Bioinformatics 1 -- Lecture 22

IMMUNOINFORMATICS:
Bioinformatics Challenges in Immunology

Most slides courtesy of
Julia Ponomarenko, San Diego Supercomputer Center
or
Oliver Kohlbacher, WSI/ZBIT, Eberhard-Karls-Universität Tübingen
The Immune Reaction

<table>
<thead>
<tr>
<th>Initial response</th>
<th>Protective Immunity</th>
<th>Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibody and effector T cells</td>
<td>Weeks</td>
<td>Years</td>
</tr>
</tbody>
</table>

- **First exposure**
- **Inapparent reinfection**
- **Mild or inapparent reinfection**
Vaccines have been made for 36 of >400 human pathogens

<table>
<thead>
<tr>
<th>Organism</th>
<th>Type</th>
<th>Vaccine Type</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variola virus</td>
<td>Virus</td>
<td>Live</td>
<td>1798</td>
</tr>
<tr>
<td>Rabies virus</td>
<td>Virus</td>
<td>Inactivated</td>
<td>1885</td>
</tr>
<tr>
<td><em>Salmonella typhi</em></td>
<td>Bacteria</td>
<td>Live</td>
<td>1896</td>
</tr>
<tr>
<td><em>Vibrio cholerae</em></td>
<td>Bacteria</td>
<td>Inactivated</td>
<td>1896</td>
</tr>
<tr>
<td><em>Yersinia pestis</em></td>
<td>Bacteria</td>
<td>Inactivated</td>
<td>1897</td>
</tr>
<tr>
<td><em>Corynebacterium diphtheriae</em></td>
<td>Bacteria</td>
<td>Toxoid</td>
<td>1923</td>
</tr>
<tr>
<td><em>Bordetella pertussis</em></td>
<td>Bacteria</td>
<td>Acellular</td>
<td>1926</td>
</tr>
<tr>
<td><em>Clostridium tetani</em></td>
<td>Bacteria</td>
<td>Toxoid</td>
<td>1927</td>
</tr>
<tr>
<td><em>Mycobacterium tuberculosis</em></td>
<td>Bacteria</td>
<td>Live</td>
<td>1927</td>
</tr>
<tr>
<td>Yellow fever virus</td>
<td>Virus</td>
<td>Live</td>
<td>1935</td>
</tr>
<tr>
<td>Influenza virus type A</td>
<td>Virus</td>
<td>Inactivated</td>
<td>1936</td>
</tr>
<tr>
<td>Influenza virus type B</td>
<td>Virus</td>
<td>Inactivated</td>
<td>1936</td>
</tr>
<tr>
<td><em>Coxiella burnetii</em></td>
<td>Bacteria</td>
<td>Inactivated</td>
<td>1938</td>
</tr>
<tr>
<td><em>Rickettsia prowazekii</em></td>
<td>Bacteria</td>
<td>Inactivated</td>
<td>1938</td>
</tr>
<tr>
<td><em>Rickettsia rickettsii</em></td>
<td>Bacteria</td>
<td>Inactivated</td>
<td>1938</td>
</tr>
<tr>
<td>Central European encephalitis virus</td>
<td>Virus</td>
<td>Inactivated</td>
<td>1939</td>
</tr>
<tr>
<td>Poliovirus types 1, 2, and 3</td>
<td>Virus</td>
<td>Inactivated/Live</td>
<td>1962</td>
</tr>
<tr>
<td>Measles virus</td>
<td>Virus</td>
<td>Live</td>
<td>1963</td>
</tr>
<tr>
<td>Mumps virus</td>
<td>Virus</td>
<td>Live</td>
<td>1967</td>
</tr>
<tr>
<td>Rubivirus</td>
<td>Virus</td>
<td>Live</td>
<td>1969</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em></td>
<td>Bacteria</td>
<td>Staphage lysate</td>
<td>1976</td>
</tr>
<tr>
<td><em>Streptococcus pneumoniae</em></td>
<td>Bacteria</td>
<td>Polysaccharide</td>
<td>1977</td>
</tr>
<tr>
<td>Human adenovirus types 4 and 7</td>
<td>Virus</td>
<td>Live</td>
<td>1980</td>
</tr>
<tr>
<td><em>Neisseria meningitidis</em></td>
<td>Bacteria</td>
<td>Polysaccharide</td>
<td>1981</td>
</tr>
<tr>
<td>Hepatitis B</td>
<td>Virus</td>
<td>Recombinant</td>
<td>1986</td>
</tr>
<tr>
<td><em>Haemophilus influenzae</em></td>
<td>Bacteria</td>
<td>Conjugate</td>
<td>1987</td>
</tr>
<tr>
<td>Hantaan virus</td>
<td>Virus</td>
<td>Inactivated</td>
<td>1989</td>
</tr>
<tr>
<td>Japanese encephalitis virus</td>
<td>Virus</td>
<td>Inactivated</td>
<td>1992</td>
</tr>
<tr>
<td>Varicella-zoster virus</td>
<td>Virus</td>
<td>Live</td>
<td>1994</td>
</tr>
<tr>
<td>Hepatitis A</td>
<td>Virus</td>
<td>Inactivated</td>
<td>1995</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>Bacteria</td>
<td>Inactivated</td>
<td>1995</td>
</tr>
<tr>
<td>Junin virus</td>
<td>Virus</td>
<td>Live</td>
<td>1996</td>
</tr>
<tr>
<td><em>Bacillus anthracis</em></td>
<td>Bacteria</td>
<td>Adsorbed</td>
<td>1998</td>
</tr>
<tr>
<td><em>Borrelia burgdorferi</em></td>
<td>Bacteria</td>
<td>Recombinant</td>
<td>1998</td>
</tr>
</tbody>
</table>

