Is there anything beyond MHC binding?

Morten Nielsen,
Department of Bio and Health Informatics, DTU
and
Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, Argentina
NO
What defines a T cell epitope?

- **MHC binding ✓**
  - MHC:peptide binding affinity
  - MHC:peptide complex stability
- Processing (Proteasomal cleavage, TAP)
- Other proteases
- T cell repertoire (similarity to self) and cross-reactivity
- Source protein abundance, cellular location and function
- ???
MHC Class I pathway
What about the other players?

Figure by Eric A.J. Reits
MHC ligands prediction
NetCTL (Larsen et al, 2005)

Immuno proteasome

TAP

MHC
How good are we at predicting T cell epitopes?
Evaluation. MHC ligands from SYFPEITHI

Sort on prediction score

Top Rank: AUC=1.0
Random Rank: AUC=0.5
ROC curves
ROC curves

AUC 0.1 = 0.85 ~ 1% FP
AUC 0.1 > 0.90 ~ 0.5% FP
AUC = 0.95 ~ 1% FP
AUC = 0.90 ~ 6% FP
Does proteasomal cleavage and TAP matter?

In 2005, we said yes (NetCTL, MHC-pathway)
2010, NetCTLpan says "not really"
MHC class I pathway co-evolution

What is the role of proteasome and TAP?

- Proteasome generates the peptides
What is the role of proteasome and TAP?

- Proteasome generates the peptides
- TAP shapes the length repertoire
What is the role of proteasome and TAP?

- Proteasome generates the peptides
- TAP shapes the length repertoire
So when is processing important?
So when is processing important?

\[ S = MHC + w_{cl} \cdot Cl_{C-term} + w_{tap} \cdot TAP \]

\[ W_{cl} = 0.225, \quad w_{tap} = 0.025 \]

<table>
<thead>
<tr>
<th>Data</th>
<th>Measure</th>
<th>NetCTLpan</th>
<th>NetMHCpan</th>
<th>( p ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train (9)</td>
<td>AUC</td>
<td>0.976</td>
<td>0.980</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>AUC(_{0.1})</td>
<td>\textbf{0.869}</td>
<td>0.852</td>
<td>0.002</td>
</tr>
<tr>
<td>Test (8/9/10/11)</td>
<td>AUC</td>
<td>0.977</td>
<td>0.979</td>
<td>0.273</td>
</tr>
<tr>
<td></td>
<td>AUC(_{0.1})</td>
<td>\textbf{0.863}</td>
<td>0.855</td>
<td>0.002</td>
</tr>
<tr>
<td>Test (HLA-C)</td>
<td>AUC</td>
<td>0.920</td>
<td>0.866</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>AUC(_{0.1})</td>
<td>\textbf{0.495}</td>
<td>0.307</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Cleavage predictions can correct inaccurate MHC binding predictions

- 2.2 training data 102 alleles, 1 HLA-C allele 3 peptides 3 binders
- 2.8 training data 150 alleles, 10 HLA-C alleles 2248 peptides 957 binders
Cleavage can correct inaccurate MHC binding predictions

Genome-wide Characterization of a Viral Cytotoxic T Lymphocyte Epitope Repertoire*S

Received for publication, July 10, 2003, and in revised form, August 1, 2003
Published, JBC Papers in Press, September 5, 2003, DOI 10.1074/jbc.M307417200

Weimin Zhong, Pedro A. Reche, Char-Chang Lai, Bruce Reinhold, and Ellis L. Reinherz;

From the Laboratory of Immunobiology and Department of Medical Oncology, Dana-Farber Cancer Institute and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115

A genome-wide search using major histocompatibility complex (MHC) class I binding and proteosome cleavage site algorithms identified 101 influenza A PR8 virus-derived peptides as potential epitopes for CD8+ T cell recognition in the H-2b mouse. Cytokine-based flow cytometry, ELISPOT, and cytotoxic T lymphocyte assays reveal that 16 are recognized by CD8+ T cells recovered directly ex vivo from infected animals, accounting for greater than 70% of CD8+ T cells recruited to lung after primary infection. Only six of the 22 highest affinity MHC class I binding peptides comprise cytotoxic T lymphocyte epitopes. The remaining non-immunogenic peptides have equivalent MHC affinity and MHC-peptide complex half-lives, eliciting T cell responses when given in adjuvant and with T cell receptor-ligand avidity comparable with their immunogenic counterparts. As revealed by a novel high sensitivity nanospray tandem mass spectrometry methodology, failure to process those predicted epitopes may contribute significantly to the absent response. These results have important implications for rationale design of CD8+ T cell-based vaccines.

