

MSc course BIOINFORMATICS



Predicting Peptide Binding to Major Histocompatibility Complex (MHC) molecules

Immunoinformatics – Computational Immunology
Πρόβλεψη πρόσδεσης πεπτιδίων/επιτόπων σε μόρια MHC

Professor Katerina Chlichlia

Email: achlichl@mbg.duth.gr

Τηλ. 25510-30630

9.11.2025

- Bioinformatics has broad applicability to Immunology →

IMMUNOINFORMATICS

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

We will apply bioinformatic tools for identifying antigenic epitopes –
our aim will be to predict peptide binding to particular MHC
molecules

Η πρόοδος των υπολογιστικών μεθόδων πρόβλεψης δέσμευσης πεπτιδίων στα μόρια MHC είχε βαθύ αντίκτυπο πέρα από το θεωρητικό επίπεδο, οδηγώντας σε **πρακτικές εφαρμογές στη βασική και στη μεταφραστική ανοσολογία.**

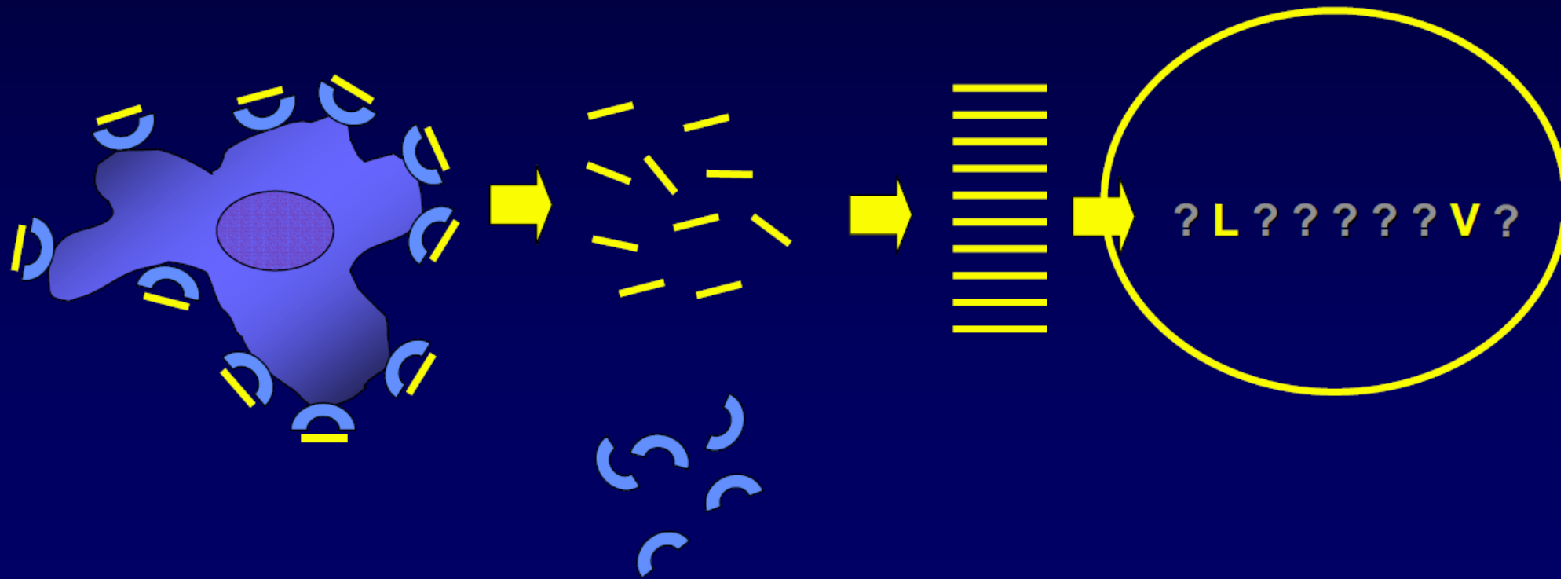
Η ικανότητα πρόβλεψης επιτόπων με υψηλή ακρίβεια επέτρεψε την ανάπτυξη εργαλείων που χρησιμοποιούνται ευρέως τόσο στην **έρευνα της ανοσολογικής απόκρισης** όσο και στον **ορθολογικό σχεδιασμό εμβολίων και ανοσοθεραπειών.**

- Οι πρώτες εφαρμογές των μοντέλων δέσμευσης MHC επικεντρώθηκαν στην **αναγνώριση T-κυτταρικών επιτόπων** από ιούς, βακτήρια και καρκινικά αντιγόνα. Η πρόβλεψη πιθανών επιτόπων με υπολογιστικά μέσα επιτρέπει τη **δραστική μείωση των πειραματικών δοκιμών**, καθώς προτεραιοποιούνται μόνο τα πεπτίδια με υψηλή πιθανότητα δέσμευσης στα MHC. Με αυτόν τον τρόπο, αυξάνεται η αποτελεσματικότητα και μειώνεται το κόστος των πειραμάτων.
- Οι προβλέψεις αυτές έχουν χρησιμοποιηθεί επιτυχώς για την **ανίχνευση επιτόπων σε πλήθος παθογόνων** (όπως HIV, SARS-CoV-2, Mycobacterium tuberculosis) αλλά και για **αντικαρκινικές μελέτες**, όπου τα μοντέλα MHC συνδυάζονται με γονιδιωματικά δεδομένα για την αναγνώριση **νεοαντιγόνων** που προκύπτουν από σωματικές μεταλλάξεις.
- Η **υπολογιστική ανοσολογία** έχει καταστεί θεμελιώδες εργαλείο στον **σχεδιασμό εμβολίων νέας γενιάς**.

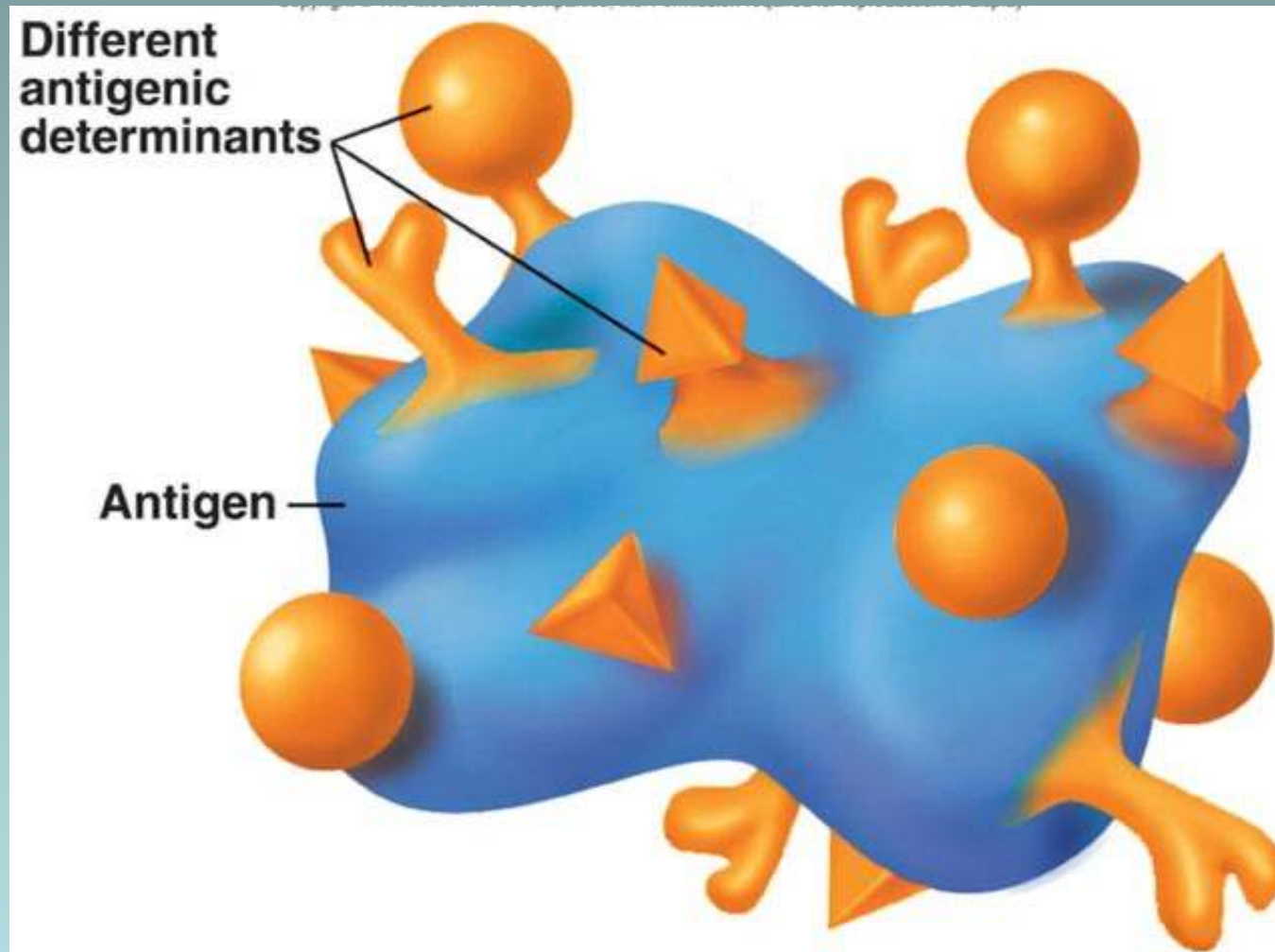
Overview of the lecture

- 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

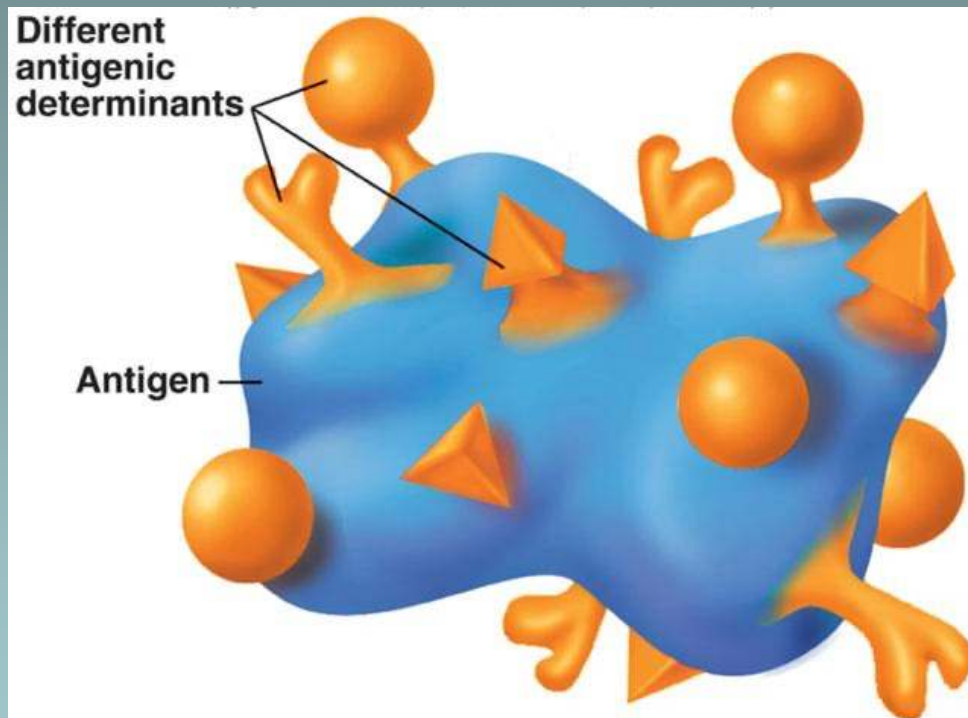
search for epitopes



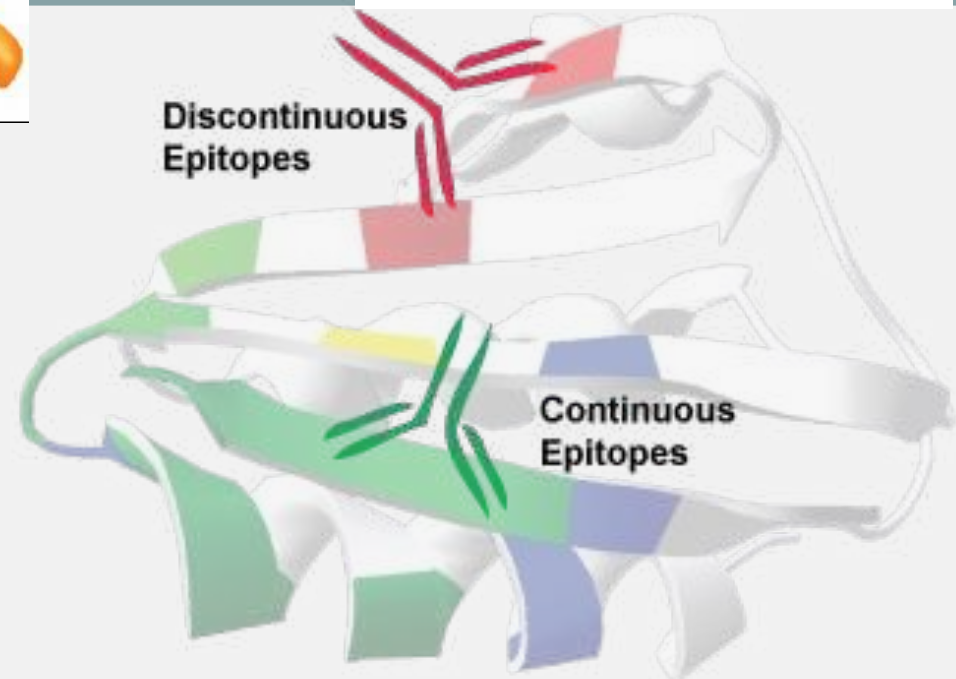
EPITOPE = antigenic determinant



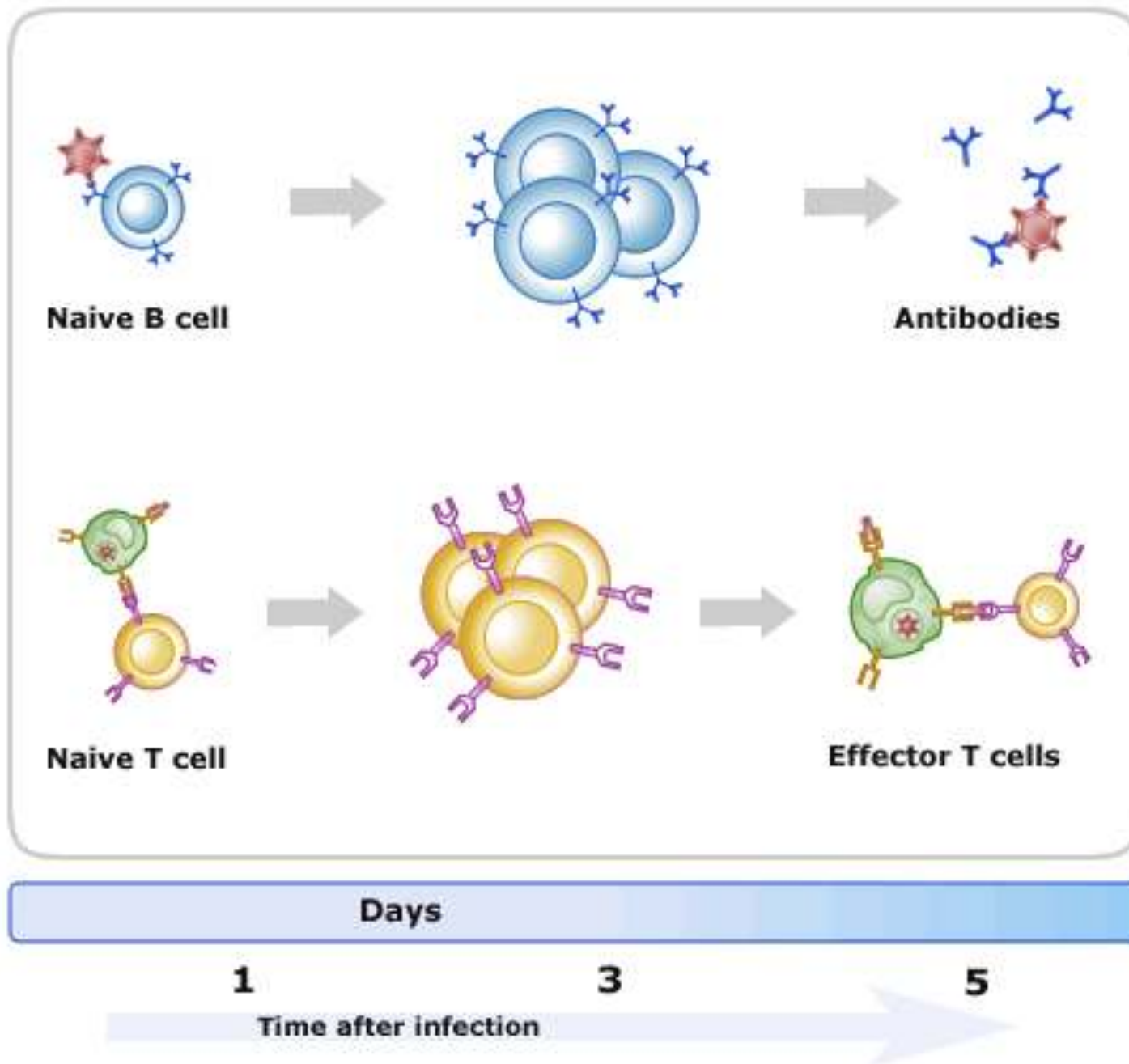
EPITOPE = antigenic determinant



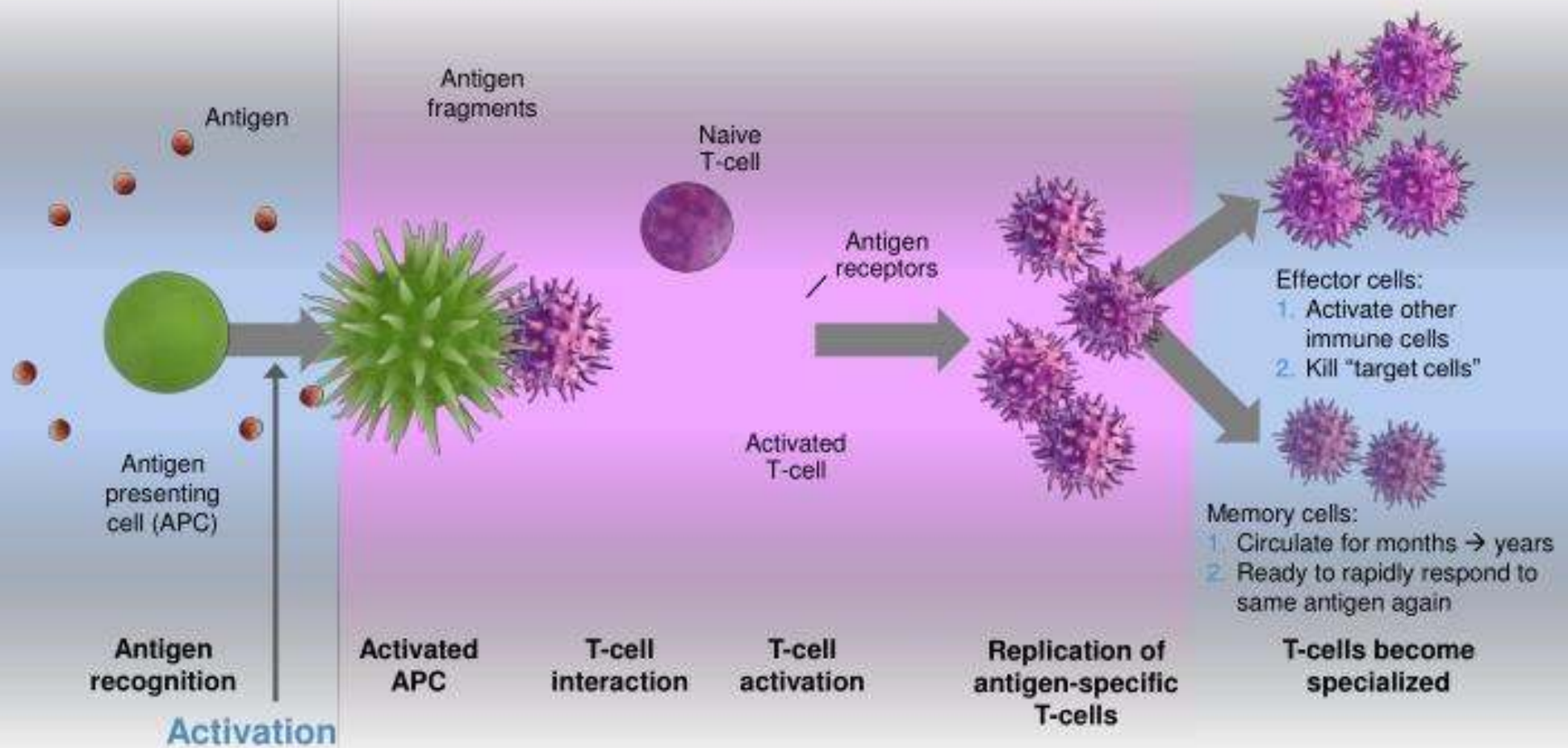
2 types of epitopes



Adaptive Immunity



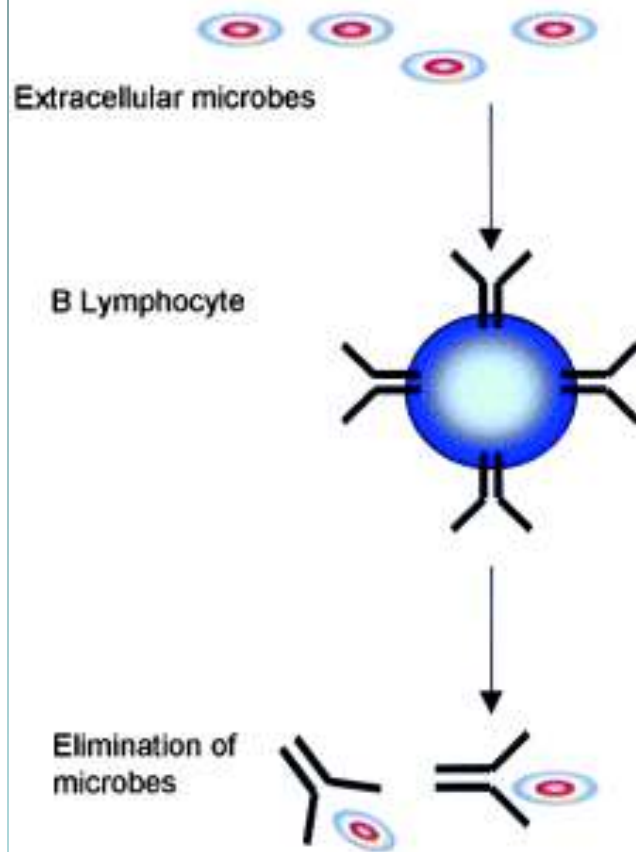
Initiation of Immune Response: Key Components



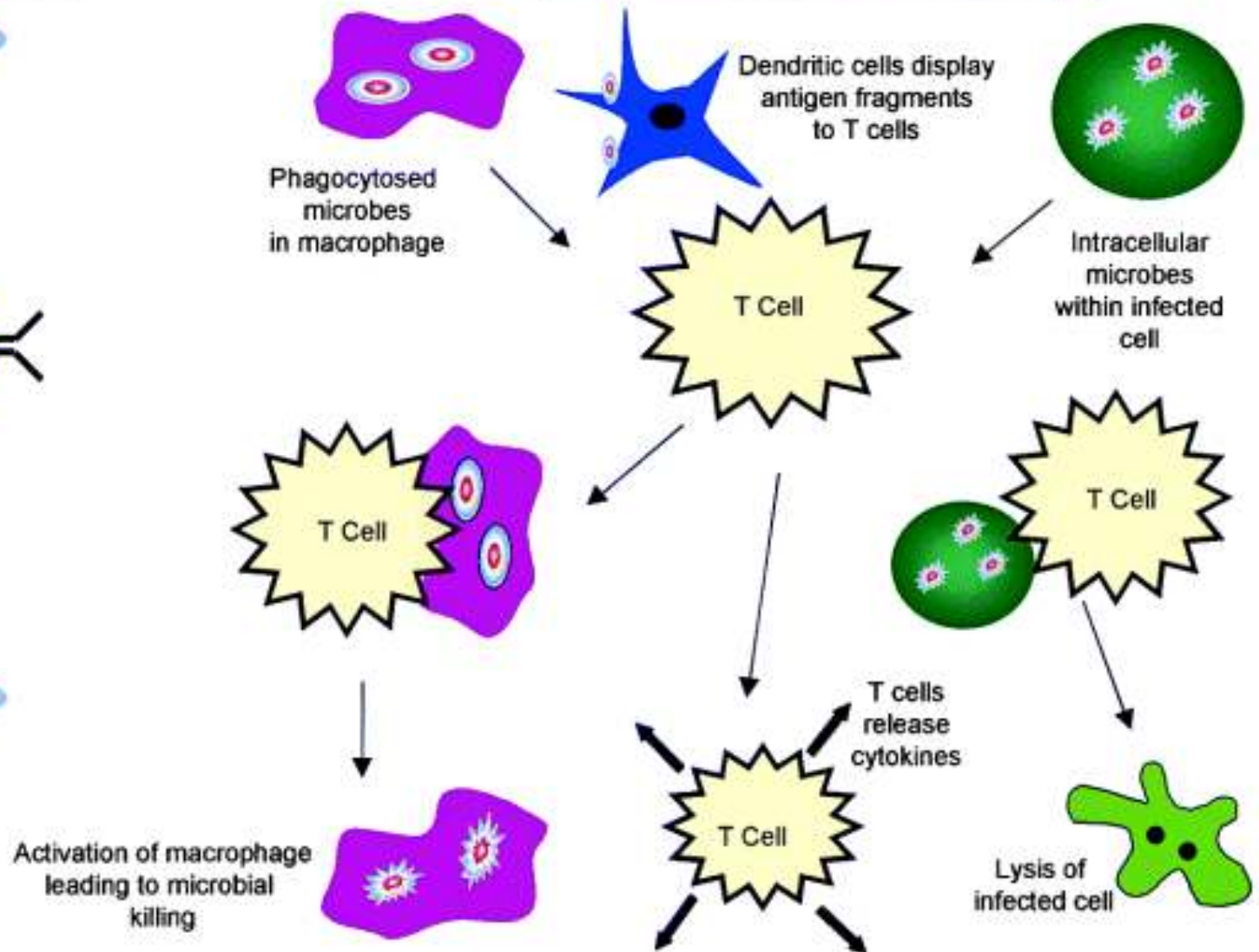
Adapted from Abbas AK, Lichtman AH.