*MIT press.*
Epitope

- Quantum unit of immunity
- Surface on which an antibody binds
- Comprises antigenic matter
- Linear epitope consists of a short AA sequence
- Conformational epitope depends on tertiary structure.
Two branches of immunity

Innate immunity

- Recognizes molecules, called pathogen-associated molecular patterns (~$10^3$), or PAMPs shared by groups of related microbes; e.g. LPS from the gram-negative cell wall, RNA from viruses, flagellin, and glucans from fungal cell walls.
- Is antigen-nonspecific.
- Immediate or within several hours response.
- Involves body defense cells that have pattern-recognition receptors:
  - Leukocytes: neutrophils, eosinophils, basophils and monocytes;
  - cells that release inflammatory mediators: macrophages and mast cells;
  - natural killer cells (NK cells); and
  - complement proteins and cytokines.
- Does not improve with repeated exposure to a given infection.

Adaptive immunity

- Recognizes epitopes; that are specific B- and T-cell recognition sites on antigens.
- Is antigen-specific.
- 3 to 10 days response.
- Involves the following:
  - antigen-presenting cells (APCs) such as macrophages and dendritic cells;
  - antigen-specific B-lymphocytes (~$10^9$);
  - antigen-specific T-lymphocytes (~$10^{12}$); and
  - cytokines.
- Improves with repeated exposure and becomes protective.
## Components of the immune system

<table>
<thead>
<tr>
<th>Innate immune system</th>
<th>Adaptive immune system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response is non-specific</td>
<td>Pathogen and antigen specific response</td>
</tr>
<tr>
<td>Exposure leads to immediate maximal response</td>
<td>Lag time between exposure and maximal response</td>
</tr>
<tr>
<td><strong>Cell-mediated and humoral components</strong></td>
<td><strong>Cell-mediated and humoral components</strong></td>
</tr>
<tr>
<td>No immunological memory</td>
<td>Exposure leads to immunological memory</td>
</tr>
<tr>
<td>Found in nearly all forms of life</td>
<td>Found only in <strong>jawed vertebrates</strong></td>
</tr>
</tbody>
</table>
B and T cells recognize different epitopes of the same protein