Of the many CTL epitopes potentially available in a viral genome, antiviral CD8+ T cell responses usually focus on only one or two immunodominant peptides (6). Apparently, the great majority of potential CTL epitopes are silent under physiological conditions. The mechanisms for this phenomenon, termed immunodominance, are poorly understood. Current rational design of CD8+ T cell vaccines is mainly focused on the induction of antiviral immunity by immunodominant CTL epitopes, assuming that protective CD8+ T cell immunity is mediated by recognition of immunodominant CTL epitopes. Such a vaccination regimen is indeed efficient in conferring immune protection against a variety of viral infections in murine models (7–9). However, many viruses, especially RNA viruses, evade immune clearance of CD8+ T cell-mediated mechanisms by mutating immunodominant CTL epitopes, as observed in a number of persistent, as well as acute, viral infections (10, 11).

To circumvent this escape mechanism, new vaccination strategies that target conserved CTL epitope regions of viral
Cleavage can correct inaccurate MHC binding predictions

What defines a T cell epitope?

- **MHC binding ✓**
- **Processing (Proteasomal cleavage, TAP) (√)**
  - Only helps to improve inaccurate MHC binding predictions
- **Other proteases ✓**
- **T cell repertoire (similarity to self) and cross-reactivity**
- **Source protein abundance, cellular location and function**
- ...
T cell epitopes lacking MHC restriction

Trimming prior to binding

BoLA Class I epitopes, Work by Ivan Morrison and co-workers
BoLA epitopes the hard way

Trimming prior to binding

BoLA Class I epitopes, Work by Ivan Morrison and co-workers
## Known BoLA class I epitopes

<table>
<thead>
<tr>
<th>Pathogen</th>
<th>Epitope</th>
<th>MHC</th>
<th>Protein</th>
<th>#pep</th>
<th>#FP</th>
<th>FR</th>
<th>Alternative</th>
<th>FR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VGYPKVKEEML</td>
<td>6*01301 (HD6)</td>
<td>Tp1</td>
<td>2138</td>
<td>34</td>
<td>0.016</td>
<td>EELKKLGMML</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>SHEELKKLGML</td>
<td>6*04101 (T2B)</td>
<td>Tp2</td>
<td>662</td>
<td>14</td>
<td>0.021</td>
<td>GFDRDALF</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>DGFDRLAF</td>
<td>6*04101 (T2B)</td>
<td>Tp2</td>
<td>662</td>
<td>143</td>
<td>0.216</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KSSHGMGKVGK</td>
<td>2*01201 (T2A)</td>
<td>Tp2</td>
<td>662</td>
<td>3</td>
<td>0.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAQSLVCVL</td>
<td>T2C</td>
<td>Tp2</td>
<td>662</td>
<td>15</td>
<td>0.023</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QSLVCVLMK</td>
<td>2*01201 (T2A)</td>
<td>Tp2</td>
<td>662</td>
<td>5</td>
<td>0.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KTSIPNPCW</td>
<td>2*01201 (T2A)</td>
<td>Tp2</td>
<td>662</td>
<td>91</td>
<td>0.137</td>
<td>KTSIPNPCW</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>TGASIQTTL</td>
<td>3*00101 (AW10)</td>
<td>Tp4</td>
<td>2282</td>
<td>7</td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SKADVIAKY</td>
<td>T5</td>
<td>Tp5</td>
<td>586</td>
<td>2</td>
<td>0.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EFISFPISL</td>
<td>T7</td>
<td>Tp7</td>
<td>2850</td>
<td>184</td>
<td>0.065</td>
<td>FISFPISL</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>CGAE MNHFL</td>
<td>3*00101 (AW10)</td>
<td>Tp8</td>
<td>1726</td>
<td>12</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AKFPGMKKSK</td>
<td>1*02301 (D18.4)</td>
<td>Tp9</td>
<td>1302</td>
<td>68</td>
<td>0.052</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ave 0.046</td>
<td></td>
<td>0.015</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pathogen</th>
<th>Epitope</th>
<th>MHC</th>
<th>Protein</th>
<th>#pep</th>
<th># FP</th>
<th>FR</th>
<th>Alternative</th>
<th>FR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SKADVIAKY</td>
<td>BoLA-1:02301</td>
<td>Ta5</td>
<td>586</td>
<td>4</td>
<td>0.007</td>
<td>RSPMFEGTL</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>QRSPMFEGTL</td>
<td>BoLA-3:00201</td>
<td>Ta9</td>
<td>1306</td>
<td>15</td>
<td>0.011</td>
<td>RSPTFGGPL</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>ERSPTFGGPL₁</td>
<td>BoLA-3:00201</td>
<td>Ta9</td>
<td>1306</td>
<td>36</td>
<td>0.028</td>
<td>SKF PKM RM</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>SKF PKM RM</td>
<td>BoLA-1:02301</td>
<td>Ta9</td>
<td>1306</td>
<td>96</td>
<td>0.074</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KRKTEGYVF</td>
<td>BoLA-2:02201</td>
<td>Ta11</td>
<td>1630</td>
<td>23</td>
<td>0.014</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ave 0.027</td>
<td></td>
<td>0.006</td>
</tr>
</tbody>
</table>

