Predicting Peptide Binding to Major Histocompatibility Complex (MHC) molecules

ImmunoInformatics - Computational Immunology

Email: achlichl@mbg.duth.gr

Ass. Prof. Katerina Chlichlia

12.11.2016
• Bioinformatics has broad applicability to Immunology → IMMUNOINFORMATICS

• Development of in-silico models of entire systems - towards a virtual immune system

TODAY we will apply bioinformatic tools for identifying antigenic epitopes - our aim will be to predict peptide binding to particular MHC molecules
Overview of the lecture

• Introduction to the adaptive immune system - antigen recognition by T cells - EPITOPES
• Major histocompatibility complex (MHC)
• Characteristics of peptides bound to MHC class I vs. MHC class II
• Computational approaches for predicting peptide binding - databases and prediction servers
• Exercise: Use of bioinformatics tools to search for epitopes - to predict peptide MHC-binding
search for epitopes
EPITOPE = antigenic determinant

Different antigenic determinants

Antigen
EPITOPE = antigenic determinant

2 types of epitopes

Discontinuous Epitopes

Continuous Epitopes
Adaptive Immunity

Naive B cell ➔ Antibodies

Naive T cell ➔ Effector T cells

Days
1 ➔ 3 ➔ 5
Time after infection
Adaptive Immunity

Humoral immunity
- Extracellular microbes
- B Lymphocyte
- Elimination of microbes
- Activation of macrophage leading to microbial killing

Cell-mediated immunity
- Dendritic cells display antigen fragments to T cells
- Phagocytosed microbes in macrophage
- Intracellular microbes within infected cell
- T cells release cytokines
- T Cell
- Lysis of infected cell
Fig. 1.12  T cells recognize antigens that originate within other cells, such as viral peptides from infected cells. They do this by binding specifically to antigenic peptides presented on the surface of the infected cells by molecules encoded by the major histocompatibility complex (MHC molecules). The T cells use their specific receptors (TCRs) to recognize the unique combination of MHC molecule plus antigenic peptide. Unlike B cells, which recognize just a portion of the antigen, a T cell recognizes residues from both the MHC molecule and the antigen peptide.
The **Major Histocompatibility Complex (MHC)** constitutes an important part of the immune system. During infection, pathogenic proteins are processed into peptide fragments by the antigen processing machinery. These **peptides** bind to MHC molecules and the MHC-peptide complex is then transported to the cell membrane where it elicits an immune response via **T-cell binding**. The molecular mechanism of this process is of great importance in determining the aetiology of various diseases and in the design of effective vaccines.
2 types of MHC molecules

MHC τάξης I $\Rightarrow$ CD8 / Tc
peptides in the cytoplasm
*endogenous proteins*

MHC τάξης II $\Rightarrow$ CD4 / $T_H$
peptides in vesicles
*exogenous proteins*
Εξωγενή Αγ  

ΜΗΣ τάξης ΙΙ

Ενδογενή Αγ  

ΜΗΣ τάξης Ι

A

APC  

T cell

CD4  

HELP

B

APC  

T cell

CD8  

KILL
The concept of MHC restriction (Nobelprize 1996)

The T cell receptor specifically recognizes both peptide and MHC molecule.
The concept of MHC restriction (Nobelprize 1996, Zinkernagel and Doherty)
The T cell receptor specifically recognizes both peptide and MHC molecule

Fig. 5.19 A mouse of the H-2^b haplotype is primed with virus and the Tc cells thus generated are isolated and tested for their ability to kill H-2^b and H-2^k cells infected with the same virus. The Tc cells kill H-2^b, but not H-2^k cells. In this instance, it is the H-2K class I gene product which is presenting the antigen to the T cells. The T cell is recognizing a specific structure produced by the association of a specific MHC molecule with a specific viral antigen.
Antigen Processing (επεξεργασία) and Presentation (παρουσίαση)

Protein antigen in cell

plasma membrane

Antigen processing by breakdown of protein

Presentation of peptide by MHC molecule

Figure 5.10 The Immune System, 3rd ed. (© Garland Science 2009)
Antigen processing and Peptide binding to MHC τάξης I
Presentation to CD8+ T cell

Endogenous Antigens
Antigen Processing for MHC class I

- Virus
- Abnormal folding
- MHC I molecule
- TAP transporter
- ER
- Ubiquitin ligase
- Ubiquitin
- Proteasome
- Peptides
Antigen processing and Peptide binding to MHC τάξης ΙΙ
Presentation to CD4⁺ T cells

Exogenous Antigens
The two classes of MHC have similar three-dimensional structures.
MHC protein HLA-A2 with HIV peptide

HIV 9mer peptide

α1

β2m

α2

α3
Strong selective pressure for pathogens to escape presentation by MHC

It is extremely difficult for the pathogens to evade MHC molecules and immune surveillance. Why?

MHC is polygenic. Several different MHC class I and II genes.

MHC is highly polymorphic. Multiple variants of the gene within the population.
MHC highly polymorphic. More than 200 alleles that occur in high frequency.

Most individuals heterozygous at MHC locus.

MHC haplotype.

Expression co-dominant.

