The Immune Epitope Database Analysis Resource:

MHC class I peptide processing and immunogenicity predictions

Bjoern Peters

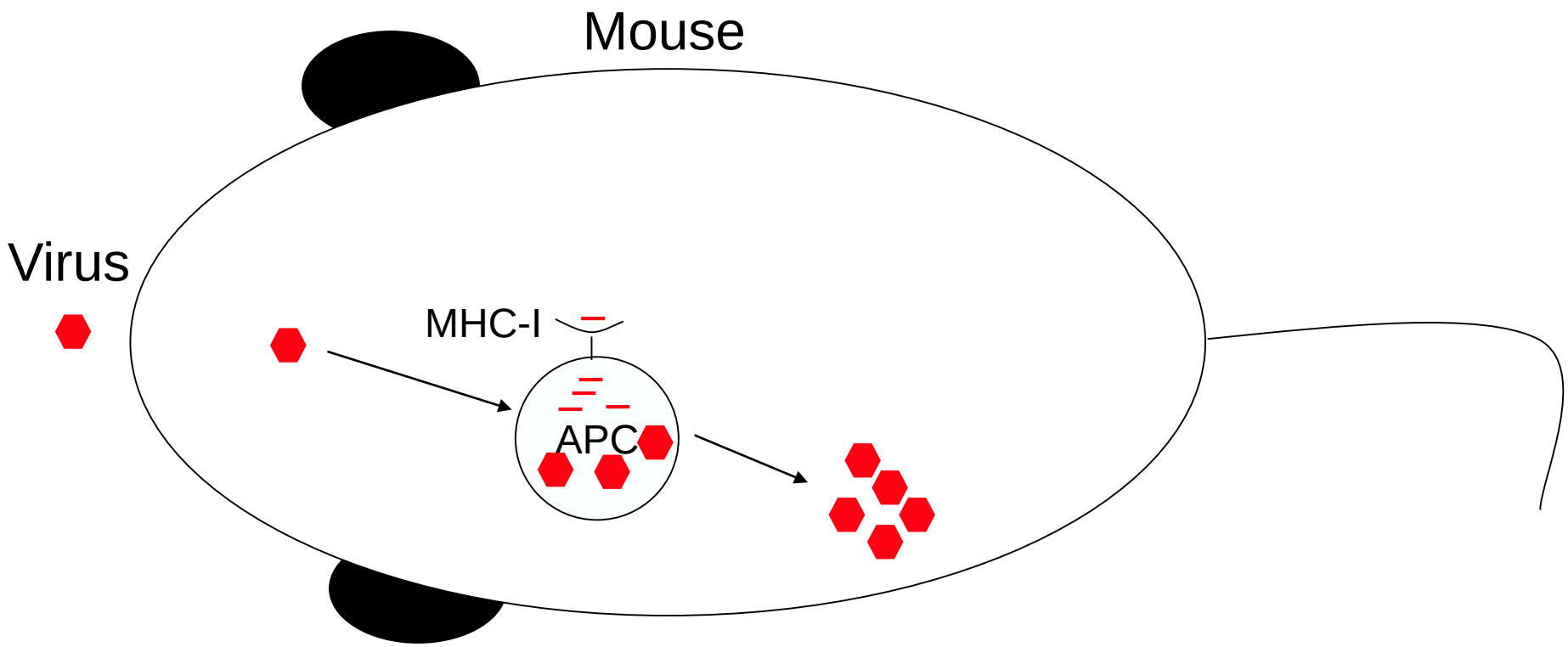
IEDB Workshop

Oct 26, 2017

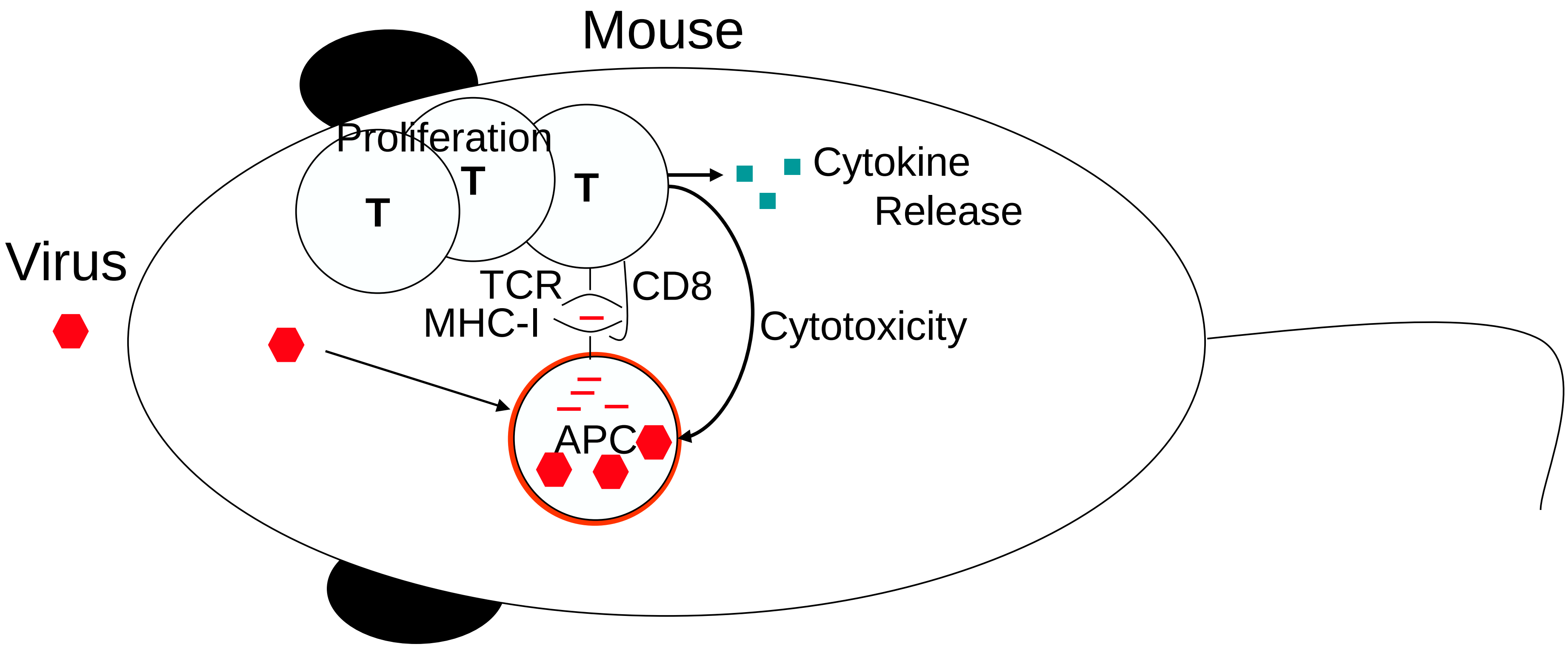
Outline

- Motivation – Factors apart from MHC binding determine what peptides are T-cell epitopes
- Processing tools in the IEDB
  - Interfaces + Prediction output
  - Performance / Caveats
- Immunogenicity tool
  - Interface + Prediction output
  - Performance / Caveats