Table 3. *In-silico* prediction of known *T. parva* and *T. annulata* CTL epitopes. The columns are; (Epitope) the CTL epitope sequence from the literature, (MHC) the corresponding restriction element, (Protein) the source protein ID, (#pep) the number of unique 8-11mer peptides contained within the source protein, (#FP) the number of false positive predictions (peptides with a predicted binding affinity value stronger than the known epitope), (FR) the false positive ratio (#FP/#pep), (Alternative) the predicted alternative epitope, and (FR) the false positive ratio for the alternative epitope. If no alternative epitope was predicted, the last column is left empty. (Ave) gives the average value for the peptide subset. Data from [3] and [27]. ¹ Allelic variant of QRSPMFEGTL.

Phil Toye and Vish Nene, ILRI
Tetramer validation

SHEELKKLGM

Tp2<sub>27-37</sub>

EELKKLGM

Tp2<sub>29-37</sub>
What defines a T cell epitope?

- **MHC binding ✓**
- **Processing (Proteasomal cleavage, TAP) (✓)**
  - Only helps to improve inaccurate MHC binding predictions
- **Other proteases ✓**
- **T cell repertoire (similarity to self) and cross-reactivity**
- **Source protein abundance, cellular location and function**
- ...
T cell cross-reactivity

Can we define the shape and size of these “Holes”?

Negative T cell selection induces “Holes” in the T cell repertoire
T cell cross reactivity to self

• We have a limited understanding of T cell cross reactivity
  – What defines the radius of cross-reactivity?

• What is self?
  – All human proteins?
  – Protein expressed in Thymus?
  – Human microme?

• Most studies are made in mice
  – Small self, Small T cell repertoire
  – Most likely not possible to extrapolate to human

• Successful case stories are limited to one or two alleles
  – Frankild et al. (HLA-A2:01), Toussaint et al. (HLA-B35:01)

• Filtering for self-similarity makes sense in theory but we have no clue how
  – Exclude identical hits, yes, but 80% identical hits, why 80?
T cell interaction models

T cell cross-reactivity to self

Frankild et al., PLoS One, 2008

T cell cross-reactivity to microbiome

Why are T cell epitope 9 amino acids long?

- The immune system must distinguish between foreign and self
- If ligands were too short, most foreign peptides would also be found as self peptides
- If ligands were too long, very few foreign peptides would be present
- Some intermediate length is optimal
- What is this length?
Why are T cell epitope 9 amino acids long?
Why are T cell epitope 9 amino acids long?

Discriminating self from nonself with short peptides from large proteomes

Burroughs, NJ; de Boer, RJ; Kesmir, C

Immunogenetics — 2004, Volume 56, Issue 5, pp. 311-319
Figure 1. Viral and bacterial self/nonself overlaps for peptides of different lengths.


http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002412
Table 1. Summary of all the average self/nonself overlaps obtained using peptides predicted to be presented on HLA molecules.