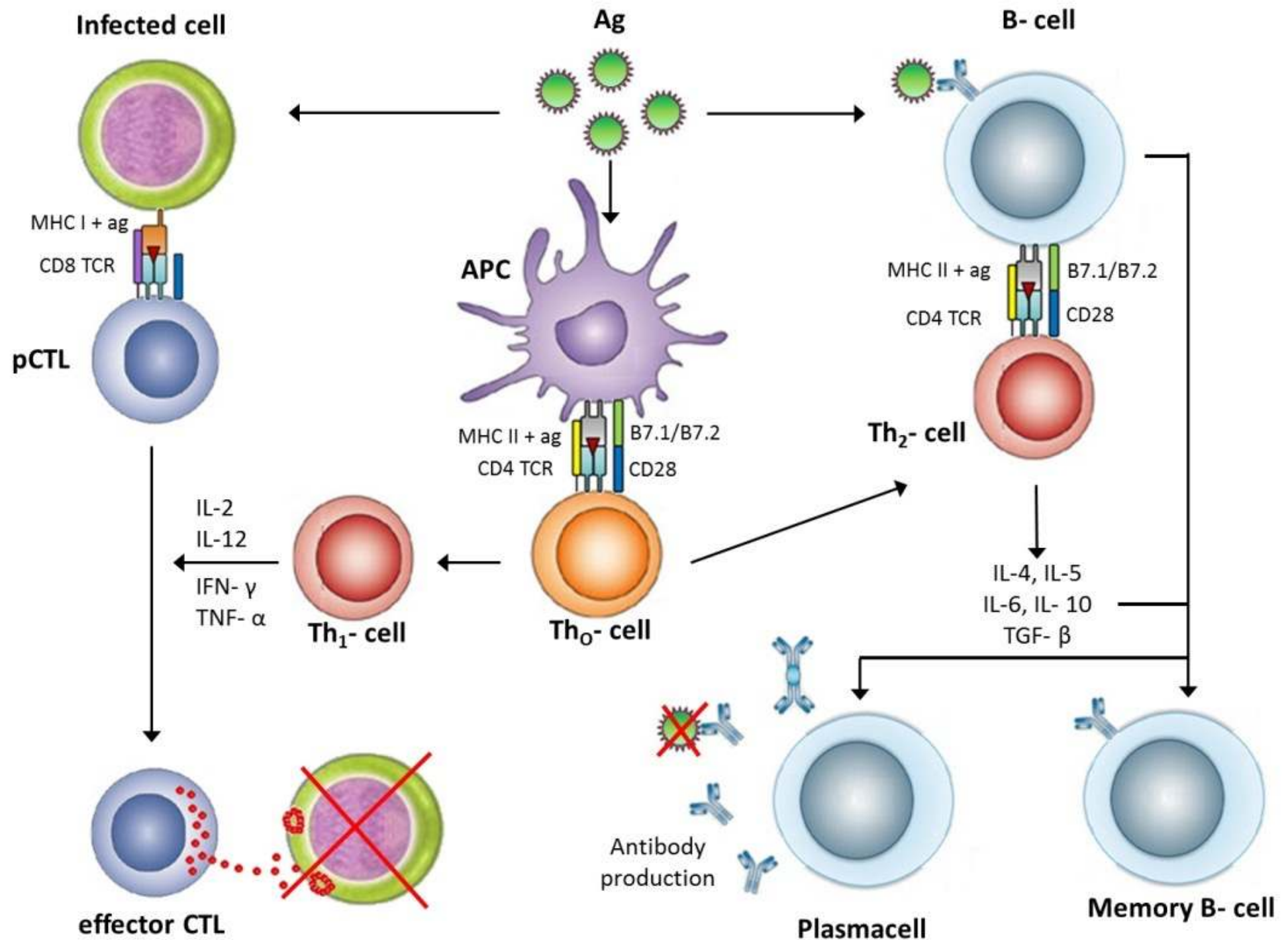
Adaptive Immunity

Humoral immunity

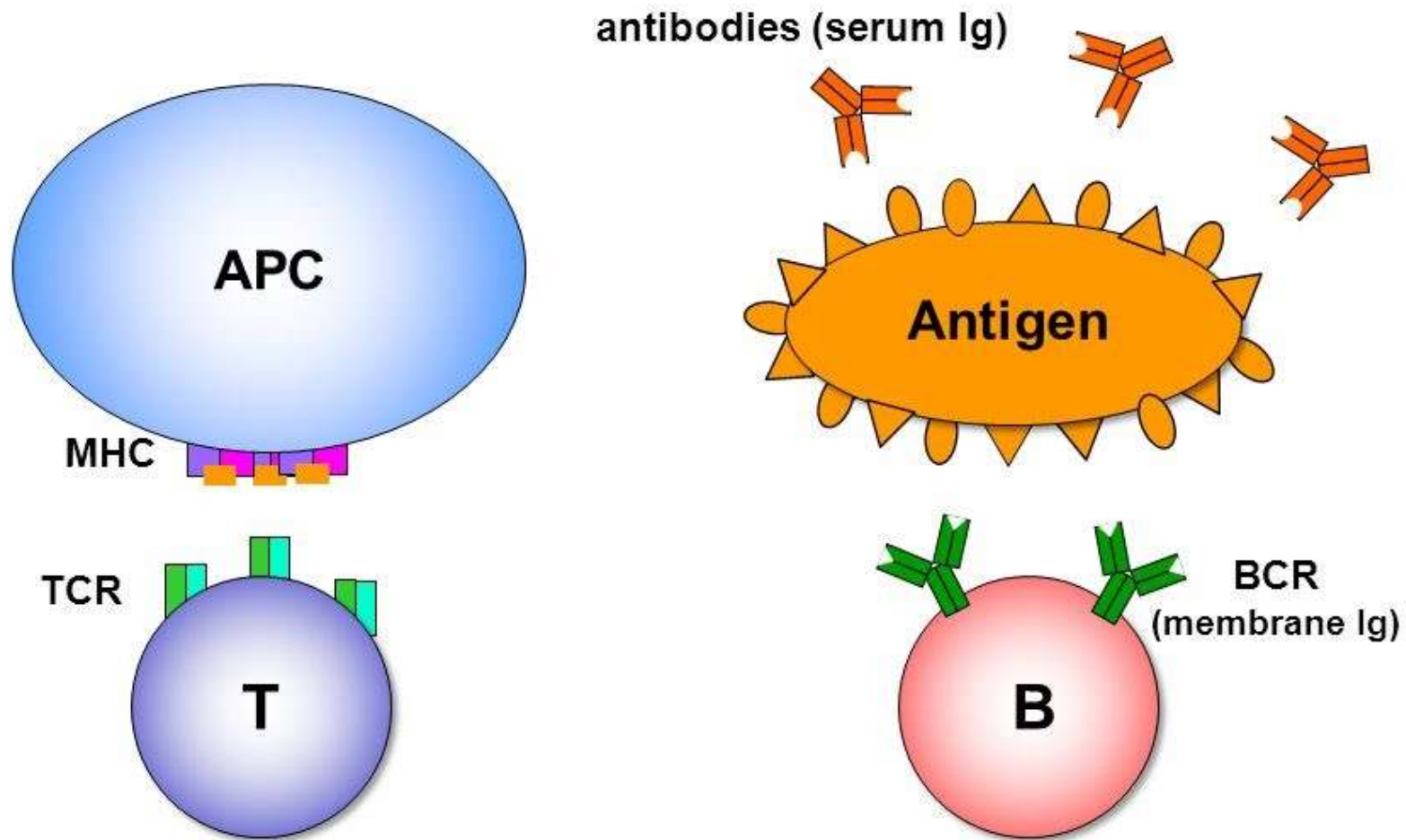


Cell-mediated immunity





ANTIGEN RECOGNITION BY LYMPHOCYTES



B cells recognise native antigens
T cells recognise processed antigens

T cell recognition of antigen peptide

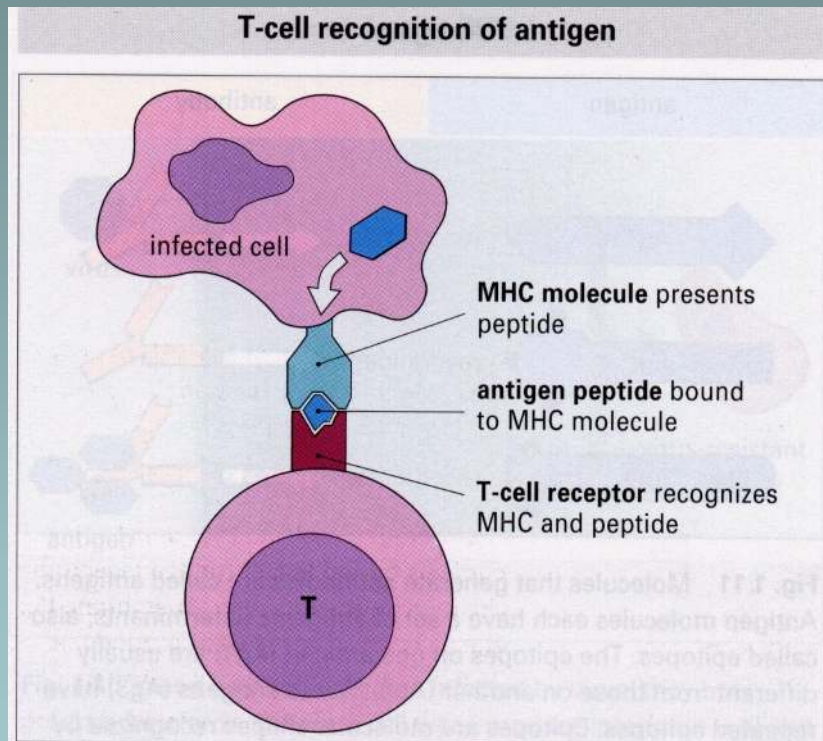
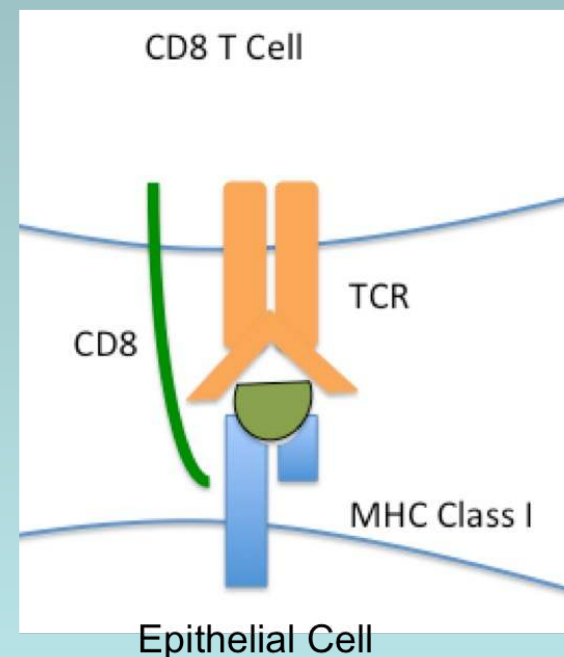
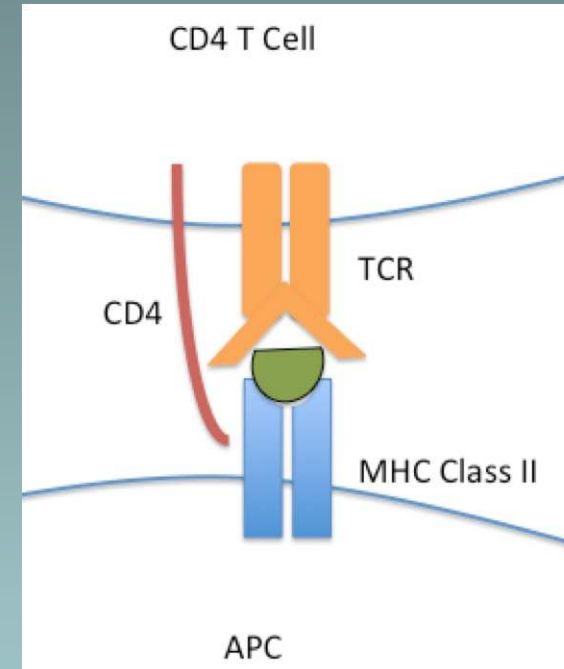
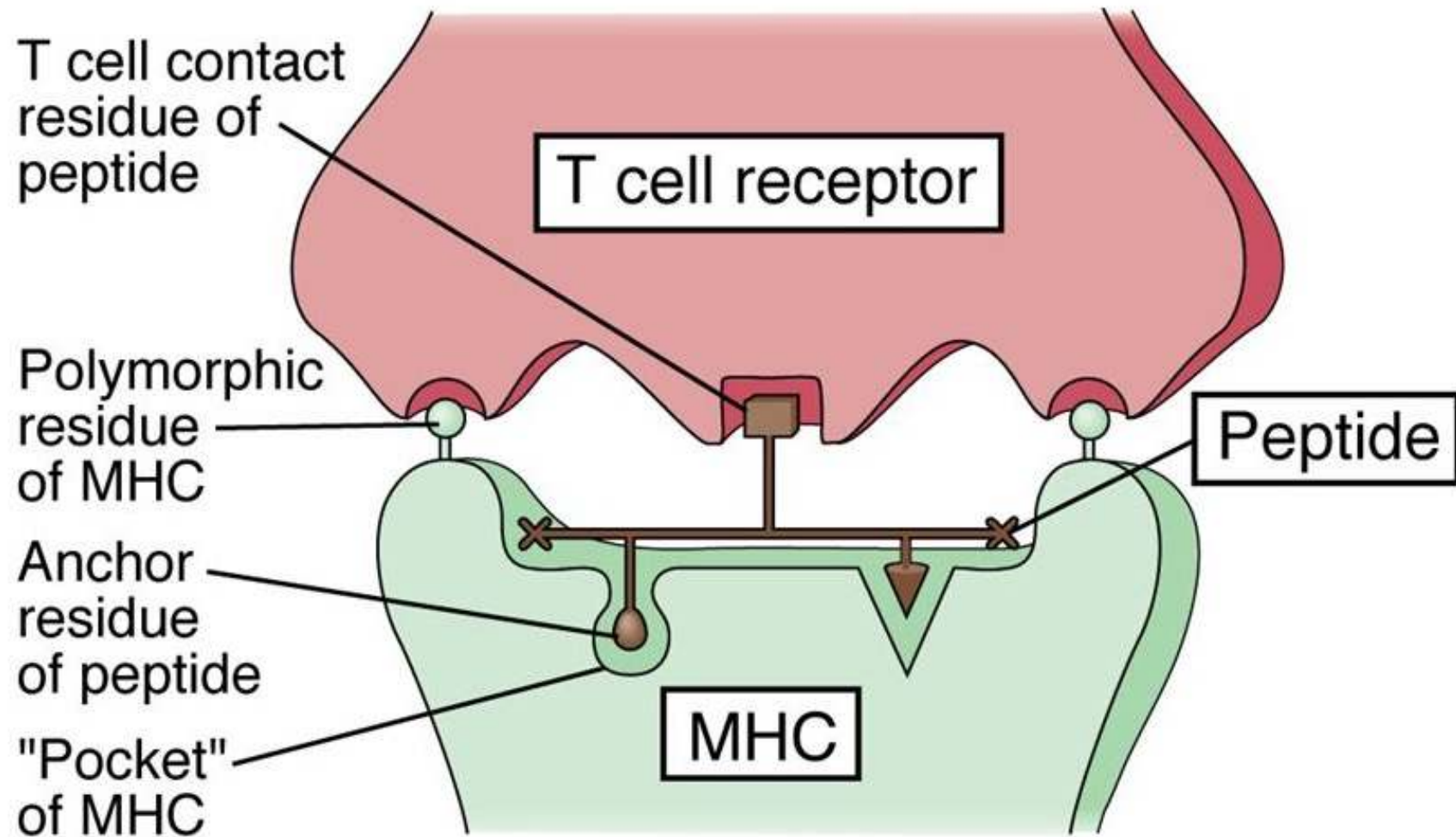


Fig. 1.12 T cells recognize antigens that originate within other cells, such as viral peptides from infected cells. They do this by binding specifically to antigenic peptides presented on the surface of the infected cells by molecules encoded by the major histocompatibility complex (MHC molecules). The 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.



Schematic model of T cell recognition of antigen



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 from 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.

Binding of peptide to MHC molecule

- Each class I or class II MHC molecule has a single peptide-binding cleft that binds one peptide at a time, but each MHC molecule can bind many different peptides.
- MHC molecules acquire their peptide cargo during their biosynthesis and assembly inside cells.
- The association of antigenic peptides and MHC molecules is a saturable interaction with a very slow off-rate.
- Very small numbers of peptide-MHC complexes are capable of activating specific T lymphocytes.

2 types of MHC molecules

MHC τάξης I → CD8 / T_c

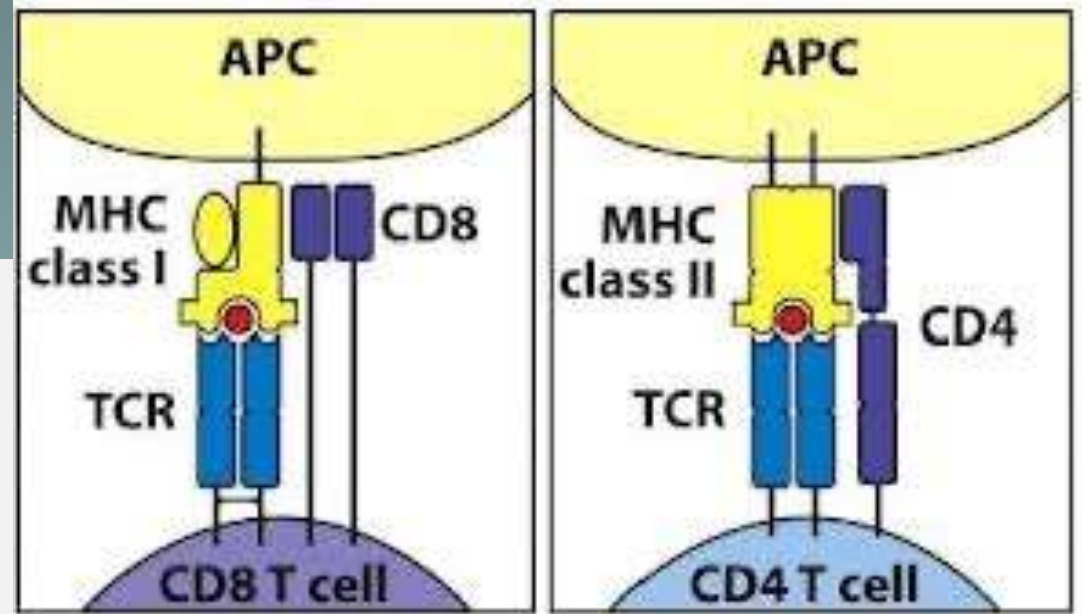
peptides in the cytoplasm

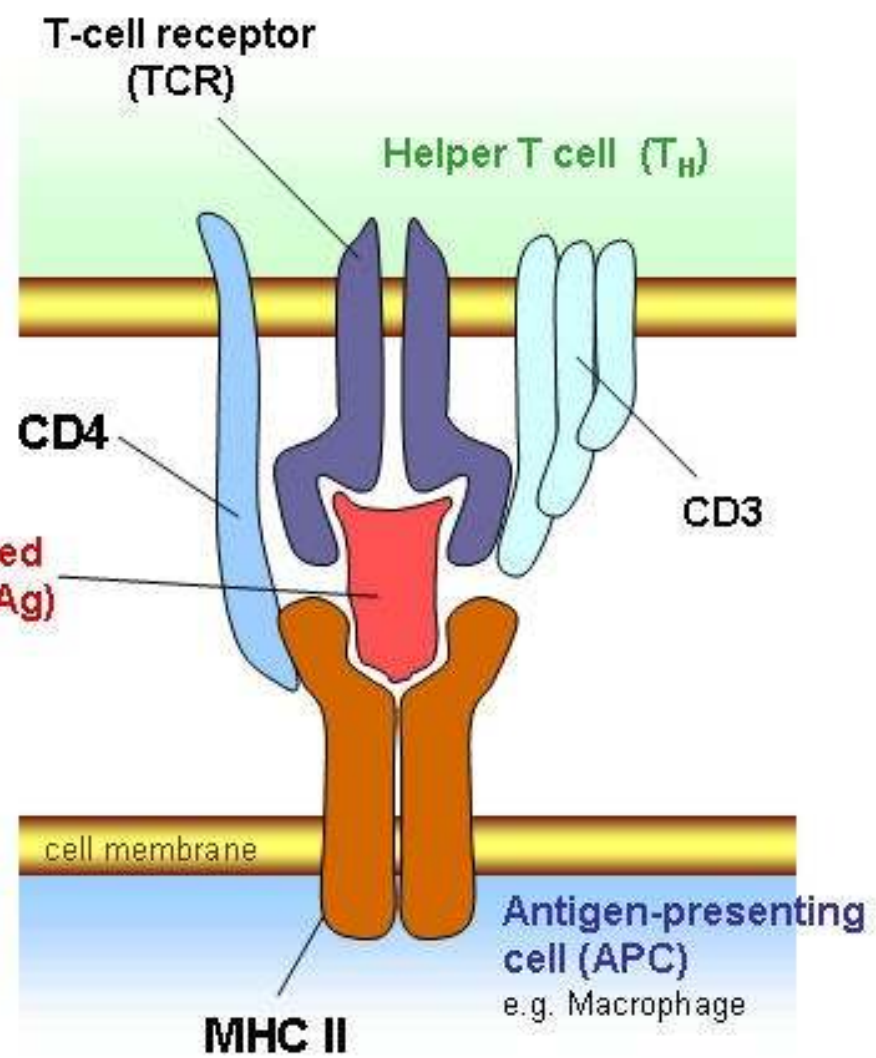
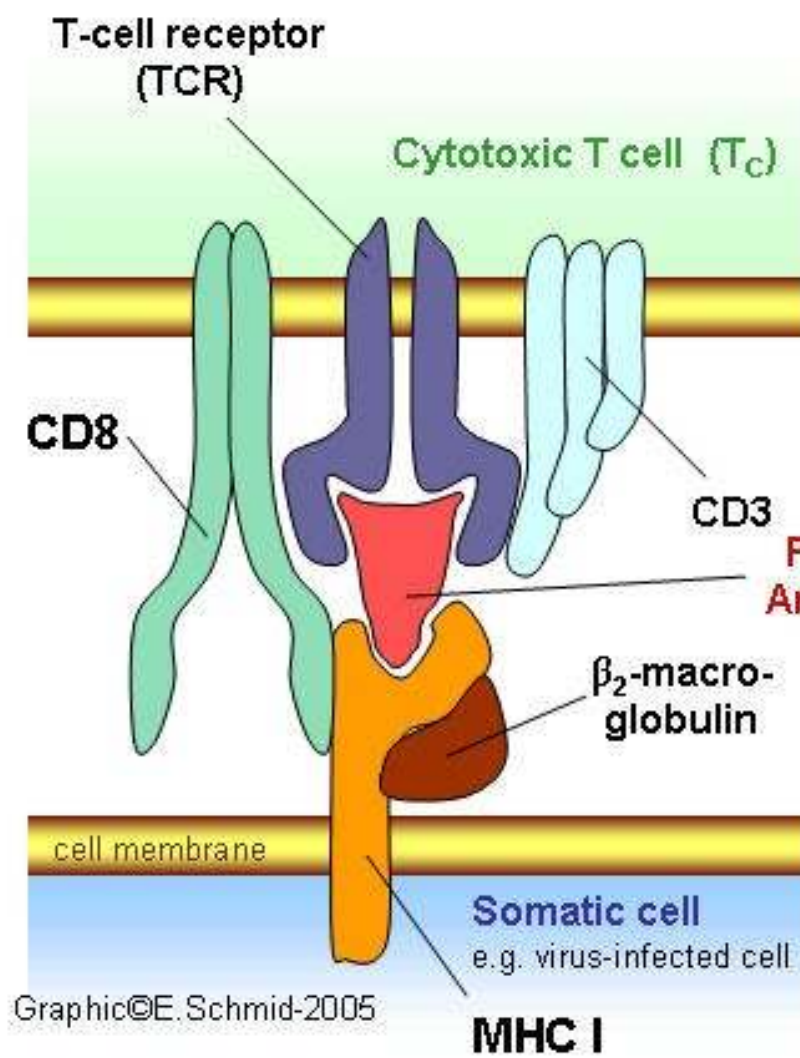
endogenous proteins