CD8+ T cell epitopes in viral infection



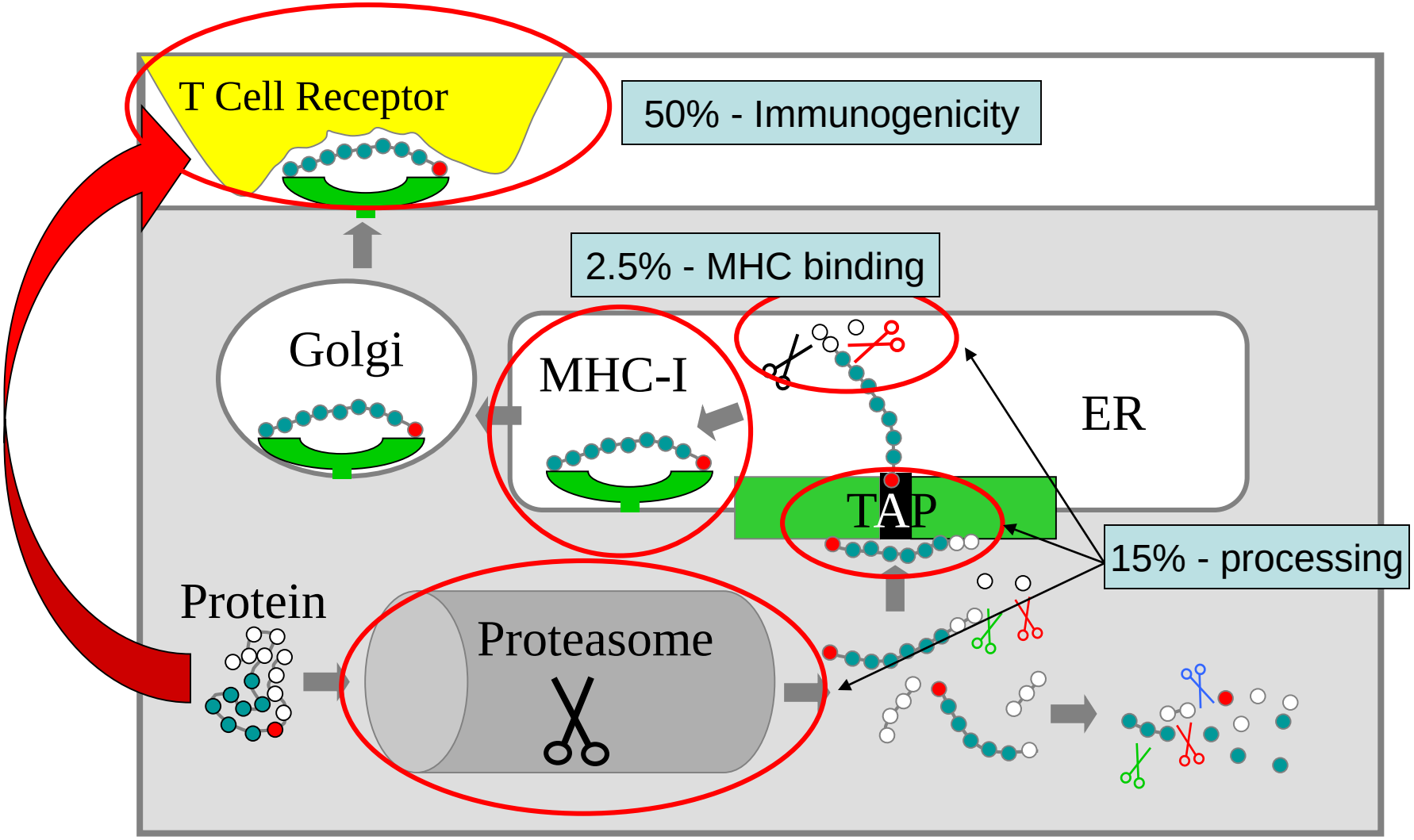
CD8^{+} T cell epitopes in viral infection



T cell epitope mapping

ORF 1	M G Q I V T M F E A L P H I **I D E V I N I V I** I V L I V I T G I K A V Y N ...
ORF 2	M G L K G P D I Y K G V Y Q F K S V E F D M S H L N L T M P N A C S A N N ...
ORF 3	M H N F C N L T S A F N K K T F D H T L M S I V S S L H L S I D G N S N Y ...
ORF 4	M S A Q S Q C R T F R G R V L D M F R T A F G G K Y M R S G W G W T G S D ...
ORF 5	M H C T Y A G P F G M S R I L L S Q E K T K F F T R R L A G T F T W T L S ...
ORF 6	M K C F G N T A V A K C N V N H D A E F C D M L R L I D Y N K A A L S K F ...
ORF 7	M L M R N H L **L D L M G V P Y C N Y** S K F W Y L E H A K T G E T S V P K C ...

MHC I - Antigen processing and presentation pathway



Peters et al, J Mol Biol 2002, Bioinformatics 2003, J Immunol.2003; CMLS 2005 ; Assarson, J Immunol 2007

Processing + immunogenicity tools available in the IEDB

- 'Combined predictor' – Combines proteasomal cleavage and TAP transport predictions, trained on specific in vitro datasets
- Neural Network based predictors (NetChop, NetCTL)
- MHC-NP: Prediction of peptides naturally processed by the MHC
- Immunogenicity predictor

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 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 spec 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-al SMM-align and Combinatorial library methods.