MHC polymorphism triggers T-cell reactions that can reject transplanted organs.
### Mouse H-2 complex

<table>
<thead>
<tr>
<th>Complex</th>
<th>I</th>
<th>II</th>
<th>III (C' proteins)</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHC class</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>K</td>
<td>IA</td>
<td>S</td>
<td>D</td>
</tr>
<tr>
<td>Gene products</td>
<td>H-2K</td>
<td>IA αβ</td>
<td>IE αβ</td>
<td>TNF-α, TNF-β</td>
</tr>
</tbody>
</table>

### Human HLA complex

<table>
<thead>
<tr>
<th>Complex</th>
<th>II</th>
<th>III (C' proteins)</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHC class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>DP</td>
<td>DQ</td>
<td>DR</td>
</tr>
<tr>
<td>Gene products</td>
<td>αβ</td>
<td>αβ</td>
<td>αβ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C4, C2, BF</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A</td>
</tr>
</tbody>
</table>

- HLA-B
- HLA-C
- HLA-A
In humans, MHC is called human leukocyte antigen (HLA).

In humans:

- There are three MHC class I genes (coding for the α chain): $\text{HLA-A, HLA-B, και HLA-C}$

- There are 3 pairs of MHC class II genes (coding for the α and β chains): $\text{HLA-DR, HLA-DP, και HLA-DQ}$
In mice, MHC is called H-2

In mice:
• There are three MHC class I genes (coding for the α chain): H-2K, H-2D, and H-2L

• There are 2 pairs of MHC class II genes (coding for the α and β chains): H-2A, H-2E
Mice

codominant expression of H-2 genes
Humans

**Codominant expression of HLA genes**

- Chromosome 6 from father
- Chromosome 6 from mother

- HLA-DP, HLA-DQ, HLA-DR
- HLA-B, HLA-C, HLA-A

- MHC II

- MHC I

- HLA-DP, HLA-DQ, HLA-DR

- Chromosome 15

- α, β

- 8-20

- Max. 6

- α

- β2 microglobulin
Most Humans are heterozygous at the MHC

HLA-A2  A32
HLA-Cw2  Cw3
HLA-B27  B65
HLA-DR1  DR11
HLA-DQ5  DQ1
HLA-DP4  DP6

MATERNAL  PATERNAL
(a) Class I MHC
Class II MHC
Anchor residues

(προσδετικά κατάλοιπα ή κατάλοιπα αγκυροβόλησης)

are residues in the peptide that bind to specific pockets on the MHC I or MHC II resulting in some specificity of interactions with MHC.
Closed peptide-binding cleft:
Conserved AA residues bind the terminal -NH2 and -COOH groups
MHC-II peptide binding

MHC τάξης II > 13 αα
## Examples

<table>
<thead>
<tr>
<th>MHC molecule</th>
<th>Amino acid sequence of peptide-binding motifs and bound peptides</th>
<th>Source of bound peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class I</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-A^*0201</td>
<td>Peptide-binding motif: [L/M] [ ] [ ] [ ] [V] [ ] [V/L]</td>
<td>HIV reverse transcriptase</td>
</tr>
<tr>
<td></td>
<td>Bound peptide: [I] [L] [K] [E] [P] [V] [H] [G] [V]</td>
<td></td>
</tr>
<tr>
<td>HLA-B^*2705</td>
<td>Peptide-binding motif: [R] [ ] [ ] [ ] [ ] [ ] [R/K]</td>
<td>Influenza A nucleoprotein</td>
</tr>
<tr>
<td></td>
<td>Bound peptide: [S] [R] [Y] [W] [A] [I] [R] [T] [R]</td>
<td></td>
</tr>
<tr>
<td><strong>Class II</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-DRB1^*0401</td>
<td>Self peptide: [G] [V] [Y] [F] [Y] [L] [Q] [W] [G] [R] [S] [T] [L] [V] [S] [S]</td>
<td>Igκ light chain</td>
</tr>
<tr>
<td>HLA-DQA1^*0501, HLA-DQB1^*0301</td>
<td>Self peptide: [I] [P] [E] [L] [N] [K] [V] [A] [R] [A] [A]</td>
<td>Transferrin receptor</td>
</tr>
</tbody>
</table>

Figure 5.30 The Immune System, 3ed. (© Garland Science 2009)
<table>
<thead>
<tr>
<th>Table 7-2 Peptide Binding by Class I and Class II MHC Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Peptide-binding domain</strong></td>
</tr>
<tr>
<td><strong>Nature of peptide-binding cleft</strong></td>
</tr>
<tr>
<td><strong>General size of bound peptides</strong></td>
</tr>
<tr>
<td><strong>Peptide motifs involved in binding to MHC molecule</strong></td>
</tr>
<tr>
<td><strong>Nature of bound peptide</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Peptide-binding domain</strong></td>
</tr>
<tr>
<td><strong>Nature of peptide-binding cleft</strong></td>
</tr>
<tr>
<td><strong>General size of bound peptides</strong></td>
</tr>
<tr>
<td><strong>Peptide motifs involved in binding to MHC molecule</strong></td>
</tr>
<tr>
<td><strong>Nature of bound peptide</strong></td>
</tr>
</tbody>
</table>
### TABLE 7-1  H-2 HAPLOTYPES OF SOME MOUSE STRAINS

