listed in the results and final result window will display the non-immunogenic substitution of each selected peptides. The default threshold is 8.5 (which is difference in the median of percentile rank from 26 reference alleles set for MHC class II). In the final result window, the tools will also take care of the fact that non-immunogenic substitution in the immunogenic peptides, should not create new immunogenic site in the neighboring peptides. Therefore, the result window will also display the effect of substitution on the neighboring peptides.

 **CD4 T cell immunogenicity prediction:**  
The server is developed to predict the allele independent CD4 T cell immunogenicity at population level. User can predict the T cell immunogenicity using 7-allele method ([Paul et. al. 2015](#)), immunogenicity method and combined method (IEDB recommended). The combined method predicts the final score that combines the predictions from 7-allele method and immunogenicity method.

# Deimmunization – background

- Wild type and engineered proteins are widely used as drugs
- Immunogenicity of protein drugs is associated with serious potency and safety issues
- A potential approach to reducing immunogenicity is based on removal of T cell epitopes (de-immunization)

# Deimmunization – approach

- Generate overlapping peptides from protein sequence
- Predict HLA class II binding regions

APPRLICDSRVLERY	1-15	38.89
ICDSRVLERYLLEAK	6-20	60.39
VLERYLLEAKEAENI	11-25	23.46
LLEAKEAENITTGCA	16-30	21.76
...		
<b>DTFRKLFRVYSNFLR</b>	<b>136-150</b>	<b>6.08</b>

- Suggest amino acid substitutions that are predicted to decrease binding
  - Also consider the effect of substitutions on neighboring peptides

# Deimmunization – example

<http://tools.iedb.org/deimmunization/>

IEDB Analysis Resource - Labs

Home Help Example Reference Contact

## Deimmunization tool

Step 1/2 (Predicting Immunogenic regions in the given protein sequence/s)

**Specify Sequence(s)**

Enter epitope sequence(s) in FASTA format

```
>sp|P01588|EPO_HUMAN Erythropoietin OS=Homo sapiens GN=EPO PE=1 SV=1 APPRLICDSRVLERYLLEAKEAENITTGCAEHC SLNENITVPDTKVNIFYAWKRMEVGQQAVEVWQGLALLSEAVLRGQALLVNSSQP WEPLQLHVDKAVSGLRSLTLLRALGAQKEAISPPDAASAAPLRTITADTFRKLFRVY SNFLRGKL KLYTGEACRTGDR
```

FASTA format detected.

Or upload epitope sequence(s) from a file  No file chosen

**Select Median Percentile Rank Threshold**

Select maximum median percentile rank threshold:

**Epitope sequence**

**Threshold**

# Deimmunization – example

<http://tools.iedb.org/deimmunization/>

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

## Deimmunization tool (Peptide mutant prediction)

The prediction may take about more than 10 minutes per peptide selected. Email address is required to ensure you receive the result.

Step 2/2 (Predicting non-immunogenic variants of selected immunogenic peptides)

### Choose Immunogenic peptides for deimmunization

Protein Number	Start Position	End Position	Median Percentile Rank	Peptide	Select Peptide/s
1	136	150	10.255	DIFRKLFRVYSNFLR	<input checked="" type="checkbox"/>
1	96	110	13.505	DKAVSGLRSLTLLR	<input type="checkbox"/>
1	101	115	14.25	GLRSLTLLRALGAQ	<input type="checkbox"/>
1	141	155	14.5	LFRVYSNFLRGKCLK	<input type="checkbox"/>

### Choose threshold for deimmunization

Select the cutoff value for the difference in median percentile Rank : [?](#)

### Enter Job Details

Enter the Job Name (Optional)

Enter your Email Address (Required)



# Deimmunization Score

<b>Immunogenicity for Neighboring peptide (1)</b>	<b>Immunogenicity for Neighboring peptide (2)</b>	<b>Score</b>
Absent	Absent	1
Reduced	Absent	2
Reduced	Reduced	3
Neutral	Absent	4
Reduced	Neutral	5
Neutral	Neutral	6
Increased	Absent	7
Reduced	Increased	8
Increased	Neutral	9
Increased	Increased	10

# Deimmunization – example

	A	B	C	D	E	F	G	H	I	J	K	L
1	Protein Number	Peptide	Peptide ID	Start Position	End Position	Median Percentile Rank	Median Difference	C terminal Neighbor 1 (Median)	C terminal Neighbor 2 (Median)	N terminal Neighbor 1 (Median)	N terminal Neighbor 2 (Median)	Deimmunization Score
2	1	DTFRKLFVYSNFLR	wild	136	150	10.255	0	14.5	49.75	47	44.25	NA
3	1	DTFRKLFVYSNFDR	L149D	136	150	32	21.745	41	68.5	NA	NA	3
4	1	DTFRKLFVYSNFGR	L149G	136	150	27.5	17.245	31.5	68.75	NA	NA	3
5	1	DTFRKEFRVYSNFLR	L141E	136	150	26.5	16.245	26.5	NA	60.75	NA	3
6	1	DTFRKPFRVYSNFLR	L141P	136	150	26.35	16.095	26.5	NA	52.5	NA	3
7	1	DTFRKQFRVYSNFLR	L141Q	136	150	26	15.745	25.5	NA	53.25	NA	3
8	1	DTFRKLFVYSNFNR	L149N	136	150	26	15.745	21.75	61.75	NA	NA	3
9	1	DTFRKLGRVYSNFLR	F142G	136	150	25.075	14.82	31	NA	50.75	NA	3
10	1	DTFRKLFDVYSNFLR	R143D	136	150	24.75	14.495	35.5	NA	49.25	NA	3
11	1	DTFRKKFRVYSNFLR	L141K	136	150	24.525	14.27	23	NA	59	NA	3
12	1	DTFRKLFVYSNFCCR	L149C	136	150	24.25	13.995	36	64	NA	NA	3
13	1	DTFRKCFRVYSNFLR	L141C	136	150	23.85	13.595	23.5	NA	73	NA	3
14	1	DTFRKLFVYSNCFR	N147C	136	150	23.75	13.495	39.75	50	NA	NA	3
15	1	DTFRKDFRVYSNFLR	L141D	136	150	23.5	13.245	27.5	NA	62.5	NA	3
16	1	DTFRKGFRVYSNFLR	L141G	136	150	23.5	13.245	27	NA	57.5	NA	3
17	1	DTFRKLFVYSNFKR	L149K	136	150	23.5	13.245	28	59	NA	NA	3
18	1	DTFRKLFVYSNFER	L149E	136	150	22.75	12.495	34	62	NA	NA	3
19	1	DTFRKLERVYSNFLR	F142E	136	150	22.7	12.445	31	NA	48	NA	3
20	1	DTFRKLCRVYSNFLR	F142C	136	150	22.4	12.145	31	NA	50.25	NA	3
21	1	DTFRKLFVYSNFLR	F142T	136	150	22.175	11.92	30	NA	48.75	NA	3

# Analysis Tools Recap

- Help to examine existing sets of epitopes and gain new knowledge across a broad array of applications

## Population Coverage

Analyze T cell epitopes with known HLA restriction that are recognized in a population based on HLA frequencies

## Conservancy

Investigate epitope conservancy across different protein sequences.

## Cluster

Cluster epitopes on the basis of homology

## RATE

(Restrictor Analysis)

Infer HLA restrictions for epitopes of T cell response frequency in HLA typed subjects

## Deimmunization

Identify immunodominant regions in a given protein, and suggest amino-acid substitutions that create non-immunogenic versions of the protein.