<table>
<thead>
<tr>
<th></th>
<th>Recognized peptide positions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self</td>
</tr>
<tr>
<td>percentage</td>
<td>100</td>
</tr>
<tr>
<td>Exact</td>
<td>100</td>
</tr>
<tr>
<td>Degenerate</td>
<td>100</td>
</tr>
</tbody>
</table>

Overlaps were determined using all positions of the peptide (P1–9), the non-anchor positions (P1 and P3–8) or the middle positions between the anchors (P3–8). Further, overlaps were determined as exact, i.e., every position should be identical, or as degenerate, i.e., with 1 or 2 substitutions being allowed to mimic T-cell recognition (see Methods). Finally, overlaps with 100% or (a randomly chosen) 50% of the human proteome are shown. Self/nonself overlaps indicated with a star (*) are shown per HLA molecule in Figure 2.

doi:10.1371/journal.pcbi.1002412.t001
Figure 3. TCR interactions per peptide position.


http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002412
Table 1. Summary of all the average self/nonself overlaps obtained using peptides predicted to be presented on HLA molecules.

<table>
<thead>
<tr>
<th>Recognized peptide positions</th>
<th>Self percentage</th>
<th>P1–9 (complete)</th>
<th>P1 and P3–8 (non-anchor)</th>
<th>P3–8 (middle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact</td>
<td>100</td>
<td>0.15%*</td>
<td>0.41%</td>
<td>2.7%*</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.09%</td>
<td>0.25%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Degenerate</td>
<td>100</td>
<td>0.7%</td>
<td>5.2%</td>
<td>29%*</td>
</tr>
</tbody>
</table>

Overlaps were determined using all positions of the peptide (P1–9), the non-anchor positions (P1 and P3–8) or the middle positions between the anchors (P3–8). Further, overlaps were determined as exact, i.e. every position should be identical, or as degenerate, i.e. with 1 or 2 substitutions being allowed to mimic T-cell recognition (see Methods). Finally, overlaps with 100% or (a randomly chosen) 50% of the human proteome are shown. Self/nonself overlaps indicated with a star (*) are shown per HLA molecule in Figure 2.

doi:10.1371/journal.pcbi.1002412.t001

http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002412
Conserved anchors

Copyright © 2003, Elsevier Science (USA). All Rights Reserved.
T cell interaction models

T cell propensity model

created by Seq2Logo

Calis et al., Plos Compu, 2013
T cell interaction models

T cell propensity model

NetTepi

Calis et al., Plos Compu, 2013

Trolle et al. Accepted, Immunogenetics, 2014
NetTepi 1.0 Server

NetTepi 1.0 predicts T-cell epitopes from protein sequences. The method integrates three prediction types, peptide-MHC binding affinity, peptide-MHC stability and T-cell propensity.

The server allows for predictions of T-cell epitopes restricted to 13 different Human MHC (HLA) alleles representing 11 of the 12 common HLA-A and B Supertypes as defined by Lund et al (2004).

Predictions of lengths 8-14: Predictions can be made for lengths between 8 and 14 for all alleles using an approximation algorithm. Note that only lengths 9 and 10 have been thoroughly benchmarked. Caution should be taken with predictions for all other lengths.

Prediction values are calculated as a weighted sum of binding affinity, stability and T-cell propensity prediction scores. A % Rank score based on predictions for 200,000 random natural peptides is also provided.

Peptide-MHC binding affinity predictions are obtained using the NetMHCoons method. Peptide-MHC stability predictions are obtained using the NetMHCstab method. T-cell propensity is predicted using the immunogenicity model described by Calis et al (2013).

SUBMISSION

Paste a single sequence or several sequences in FASTA format into the field below:

or submit a file in FASTA format directly from your local disk:

Peptide length (several lengths are possible):
- 8mer peptides
- 9mer peptides
- 10mer peptides
- 11mer peptides

Select loci/species
- HLA-A

Select Allele(s)
- HLA-A*01:01
- HLA-A*02:01

Allows prediction for 13 different HLA molecules
What defines a T cell epitope?