<table>
<thead>
<tr>
<th>Prototype strain</th>
<th>Other strains with the same haplotype</th>
<th>Haplotypes</th>
<th>H-2 alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBA</td>
<td>AKR, C3H, B10.BR, C57BR</td>
<td>k</td>
<td>K: k, IA: k, IE: k, S: k, D: k</td>
</tr>
<tr>
<td>DBA/2</td>
<td>BALB/c, NZB, SEA, YBR</td>
<td>d</td>
<td>K: d, IA: d, IE: d, S: d, D: d</td>
</tr>
<tr>
<td>C57BL/10 (B10)</td>
<td>C57BL/6, C57L, C3H.SW, LP, 129</td>
<td>b</td>
<td>K: b, IA: b, IE: b, S: b, D: b</td>
</tr>
<tr>
<td>A</td>
<td>A/He, A/Sn, A/Wy, B10.A</td>
<td>a</td>
<td>K: a, IA: k, IE: k, S: d, D: d</td>
</tr>
<tr>
<td>A.SW</td>
<td>B10.S, SJL</td>
<td>s</td>
<td>K: s, IA: s, IE: s, S: s, D: s</td>
</tr>
<tr>
<td>A.TL</td>
<td></td>
<td>tL</td>
<td>K: s, IA: k, IE: k, S: k, D: d</td>
</tr>
<tr>
<td>DBA/1</td>
<td>STOLI, B10.Q, BDP</td>
<td>q</td>
<td>K: q, IA: q, IE: q, S: q, D: q</td>
</tr>
</tbody>
</table>
(a) Mating of inbred mouse strains with different MHC haplotypes

Homologous chromosomes with MHC loci

- H-2<sup>b</sup> parent: b/b
- H-2<sup>k</sup> parent: k/k

F<sub>1</sub> progeny (H-2<sup>b/k</sup>)
(b) Skin transplantation between inbred mouse strains with same or different MHC haplotypes

Parental recipient

$k/k$

Skin graft donor

Parent

Progeny recipient

$b/k$

$b/b$

$k/k$

$b/k$

$b/b$

$k/k$

$b/k$

$b/b$

$k/k$

$b/k$
How to search for T cell epitopes?
Understanding the specificity and sensitivity of the binding process of peptides to the respective MHC is challenging. Since there are over 512 billion potential binding peptides for each MHC molecule, an empirical approach is not feasible.

Computational approaches offer the promise of predicting peptide binding, thus dramatically reducing the number of peptides proceeding to experimental verification.
Immunology Databases

- Management and analysis of immunological data
- Improve the efficiency of immunological research
Databases and Prediction Servers

**MHC-binding peptide databases**
- SYFPEITHI
- MHCPEP
- JenPep
- FIMM
- MHCBN
- HLALigand/Motif database
- HIV Molecular Immunology database
- EPIMHC

**Prediction of MHC binding**
- BIMAS
- SYFPEITHI
- PREDEPP
- Epipredict
- Predict
- Propred
- MHCPred
- NetMHC
Peptide binding prediction methods can be categorised into three major groups:

1. **motif- and scoring matrix-based methods** (sequence-based approach)
2. **artificial intelligence-based methods** (sequence-based approach)
3. **structure-based methods** (based on structural features and the distribution of energy between the binding peptide and the MHC molecule)

1+2 are generally based on common sequence motifs in peptides known to bind to MHC molecules

3 the development of structure-based methods has been relatively slow compared to sequence-based methods

Comparisons of various methods showed that the best sequence-based methods significantly outperform structure-based methods

Like T cell epitope prediction algorithms, there are also B cell epitope prediction algorithms
B cell epitope prediction algorithms:
• Hopp and Woods - 1981
• Welling et al - 1985
• Parker & Hodges - 1986
• Kolaskar & Tongaonkar - 1990
• Kolaskar & Urmila Kulkarni - 1999, 2005
• Haste et al., 2006

T cell epitope prediction algorithms:
• Margalit, Spouge et al - 1987
• Rothbard & Taylor - 1988
• Stille et al - 1987
• Tepitope - 1999
• Rammensee et al. 1999

REVIEW

Hans-Georg Rammensee · Jutta Bachmann
Niels Philipp Nikolaus Emmerich
Oskar Alexander Bachor · Stefan Stevanović

SYFPEITHI: database for MHC ligands and peptide motifs

www.syfpeithi.de
SYFPEITHI [22] is the name of a database of MHC ligands and peptide motifs of humans and other vertebrate species. The database facilitates search for peptides as well as prediction of T-cell epitopes. The prediction of T-cell epitopes is based on an algorithm that takes into account the position of amino acids in the peptide, such as the anchor position, unusual anchor position, and auxiliary anchor position. Preferred amino acids as well as amino acids whose presence at particular positions is undesirable for peptide binding are also taken into account and are scored accordingly.