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 ea 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 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 natur processed and binds to a given MHC molecule. This tool was the winner of the 2nd Machine Learning Competition in Immunol

T Cell Epitopes - Immunogenicity Prediction

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

T cell pMHC Immunogenicity predictor

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

MHC-I processing predictions

Prediction Method Version	
Prediction Method Version	2009-09-01
Specify Sequence(s)	
Enter protein sequence(s) (Browse for sequences in NCBI)	
Or select file containing sequence(s)	Choose File No file chosen
Choose sequence format	auto detect format
MHC binding predictions	
Prediction Method	IEDB recommended Help on prediction method selections
MHC source species	human
Show only frequently occuring alleles: ☑ Select MHC allele(s)	Allele Length Upload allele file
Proteasomal cleavage prediction	
Specify proteasome type	immuno
TAP transport predictions	
Maximum precursor extension	1

Proteasomal cleavage

Proteasomal cleavage prediction	
Specify proteasome type	immuno
TAP transport predictions	
Maximum precursor extension	1
Alpha factor	0.2

- Proteasomes create the C-terminal end of peptides. The prediction looks for a sequence motive up and downstream of a potential cleavage site
- Cleavage sequence motif was determined based on in vitro protein digests by proteasomes
- Choice between two types of proteasomes with slightly different motif constitutive or immuno (should be default choice)

TAP transport

Proteasomal cleavage prediction	
Specify proteasome type	immuno
TAP transport predictions	
Maximum precursor extension	1
Alpha factor	0.2

- TAP transports peptides into the ER that can be further N-terminally trimmed before binding to MHC.
- The TAP transport efficiency of peptides is sequence dependent, and a motif was derived based on in vitro assays
- The overall TAP transport efficiency of a presented MHC ligand can be the result of a collection of precursors.
- The parameters shown describe that collection. Unless you read the paper and know something about the precursor length distribution, **keep parameters unchanged**

Difference in prediction output

Allele	#	Start	End	PepLength	Sequence	Proteasome Score	TAP Score	MHC Score	Processing Score	Total Score	MHC IC50[nM]
H-2-Kb	1	1	10	10	MGQIVTMFEA	0.91	-0.29	-4.36	0.62	-3.74	22777.84
H-2-Kb	1	2	11	10	GQIVTMFEAL	1.51	0.42	-3.93	1.93	-2.00	8485.76
H-2-Kb	1	3	12	10	QIVTMFEALP	0.65	0.13	-4.49	0.77	-3.72	31246.67
H-2-Kb	1	4	13	10	IVTMFEALPH	0.65	-0.20	-4.11	0.45	-3.67	12949.50
H-2-Kb	1	5	14	10	VTMFEALPHI	1.24	0.28	-3.59	1.52	-2.06	3850.57
H-2-Kb	1	6	15	10	TMFEALPHII	1.06	0.34	-3.52	1.40	-2.11	3273.98
H-2-Kb	1	7	16	10	MFEALPHIID	1.13	-0.75	-4.23	0.37	-3.85	16798.51

- Higher scores = higher efficiency for MHC-I presentation
- MHC binding score = –log10(IC50) (→sign change)
- Combined scores are additive
  - Processing = proteasome + TAP
  - Total = proteasome + TAP + MHC
- Different variance in scores reflects different selectivity
  - Proteasome (1.7) < TAP (2.8) < MHC (6.7)

Caveats / performance of processing predictions

- Processing predictions beat MHC binding predictions when predicting **eluted peptides**
- So far, there is no clear evidence that processing predictions are better at predicting **epitopes**
- Issues are:
  - All data has been derived for *human* proteasome and TAP; most well defined epitopes are mapped in mice (which has different TAP specificity)
  - Eluted peptides may over represent ‘best possible’ ligands, and the difference in processing may not be relevant in practice
- Recommendation: Use MHC binding predictions alone by default. If resources require limiting the number of peptides considered, use total score of processing predictions as an additional filter.