MHC τάξης II → CD4 / T_H

peptides in vesicles

exogenous proteins

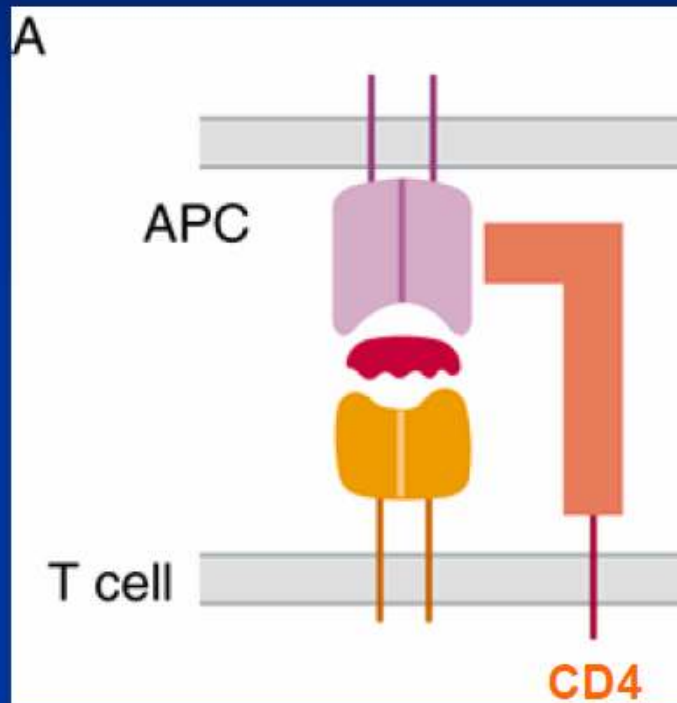




Graphic©E. Schmid-2005

εξωγενή Ag

ΜHC τάξης II

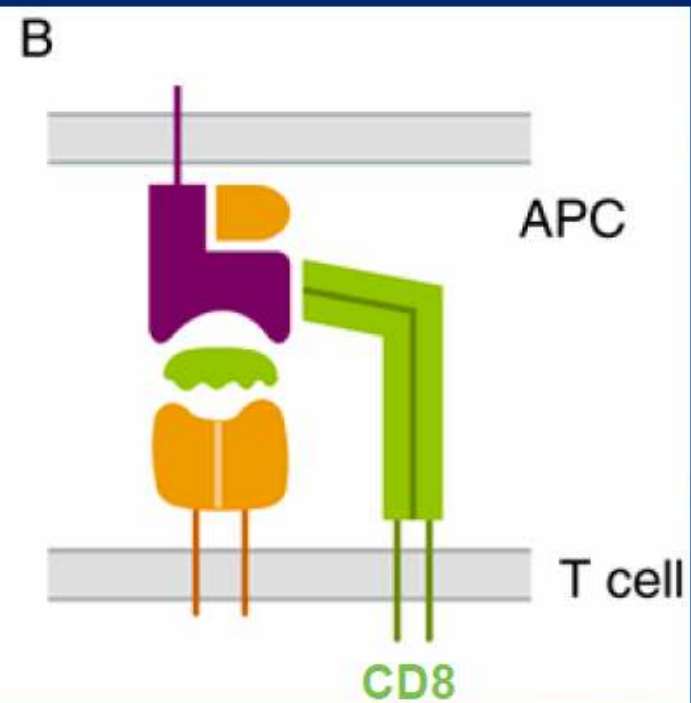


CD4

↓
HELP

ενδογενή Ag

ΜHC τάξης I



CD8

↓
KILL

Antigen processing and presentation to T lymphocytes

MHC class I pathway (cytosolic source)

Present antigen to **CD8** T cells

Virus and intracellular bacteria

Mutated tumor antigen

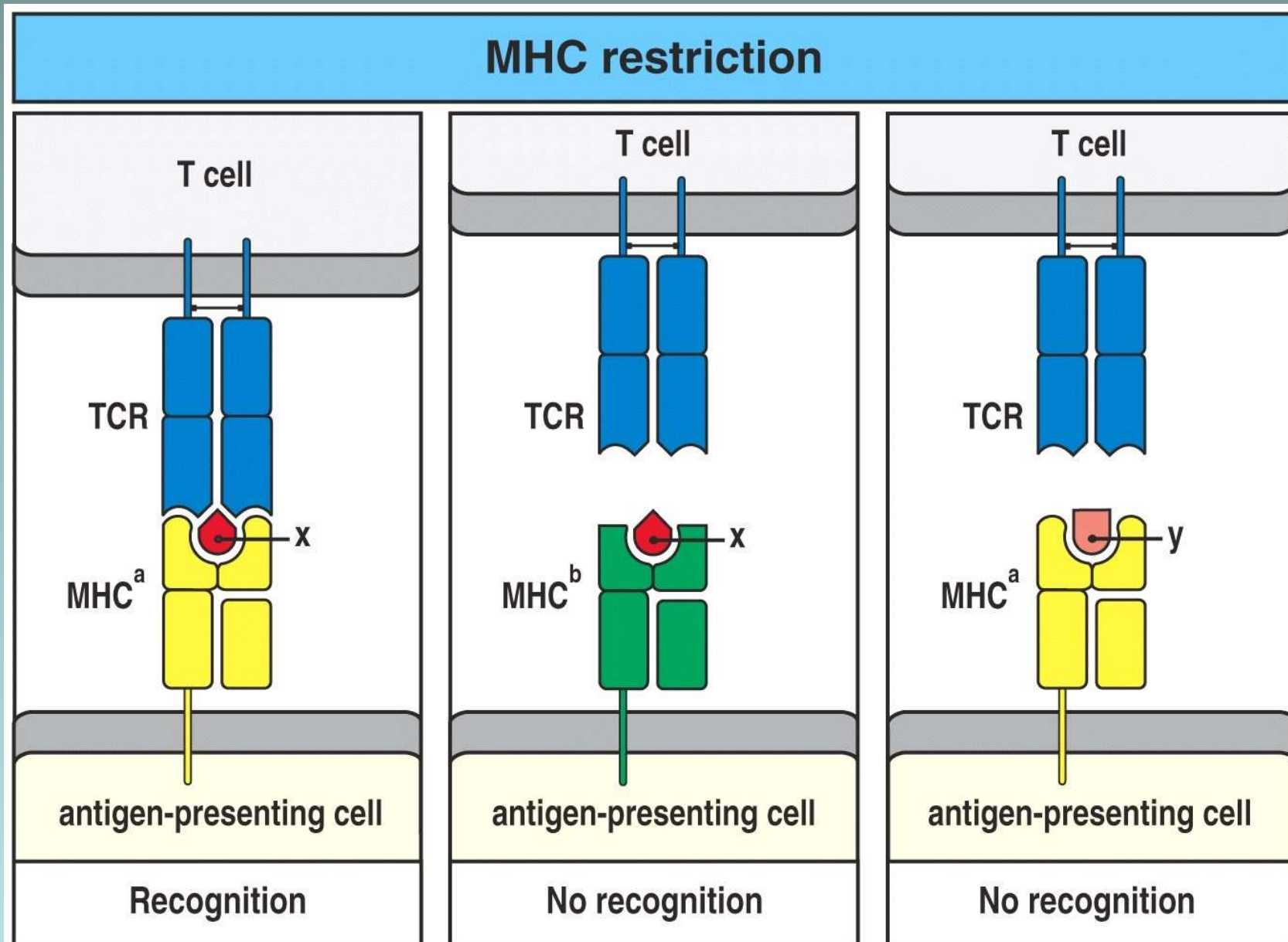
MHC class II pathway (endosomal source)

Present antigen to **CD4** T cells

Bacteria

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

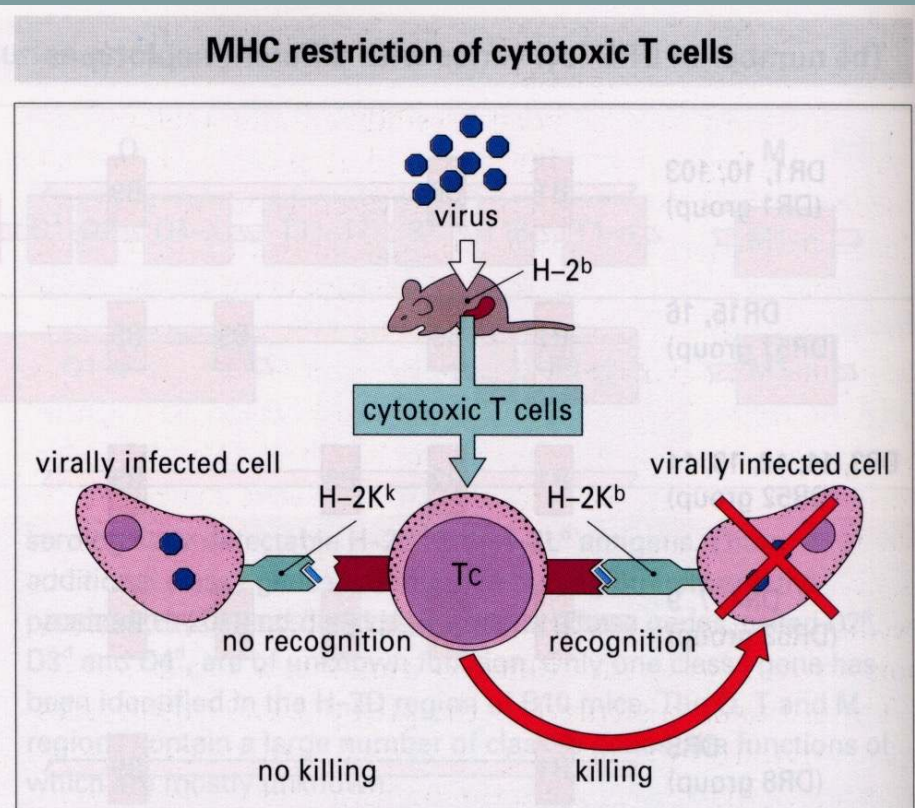


Fig. 5.19 A mouse of the H-2^b haplotype is primed with virus and the T_c cells thus generated are isolated and tested for their ability to kill H-2^b and H-2^k cells infected with the same virus. The T_c cells kill H-2^b, but not H-2^k cells. In this instance, it is the H-2K class I gene product which 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.

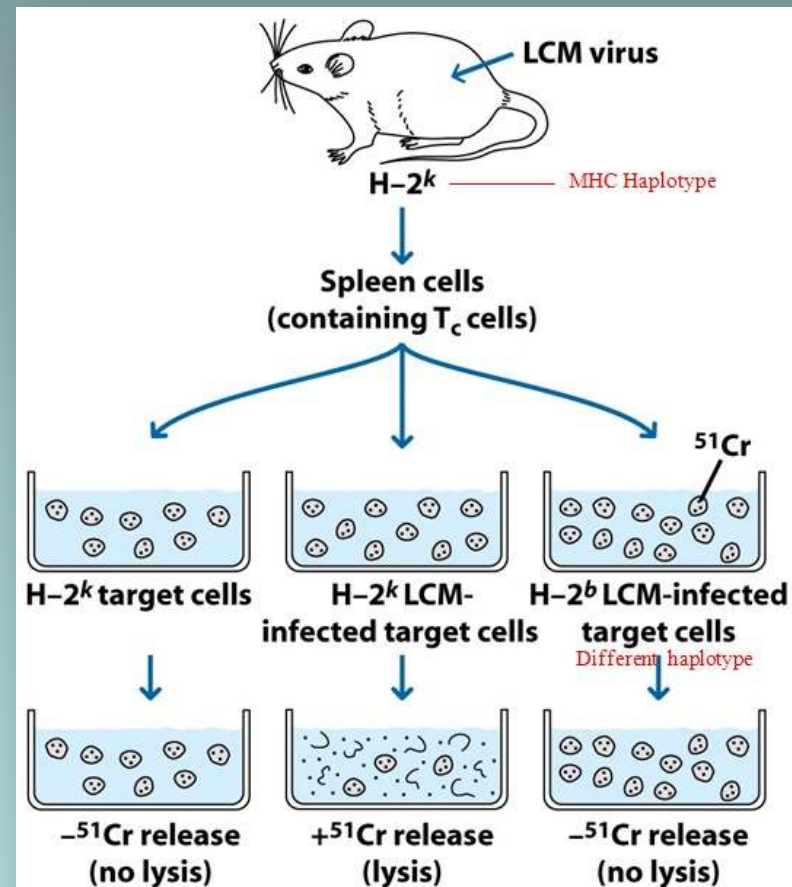
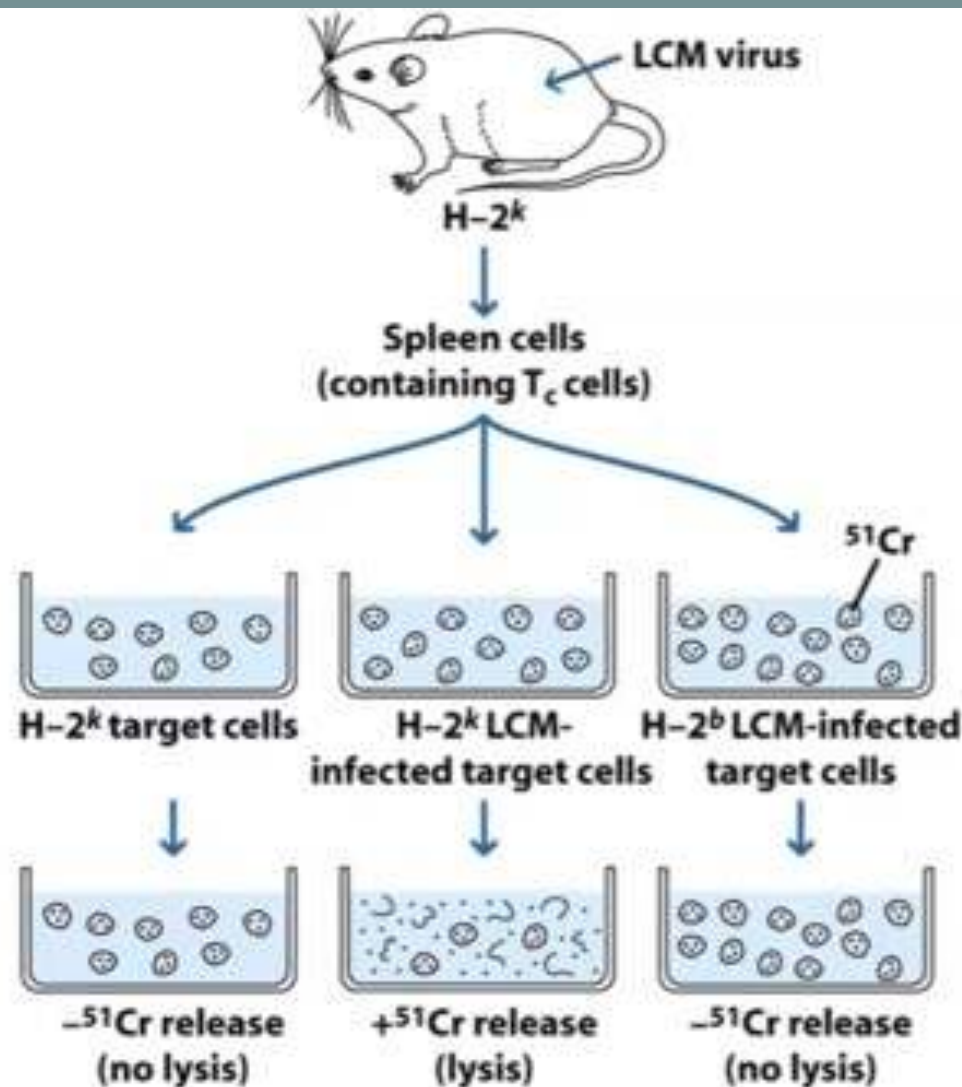


Figure 8-15
Kuby IMMUNOLOGY, Sixth Edition
© 2007 W.H. Freeman and Company

Shows that CTLs can only kill cells infected with LCM and have same haplotype



The concept of MHC restriction-Experimental evidence that antigen recognition by T cells depends on the presence of specific SELF MHC molecules in APCs is known as a phenomenon of MHC restriction.

CD8⁺ T cells are MHC class I restricted.

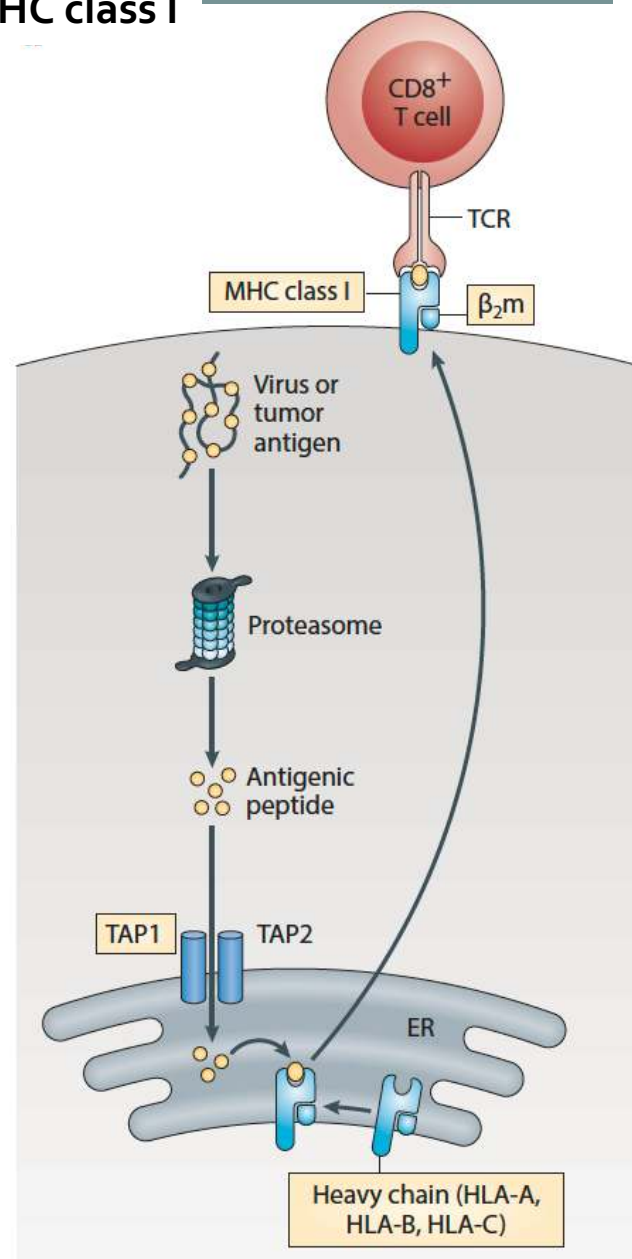
Doherty and Zinkernagel (1975, JEM) Nobel prize in 1996

Figure 8-15
Kuby IMMUNOLOGY, Sixth Edition

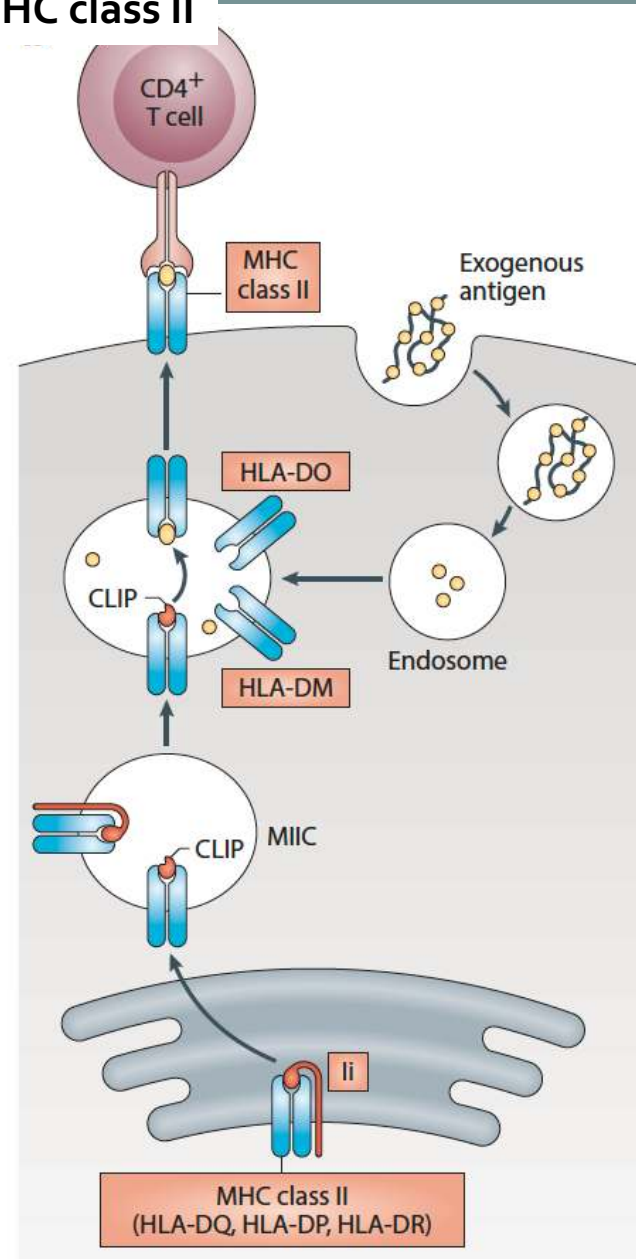
- **CD8+ T_c cells are MHC class I restricted**
 - Can only recognize antigen presented by MHC class I molecules
 - All nucleated cells express MHC class I
 - Cells with MHC class I are “target cells” and can be killed by cytotoxic T cells
- **CD4+ T_H cells are MHC class II restricted**
 - Cells with MHC class II are antigen-presenting cells (APCs)

Antigen Processing (επεξεργασία) and Presentation (παρουσίαση)