**T-cell epitope**
- Denatured antigen
- Linear (often) peptide 8–37 aa
- Internal (often)

**Binding to T cell receptor:**
- $K_d \ 10^{-5} - 10^{-7} \text{ M (low affinity)}$
- Slow on-rate, slow off-rate
  (once bound, peptide may stay associated for hours to many days)

**B-cell epitope**
- Native or denatured (rare) antigen
- Sequential (continuous) or conformational (discontinuous)
- Accessible, hydrophilic, mobile, usually on the surface or could be exposed as a result of physicochemical change

**Binding to antibody:**
- $K_d \ 10^{-7} - 10^{-11} \text{ M (high affinity)}$
- Rapid on-rate, variable off-rate
B cell (magenta, orange) and T cell epitopes (blue, green, red) of hen egg-white lysozyme

PDB: 1dpx
MHC class I pathway

Intracellular pathogen (virus, mycobacteria)

- Cytosolic protein
- Proteasome
- TAP
- ER
- MHC I
- CD8 epitope
- TCR
- CD8
- CTL (T_{CD8+})
MHC class I pathway

Intracellular pathogen (virus, mycobacteria)

- Cytosolic protein
- Proteasome
- TAP
- ER
- MHC
- CD8 (T<sub>CD8+</sub>)
- CTL
- TCR
- V<sub>β</sub>
- β-2-Microglobulin
- MHC class I
- HLA-A2.1

Xenoreactive Complex AHIII 12.2 TCR bound to P1049 (ALWGFFPVLS)/HLA-A2.1
MHC class I pathway

Intracellular pathogen (virus, mycobacteria)

Bioinformatics approaches at epitope prediction:

(1) Prediction of proteosomal cleavage sites (several methods exist based on small amount of in vitro data).

(2) Prediction of peptide–TAP binding (ibid.).

(3) **Prediction of peptide–MHC binding.**

(4) Prediction of pMHC–TCR binding.
MHC class I epitope prediction: Challenges

- High rate of pathogen mutations. Pathogens evolve to escape:
  - Proteosomal cleavage (HIV);
  - TAP binding (HIV, HSV type I);
  - MHC binding.

- MHC genes are highly polymorphic (2,292 human alleles/1,670 – 2 years ago).

- MHC polymorphism is essential to protect the population from invasion by pathogens. But it generates problem for epitope-based vaccine design: a vaccine needs to contain a unique epitope binding to each MHC allele.

- Every normal (heterozygous) human expresses six different MHC class I molecules on every cell, containing α-chains derived from the two alleles of HLA-A, HLA-B, HLA-C genes that inherited from the parents.

- Every human has ~$10^{12}$ lymphoid cells with a T-cell receptors repertoire of ~$10^7$, depending on her immunological status (vaccinations, disease history, environment, etc.).
Prediction of MHC class I binding peptide – potential epitopes

- MHC allele or allele supertype (similar in sequences alleles bind similar peptides) specific.
- Peptide length (8-, 9-, 10-, 11-mers) specific.

**Sequence-based approaches:**
- Gibbs sampling (when the training peptides are of different lengths)
- Hidden Markov Models (ibid.)
- Sequence motifs, position weight matrices
- Artificial Neural Networks (require a large number of training examples)
- SVM*

Peptides known to bind to the HLA-A*0201 molecule.
### Performance measures for prediction methods

<table>
<thead>
<tr>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Score</th>
<th>threshold</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>TP+FN</th>
<th>TP / (TP + FN)</th>
<th>6/7 = 0.86</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN+FP</td>
<td>FP / (FP + TN)</td>
<td>6/8 = 0.75</td>
</tr>
</tbody>
</table>

**Sensitivity** = \( \frac{TP}{TP + FN} = \frac{6}{7} = 0.86 \)

**Specificity** = \( \frac{TN}{TN + FP} = \frac{6}{8} = 0.75 \)

**ROC curve**

- True positive rate, \( \frac{TP}{TP + FN} \)
- False positive rate, \( \frac{FP}{FP + TN} \)

**AUC** or **A\_ROC**
Performance measures for prediction methods (cont)

TP+FN – actual binders (based on a defined threshold on actual binding affinity values $a_i$)

TN+ FP – actual non–binders (ibid.)