The scoring system of the algorithm evaluates every amino acid within a given peptide. The values are assigned to the amino acids at various positions in a peptide based on the frequency of occurrence of the respective amino acids in natural ligands, T-cell epitopes or binding peptides. The value of an amino acid can vary from a high positive value, say 15, the highest value that is attributed to ideal/optimal anchor residues to a low positive value of 1, which is attributed to amino acids that are only slightly preferred to a negative value which is attributed to amino acids that are disadvantageous to peptide binding at a particular position in the peptide. The values at each position are summed up to assign a final score for the peptide that acts as a T-cell epitope.
SYFPEITHI: database for MHC ligands and peptide motifs

Abstract The first version of the major histocompatibility complex (MHC) databank SYFPEITHI: database for MHC ligands and peptide motifs, is now available to the general public. It contains a collection of MHC class I and class II ligands and peptide motifs of humans and other species, such as apes, cattle, chicken, and mouse, for example, and is continuously updated. All motifs currently available are accessible as individual entries. Searches for MHC alleles, MHC motifs, natural ligands, T-cell epitopes, source proteins/organisms and references are possible. Hyperlinks to the EMBL and PubMed databases are included. In addition, ligand predictions are available for a number of MHC allelic products. The database content is restricted to published data only.
Table 4  Peptide motif and natural ligands of HLA-B*1510 (Seeger and co-workers, in press)

<table>
<thead>
<tr>
<th>Position</th>
<th>Source</th>
<th>Accession No.</th>
<th>EMBL database</th>
</tr>
</thead>
<tbody>
<tr>
<td>H L</td>
<td>HLA-DP α-chain (220–229)</td>
<td>X00457</td>
<td></td>
</tr>
<tr>
<td>I E P A V V V F</td>
<td>Cytochrome C reductase (66–73)</td>
<td>M36647</td>
<td></td>
</tr>
<tr>
<td>Y A D G I R R M</td>
<td>Heat shock protein 90 β (440–450)</td>
<td>M16660</td>
<td></td>
</tr>
<tr>
<td>T S E P M P E</td>
<td>Elongation factor 2 (489–497)</td>
<td>M19997</td>
<td></td>
</tr>
<tr>
<td>G G N K T</td>
<td>60 S acidic rib. protein PQ (67–75)</td>
<td>M17885</td>
<td></td>
</tr>
<tr>
<td>E K R A</td>
<td>Chaperonin cont. TCP-1 η (282–290)</td>
<td>AF026292</td>
<td></td>
</tr>
<tr>
<td>Q V</td>
<td>60 S ribosomal protein L8 (49–58)</td>
<td>Z28407</td>
<td></td>
</tr>
<tr>
<td>G H D P R A Q G T L</td>
<td>Septin 2 homologue (70–78)</td>
<td>D50918</td>
<td></td>
</tr>
<tr>
<td>D H C V A H K L</td>
<td>Transcription activator SNF2L4 (899–907)</td>
<td>U29175</td>
<td></td>
</tr>
<tr>
<td>I H E D S T N R R L</td>
<td>Human EST</td>
<td>T6718</td>
<td></td>
</tr>
<tr>
<td>E H A H N M R V M</td>
<td>Actin-related protein Arp2 (402–410)</td>
<td>AF006082</td>
<td></td>
</tr>
<tr>
<td>G H L E N N P A L</td>
<td>Cyclin-dep. kin. reg. subunit 1 (59–67)</td>
<td>X54941</td>
<td></td>
</tr>
<tr>
<td>H H S G A K V V L</td>
<td>DNA repl. lic. factor MCM4 (694–702)</td>
<td>X74794</td>
<td></td>
</tr>
<tr>
<td>I H D P G R G A P L</td>
<td>HBV X interacting protein (40–48)</td>
<td>AF029890</td>
<td></td>
</tr>
<tr>
<td><strong>Actin-related protein 2 (ARP2; human)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA pos.</td>
<td>Score</td>
<td>Sequence</td>
<td></td>
</tr>
<tr>
<td>378</td>
<td>19</td>
<td>YOEGKVRVL</td>
<td></td>
</tr>
<tr>
<td>227</td>
<td>15</td>
<td>IESEQKLAL</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>14</td>
<td>EHIFPALVG</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>14</td>
<td>EASELSVML</td>
<td></td>
</tr>
<tr>
<td>164</td>
<td>14</td>
<td>THICPVPYEG</td>
<td></td>
</tr>
</tbody>
</table>

| **Septin 2 homologue (SEP2; human)** |   |   |
| AA pos. | Score | Sequence |
| 70      | 25    | THTOIPGVO |
| 156     | 22    | GHSLLKSLDL |
| 371     | 22    | LHQDEKELKL |
| 129     | 17    | EELKIRRVL |
| 84      | 15    | DLQESNVRL |

| **60 S acidic ribosomal protein RQ (RLA0; human)** |   |   |
| AA pos. | Score | Sequence |
| 67      | 25    | GHLENPPAL |
| 80      | 19    | PHIRGNVGF |
| 241     | 18    | IINGYKVVL |
| 46      | 15    | SLRGKAVVL |
| 196     | 15    | GSIYNPEVL |