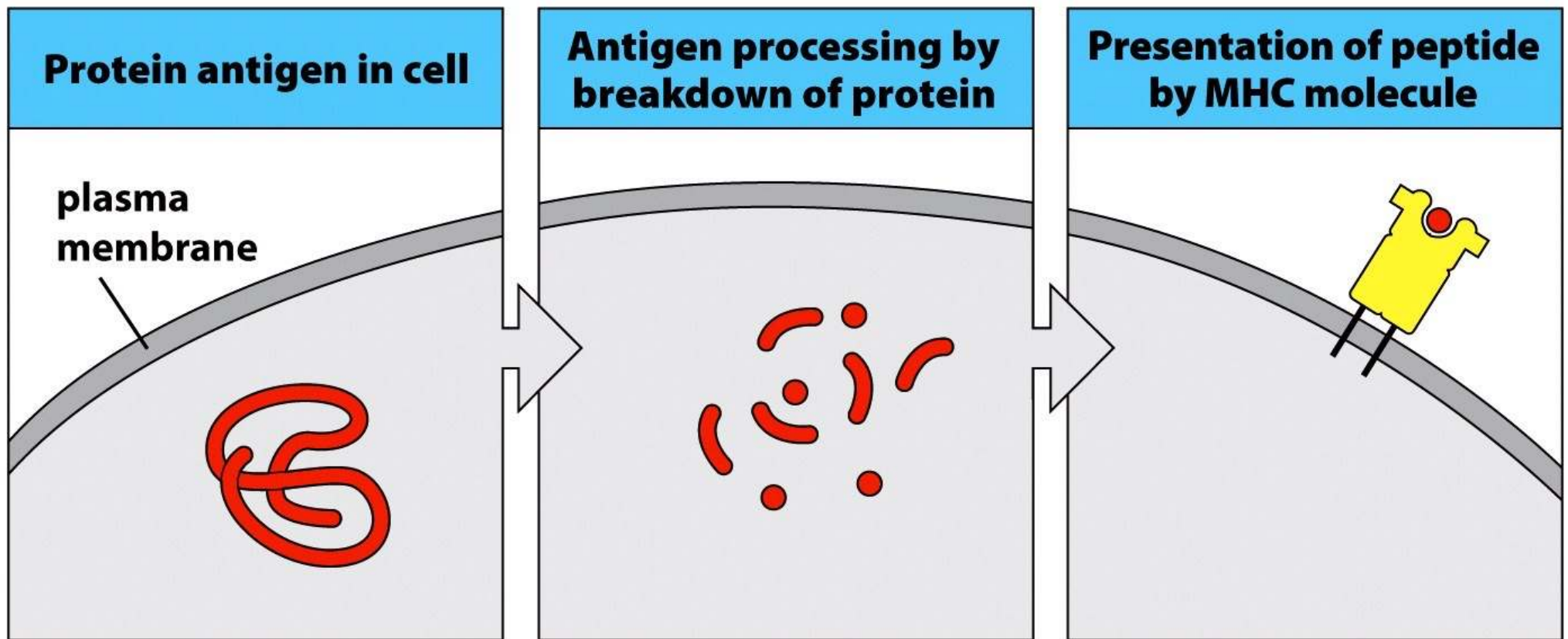
MHC class I



MHC class II



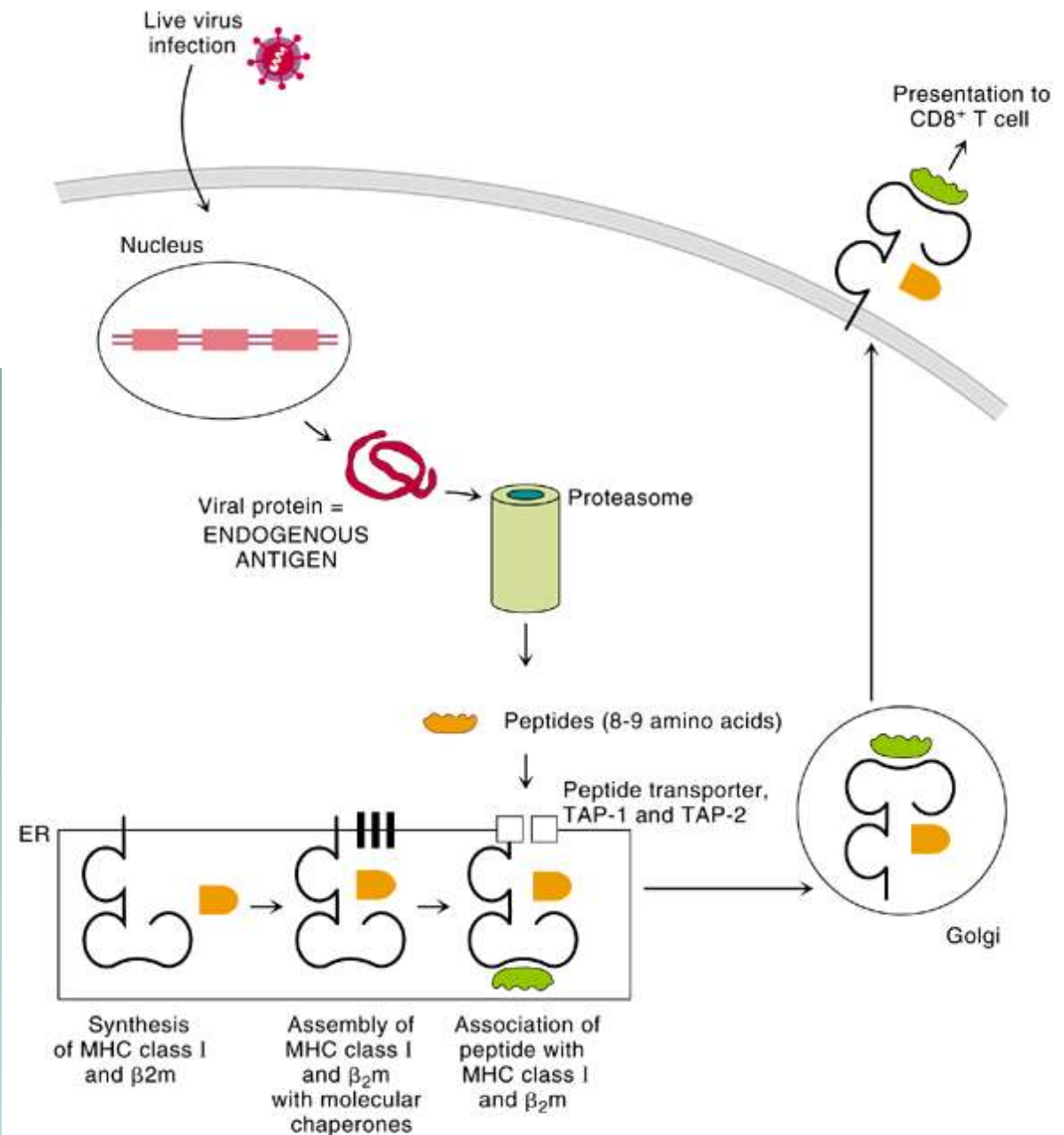
Antigen Processing (επεξεργασία) and Presentation (παρουσίαση)



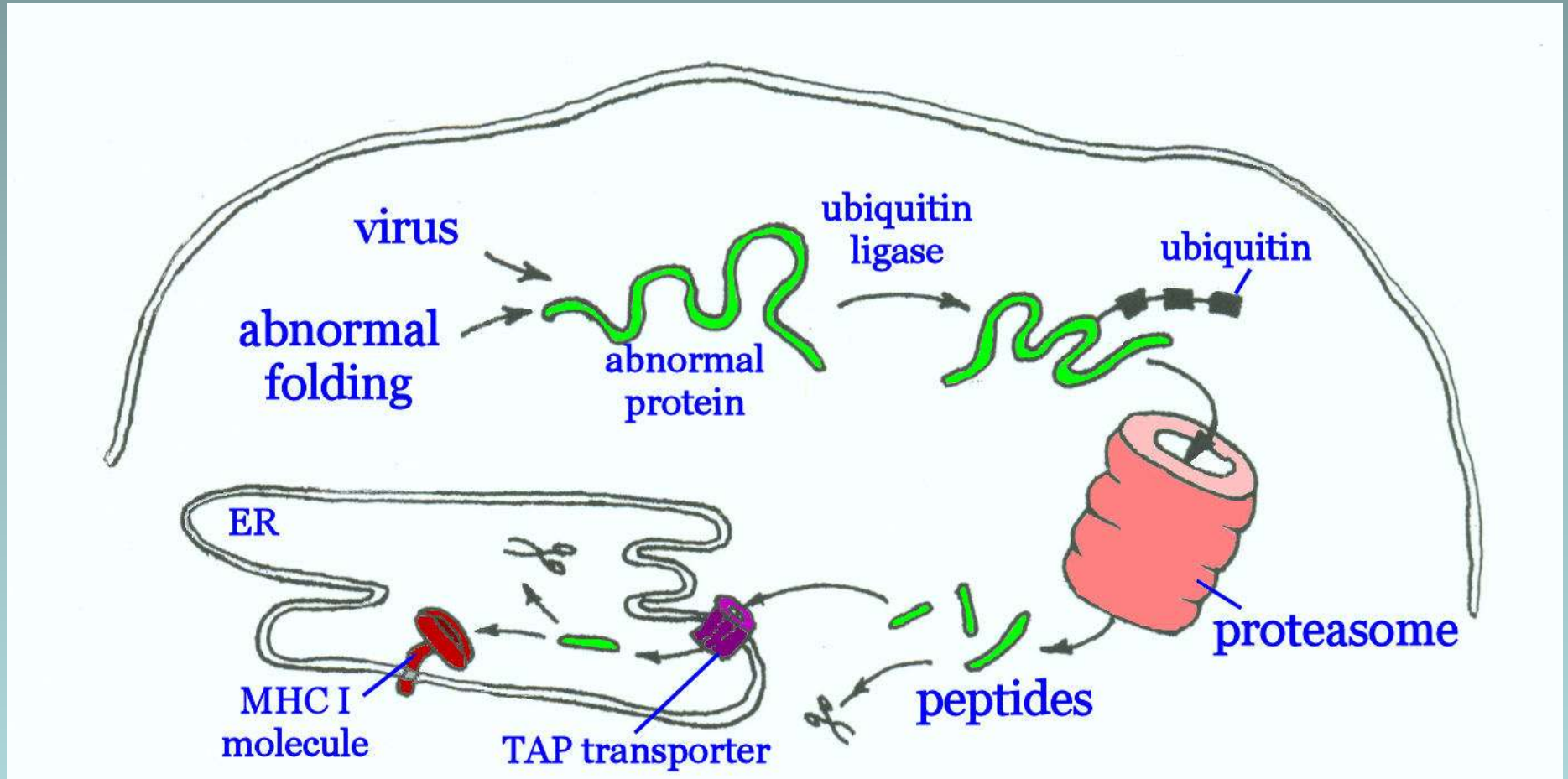
Antigen processing and Peptide binding to MHC τάξης I

Presentation to CD8⁺ T cells

Endogenous Antigens

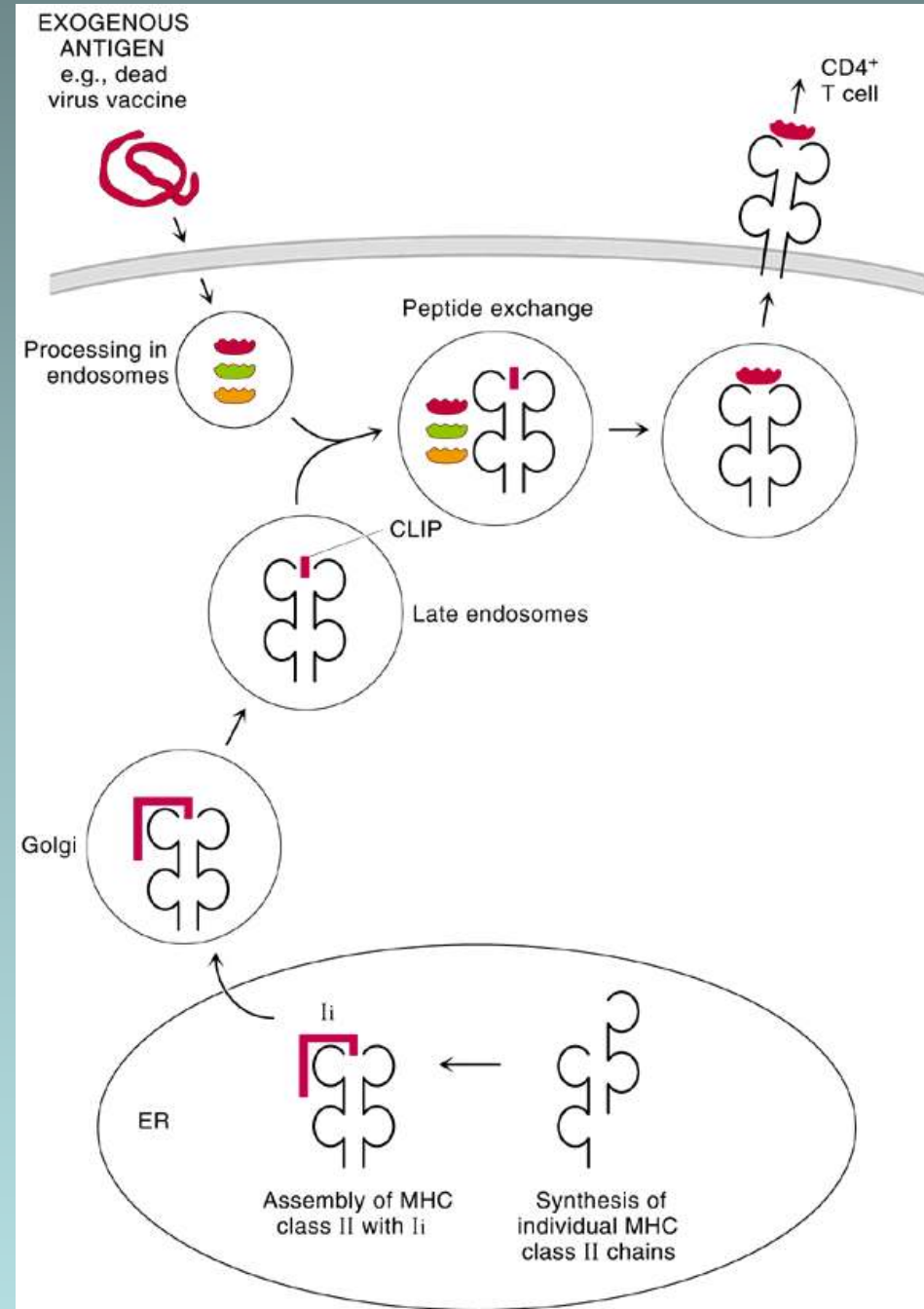


Antigen Processing for MHC class I



**Antigen processing and Peptide
binding to MHC τάξης II
Presentation to CD4⁺ T cells**

Exogenous Antigens

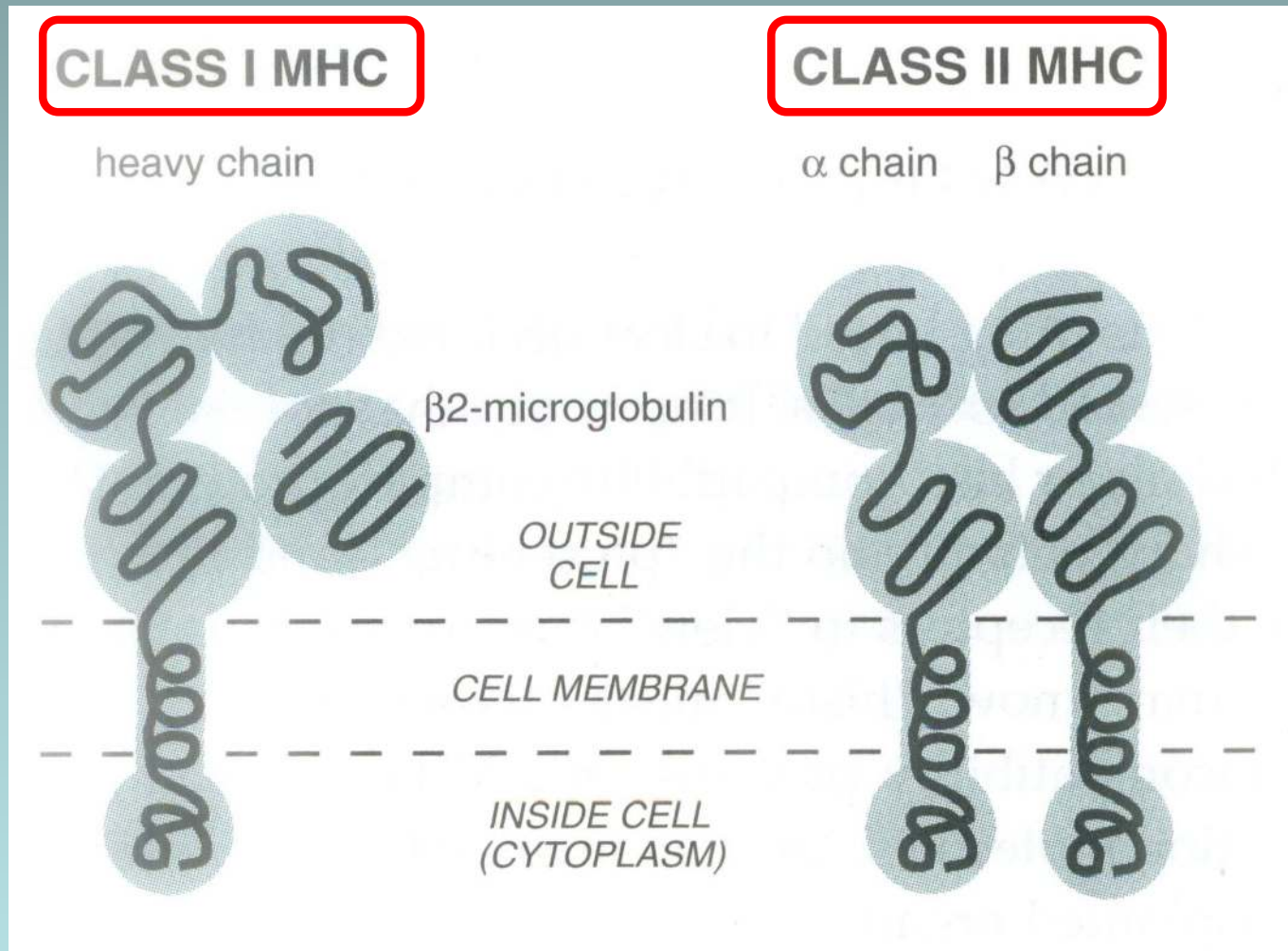


Antigen Processing Pathways

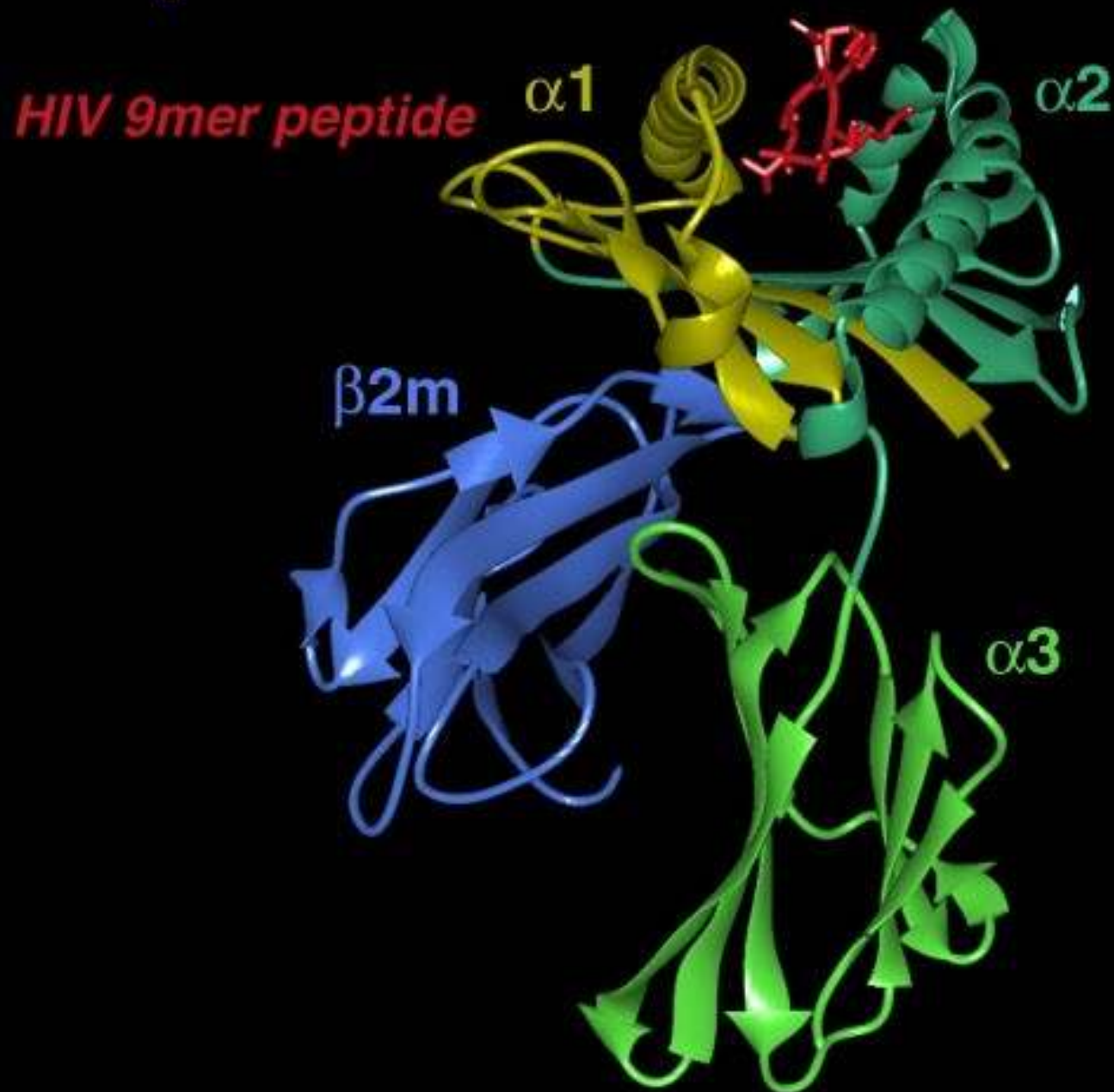
- MHC class II
 - Exogenous protein Ags
 - Peptides made in acidic vesicles by proteases
 - No equivalent
 - Peptides bind class II in acidic vesicles
 - Peptide:class II complexes presented to CD4⁺ T cells
- MHC class I
 - Endogenous protein Ags
 - Peptides made in cytosol by proteasome
 - TAP transports peptides to ER lumen
 - Peptides bind class I in ER lumen
 - Peptide:class I complexes presented to CD8⁺ T cells

https://www.youtube.com/watch?v=LwLYGTS_3EI

The two classes of MHC have similar
three-dimensional structures



MHC protein HLA-A2 with HIV peptide



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.

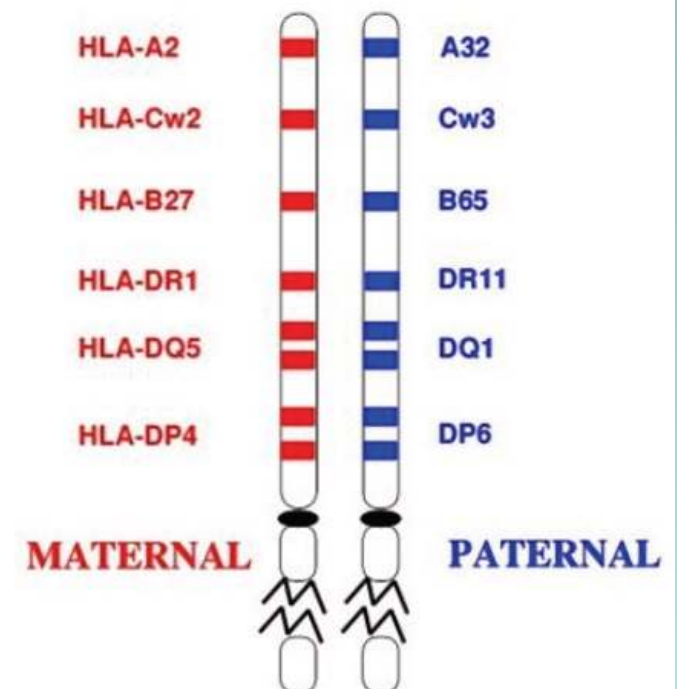
Most individuals heterozygous at MHC locus.

MHC haplotype.

Expression co-dominant.

MHC polymorphism triggers T-cell reactions that can reject transplanted organs.