Sensitivity = $TP / (TP + FN) = 6/7 = 0.86$

Specificity = $TN / (TN + FP) = 6/8 = 0.75$
MHC class I pathway

Intracellular pathogen (virus, mycobacteria)

Cytosolic protein → Proteasome → ER → MHC I → TAP → CD8 epitope → CTL (T\textsubscript{CD8+})

Any cell

MHC class II pathway

Extracellular protein

Endosome → CD4 epitope → CD4 → TCR

B-cell, macrophage, or dendritic cell
MHC class I pathway

Any cell

Peptides

ER

T

CD8+

Intracellular pathogen (virus, mycobacteria)

Proteasome

Cytosolic protein

MHC I

MHC class II pathway

Extracellular protein

Endosome

MHC II

TAP

Complex Of A Human TCR, Influenza HA Antigen Peptide (PKYVKQNTLKLAT) and MHC Class II

T–Cell Receptor

MHC class II α

MHC class II β

α

MHC class II β

V

β

Vα

B–cell, macrophage, or dendritic cell

CD4

TCR

CD4+ T

TCD4+
Epitope, or antigenic determinant, is defined as the site of an antigen recognized by immune response molecules (antibodies, MHC, TCR). T cell epitope – a short linear peptide or other chemical entity (native or denatured antigen) that binds MHC (class I binds 8-10 ac peptides; class II binds 11-25 ac peptides) and may be recognized by T-cell receptor (TCR). T cell recognition of antigen involves tertiary complex "antigen-TCR-MHC".

Complex Of A Human TCR, Influenza HA Antigen Peptide (PKYVKQNTLKLAT) and MHC Class II

Xenoreactive Complex AHIII 12.2 TCR bound to P1049 (ALWGFFPVLS) /HLA-A2.1

MHC class I

T-Cell Receptor

Vβ

Vα

MHC class II α

β-2-

Microglobulin

1fyt

1lp9
MHC class II epitope prediction: Challenges

- MHC class II genes are as highly polymorphic as MHC class I (1,012 human alleles for today).
- The repertoire of T-cell receptors is ~10⁷ and depends on an individual’s immunological status (vaccinations, disease history, environment, etc.).
- The epitope length 9–37 aa.
- The peptide may have non-linear conformation.
- The MHC binding groove is open from both sides and it is known that residues outside the groove effect peptide binding.

Complex Of A Human TCR, Influenza HA Antigen Peptide (PKYVKQNTLKLA) and MHC Class II
The processing of MHC class II epitopes is still a mystery and likely depends on the antigen structure, the cell type and other factors.
Prediction of MHC class II binding peptide – potential epitopes

- MHC allele or allele supertype (similar in sequences alleles bind similar peptides) specific.
- Predictions for peptides of length 9 aa (the peptide–MHC binding core)
- Sequence–based approaches:
  - Gibbs sampling
  - Sequence motifs, position weight matrices
  - Machine learning: SVM, HMM, evolutionary algorithms

Fig. 5. An alignment generated by the Gibbs sampler for the DR4(B1*0401) binding motif. In the left panel are shown the unaligned sequences, and in the right panel the aligned sequences. The core motif is shown underlined and in italic.

Nielsen et al., Bioinformatics 2004
Benchmarking predictions of peptide binding to MHC II

- Data: pairs {peptide – affinity value in terms of IC$_{50}$ nM} for a given MHC allele
- 16 different mouse and human MHC class II alleles.
- 10,017 data points.
- 9 different methods were evaluated: 6 matrix-based, 2 SVM, 1 QSAR-based.
- AUC values varied from 0.5 (random prediction) to 0.83, depending on the allele.
- Comparison with 29 X-ray structures of peptide–MHC II complexes (14 different alleles):
  - The success level of the binding core recognition was 21%–62%,
  - with exception of TEPETOPE method (100%) that is based on structural information and measured affinity values for mutant variants of MHC class II and peptides (Sturniolo et al., Nature Biotechnology, 1999).