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 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 spec 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-al SMM-align and Combinatorial library methods.

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 ea 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 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 natur processed and binds to a given MHC molecule. This tool was the winner of the 2nd Machine Learning Competition in Immunol

T Cell Epitopes - Immunogenicity Prediction

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

T cell pMHC Immunogenicity predictor
: This tool uses amino acid properties as well as their position within the peptide to predict the immunogenicity of a peptide MH (pMHC) complex.

Additional processing predictions

- NetChop (proteasomal cleavage)
- NetCTL (combines NetChop, TAP transport, NetMHC)
- NetCTLpan (combines NetChop, TAP transport, NetMHCpan)

→ Key difference is the use of NetChop

IEDB Analysis Resource

Home | Help | Example | Reference | Contact

NetChop/NetCTL/NetCTLpan

Choose a Prediction Method

Prediction Method	NetCHOP

Specify Sequence(s)

Enter protein equence(s) in FASTA format	>BHB191648|gi:90572034|gb:CY010133|UniProtKB:Q1WPY8|Gene Symbol:M2|Protein Name:Matrix protein 2|Organism:Influenza A Virus A/Canterbury/100/2000|Segment:7|Subtype:H1N1|Host:Human MSLLTEVETPIRNEWGCRCNDSSDPLVVAASIIGIVHLILWIIDRLFSKSIYRIFKHGLKR TEGVPESMREEYREEQQNAVDADDGHFVSIELE >BHB191653|gi:90572040|gb:CY010136|UniProtKB:Q1WPY3|Gene Symbol:NS1|Protein Name:Nonstructural protein
Or select file containing sequence(s)	Choose File No file chosen

Method Specific Options

Method	C term 3.0
Threshold	0.5

Submit

NetChop 3.0

- Predicts C-terminal cleavage based on two approaches
  - *C-term 3.0*: C-terminal residues found for MHC ligands
  - *20S 3.0*: Cleavage sites based on in vitro protein digests
- C-term 3.0 is not truly a proteasome predictor but performs better.
- NetCTL and NetCTLPan use C-term 3.0 by default

References

- Peters et al, JMB 2002 (proteasome)
- Peters et al, J Immunol 2003 (TAP)
- Tenzer et al, CMLS, 2005 (combined)
- Nielsen, Immunogenetics, 2005 (NetChop)
- Larsen, BMC Bioinformatics, 2007 (NetCTL)
- Stranzl, Immunogenetics, 2010 (NetCTLPan)

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 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 spec 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-al SMM-align and Combinatorial library methods.

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 ea 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 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 natur processed and binds to a given MHC molecule. This tool was the winner of the 2nd Machine Learning Competition in Immunol

T Cell Epitopes - Immunogenicity Prediction

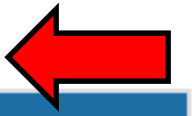
This tool predicts the relative ability of a peptide/MHC complex to elicit an immune response.

T cell pMHC Immunogenicity predictor

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

MHC-NP: Prediction of Peptides Naturally Processed by the MHC

Developed by: Sébastien Giguère Alexandre Drouin, Alexandre Lacoste, Mario Marchand, Jacques Corbeil and François Laviolette



Specify Sequence(s)

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

Or select file containing sequence(s)

Choose File | No file chosen

Choose sequence format: auto detect format

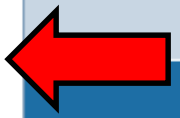
Specify what to make binding predictions for

Select MHC allele(s)

Allele | Length

Upload allele file

- H-2-Db
- H-2-Kb
- HLA-A*02:01
- HLA-B*07:02
- HLA-B*35:01
- HLA-B*44:03
- HLA-B*53:01
- HLA-B*57:01



Specify Output

Output format

Submit | Reset

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 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 spec 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-al SMM-align and Combinatorial library methods.