**Most Humans are heterozygous
at the MHC**



HLA Polymorphism & Epitope Recognition

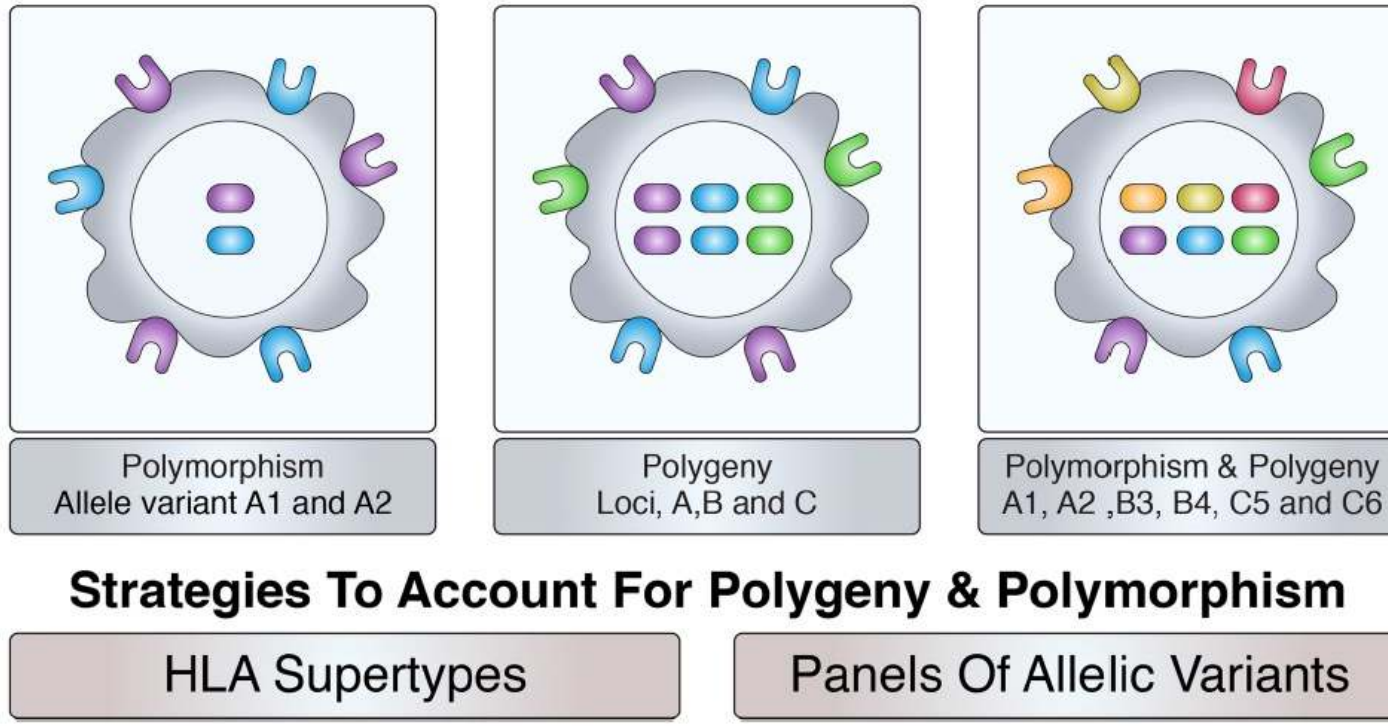
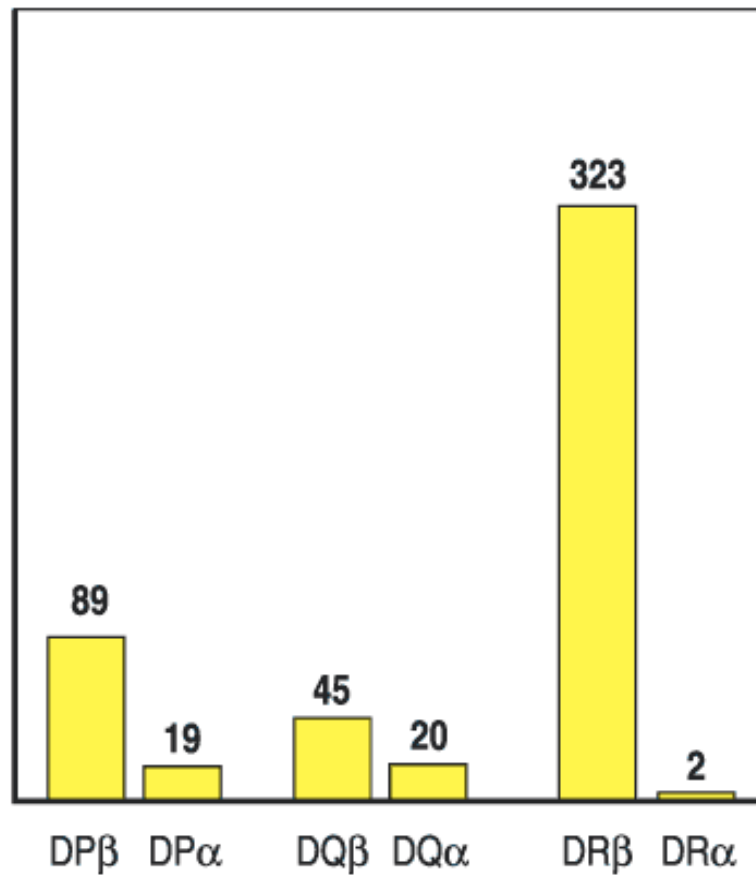
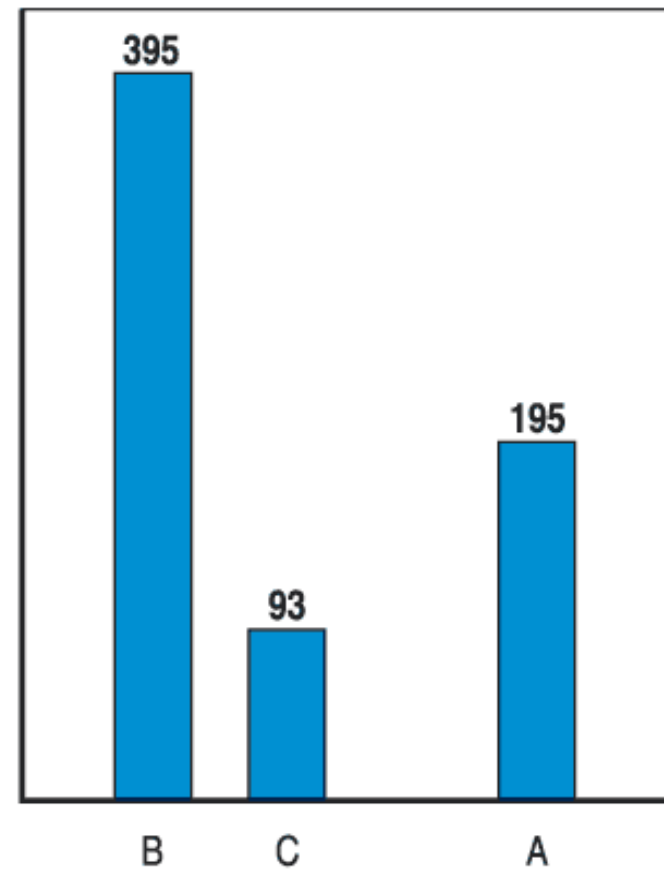


Fig. 5. HLA polymorphism and polygeny. The HLA region is highly polymorphic, with thousands of different allelic variants expressed at each locus. With heterozygosity, each individual will express up to two different alleles at each locus. Further, with class I and II presentation across several loci (polygeny), up to six different class I (2 HLA-A, 2 HLA-B and 2 HLA-C) and eight different class II (2 HLA-DRB1, 2 HLA-DRB3/4/5, 2 DQ and 2 DP) molecules may be expressed by an individual.

MHC class II



MHC class I



Mouse H-2 complex

Complex	H-2						
MHC class	I	II		III		I	
Region	K	IA	IE	S		D	
Gene products	H-2K	IA $\alpha\beta$	IE $\alpha\beta$	C' proteins	TNF- α TNF- β	H-2D	H-2L

Human HLA complex

Complex	HLA							
MHC class	II			III		I		
Region	DP	DQ	DR	C4, C2, BF		B	C	A
Gene products	DP $\alpha\beta$	DQ $\alpha\beta$	DR $\alpha\beta$	C' proteins	TNF- α TNF- β	HLA-B	HLA-C	HLA-A

Human Leukocyte Antigen (HLA)

In humans, MHC is called human leukocyte antigen (HLA)

In humans:

- There are three MHC class I genes (coding for the a chain):

HLA-A, HLA-B, και HLA-C

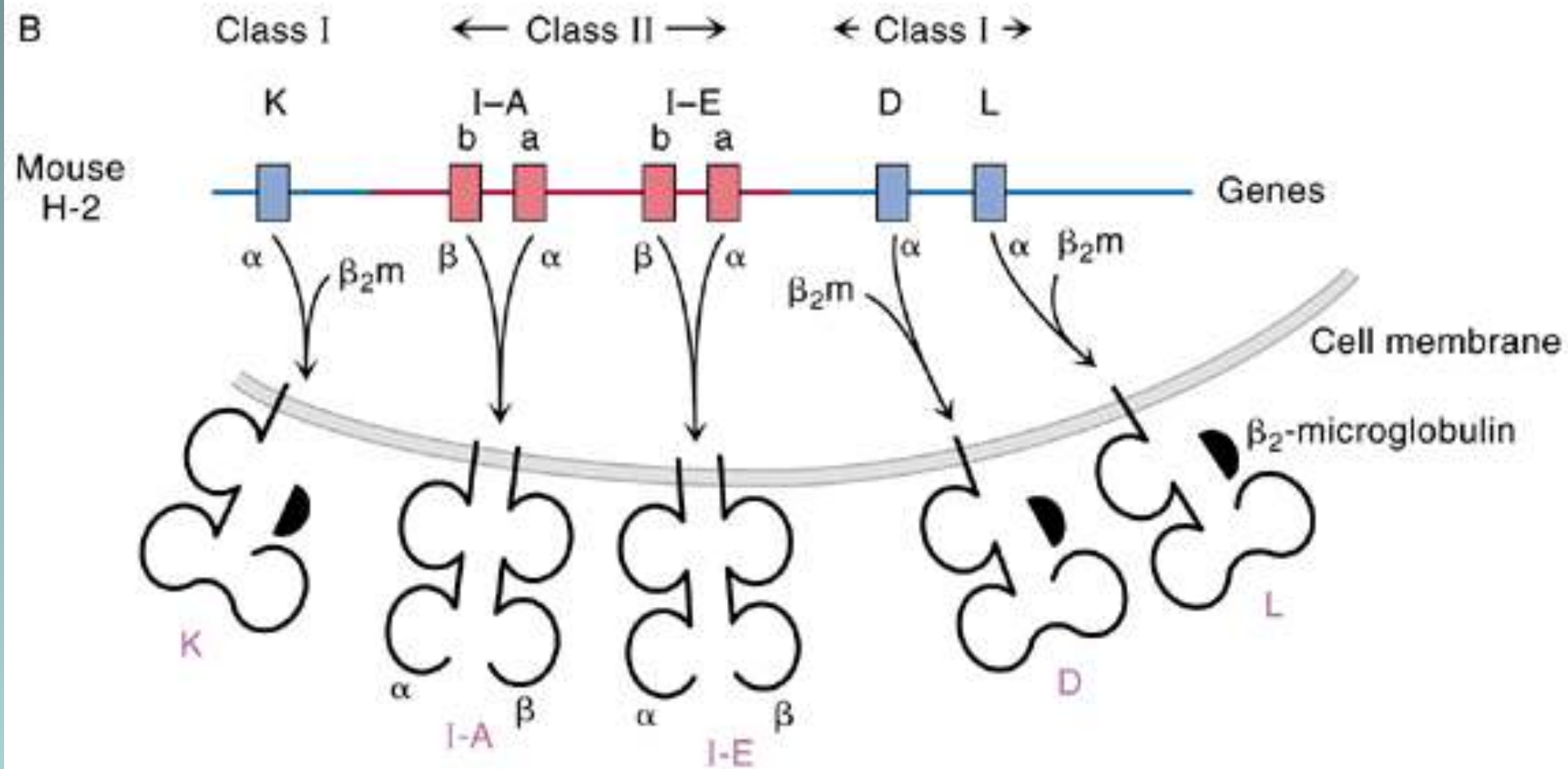
- There are 3 pairs of MHC class II genes (coding for the a and b chains): **HLA-DR, HLA-DP, και HLA-DQ**

H-2 (mouse)

In mice, MHC is called H-2

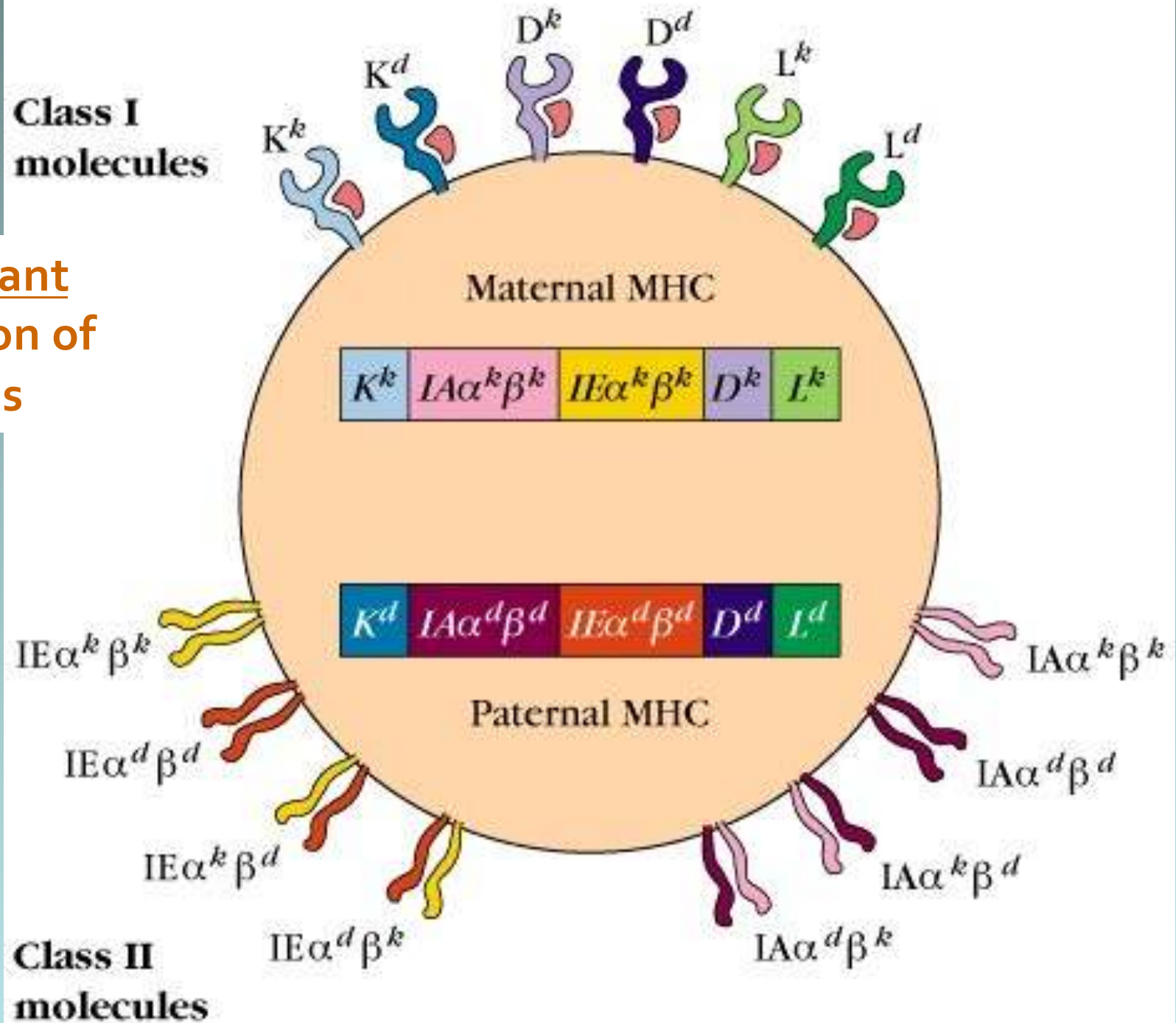
In mice:

- There are three MHC class I genes (coding for the a chain): **H-2K, H-2D, and H-2L**
- There are 2 pairs of MHC class II genes (coding for the a and b chains): **H-2A, H-2E**



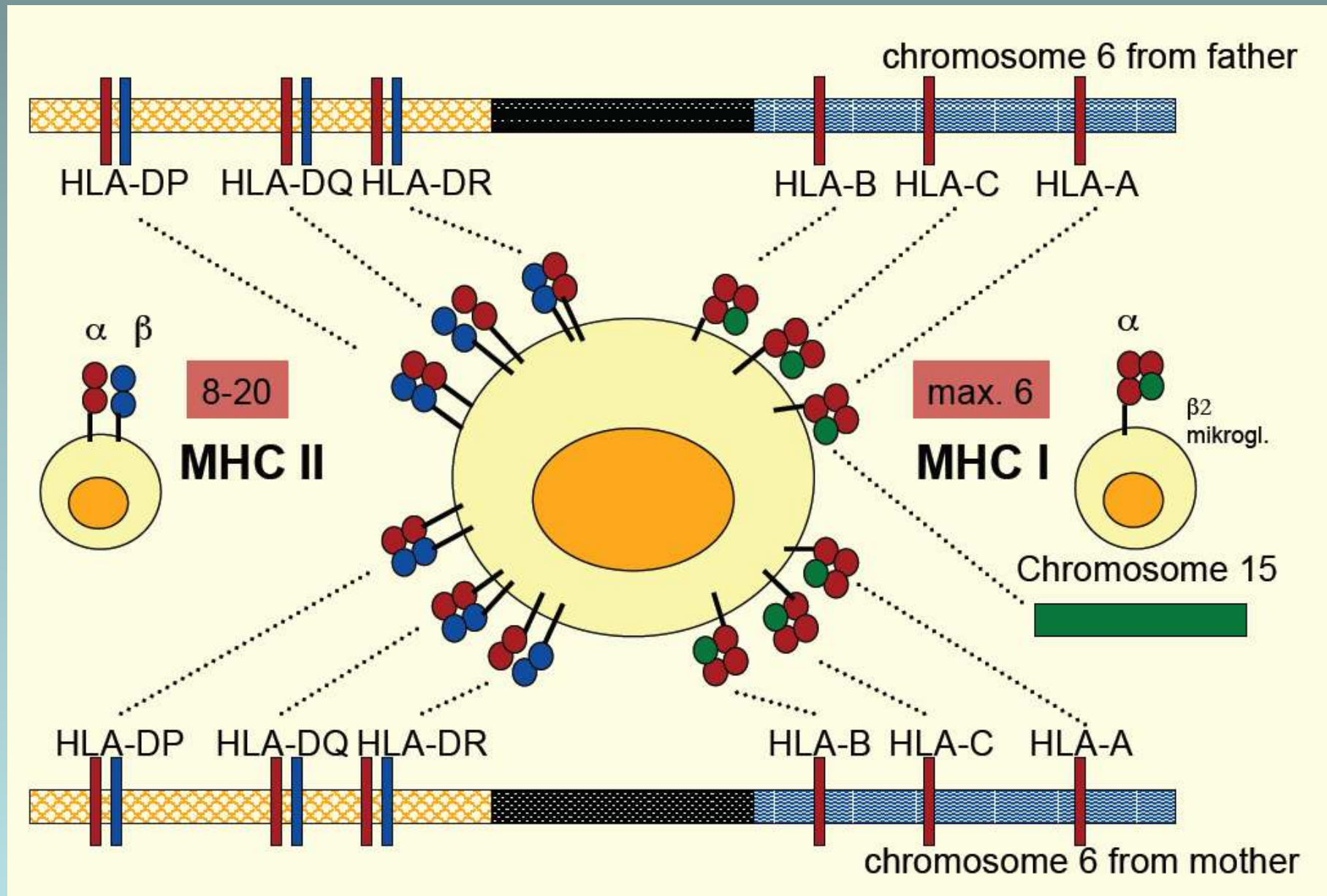
Mice

codominant
expression of
H-2 genes

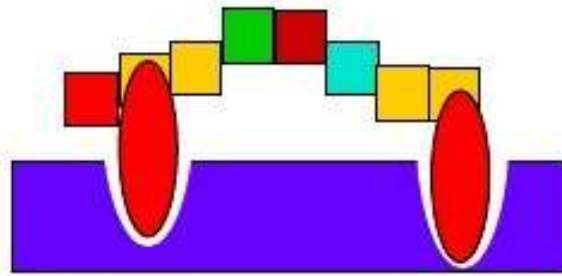


Humans

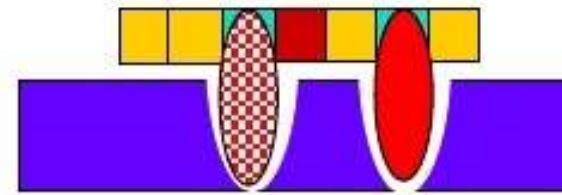
Codominant expression of HLA genes



Polymorphism in the MHC affects peptide antigen binding



MHC allele A



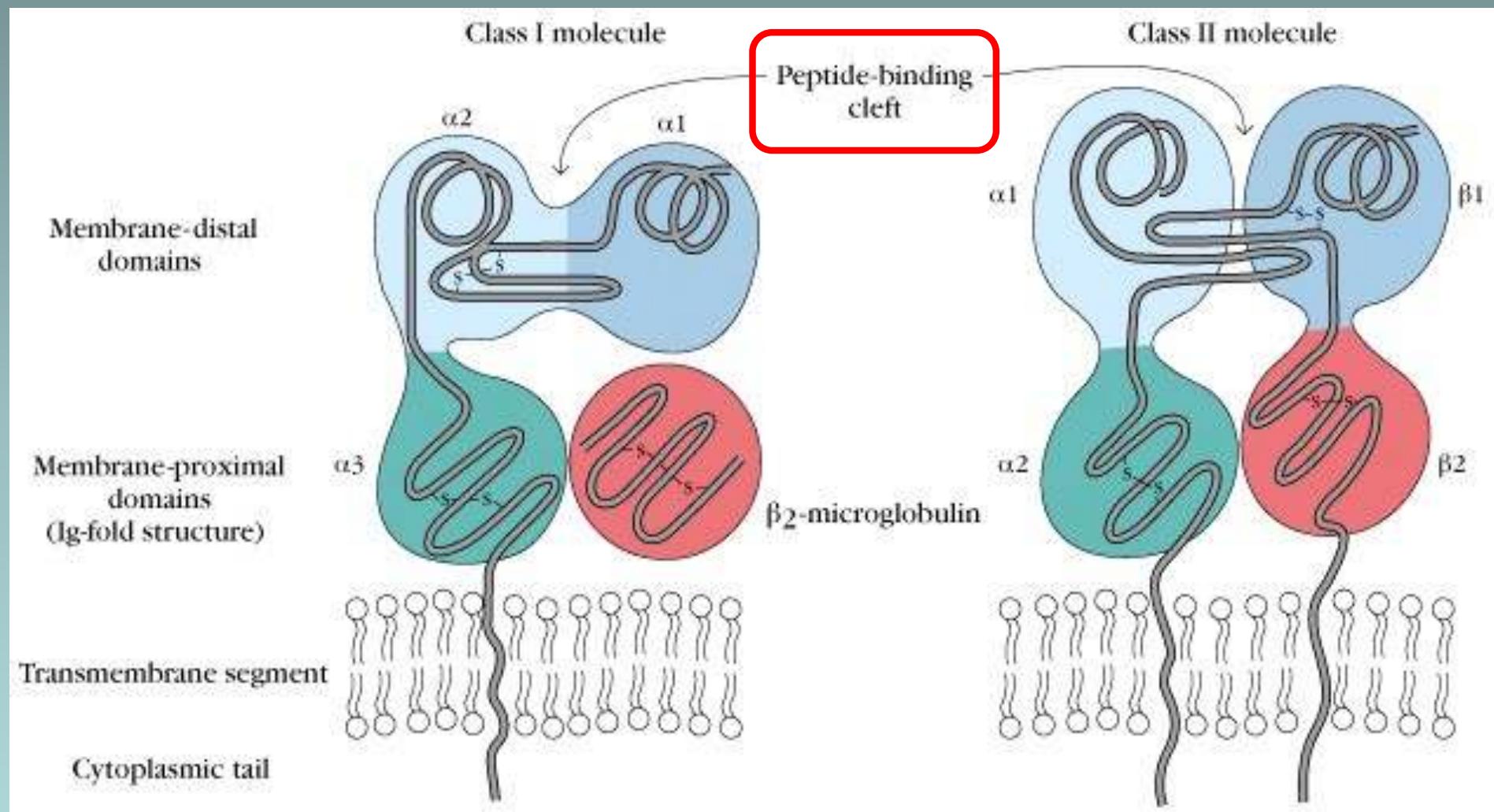
MHC allele B

Changes in the pockets, walls and floor of the peptide binding cleft alter peptide MHC interactions and determine which peptides bind.