<p>| <strong>Elongation factor 2 (EF2; human)</strong> |   |   |
| AA pos. | Score | Sequence |
| 489     | 24    | EHAHNMRVM |
| 356     | 18    | IHLPSPVTA |
| 147     | 17    | IAERIKPVL |
| 491     | 17    | AHNMRVMKF |
| 843     | 16    | GLKEGIPAL |</p>
<table>
<thead>
<tr>
<th>Position</th>
<th>Source protein</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (2877–2894)</td>
<td>Malcherek et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (2877–2893)</td>
<td>Malcherek et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (2877–2892)</td>
<td>Malcherek et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>α1-Antitrypsin (149–164)</td>
<td>Malcherek et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>LDL receptor (518–532)</td>
<td>Malcherek et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>IG2a (384–395)</td>
<td>Malcherek et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Unknown</td>
<td>Malcherek et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Unknown</td>
<td>Malcherek et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Transferin receptor (618–632)</td>
<td>Malcherek et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Invariant chain (97–111)</td>
<td>Ribery et al. 1992; Chicz et al. 1993; Sette et al. 1992</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Invariant chain (97–119)</td>
<td>Ribery et al. 1992; Chicz et al. 1993; Sette et al. 1992</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Invariant chain (97–120)</td>
<td>Ribery et al. 1992; Chicz et al. 1993; Sette et al. 1992</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Invariant chain (98–113)</td>
<td>Ribery et al. 1992; Chicz et al. 1993; Sette et al. 1992</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Invariant chain (98–117)</td>
<td>Ribery et al. 1992; Chicz et al. 1993; Sette et al. 1992</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Invariant chain (99–116)</td>
<td>Ribery et al. 1992; Chicz et al. 1993; Sette et al. 1992</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Invariant chain (99–119)</td>
<td>Ribery et al. 1992; Chicz et al. 1993; Sette et al. 1992</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>HLA-A30 (52–67)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Invariant chain (131–149)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>ACh receptor (289–304)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>ICAM-2 (64–76)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>IFNγ receptor (128–147)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>IFNγ receptor (128–148)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>CytB5 (155–172)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (1207–1224)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (1276–1295)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (1276–1292)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (1276–1291)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (1273–1290)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (1273–1291)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (1273–1290)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (1273–1289)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (1276–1290)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (1276–1290)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Apolipoprotein B-100 (1294–1310)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>HLA-A2 (103–117)</td>
<td>Chicz et al. 1993</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>Lol pol. P1 (171–190)</td>
<td>Geluk et al. 1994</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>HSP65 (cattle) (3–13)</td>
<td>Hawes et al. 1995</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>M. tuberculosis 30/31 kD protein</td>
<td>Geluk et al. 1997 (56–65)</td>
</tr>
<tr>
<td>2 3 4 5 6 7 8 9</td>
<td>M. tuberculosis HSP70 (257–269)</td>
<td>Geluk et al. 1997</td>
</tr>
</tbody>
</table>
Epitope prediction

This page allows you to find out the ligation strength to a defined HLA type for a sequence of amino acids. The algorithms used are based on the book "MHC Ligands and Peptide Motifs" by H.G. Rammensee, J. Bachmann and S. Stevanovic. The probability of being processed and presented is given in order to predict T-cell epitopes.

1. Select MHC type
If you chose "all", max. sequence length is 100 amino acids (letters)!

   all
   H2-Ad
   H2-Ak
   H2-Db
   H2-Ed

   Hold down ctrl key when clicking
to select multiple items

2. Choose a mer

   octamers (8 aa)
   nonamers (9 aa)
   decamers (10 aa)
   endecamers (11 aa)
   15-mers (15 aa) for MHC Type II only
   all mers

3. Paste your sequence here:
Max. input 2048 amino acids (letters)!
Letters only, no numbers or non-ASCII-symbols please.
You may use 'SYFPEITHI' with H2-Kd to see an example.

SYFPEITHI

4. Choose Run to start analysis

Run   Reset   Home
Immune epitope database analysis resource (IEDB-AR)

Qing Zhang¹, Peng Wang¹, Yohan Kim¹, Pernille Haste-Andersen², John Beaver³, Philip E. Bourne³, Huynh-Hoa Bui¹, Soren Buus², Sune Frankild², Jason Greenbaum¹, Ole Lund², Claus Lundegaard², Morten Nielsen², Julia Ponomarenko³, Alessandro Sette¹, Zhanyang Zhu³ and Bjoern Peters¹,*

¹Immune Epitope Database and Analysis Resource (IEDB-AR), La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA, ²Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark, DK-2800 Lyngby, Denmark and ³San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA, USA

Received January 31, 2008; Revised April 14, 2008; Accepted April 20, 2008

www.iedb.org

ABSTRACT

We present a new release of the immune epitope database analysis resource (IEDB-AR, http://tools.immuneepitope.org), a repository of web-based tools for the prediction and analysis of immune epitopes. New functionalities have been added to most of the previously implemented tools, and a total of eight new tools were added, including two B-cell epitope prediction tools, four T-cell epitope prediction tools and two analysis tools.
Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion

Martin Christen Frølund Thomsen and Morten Nielsen*

Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Received January 6, 2012; Revised April 30, 2012; Accepted May 2, 2012
http://www.cbs.dtu.dk/biotoools/Seq2Logo/