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 ea 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 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 natur processed and binds to a given MHC molecule. This tool was the winner of the 2nd Machine Learning Competition in Immunol

T Cell Epitopes - Immunogenicity Prediction

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

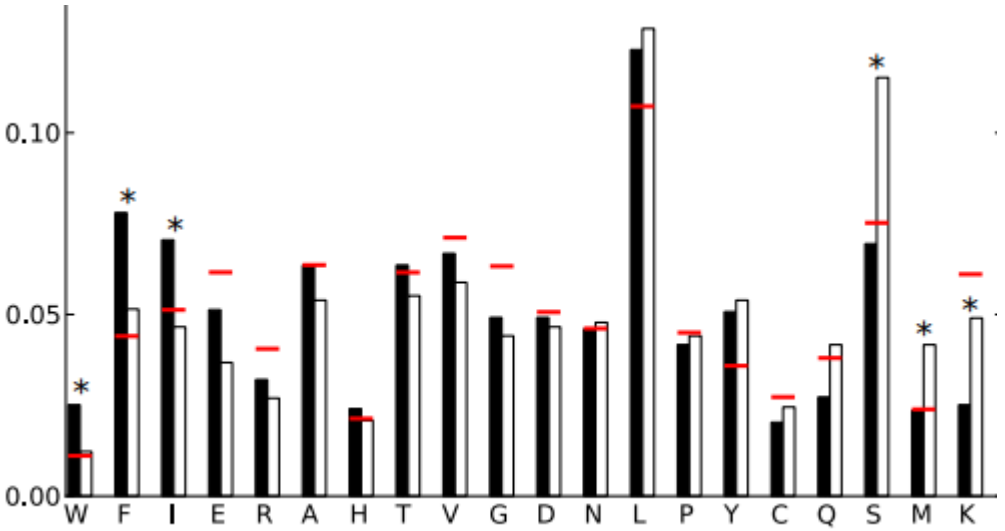
T cell pMHC Immunogenicity predictor
This tool uses amino acid properties as well as their position within the peptide to predict the immunogenicity of a peptide MH (pMHC) complex.

Immunogenicity prediction

- Approach: Assemble two datasets of peptides with similar MHC binding affinity, that are
  1) recognized or
  2) not recognized
  by T cells

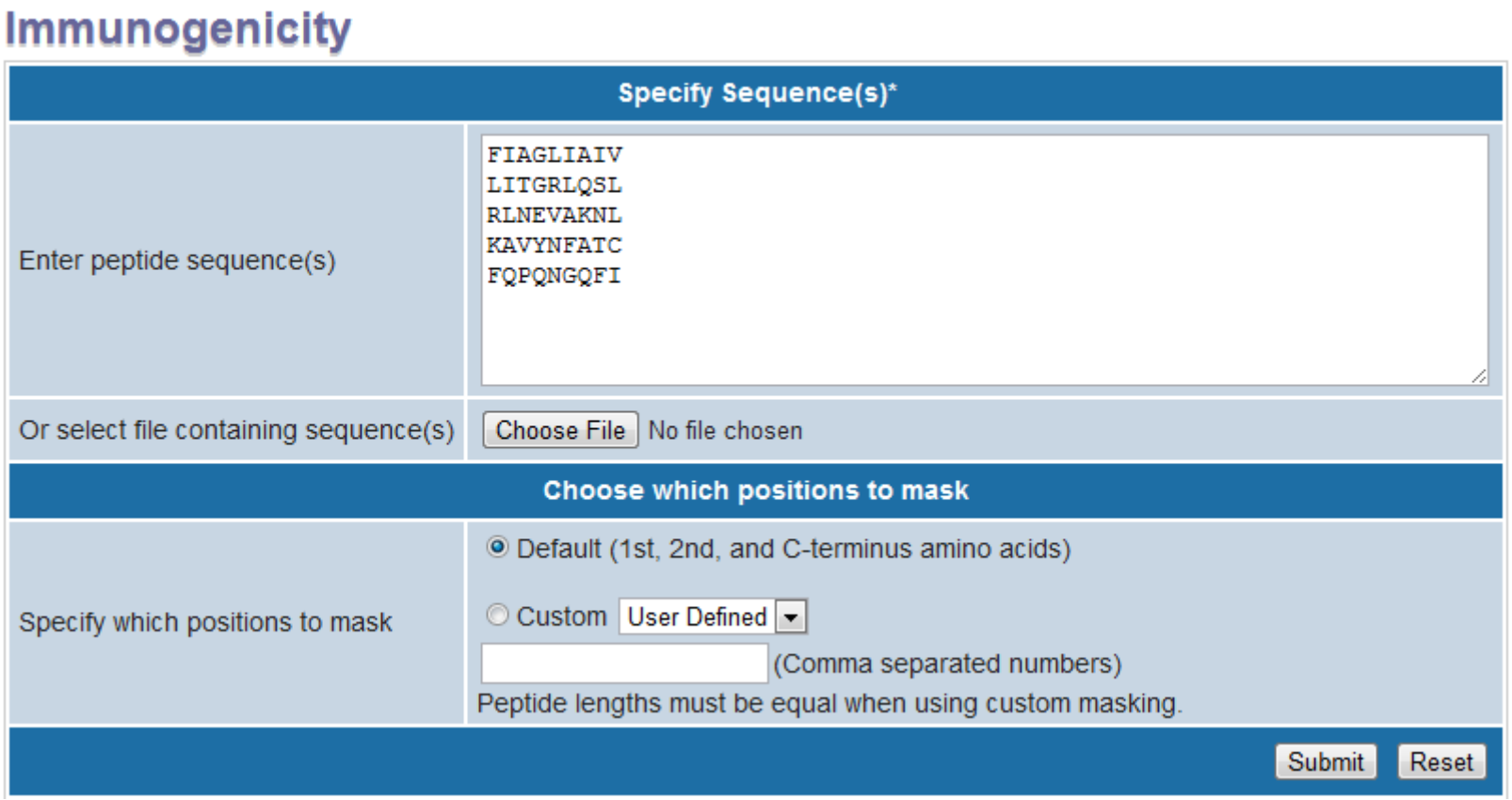
→ Enrichment of W,F,I and depletion of S,M,K in immunogenic peptides