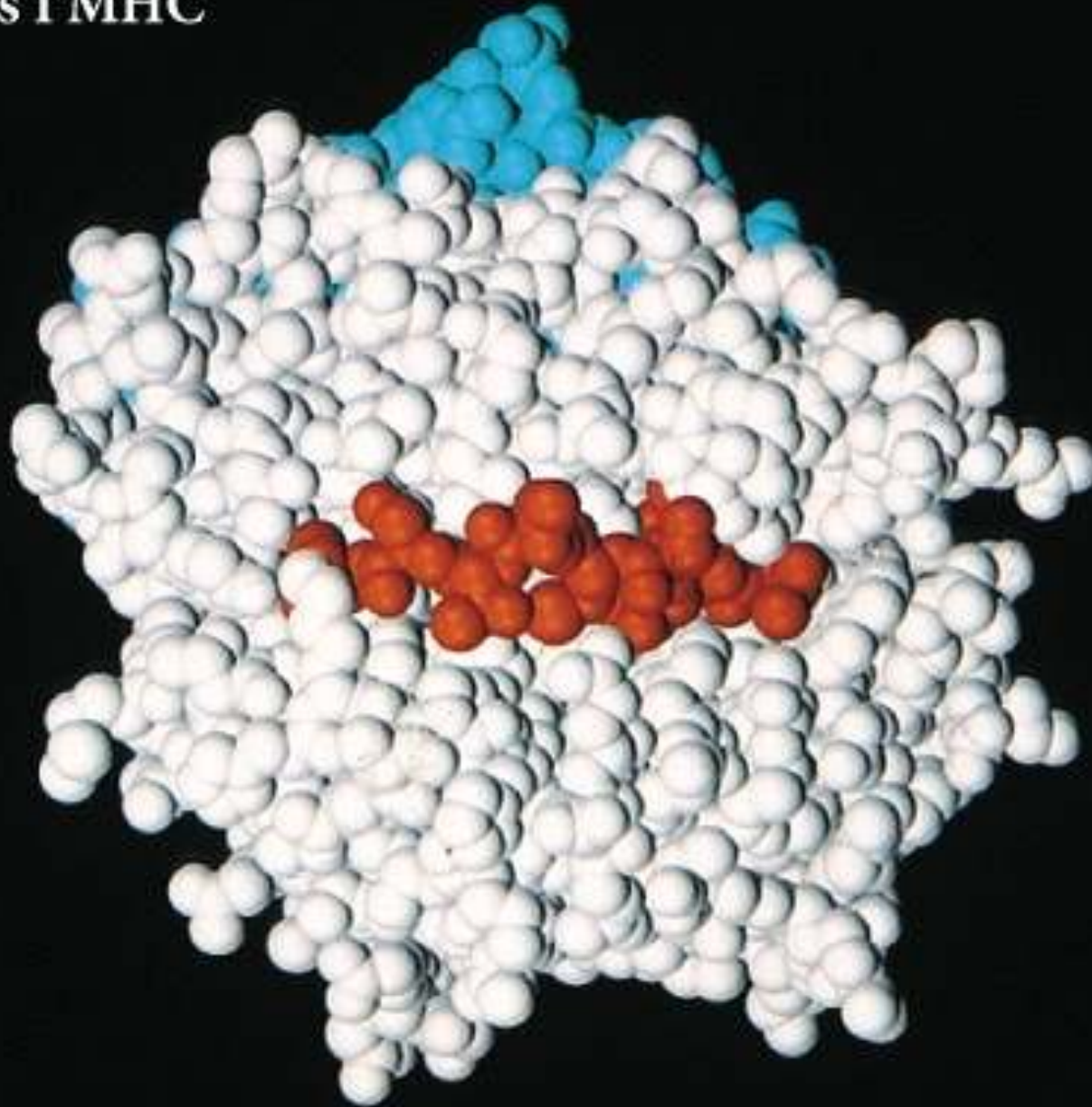
S Y I P S A K I MHC allele A

R G Y V Y Q Q L MHC allele B

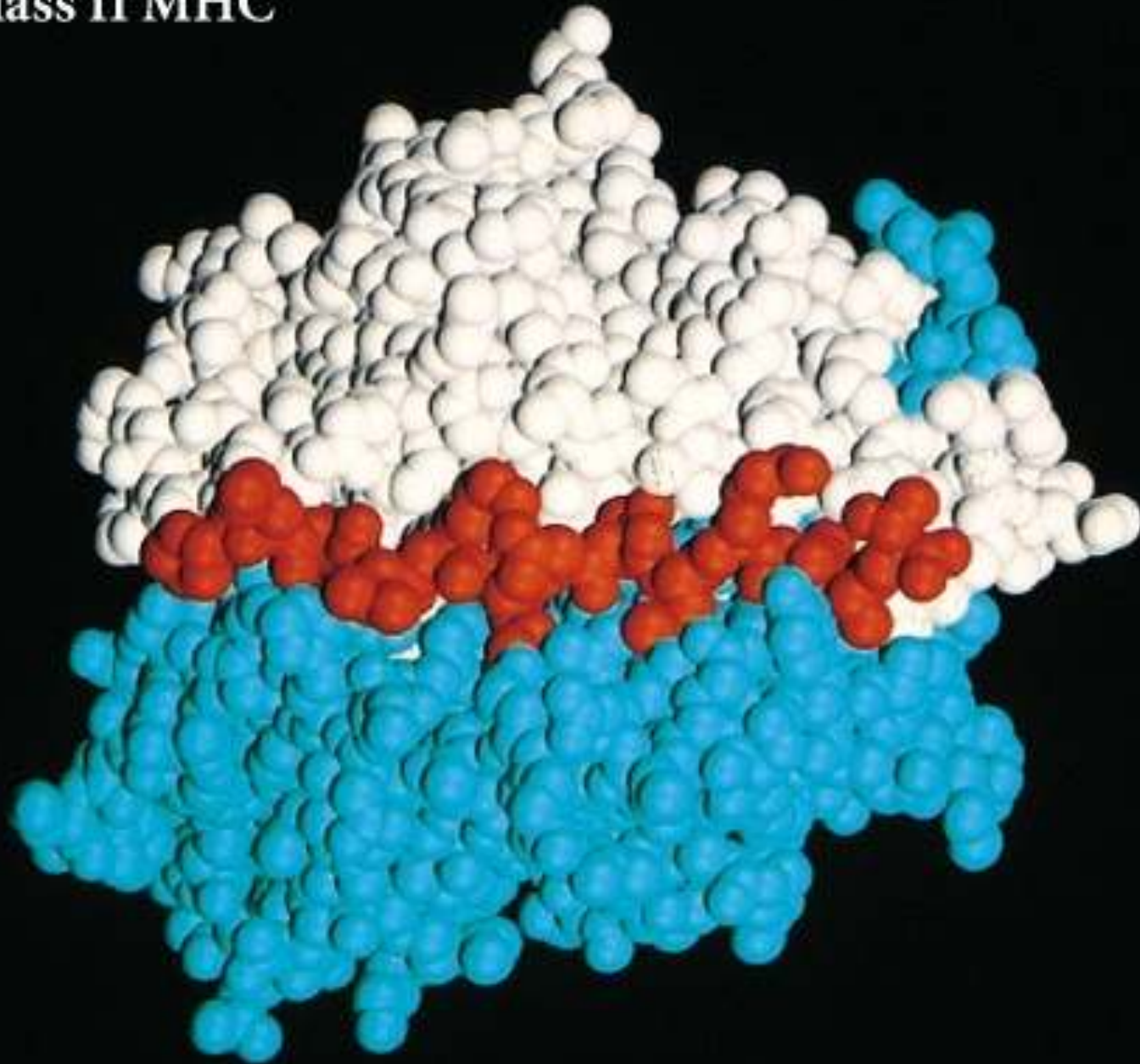
Products of different MHC alleles bind a different repertoire of peptides



(a) Class I MHC

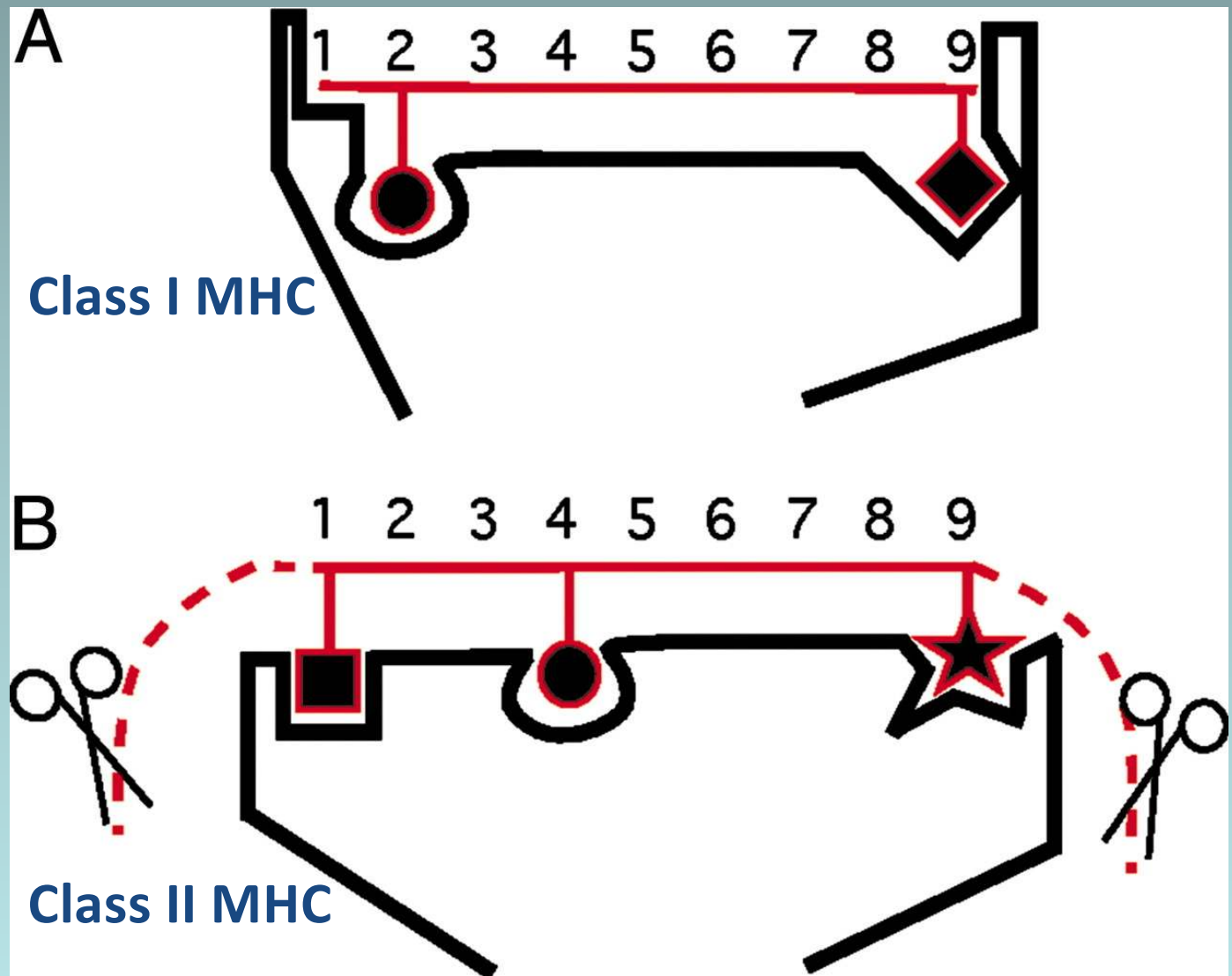


(b) Class II MHC

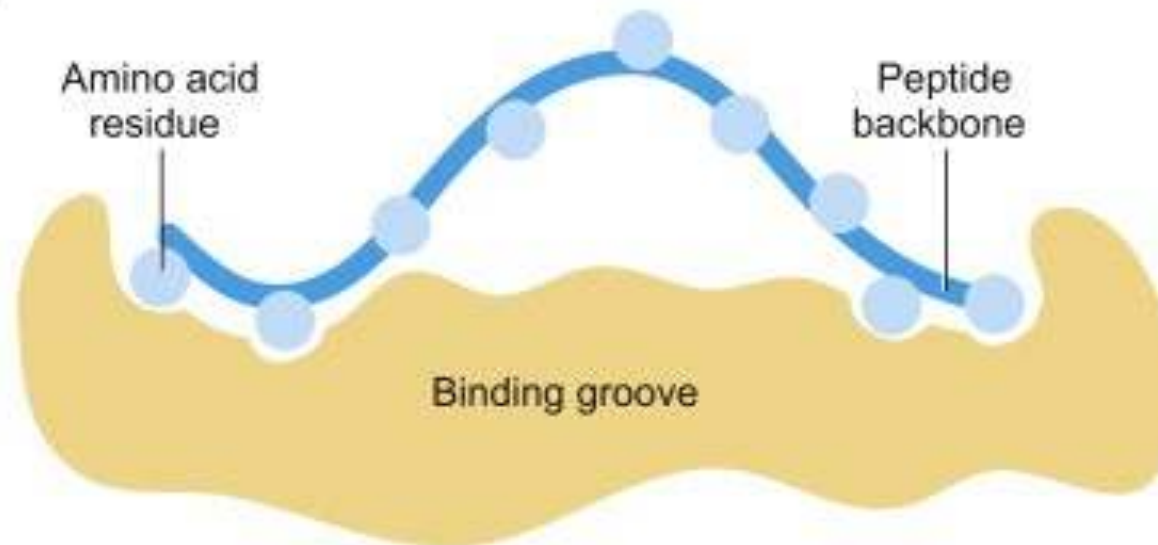


Anchor residues (προσδεδικά κατάλοιπα ή κατάλοιπα αγκυροβόλησης)

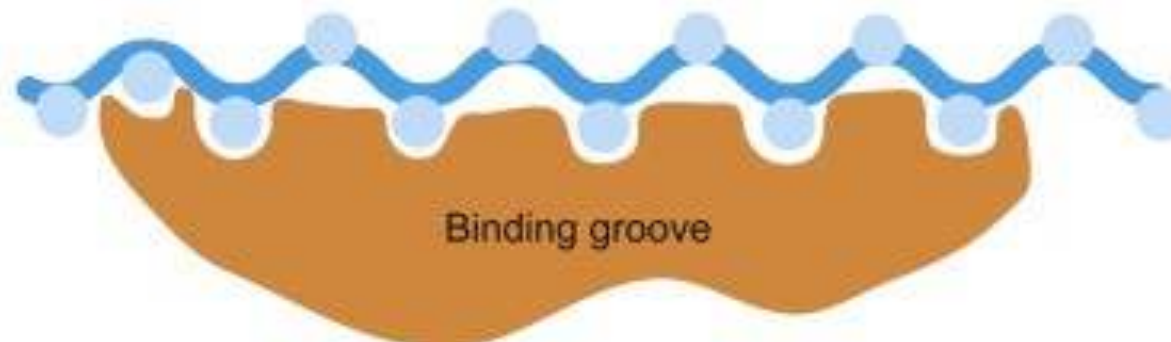
are residues in the peptide that bind to specific pockets on the MHC I or MHC II resulting in some specificity of interactions with MHC.



(A) Peptide in MHC Class I Binding Groove



(B) Peptide in MHC Class II Binding Groove



Eluted peptides from MHC molecules have different sequences but contain motifs

Peptides bound to a particular type of MHC class I molecule have conserved patterns of amino acids

A common sequence in a peptide antigen that binds to an MHC molecule is called a **MOTIF**

Amino acids common to many peptides tether the peptide to structural features of the MHC molecule
ANCHOR RESIDUES

Tethering amino acids need not be identical but must be related

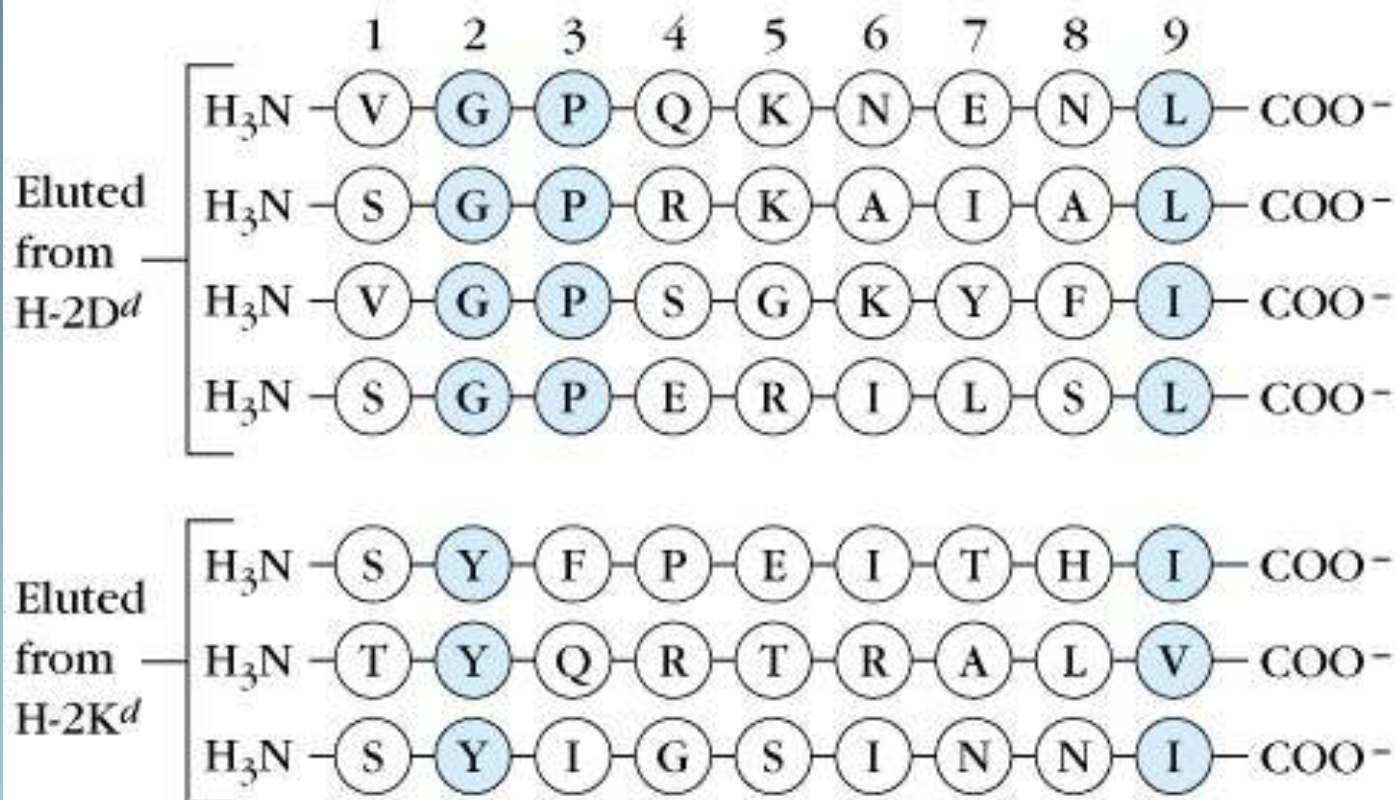
Y & F are aromatic

V, L & I are hydrophobic

Side chains of anchor residues bind into **POCKETS** in the MHC molecule

Different types of MHC molecule bind peptides with different patterns of conserved amino acids





A = alanine

E = glutamic acid

F = phenylalanine

G = glycine

H = histidine

I = isoleucine

K = lysine

L = leucine

N = asparagine

P = proline

Q = glutamine

R = arginine

S = serine

T = threonine

V = valine

Y = tyrosine

MHC-I peptide binding

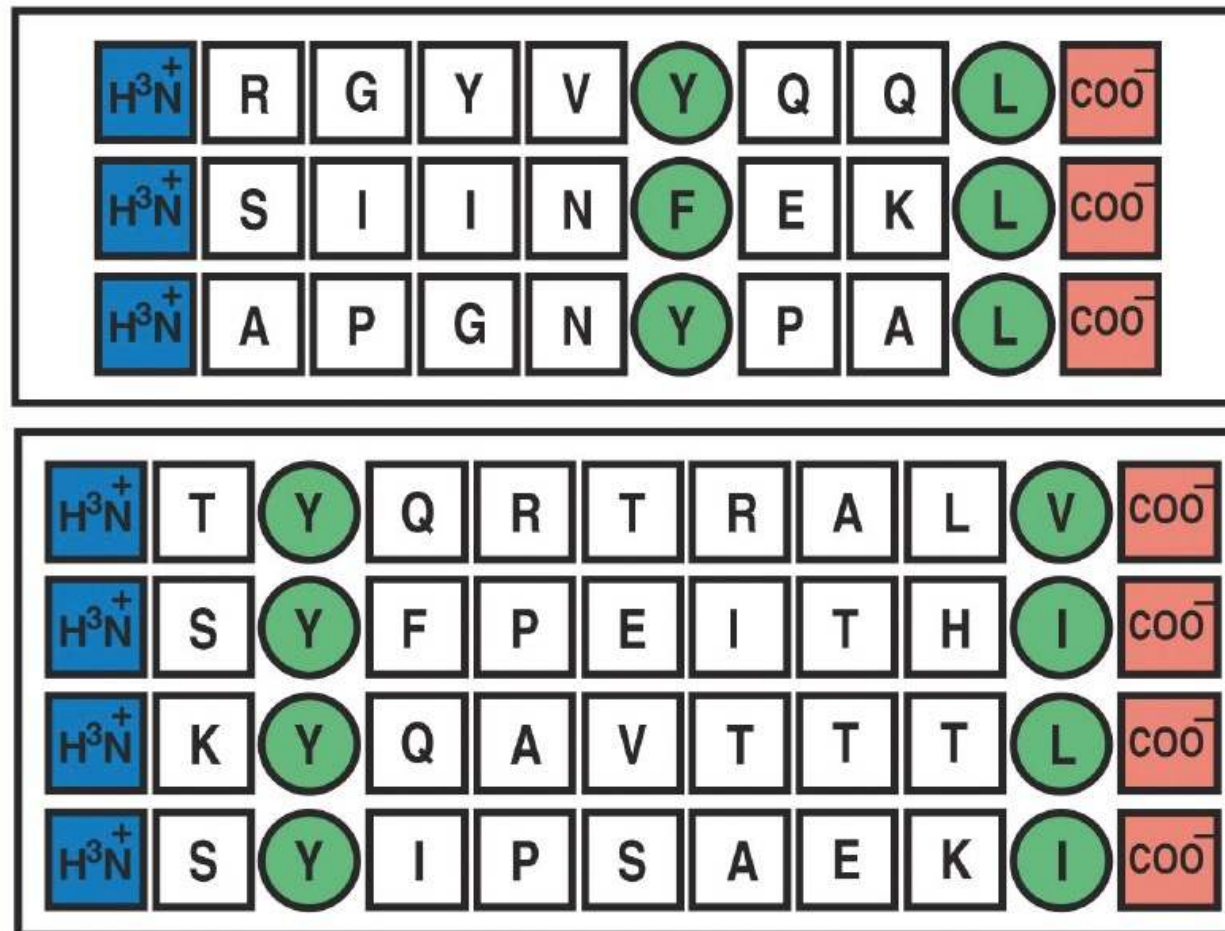
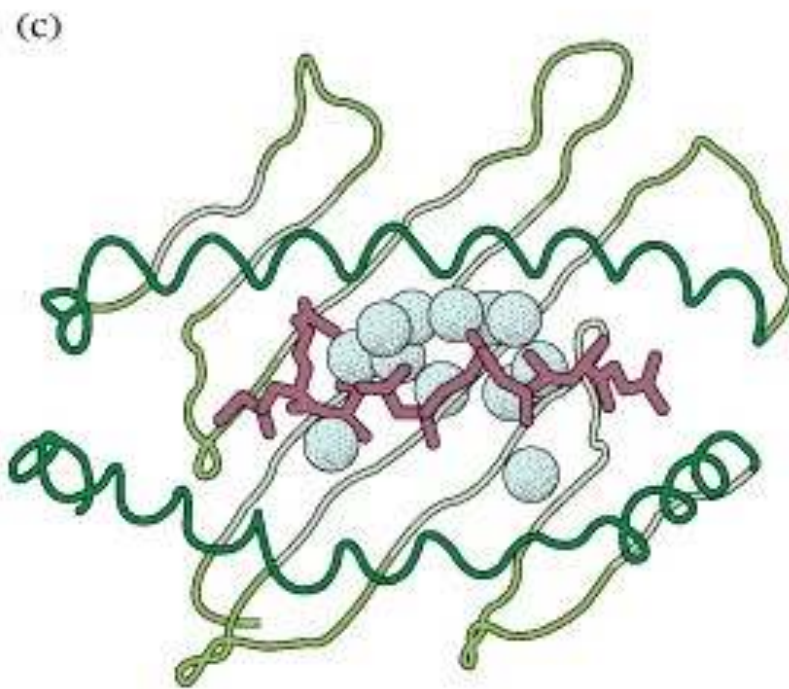
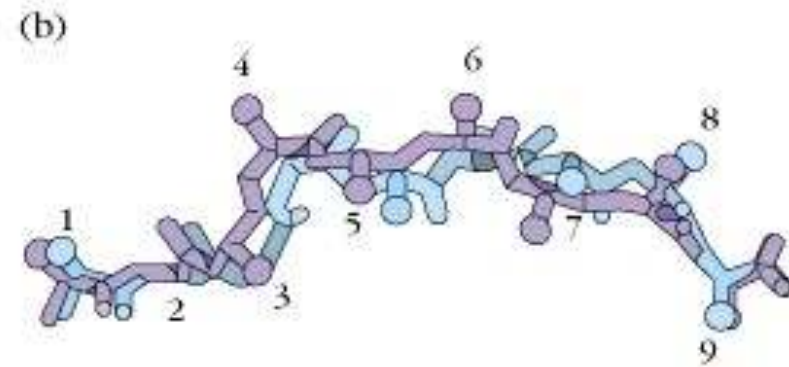
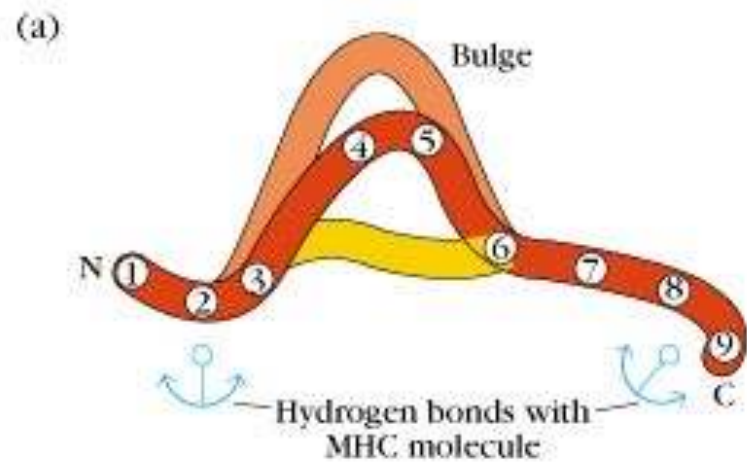
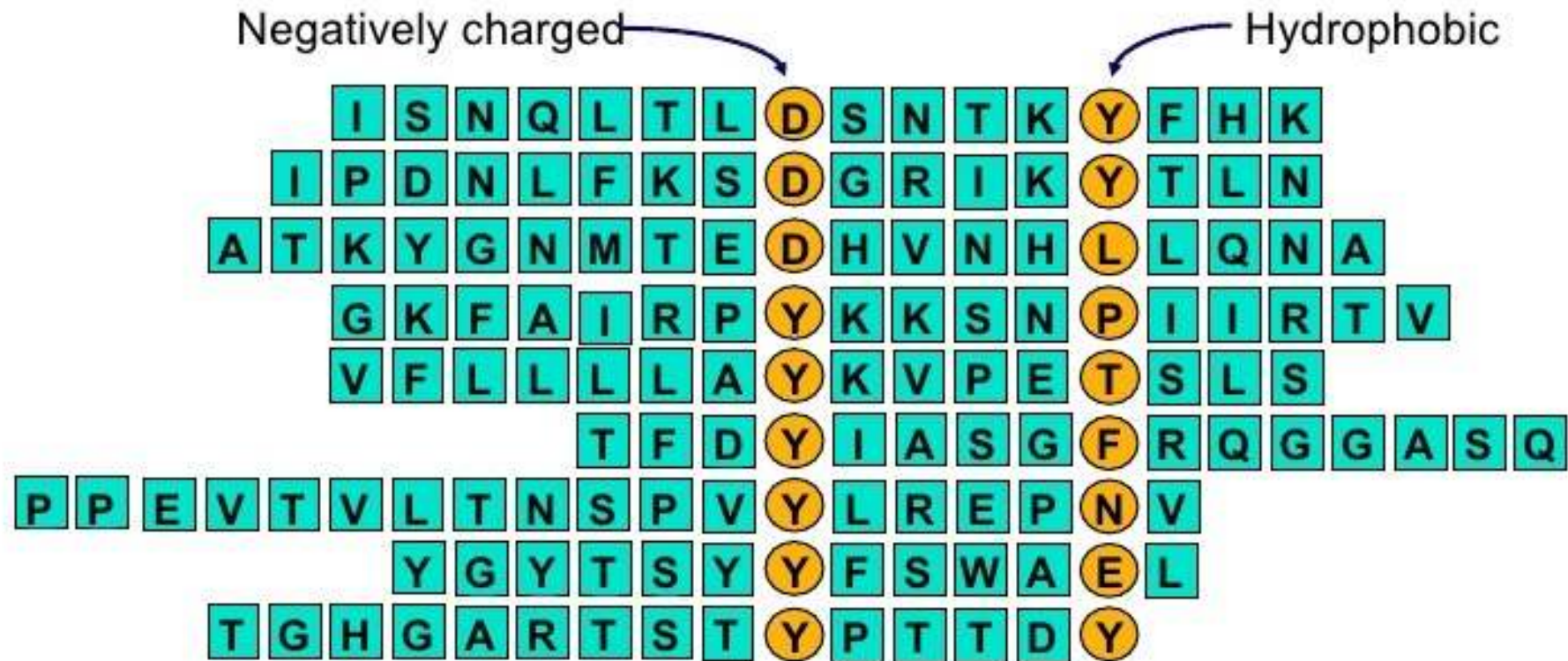


Figure 3-24 Immunobiology, 6/e. (© Garland Science 2005)

Closed peptide-binding cleft:
Conserved AA residues bind the terminal $-\text{NH}_2$ and $-\text{COOH}$ groups



Peptide antigen binding to MHC class II molecules



- Anchor residues *are not* localised at the N and C termini
- Ends of the peptide are in extended conformation and may be trimmed
- Motifs are less clear than in class I-binding peptides
- Pockets are more permissive

MHC-II peptide binding

ΜΗC τάξης II > 13 αα

 $I-A^k$

Diagram illustrating a protein structure with a sequence of amino acids. The sequence is: Q A T N R N T D G S T D Y G I L Q I N S R W W C N D G R. The sequence is shown in a 3D representation with a blue box labeled 'I-A^k' indicating a specific region.