ABSTRACT

Seq2Logo is a web-based sequence logo generator. Sequence logos are a graphical representation of the information content stored in a multiple sequence alignment (MSA) and provide a compact and highly intuitive representation of the position-specific amino acid composition of binding motifs, active sites, etc. in biological sequences. Accurate generation of sequence logos is often compromised by sequence redundancy and low number of observations. Moreover, most methods available for sequence logo generation focus on displaying the position-specific enrichment of amino acids, discarding the equally valuable information related to amino acid depletion. Seq2Logo aims at resolving these issues allowing the user to include sequence weighting to correct for data redundancy, pseudo counts to correct for low number of observations and different logotype representations each capturing different aspects related to amino acid enrichment and depletion. Besides allowing input in the format of peptides and MSA, Seq2Logo accepts input as Blast sequence profiles, providing easy access for non-expert end-users to characterize and identify functionally conserved/variable amino acids in any given protein of interest. The output from the server is a sequence logo and a PSSM. Seq2Logo is available at http://www.cbs.dtu.dk/biotoools/Seq2Logo (14 May 2012, date last accessed).
Prediction of T cell epitopes

Use of bioinformatics tools to predict peptide MHC-binding

Exercise:

1. Identification of MHC-binding motif and identify potential epitopes
   Search the SYFPEITHI or IEDB database to characterize the binding motif for different MHC alleles - use the protein sequence from selected antigens to identify potential epitopes

2. Visualize the binding motif using sequence logos
FASTA format
Για την εκκίνηση των υπολογιστών:
- Ανοίξτε τις οθόνες των υπολογιστών
- Κάντε δεξί κλικ στην εντολή mozilla, και κλικ για άνοιγμα παράθυρου
- Πληκτρολογήστε την εντολή mozilla, και κλικ για άνοιγμα παραθύρου
- Επαναλάβατε τη διαδικασία για να ανοίξετε και ένα δεύτερο παράθυρο
- Στο πρώτο παράθυρο εμφανίστε την ηλεκτρονική διεύθυνση SwissProt και βρείτε την πρωτεϊνική αλληλουχία του γονίδιου επιλογής σε μορφή FASTA
**SYFPEITHI: MHC Ligands and peptide motifs**

- Κάντε «Αντιγραφή» της αλληλουχίας FASTA
- Στο δεύτερο παράθυρο ανοίξτε την βάση δεδομένων SYFPEITHI (www.syfpeithi.de)

and

**IEDB: ImmunoEpitope Database and Analysis Motifs**

- Κάντε «Αντιγραφή» της αλληλουχίας FASTA
- Στο δεύτερο παράθυρο ανοίξτε την βάση δεδομένων IEDB (www.iedb.org)

→ T-cell epitope prediction
→ Peptide binding to MHC class I molecules
→ Enter the protein sequence in fasta format .....  

*Results: low percentile rank = good binders*
**ΠΗΓΑΔΙΑ ΕΠΙΛΟΓΩΝ**

**ΑΣΚΗΣΕΙΣ:**
- Να γράψετε τα 3-5 πεπτίδια με το μεγαλύτερο σκορ.
- Να προτείνετε ποιον επίτοπο θα διαλέγατε για την πραγματοποίηση πειράματος (π.χ. σχεδιασμός εμβολίου). Να δικαιολογήσετε την απάντησή σας.
- Να συγκρίνετε τα αποτελέσματα που προκύπτουν από τις δύο βάσεις δεδομένων (SYFPEITHI & IEDB) για τις ίδιες πρωτεΐνες.

**EpCAM**, epithelial cell adhesion molecule, **survivin or BIRC5**, baculoviral IAP repeat containing 5, **MTA1**, metastasis associated antigen 1, **NPTN**, neuroplastin.

<table>
<thead>
<tr>
<th>cDNA</th>
<th>Species</th>
<th>MHC type</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPCAM</td>
<td>HUMAN</td>
<td>HLA-A*01</td>
<td>A</td>
</tr>
<tr>
<td>BIRC5</td>
<td>HUMAN</td>
<td>HLA-B*08</td>
<td>B</td>
</tr>
<tr>
<td>MTA1</td>
<td>MOUSE</td>
<td>H2-Db</td>
<td>C</td>
</tr>
<tr>
<td>NPTN</td>
<td>MOUSE</td>
<td>H2-Kd</td>
<td>D</td>
</tr>
</tbody>
</table>
Ανοίξτε τη βάση δεδομένων SWISSPROT και αναζητήστε την πρωτεϊνική αλληλουχία του γονιδίου σε μορφή fasta
>sp|P164221|EPCAM_HUMAN Epithelial cell adhesion molecule OS=Homo sapiens GN=EPCAM PE=1 SV=2
MAPPOVLAFGLLLAAATATFAAAAQQECECVENYKLAVNCFYXNNRQCQCTSVGAQKTVICS
KLAACKLVMKAEMNGSKLGRKPEKAGLQNNNDGLYDPDCDESGLPKAKQCNNGTSMCWVN
TAGVRRTEKDTEITCSERVETYWIIIELKSKAREKPYDSKSLRTALQELIITTFYQLDPKF
ITSYENNVITIDIVQNSQKTQNDVDIALVAYFEBKQVQRGELSEFHSKRMILTVNCEQL
DLDPQTLYYVDEKAPEFSMQGLKAGVIAVIVVVVIVAVVAGIVVLVRSRSKRAMYKEKA
RTKRMGEXHERFNA
Κάντε «Αντιγραφή» της αλληλουχίας FASTA
Στο δεύτερο παράθυρο ανοίξτε την βάση δεδομένων SYFPEITHI