=> Structural information together with peptide–MHC binding data should improve the prediction.
Antigen Processing and Presentation: CBS tools
Class I Ag Processing

- Protein digestion in the cytosol
- Amino acids and peptides
- ATP, ADP, and Pi involvement
- Class I MHC loading
- TAP transport
- Calreticulin, Tapasin, RER lumen
- Class I α chain
- Calnexin interaction
The Proteasome

- 20S proteasome consists of four rings of protein subunits with a central channel

20S

PDB code 1IRU 113 Å

26S

19S

148 Å
The 20S Proteasome - Cleavage Sites

- There is a total of six major catalytic sites in the proteasome, described as:
  - Trypsin-like (Arg, Lys)
  - Chymotrypsin-like (Phe, Tyr, Trp)
  - Peptidylglutamyl-peptide hydrolyzing (PGPH) (Asp, Glu)

http://www.uni-stuttgart.de/ibc/wolf/proteasome.html
Serine Protease Mechanism
Serine Protease Mechanism

There is a nucleophilic attack of the O from a Ser on the carbonyl C of the peptide bond to be broken.
Serine Protease Mechanism

The oxyanion intermediate state.
Serine Protease Mechanism

The amine product leaves the complex and we have an acyl-enzyme (ester) intermediate.
Water joins the game and another oxyanion intermediate state is formed.
Serine Protease Mechanism

The peptide bond is cleaved and the enzyme is restored to its initial state.
Determinant of Cleavage Sites

- Experimental identification of proteasomal cleavage sites entails the following steps:
  - Purification of proteasomes
  - Isolation of a protein to be cleaved
  - Degradation experiment - incubation of the protein to be cleaved with proteasomes
  - Separation and analysis of cleavage products

- All published reports on whole protein digestion experiments involve people from Tübingen
Analysis of Cleavage Products

- HPLC is used to separate the cleavage products
- MS can be used to identify peptide fragments and Edman degradation to quantify the amounts of a certain cleavage product/peptide
- Identified cleavage products can then be mapped back onto the protein sequence to determine the actual cleavage patterns
Determination of Cleavage Sites

- Tenzer et al. determined the cleavage pattern for both the constitutive(c20S) and the immuno-proteasome(i20S) on a prion protein (210 aa)
- For the c20S (i20S) they identified 104 (113) different cleavage fragments with an average length of 20.2 (17.5) aa
- They observed different cleavage patterns for the two proteasome types

Prion Protein Cleavages

A

VYKSHSGW LVLIVWNSD VQG \text{...}

B

DCVNTVQN FYTITTINGEN FETEDEEine AVVEQGCM TO YORESSAVYG RGA

Tenser et al., Quantitative analysis of prion-protein degradation by constitutive and immuno-20S proteasomes indicates differences correlated with disease susceptibility, 2004, Journal of Immunology, 172: 1083-1091


c20S

i20S
Prion Protein Cleavages by i20S

Tenzer et al., Quantitative analysis of prion-protein degradation by constitutive and immuno-20S proteasomes indicates differences correlated with disease susceptibility. 2004, Journal of Immunology, 172: 1083-1091
Prediction of Proteasomal Cleavage

- There is a number of methods for cleavage prediction available at the moment (standard methods, all of them)
- Key problem: scarcity of data available
- Some of the benchmarking procedures used are not state of the art
- Method performances are not as good as one would hope
MAPPP Cleavage Prediction

Enter or paste the sequences to analyze:

Minimum length of fragments

Min. possibility for cleavage after a single residue

For cleavage of a fragment

Separator between proteins

Comments on the output:

Sequence and Fragments
The sequence and its predicted peptide fragments can be found at the bottom of the page.
- For every single residue with a cleavage probability higher than the specified minimum, the value of the probability is shown by moving the mouse pointer over the residue. It is represented as a (non-functional) WWW-link.
- Directly beneath the residues one can find the number of peptide fragments a single residue takes places in.
- The table beneath the sequence represents the fragments with a cleavage probability higher than the specified minimum and its places within the sequence.