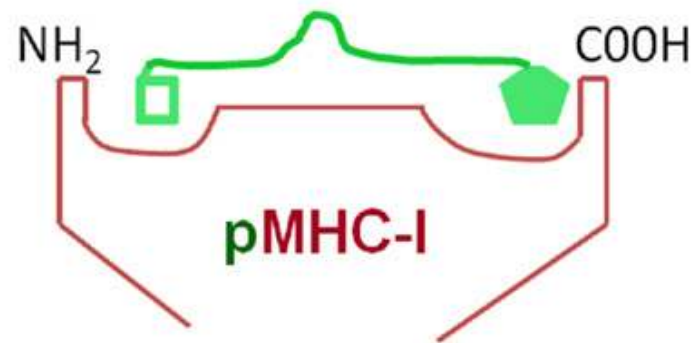
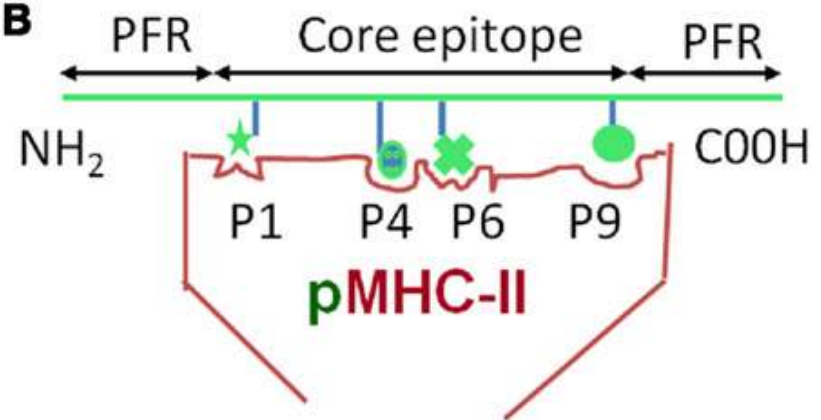
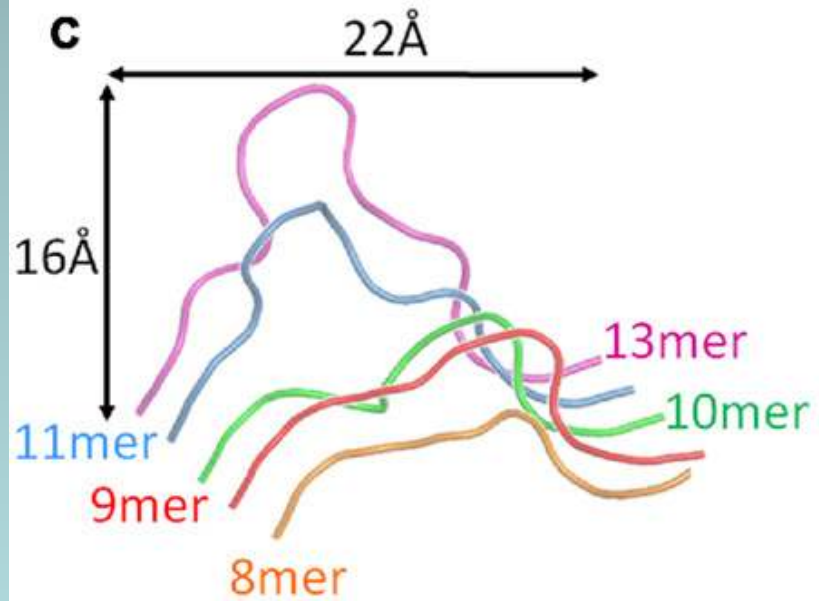
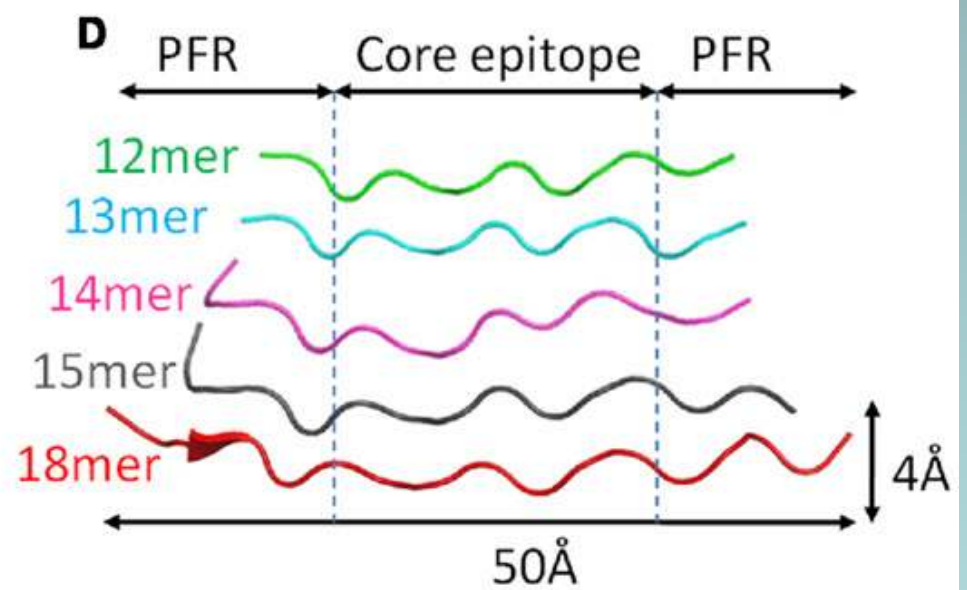
HLA-DR3

			I	S	N	Q	L	T	L	D	S	N	T	K	Y	F	H	K	L	N
				V	D	T	F	L	E	D	V	K	N	L	Y	H	S	E	A	
			K	P	R	A	I	V	V	D	P	V	H	G	F	M	Y			
				K	Q	T	I	S	P	D	Y	R	N	M	I					
				Y	P	D	F	I	M	D	P	K	E	K	D	K	V			
G	P	P		K	L	D	I	R	K	E	E	K	Q	I	M	I	D	I	F	H
			G	F	K	A	I	R	P	D	K	K	S	N	P	I	I	R	T	V
I	P	D		N	L	F	L	K	S	D	G	R	I	K	Y	T	L	N	K	N
				V	T	T	L	N	S	D	L	K	Y	N	A	L	D	L	T	N

Examples

	MHC molecule	Amino acid sequence of peptide-binding motifs and bound peptides	Source of bound peptide
		Position in peptide sequence N—1 2 3 4 5 6 7 8 9—C	
Class I	HLA-A*0201	Peptide-binding motif □ L/M □ □ □ V □ □ V/L Bound peptide I L K E P V H G V	HIV reverse transcriptase
	HLA-B*2705	Peptide-binding motif □ R □ □ □ □ □ □ R/K Bound peptide S R Y W A I R T R	Influenza A nucleoprotein
Class II	HLA-DRB1*0401	Self peptide G V Y F Y L Q W G R S T L V S V S	Igκ light chain
	HLA-DQA1*0501 HLA-DQB1*0301	Self peptide I P E L N K V A R A A A	Transferrin receptor

Figure 5.30 The Immune System, 3ed. (© Garland Science 2009)

A**B****C****D**

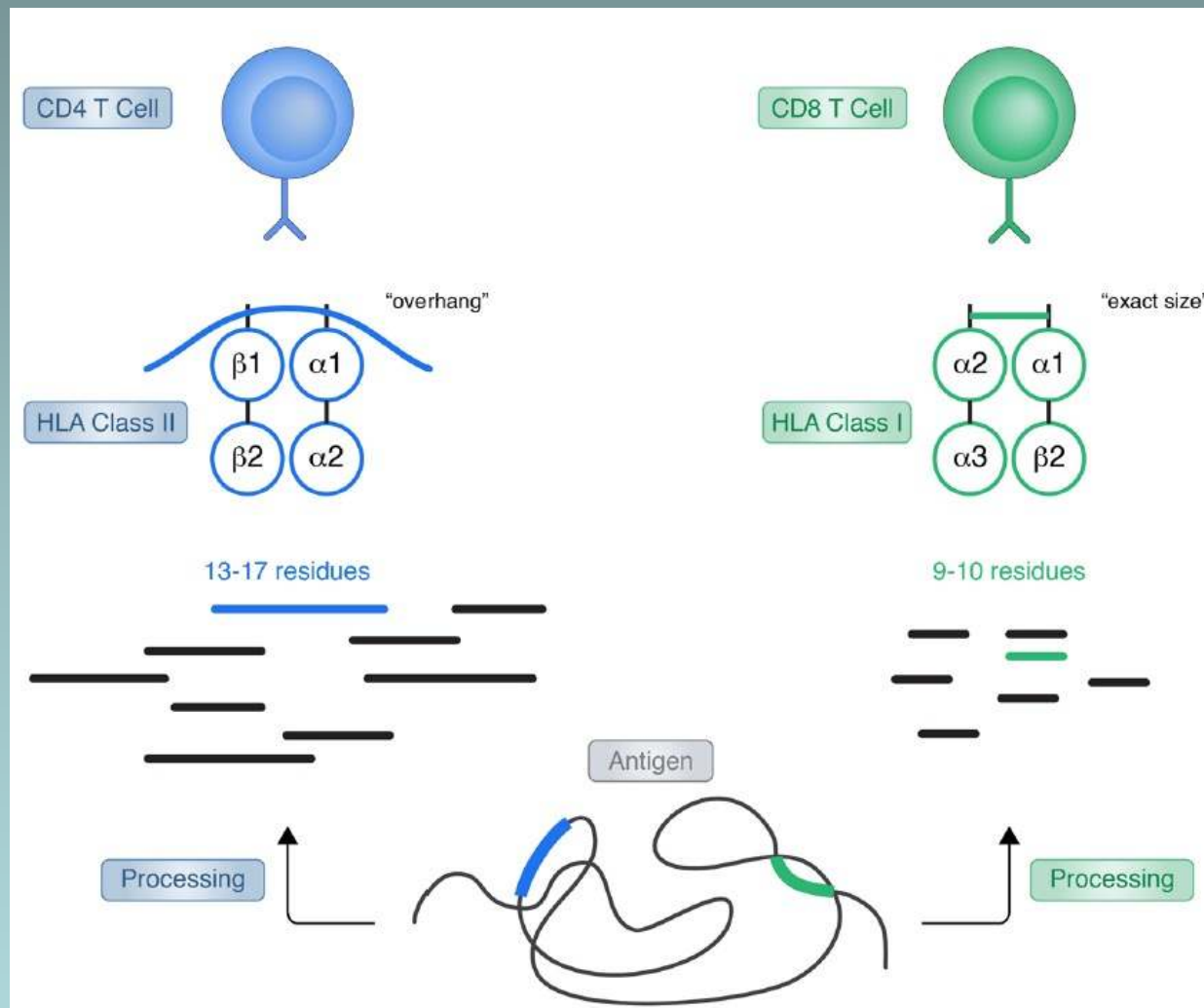


Fig. 1. Epitope definition. Presentation of peptide ligands by HLA class I and class II molecules is depicted. A) HLA class I and II ligands are generated by proteolytic processing of endogenously expressed proteins (antigen) (class I) or from proteins degraded in endocytic compartments (class II). Longer class II peptides typically overhang the open ends of the HLA class II binding groove, while shorter class I ligands are size constrained due to the closed end of the class I binding groove. With binding, and subsequent presentation on the cell surface, HLA-ligand complexes are available for scrutiny by CD4+ (class II) or CD8+ (class I) T cells. B) Ligands processed from longer protein antigens bind HLA using, in general, a nine-mer core region, where the main energy of binding is provided by interaction of some, but not all, peptide residues with residues forming the main pockets of the HLA binding groove. Definition of partial epitopes reflects that not all residues within an epitope region are necessarily important for HLA binding or T cell recognition, while mutation of other residues may ameliorate or abrogate specific immunity.

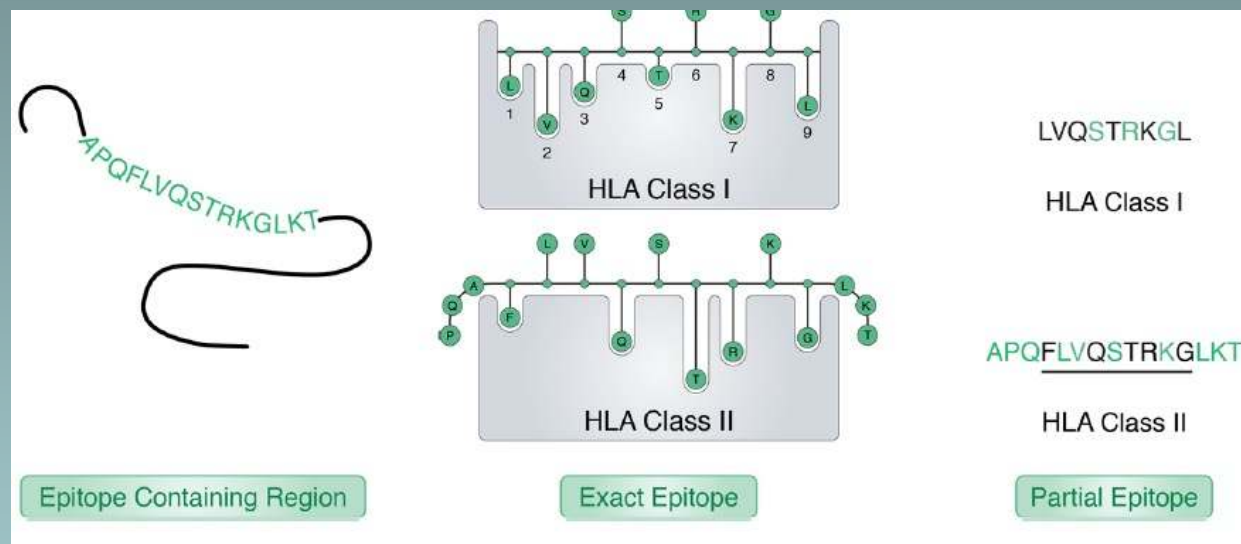


Fig. 1. Epitope definition. Presentation of peptide ligands by HLA class I and class II molecules is depicted. A) HLA class I and II ligands are generated by proteolytic processing of endogenously expressed proteins (antigen) (class I) or from proteins degraded in endocytic compartments (class II). Longer class II peptides typically overhang the open ends of the HLA class II binding groove, while shorter class I ligands are size constrained due to the closed end of the class I binding groove. With binding, and subsequent presentation on the cell surface, HLA-ligand complexes are available for scrutiny by CD4+ (class II) or CD8+ (class I) T cells. B) Ligands processed from longer protein antigens bind HLA using, in general, a nine-mer core region, where the main energy of binding is provided by interaction of some, but not all, peptide residues with residues forming the main pockets of the HLA binding groove. Definition of partial epitopes reflects that not all residues within an epitope region are necessarily important for HLA binding or T cell recognition, while mutation of other residues may ameliorate or abrogate specific immunity.

Important Features of Some Human MHC Gene Products

	Class I	Class II
Genetic loci (partial list)	HLA-A, -B, and -C	HLA-DP, -DQ, and -DR
Polypeptide composition	MW 45,000 + β_2M (MW 12,000)	α chain, β chain, and Ii chain
Cell distribution	All nucleated somatic cells	Antigen- presenting cells, activated T cells
Present peptide antigens to	CD8+ T cells	CD4+ T cells
Size of peptide bound	8 – 11 residues	10 – 30 or more residues

TABLE 7-2 PEPTIDE BINDING BY CLASS I AND CLASS II MHC MOLECULES

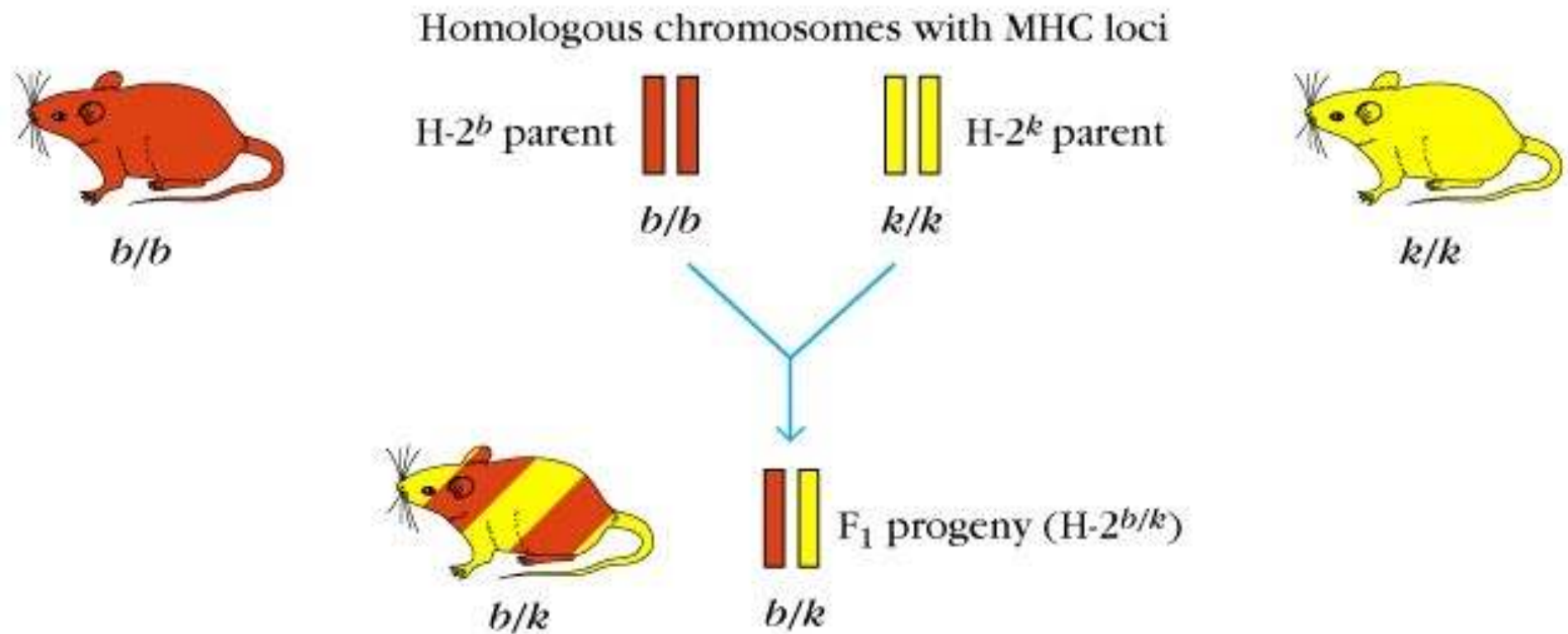
	Class I molecules	Class II molecules
Peptide-binding domain	$\alpha 1/\alpha 2$	$\alpha 1/\beta 1$
Nature of peptide-binding cleft	Closed at both ends	Open at both ends
General size of bound peptides	8–10 amino acids	13–18 amino acids
Peptide motifs involved in binding to MHC molecule	Anchor residues at both ends of peptide; generally hydrophobic carboxyl-terminal anchor	Anchor residues distributed along the length of the peptide
Nature of bound peptide	Extended structure in which both ends interact with MHC cleft but middle arches up away from MHC molecule	Extended structure that is held at a constant elevation above the floor of MHC cleft

PEPTIDE BINDING BY CLASS-I AND CLASS-II MHC MOLECULES		
	CLASS-I MOLECULE	CLASS-II MOLECULE
Peptide binding domain	$\alpha 1/\alpha 2$	$\alpha 1/\beta 2$
Nature of peptide binding cleft	Closed at both ends	Open at both ends
General size of bound peptides	8-10 amino acids	13-18 amino acids
Peptide motifs involved in binding of MHC molecule	Anchor residues at both ends of peptide	Anchor residues distributed along the length of the peptide
Nature of bound peptide	Extended structure in which both ends interact with MHC cleft but middle arches up away from MHC molecule	Extended structure that is held at a constant elevation above the floor of MHC cleft