( www.syfpeithi.de )

supported by
DFG-Sonderforschungsbereich 635 and the European Union:
EU BIOMED CT95-1627, BIOTECH CT95-0263,
and EU QLQ-CT-1999-00713

SYFPEITHI

A DATABASE OF MHC LIGANDS
AND PEPTIDE MOTIFS (Ver. 1.0)

SYFPEITHI is a database comprising
more than 7000 peptide sequences
known to bind class I and class II MHC
molecules. The entries are compiled
from published reports only.
Welcome to SYFPEITHI

This Database contains information on:
- Peptide sequences
- anchor positions
- MHC specificity
- source proteins, source organisms
- publication references

Links with sequence databases and 'MedLine' are available online

Epitope prediction and retrieval of sequences according to their molecular mass is also possible

The following search options are available:
- FIND YOUR MOTIF, LIGAND OR EPITOPE
- EPITOPE PREDICTION
- INFORMATION
Epitope prediction

This page allows you to find out the ligation strength to a defined HLA type for a sequence of aminoacids. The algorithm used are based on the book "MHC Ligands and Peptide Motifs" by H.G. Rammensee, J. Bachmann and S. Stevanovic. The probability of being processed and presented is given in order to predict T-cell epitopes.

1. Select MHC type
   If you chose "all", max. sequence length is 100 aminoacids (letters)!
   - all
   - H2-Ad
   - H2-Ak
   - H2-Db
   - H2-Ed
   Hold down ctrl key when clicking to select multiple items

2. Choose a mer
   - octamers (8 aa)
   - nonamers (9 aa)
   - decamers (10 aa)
   - endecamers (11 aa)
   - 15 -mers (15 aa) for MHC Type II only
   - all mers

3. Paste your sequence here:
   Max. input 2048 aminoacids (letters)!
   Letters only, no numbers or non-ASCII-symbols please.
   You may use "SYFPEITHI" with H2-Kd to see an example.
   syfpeithi

4. Choose Run to start analysis
   Run
   Reset
   Home
### Αποτελέσματα

Your search Results

Return to search conditions
H2-Kd nonamers

<table>
<thead>
<tr>
<th>Pos</th>
<th>H2-Kd nonamers</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>92</td>
<td>AYSNGNVSY</td>
<td>24</td>
</tr>
<tr>
<td>341</td>
<td>PFLGILAEI</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>SSFLGALAL</td>
<td>21</td>
</tr>
<tr>
<td>128</td>
<td>RQNFSTIW</td>
<td>31</td>
</tr>
<tr>
<td>14</td>
<td>LLLYSGSL</td>
<td>30</td>
</tr>
<tr>
<td>214</td>
<td>EYHCUYKHFV</td>
<td>20</td>
</tr>
<tr>
<td>15</td>
<td>SLLVSGLS</td>
<td>18</td>
</tr>
<tr>
<td>149</td>
<td>RIUTSERVI</td>
<td>18</td>
</tr>
<tr>
<td>273</td>
<td>KENGVFEEL</td>
<td>18</td>
</tr>
<tr>
<td>291</td>
<td>TNKENYTEL</td>
<td>18</td>
</tr>
<tr>
<td>322</td>
<td>GSASVSTYL</td>
<td>18</td>
</tr>
<tr>
<td>328</td>
<td>TVLVRSEL</td>
<td>18</td>
</tr>
<tr>
<td>35</td>
<td>KSPHSETEL</td>
<td>17</td>
</tr>
<tr>
<td>42</td>
<td>KLTGDAFEL</td>
<td>17</td>
</tr>
<tr>
<td>95</td>
<td>SNGVSULRI</td>
<td>17</td>
</tr>
<tr>
<td>201</td>
<td>EYRINKERA</td>
<td>17</td>
</tr>
</tbody>
</table>
Εναλλακτικά:

Ανοίξτε στο δεύτερο παράθυρο την βάση δεδομένων IEDB.
2. Επιλέξτε «T cell epitope prediction tools» → «MHC binding prediction» ή «B cell epitope prediction tools» → «From protein sequence»

και ακολουθήστε τις οδηγίες που παρουσιάζονται και παρακάτω.

Όπου σας ζητάει η αλληλουχία FASTA πριν την επικολλήσετε προσθέστε στο πλαίσιο το σύμβολο >. Όπου σας ζητάει τον κωδικό SwissProt να συμπληρώνετε τον πλήρη κωδικό κάθε συγκεκριμένης ισομορφής.

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 for recognition by T cells.

- Peptide binding to MHC class I molecules
  This tool will take in an amino acid sequence, or set of sequences and determine each subsequence's ability to bind to specific MHC class I molecule.

- Peptide binding to MHC class II molecules
  This tool employs a consensus approach to predict MHC Class II epitopes based upon Sturmlo, ARB, and others.