→ Use enrichments to calculate propensity scores



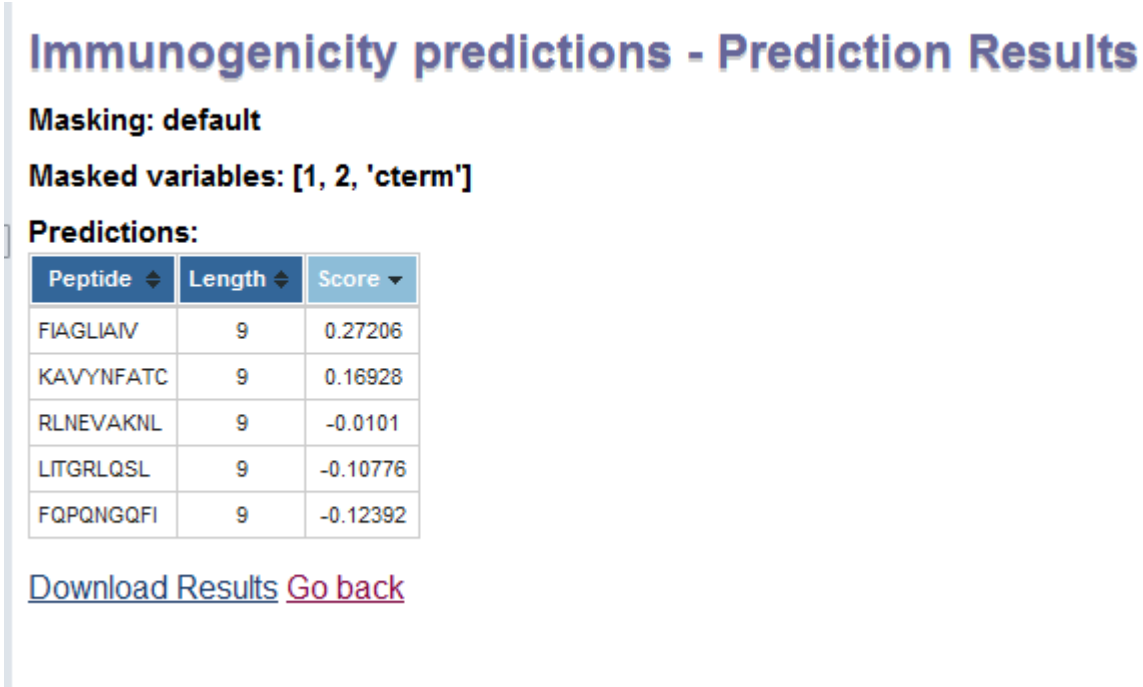
Calis et al, PLoS Comp Biol, 2013

Immunogenicity prediction - interface



- Mask positions that are MHC anchors
- Tool only validated for 9-mers

Immunogenicity prediction - output



Immunogenicity predictions - Prediction Results

Masking: default

Masked variables: [1, 2, 'cterm']

Predictions:

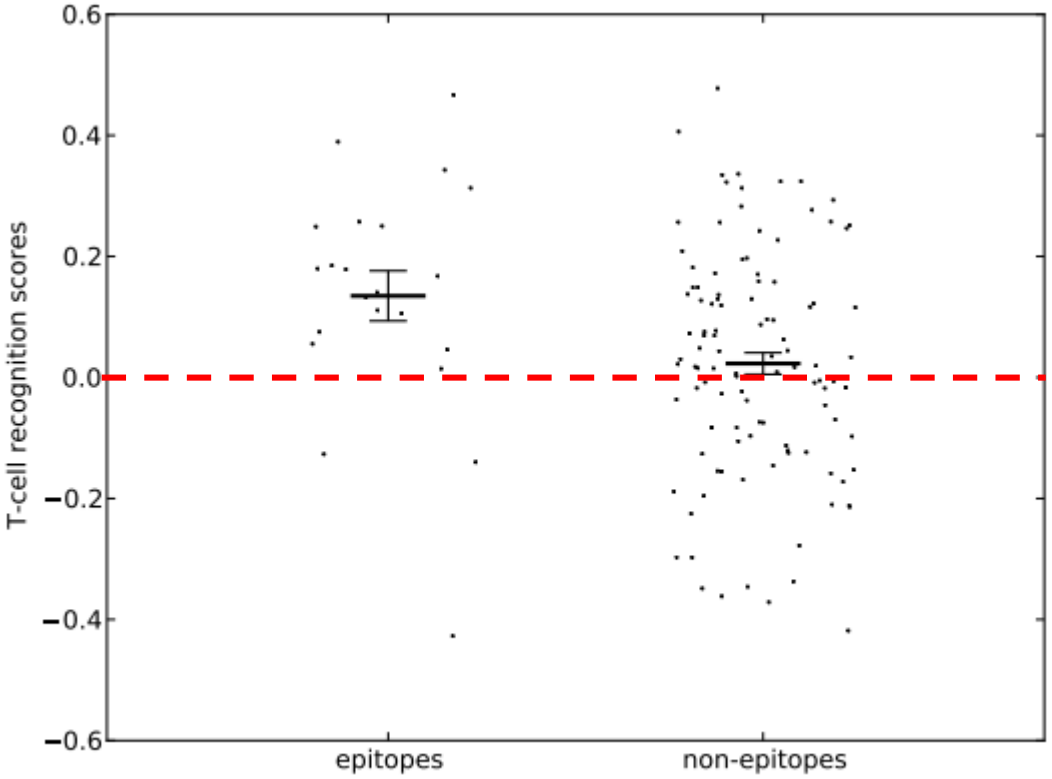
Peptide	Length	Score
FIAGLIAIV	9	0.27206
KAVYNFATC	9	0.16928
RLNEVAKNL	9	-0.0101
LITGRLQSL	9	-0.10776
FQPQNGQFI	9	-0.12392

Download Results Go back

- Scores are sums of propensity scores at all unmasked predictions
- High scores = peptide is more likely to be immunogenic

Caveats / Prediction performance

- Experimentally, many MHC binding peptides can be immunogenic (~50%)
- Cross validation gave AUC values ~ 0.65. Test on independent blind set gave AUC = 0.69
- Recommendation: Use as filter (cutoff 0) if high specificity is desired. Suggested cutoff is 0.



Questions?