Example
Example output for the default values with p60

Based on FRAGPREDICT developed by H.-G. Holzhüter et. al.

http://www.mpiib-berlin.mpg.de/MAPPP/cleavage.html
# MAPPP Cleavage Prediction

## Query results

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of sequence</td>
<td>47</td>
</tr>
<tr>
<td>Range for length of peptide fragments</td>
<td>9.11</td>
</tr>
<tr>
<td>Min. probability for cleavage after a single residue</td>
<td>0.5</td>
</tr>
<tr>
<td>Min. probability for cleavage of a fragment</td>
<td>0.5</td>
</tr>
</tbody>
</table>

## Prediction results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of possible fragments</td>
<td>17</td>
</tr>
<tr>
<td>Total no. of fragments</td>
<td>111</td>
</tr>
<tr>
<td>Percentage (fragments / total fragments)</td>
<td>15.31%</td>
</tr>
<tr>
<td>Highest fragment cleavage probability (at pos. 0.301)</td>
<td>1</td>
</tr>
<tr>
<td>Highest residue cleavage probability (at pos. 2.001)</td>
<td>0.9562</td>
</tr>
</tbody>
</table>

```
SSAPPWEPPRPEPTPYMKDSAMDKMASDCCVDDSLQLSLPIISY
11111111111112246665571110477777643223333222222
099
SSAPPWEPPR(1)
REPTPYMKD(0.0493)
PymkDSaMdk(1)
YmkDSaMdk(0.0224)
YmkDSaMdkMA(0.3519)
MkDSaMdk(0.5257)
MkDSaMdkMA(0.2489)
MDKMASDCCV(0.5714)
MDKMASDCCVD(0.9991)
DKMASDCCVD(0.9999)
DKMASDCCVDD(0.9999)
ASDCCVDD(0.3015)
SDCCVDDLQ(1)
DDSLQLLSL(0.6041)
LSLPIISY(1)
```
Proteasomal Splicing

- Two papers published in 2004 put an extra twist in the tale
- Hanada et al. found a peptide that was generated from two non-contiguous parts of its source proteins
- Vigneron et al. even proved that a specific peptide could be produced from longer fragments by means of proteasomal splicing

TAP

- TAP - the transporter associated with antigen processing
- TAP transports the peptides created by the proteasome into the ER
- TAP is an active transporter and requires energy to transport the peptides
- This energy is provided by ATP
  - ATP is the main energy “currency” in a cell
  - ATP is cleaved into ADP and P_i
  - Breaking the bond provides significant energy, which can be used for transport
ABC Transporters

- TAP is member of the ABC transporter family (ABC = ATP binding cassette)
- ABC transporters have a wide range of roles: from transport of vitamins, over drug-resistance transporters, to their role in Ag processing
- ABC transporters contain
  - Transmembrane domains (TMDs)
  - Nucleotide-binding domains (NBDs)
- ABC transporters are usually dimeric, either
  - Homodimers: (TMD-NBD)$_2$
  - Heterodimers: TMD$^A$-NBD$^A$:TMD$^B$-NBD$^B$
- TAP is a heterodimer formed by TAP1 and TAP2 chains
Structure of Vitamin B$_{12}$ Transporter

PDB: 1L7V
TAP Structure

- Known details of the architecture
  - NBD pointing into cytosol
  - TMDs possess probably eight (TAP1) resp. seven (TAP2) transmembrane segments
  - NBD structure known
  - Dimerization of TAP1,2 NBDs probably required for ATP cleavage
  - Interface between TMDs is responsible for peptide binding

Vos et al., J. Immunol. (1999), 159, 554
TAP in Action
TAP in Action

- ATP is required for transport, not for binding
- Both subunits are involved in binding
TAP in Action

- C & N terminus crucial for binding specificity
- Peptide binding induces ATP cleavage
TAP in Action
TAP in Action

- ATP cleavage liberates peptide
- Exact order of events uncertain
TAP in Action
Mechanistic Details
TAP and Immune Evasion