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 and rank each peptide's intrinsic potential of being a T cell epitope.

- Neural network based prediction of proteasomal cleavage sites (NetChop) and T cell epitopes (NetCTL and NetCTLpan)
  NetChop is a predictor of proteasomal processing based upon a neural network. NetCTL and NetCTLpan are predictors of T cell epitopes along a protein sequence. It also employs a neural network architecture.
B Cell Epitope Prediction Tools

The tools here are intended to predict regions of proteins that are likely to be recognized as epitopes in the context of B cell activation. They include:

**Prediction of linear epitopes from protein sequence**
A collection of methods to predict continuous linear B cell epitopes by amino acid scales and discontinuous protein 3D structures.

**Discotope - Prediction of epitopes from protein structure**
This method incorporates solvent-accessible surface area calculations, as well as contact distances into 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.

This site is best viewed with Microsoft Internet Explorer or Mozilla Firefox.
## MHC-I Binding Prediction Results

### Input Sequences

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sequence 1</td>
<td>M3G33LPGALALSLLLVGSSSLPGFAGAQAAGFVSQMSEMTENKEDGATFLYCDVVSPTPE1QWYTYAEVINRAESFRQLWDGARKRRVIVNTAYGSGNVSVSLRITIRILSGETVTECRASNPDKRNDLRQNPSITWIARATITISLQKPRFIVTSEEVIIRESLLEFLVLQCNLTSSHTLYWTRNVGELTATRKNASNMEYPRKIPRAEDGSEYHCYTHFVSAKPANATIEVKAAPDITGHKRESNKNSQGDAAMTKCYSV1P1NEWWRKRENQFETZSNZG9SFITTENYTELSTVNLQITTEDFGEYECDNAINN1GSSAVSTVTLRVRSHLALPHF1GILAEI1ILVIIIYVEKRFREPOLPDEPV1DDDEPAEGMKINSTNNHHLKDLNQVRNTN</td>
</tr>
</tbody>
</table>

### Prediction method: IEDB recommended | Low percentile_rank = good binders

<table>
<thead>
<tr>
<th>Allele</th>
<th>#</th>
<th>Start</th>
<th>End</th>
<th>Length</th>
<th>Peptide</th>
<th>Method used</th>
<th>Percentile rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-2-Kd</td>
<td>1</td>
<td>92</td>
<td>101</td>
<td>10</td>
<td>AYSGNGVSVL</td>
<td>Consensus (ann/smm)</td>
<td>0.65</td>
</tr>
<tr>
<td>H-2-Kd</td>
<td>1</td>
<td>312</td>
<td>321</td>
<td>10</td>
<td>EYECNAINSI</td>
<td>Consensus (ann/smm)</td>
<td>0.65</td>
</tr>
<tr>
<td>H-2-Kd</td>
<td>1</td>
<td>181</td>
<td>190</td>
<td>10</td>
<td>SYWTRNGVEL</td>
<td>Consensus (ann/smm)</td>
<td>0.85</td>
</tr>
<tr>
<td>H-2-Kd</td>
<td>1</td>
<td>214</td>
<td>221</td>
<td>8</td>
<td>EYHCYVF</td>
<td>Consensus (ann/smm)</td>
<td>0.9</td>
</tr>
<tr>
<td>H-2-Kd</td>
<td>1</td>
<td>328</td>
<td>336</td>
<td>9</td>
<td>TVLRVRSVL</td>
<td>Consensus (ann/combiblb_sidency2008/smm)</td>
<td>1.3</td>
</tr>
<tr>
<td>H-2-Kd</td>
<td>1</td>
<td>320</td>
<td>329</td>
<td>10</td>
<td>SGGASAVSTIV</td>
<td>Consensus (ann/smm)</td>
<td>1.35</td>
</tr>
<tr>
<td>ΓΟΝΙΔΙΟ</td>
<td>ΠΡΟΕΛΕΥΣΗ</td>
<td>ΜΗΣ</td>
<td>ΟΜΑΔΑ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPCAM</td>
<td>HUMAN</td>
<td>HLA-A*01</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIRC5</td>
<td>HUMAN</td>
<td>HLA-B*08</td>
<td>Β</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTA1</td>
<td>MOUSE</td>
<td>H2-Db</td>
<td>Γ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPTN</td>
<td>MOUSE</td>
<td>H2-Kd</td>
<td>Δ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

EpCAM, epithelial cell adhesion molecule, survivin or BIRC5, baculoviral IAP repeat containing 5, MTA1, metastasis associated antigen 1, NPTN, neuroplastin.
ΑΣΚΗΣΕΙΣ:

1. Να γράψετε τα 3-5 πεπτίδια με το μεγαλύτερο σκορ.
2. Να προτείνετε ποιον επίτοπο θα διαλέγατε για την πραγματοποίηση πειράματος (π.χ. σχεδιασμός εμβολίου). Να δικαιολογήσετε την απάντηση σας.
3. Να συγκρίνετε τα αποτελέσματα που προκύπτουν από τις δύο βάσεις δεδομένων (SYFPEITHI & IEDB) για τις ίδιες πρωτεϊνές.