- MHC binding ✓
- Processing (Proteasomal cleavage, TAP) ✓
- Other proteases ✓
- T cell repertoire (similarity to self) and cross-reactivity ✓
- Source protein abundance, cellular location and function

• ...

But, can we find the haystack?
MTB (mycobacterium tuberculosis)

• Bacterial genome coding for more than 4000 proteins
• 700 known epitopes, found in only 30 proteins (ORFs)
MTB (mycobacterium tuberculosis)

• Bacterial genome coding for more than 4000 proteins
• 700 known epitopes, found in only 30 proteins (ORFs)
• Is this biology, or history?
  - More than 500,000 unique 9mer peptides
  - Where to start?
    • Each HLA allele will binding ~5000 of these peptides.
Table 2. Proliferative CD8+ T cell responses against peptides selected to be restricted by each of the 3 supertypes HLA-A2, -A3 and -B7. Results are shown for each of the 8 antigen selections.

<table>
<thead>
<tr>
<th>Selection</th>
<th>A2 peptides</th>
<th>A3 peptides</th>
<th>B7 peptides</th>
<th>all 3 supertypes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># tested</td>
<td>#+ peptide</td>
<td>% reactivity</td>
<td># tested</td>
</tr>
<tr>
<td>TB-VAC</td>
<td>8</td>
<td>1</td>
<td>13%</td>
<td>9</td>
</tr>
<tr>
<td>TB-CD8</td>
<td>17</td>
<td>1</td>
<td>6%</td>
<td>23</td>
</tr>
<tr>
<td>BestPred</td>
<td>16</td>
<td>1</td>
<td>6%</td>
<td>20</td>
</tr>
<tr>
<td>Cons</td>
<td>18</td>
<td>1</td>
<td>6%</td>
<td>22</td>
</tr>
<tr>
<td>DOS/LAG</td>
<td>19</td>
<td>9</td>
<td>47%</td>
<td>23</td>
</tr>
<tr>
<td>Bepi</td>
<td>18</td>
<td>9*</td>
<td>50%</td>
<td>10</td>
</tr>
<tr>
<td>Secret</td>
<td>19</td>
<td>14</td>
<td>74%</td>
<td>18</td>
</tr>
<tr>
<td>PredSecret</td>
<td>15</td>
<td>8</td>
<td>53%</td>
<td>14</td>
</tr>
<tr>
<td>TOTAL</td>
<td>130</td>
<td>44</td>
<td>34%</td>
<td>139</td>
</tr>
</tbody>
</table>

# tested, number of peptides tested for this supertype; #+ positive, number of peptides which gave a positive proliferative CD8+ T cell response; % reactivity, percentage of positive peptides with a proliferative CD8+ T cell response.

All selections equally good, except for conservation

Figure 1. Flow diagram of the approach used in this study. The complete genome sequence of *Mycobacterium tuberculosis* strain H37Rv was used for CTL epitope prediction. Mycobacterial proteins were included based on 8 different selection criteria: 1. TB-VAC: proteins used in vaccine trials, 2. TB-CD8: proteins with known CD8 epitopes, 3. BestPred: Proteins containing peptides with the best prediction values, 4. Cons: Conserved proteins, 5. DOS/LAG: Proteins encoded in the *Mtb* dos regulon, 6. Bepi: Proteins with B cell epitopes, 7. Secret: Secreted proteins, 8. PredSecret: Proteins predicted to be secreted. (see Material and Methods for details). The number of selected peptides in each selection is given in parenthesis. Epitope predictions were done for HLA-A2, -A3, and -B7. A total of 432 peptides were synthesized and binding affinities were measured on an *in vitro* biochemical peptide-HLA class I binding assay (157). CD8 T cell proliferative responses were performed using a proliferation assay on PBMC from PPD+ donors and flow cytometry analysis.
Conclusions

- Rational epitope discovery is feasible
  - Prediction methods are an important guide for epitope identification
  - Given a protein sequence and an HLA molecule, we can predict the peptide binders (find the needle in the haystack)
- Pan-specific MHC prediction method can deal with the immense MHC polymorphism
- All T cell epitopes have specific MHC restrictions matching their host
  - There is no such thing as a non-binding T cell epitope
- Processing have little impact in predicting of CTL epitopes
- We begin to have some clues about what features other than MHC binding it takes to become a T cell antigen