- TAP is also a target for immune evasion for some viruses
- Inhibiting TAP or reducing its expression seriously affects Ag presentation and thus avoids detection of the virus infected cell
- Example:
  - Adenoviruses possess a protein (E3/19K) binding to - and thus inactivating - TAP
  - Eppstein-Barr Virus (EBV) expresses LMP-1, which in turn induces expression of TAP2 while down-regulating TAP1
  ⇒ disequilibrium leads to few active TAP1/2 complexes
Affinity vs. Transport

- TAP transport is difficult to measure
- Simpler is a measurement of TAP affinity:
  - If no ATP is present, no transport is observed
  - Binding to the TAP heterodimer occurs nevertheless
  - Binding affinity was found to be correlated to transport efficiency
  - Binding studies can be done on a larger scale

Van Endert et al., Immunity (1994), 1, 491
TAP Specificity: Binding

- Binding assays
  - Radio-labeled peptides
  - TAP-overexpressed microsomes
  - ATP depletion, low temperature
    ⇒ no transport, but binding (1)
  - Centrifugation (2), counting to determine concentration of unbound peptide
Proteasome, TAP & MHC

- Length distributions of peptides created by proteasome, transported by TAP, and bound by MHC overlap significantly
- Length preferences of proteasome and TAP create and transport the full range of MHC peptides
- Proteasome
  - 3-22 aa long
- TAP
  - 6-16 aa long
- MHC
  - 8-11 aa long - after cleavage by ER peptidases
**TAP Specificity**

- Specificity is mostly determined by C-terminus (9) and three N-terminal positions (1-3)
- TAP favors peptides with basic, aromatic or hydrophobic C-termini
  - Phe, Tyr, Trp
  - Ile, Leu, Val
  - Arg
- Pro at position 2 is very unfavorable

Lankat-Buttgeritt, Tampé, Physiol Rev. (2002), 82, 187
Binding Mode

![Graph showing variance across positions 1 to 9 with bars indicating data points.]

- Variance
- Position

H$_3$N$^+$

1 2 3 4 5 6 7 8 9

COO$^-$
Binding Mode
Outstanding problems in TAP/Proteosome/MHC peptide prediction

- High throughput data generation
  - Currently few verified peptide sequences.
  - Overfitting likely

- Something better than the usual machine learning approach?
  - Structure-based motifs, length preferences?
  - Combine multiple motifs, reflecting multiple enzymes?

- Vaccine design to incorporate cleavage motifs
  - Optimize recombinant vaccine for cleavage?
Cellular immunity
How does our immune system detect infections?
Epitope-based vaccines

- A major reason for analyzing and predicting epitopes is because they may lead to the development of peptide-based synthetic vaccines.
- Thousands of peptides have been pre-clinically examined; over 100 of them have progressed to phase I clinical trials and about 30 to phase II, including vaccines for foot-and-mouth virus infection, influenza, HIV.
- However, not a single peptide vaccine has passed phase III and became available to the public.
- The only successful synthetic peptide vaccine has been made against canine parvovirus (causing enteritis and myocarditis in dogs and minks). It consists of several peptides from the N-terminal region of the viral VP2 protein (residues 1–15, 7–1, and 3–19) coupled to a carrier induced an immune response in dogs and minks. However, it is expensive and has lower than the conventional vaccine (attenuated virus) coverage.
Immunoinformatics

- The goals:
  - Modeling of the immune system at the population and individual levels (in silico immune system).
  - Design of medical diagnostics and therapeutic/prophylactic vaccines for cancers, allergies, autoimmune and infectious diseases.
- **Epitope discovery**: how the antigens are recognized by the cells of the immune system.
- Data collection and analysis: IEDB, HIV database, AntiJen (UK), IMGT (France)
- Evolution of the adaptive and innate immune system: Gary Litman (FL), Louis Du Pasquier (Switzerland)
- Evolution of pathogens and co-evolution of host and pathogen.
- Modeling of host–pathogen interactions: Leor Weinberger (UCSD)
- Deciphering regulatory networks in APCs, lymphocytes and other cells.