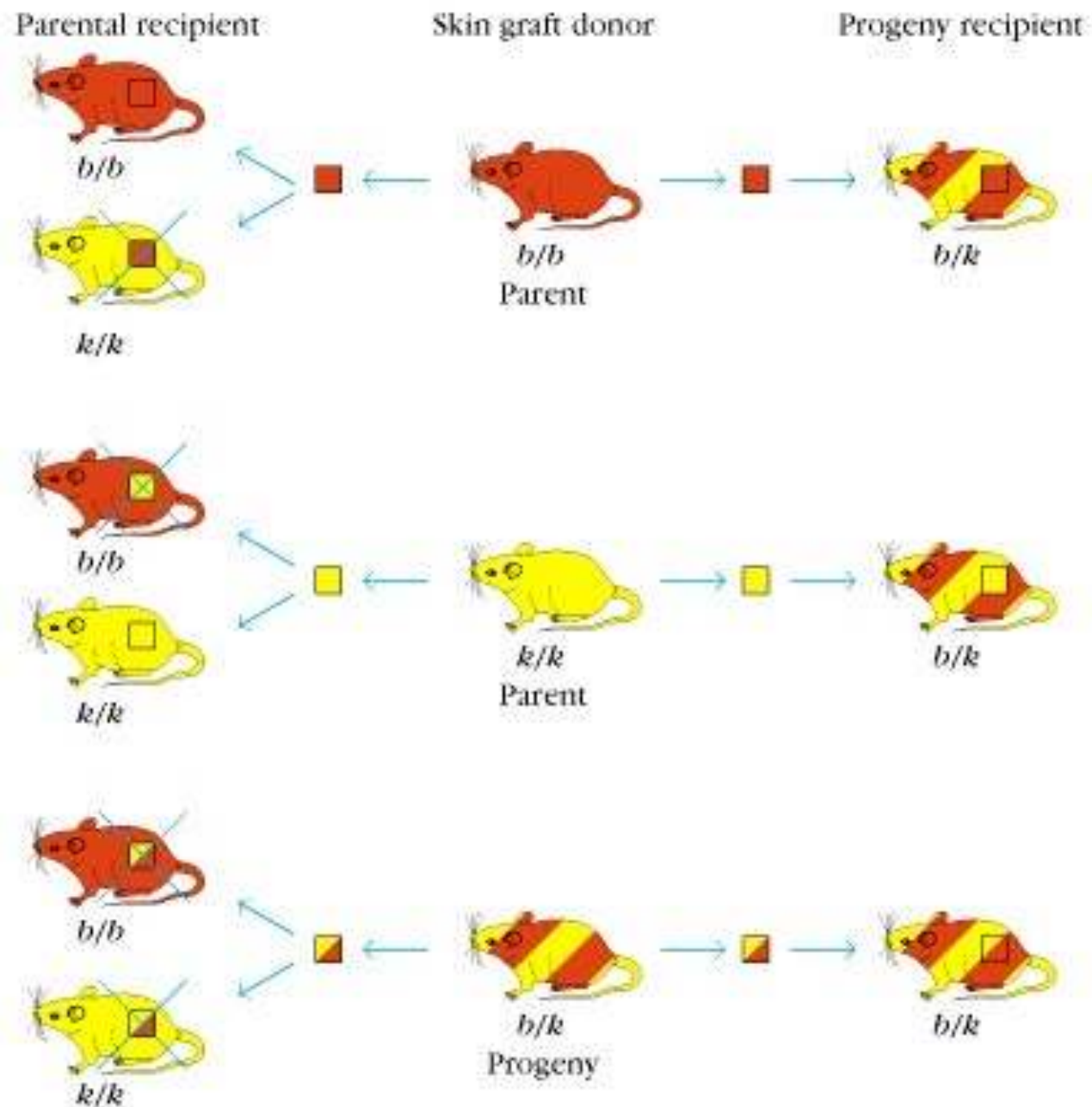
TABLE 7-1 H-2 HAPLOTYPES OF SOME MOUSE STRAINS

Prototype strain	Other strains with the same haplotype	Haplotype	H-2 alleles				
			<i>K</i>	<i>I</i> A	<i>I</i> E	<i>S</i>	<i>D</i>
CBA	AKR, C3H, B10.BR, C57BR	<i>k</i>	<i>k</i>	<i>k</i>	<i>k</i>	<i>k</i>	<i>k</i>
DBA/2	BALB/c, NZB, SEA, YBR	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
C57BL/10 (B10)	C57BL/6, C57L, C3H.SW, LP, 129	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
A	A/He, A/Sn, A/Wy, B10.A	<i>a</i>	<i>k</i>	<i>k</i>	<i>k</i>	<i>d</i>	<i>d</i>
A.SW	B10.S, SJL	<i>s</i>	<i>s</i>	<i>s</i>	<i>s</i>	<i>s</i>	<i>s</i>
A.TL		<i>tl</i>	<i>s</i>	<i>k</i>	<i>k</i>	<i>k</i>	<i>d</i>
DBA/1	STOLI, B10.Q, BDP	<i>q</i>	<i>q</i>	<i>q</i>	<i>q</i>	<i>q</i>	<i>q</i>

(a) Mating of inbred mouse strains with different MHC haplotypes



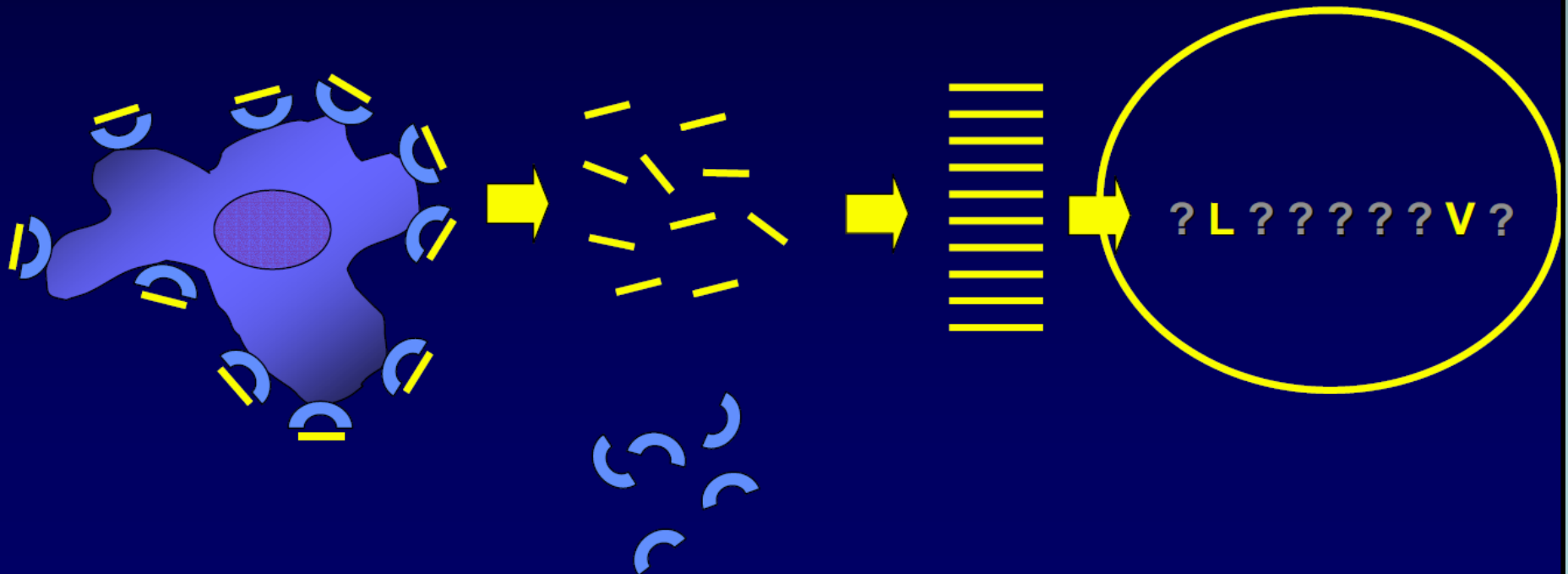
(b) Skin transplantation between inbred mouse strains with same or different MHC haplotypes





How to search for T cell epitopes?

search for 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**

MHC binding predictions

- Experimental characterization of peptide–MHC interactions is highly **cost-intensive**
- Prediction methods facilitate selection of potential epitopes from a pool of peptides

Peptide binding data HLA-A*01:01

Peptide	IC ₅₀ (nM)
ASFCGSPY	51.4
LTDFGLSK	739.3
FTSFFYRY	1285.0
KSVFNSLY	1466.0
RDWAHNSL	1804.6
FSSCPVAY	1939.4
RNWAHSSL	2201.7
LSCAASGF	2830.1
LASIDLKY	3464.0

+

Machine learning algorithms



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

Annual Review of Biomedical Data Science
Immunoinformatics: Predicting
Peptide–MHC Binding

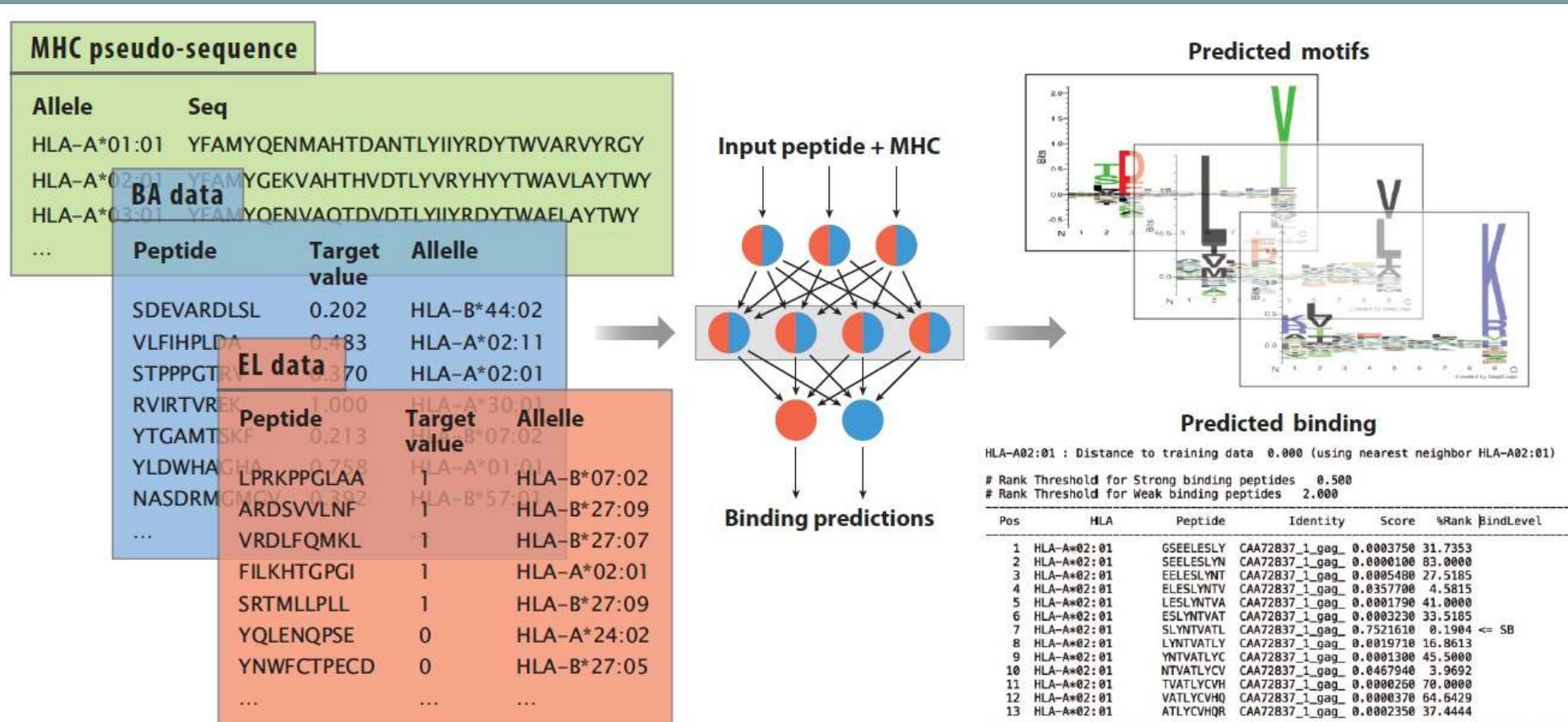


Figure 3

The NNAlign machine learning framework. Different kinds of peptide–MHC binding data are integrated in a machine learning framework leveraging information between multiple MHC molecules and peptide length, resulting in a pan-length, pan-MHC prediction method that captures individual binding motifs and allows for accurate epitope prediction. Abbreviations: BA, binding affinity; EL, eluted ligand.

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

Sequence based

Structure based

T cell epitope prediction algorithms :

- Margalit, Spouge et al - 1987
- Rothbard & Taylor – 1988
- Stille et al –1987
- Tepitope -1999
- Rammensee et al. 1999

Table 1 Examples of databases offering immunological data

Database	Content	Reference
SYFPEITHI	MHC ligands, T-cell epitopes	[14]
IEDB	Epitopes, epitope–MHC/BCR complexes	[15]
IMGT	Antibodies, T-cell receptors	[18]
IMGT/HLA	HLA alleles	[18]
MHCBN 4.0	MHC peptides, TAP-interacting peptides	[16]
AntiJen	MHC ligands, TCR–MHC complexes, T-cell epitopes, TAP, B-cell epitopes, protein–protein interactions	[17]
Dana-Farber Repository	MHC ligands for machine learning	[21]

Abbreviations: BCR B-cell receptor, HLA human leukocyte antigen, IEDB Immune Epitope Database, IMGT International ImMunoGeneTics information system, MHC major histocompatibility complex, MHCBN MHC binding and non-binding, TAP transporter associated with antigen processing, TCR T-cell receptor

Table 2 Methods for analyzing steps in the antigen-processing pathway and for HLA typing

Predictor/tool	Key method	Reference
HLA class I binding		
Allele-specific		
SYFPEITHI	PSSM	[14]
RANKPEP	PSSM	[27]
BIMAS	PSSM	[28]
SVMHC	SVM	[7]
netMHC	ANN	[29]
Pan-specific		
MULTIPRED	HMM/ANN	[39]
netMHCpan	ANN	[40]
PickPocket	PSSM	[41]
TEPITOPEpan	PSSM	[42]
ADT	Threading	[43]
UniTope	SVM	[44]
KISS	SVM	[45]

HLA class II binding

Allele-specific

SYFPEITHI	PSSM	[14]
netMHCII/SM-align	PSSM/ANN	[48, 49]
ProPred	PSSM	[50]
RANKPED	PSSM	[27]
TEPITOPE	PSSM	[51]
SVRMHC	SVM	[8]
MHC2MIL	Multi-instance learning	[52]
MHC2pred	SVM	–

Pan-specific

MULTIPRED	HMM/ANN	[39]
MHCIIMulti	Multi-instance learning	[55]
TEPITOPEpan	PSSM	[42]
netMHCIIpan	ANN	[56, 90]

Consensus methods

CONSENSUS	–	[57]
netMHCcon	–	[56]

Table 1 MHC binding prediction methods available and described in this review

Method	URL	Pan-specific	Includes EL data
Class I			
NetMHC	http://www.cbs.dtu.dk/services/NetMHC	No	No
NetMHCpan	http://www.cbs.dtu.dk/services/NetMHCpan	Yes	Yes
MixMHCpred	https://github.com/GfellerLab/MixMHCpred	No	Yes
MHCflurry	https://github.com/openvax/mhcflurry	No	Yes
BIMAS	Decommissioned on March 8, 2019	No	No
SYFPEITHI	http://www.syfpeithi.de	No	Yes
SMM	http://tools.iedb.org/mhci	No	No
Class II			
NetMHCIIpan	http://www.cbs.dtu.dk/services/NetMHCIIpan	Yes	No

Abbreviation: EL, eluted ligand.

Table 1

Comprehensive list of T cell epitope prediction servers.

Server name	Link	Predictive server for		Predictive method
		MHC I	MHC II	
EpiJen	http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm	24		Multi-step algorithm
SYFPEITHI	http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm	42	7	Published motifs
ANNPRED	http://www.imtech.res.in/raghava/nhlapred/neural.html	30		ANN-regression
BIMAS	http://www-bimas.cit.nih.gov/molbio/hla_bind/	41		Published coefficient tables
ProPred I	http://www.imtech.res.in/raghava/propred1/	47		Quantitative matrix
ProPred	http://www.imtech.res.in/raghava/propred/		51	Quantitative matrix
MHCPred	http://www.ddg-pharmfac.net/mhcpred/MHCPred/	14	11	Additive method
MHC2Pred	http://www.imtech.res.in/raghava/mhc2pred/		42	SVM-based method
NetMHC	http://www.cbs.dtu.dk/services/NetMHC/	57		ANN based method
PREDEP	http://margalit.huji.ac.il/Teppred/mhc-bind/index.html	13		Published coefficient tables
RANKPEP	http://bio.dfci.harvard.edu/RANKPEP/	118	62	PSSM
SVMHC	http://abi.inf.uni-tuebingen.de/Services/SVMHC	33	51	SVM-based method
IEDB binding	http://tools.immuneepitope.org/analyze/html/mhc_processing.html	77		ANN and SMM method
EpiVax	http://www.epivax.com/	6	8	Epimatrix algorithm
MMBPred	http://www.imtech.res.in/raghava/mmbpred/	46		Quantitative matrix
NetCTL	http://www.cbs.dtu.dk/services/NetCTL	12		ANN-regression
nHLAPred	http://www.imtech.res.in/raghava/nhlapred/	67		Artificial Neural Networks
KISS	http://cbio.ensmp.fr/kiss/	64		SVM based method
SVRMHC	http://svrmhc.biolead.org/	36	6	SVM-basedmethod
IMTECH	http://www.imtech.res.in/raghava/mhc		3	Quantitative matrix

Table 3

Comprehensive list of B cell epitope prediction servers.

Server name	Link	Type
Bcepred	http://www.imtech.res.in/raghava/bcepred/	Prediction of continuous B-cell epitopes
BepiPred	http://www.cbs.dtu.dk/services/BepiPred/	Prediction of continuous B-cell epitopes
ABCPred	http://www.imtech.res.in/raghava/abcpred/	Prediction of continuous B-cell epitopes
BEST	http://biomine.ece.ualberta.ca/BEST/	Prediction of continuous B-cell epitopes
EPCEs	http://sysbio.unl.edu/services/EPCEs/	Prediction of discontinuous B-cell epitopes
DiscoTope	http://www.cbs.dtu.dk/services/DiscoTope/	Prediction of discontinuous B-cell epitopes
BEPro (PEPITO)	http://pepito.proteomics.ics.uci.edu/	Prediction of discontinuous B-cell epitopes
SEPPA	http://lifecenter.sgst.cn/seppa/index.php	Prediction of discontinuous B-cell epitopes
EpiSearch	http://curie.utmb.edu/episearch.html	Prediction of discontinuous B-cell epitopes
MimoPro	http://informatics.nenu.edu.cn/MimoPro	Prediction of discontinuous B-cell epitopes
MIMOX	http://immunet.cn/mimox/	Prediction of discontinuous B-cell epitopes
Pep-3D-Search	http://kyc.nenu.edu.cn/Pep3DSearch	Prediction of discontinuous B-cell epitopes
Epitopia	http://epitopia.tau.ac.il/	Prediction of continuous and discontinuous B-cell epitopes
PepSurf	http://pepitope.tau.ac.il	Prediction of continuous and discontinuous B-cell epitopes
ElliPro	http://tools.immuneepitope.org/tools/ElliPro/iedb_input	Prediction of continuous and discontinuous B-cell epitopes

SVM: support vector machine. *ANN*: artificial neural networks. *PSSM*: position-specific scoring matrix.

Table 3. Summary of ML-based prediction tools employed for current benchmarking.

MHC Class I binding predictor

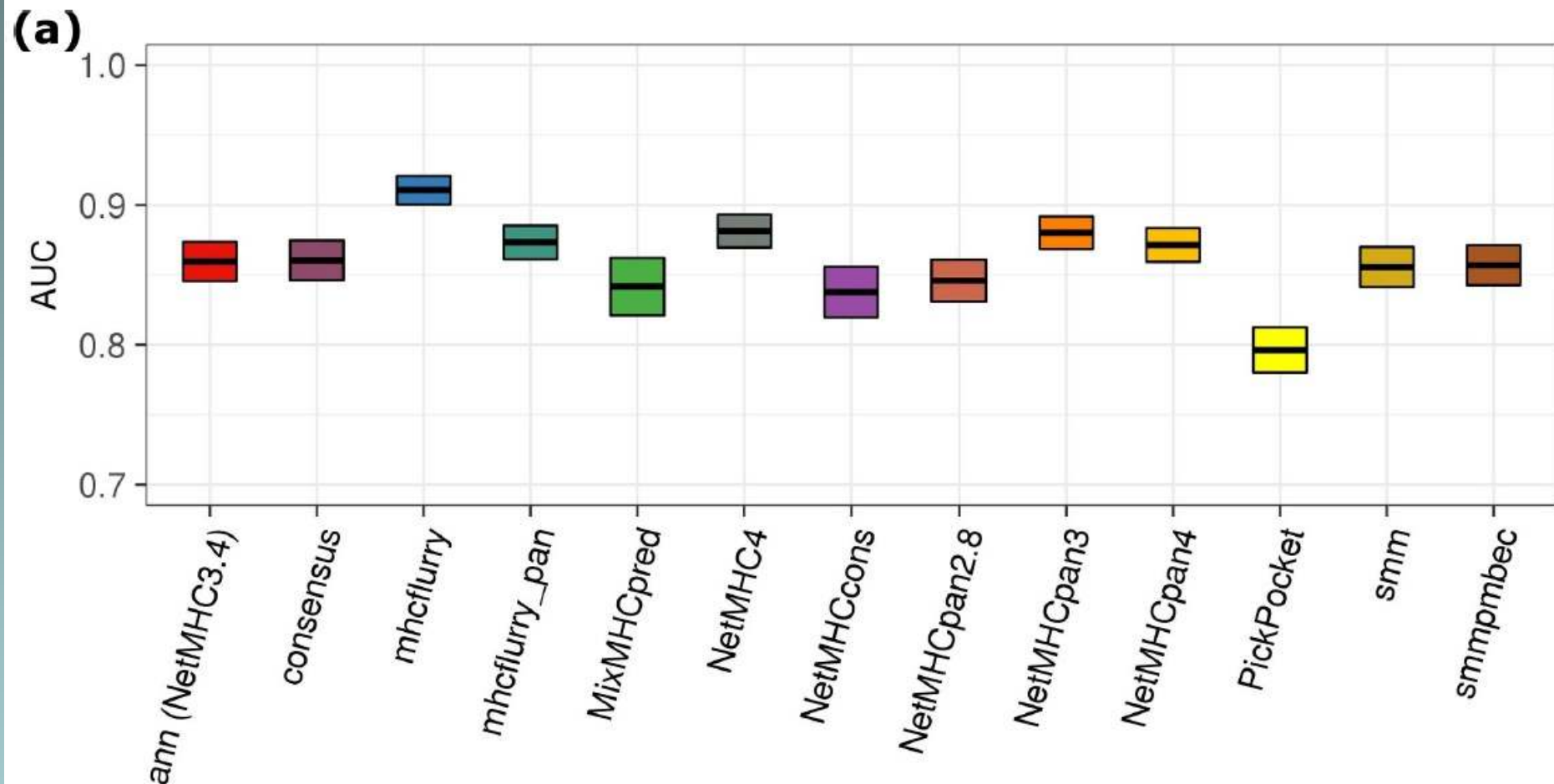
Name	Method Principle	Details	Training Data Cutoff
ann (NetMHC3.4)	ANN	2 to 10 hidden neurons; trained on 9-mer peptides	IEDB—2013
consensus	Combination	Value reported as the median of ann, smm, and PSSM	IEDB—2006
NetMHC4	ANN	5 hidden neurons; trained on all length peptides	IEDB—2014
NetMHCcons	Combination	Value reported as the best performer among NetMHC, NetMHCpan, and PickPocket	IEDB—2012
NetMHCpan2.8	ANN	Trained on 9-mer peptides; nearest neighbor searching for untrained allele	IEDB—2009
NetMHCpan3	ANN	56 or 66 hidden neurons; trained on all-mer length peptides	IEDB—2015
NetMHCpan4	ANN	Addition of MS-derived elution peptides to the training set and the prediction mode for elution probability score	IEDB—2017
PickPocket	LR	Alternative smm with binding specificity vectors of MHC pocket as additional features	IEDB—2009
smm	LR	SM with regularization term	IEDB—2005
smmpmbec	LR	smm + MHC binding pocket sequence	IEDB—2009
mhcfurry	ANN	32 or 64 hidden neurons; trained on 9-mer peptides	IEDB—2014
mhcfurry-pan	ANN	32 or 64 hidden neurons; trained on 43-mer peptides	IEDB—2014
MixMHCpred	Clustering + LR	Nearest neighbor clustering with distance calculated by PSSM	Collective HLA-peptidomics—2017

MHC Class II Binding Predictor

Name	Method Principle	Details	Training Data Cutoff
nn_align (NetMHCII2)	ANN	2 to 60 hidden neurons; trained on 9-mer binding core with additional flanking region features	IEDB—2011
NetMHCIIpan	ANN	10 to 60 hidden neurons; trained on 9-mer binding core with additional flanking region features; nearest neighbor searching for untrained allele	IEDB—2014
consensus	Combination	Value reported as the median of nn-align, smm_align, and PSSM	IEDB -2010
smm_align	LR	SM with regularization term; trained on 9-mer binding core with additional flanking region features	IEDB—2007
comblib	LR	Naïve PSSM	IEDB—2008
tepitope	LR	Naïve PSSM with binding specificity of MHC pocket as additional features	IEDB—2001
mhcflurry	ANN	32 or 64 hidden neurons; trained on 15-mer all-length peptides	IEDB—2014

*PSSM (also know as Position-weighted Matrix): the binding specificity of each residue to a given MHC protein is represented by a score contributing independently to overall binding affinity. The derivation of position-specific score of individual amino acid is by regression method similarly applied in SM, but without the regularization term.

<https://doi.org/10.1371/journal.pcbi.1006457.t003>



AUC: Area Under the Curve, αναφέρεται στην καμπύλη ROC (Receiver Operating Characteristic)

Το AUC δείχνει **πόσο καλά ένα μοντέλο μπορεί να διαχωρίσει** τις δύο κατηγορίες δεδομένων:

Θετικά δείγματα (π.χ. πεπτίδια που δεσμεύονται στο MHC) και **αρνητικά δείγματα** (π.χ. πεπτίδια που δεν δεσμεύονται). Όσο πιο μεγάλο είναι το AUC, τόσο **καλύτερη είναι η ικανότητα διαχωρισμού** του μοντέλου.

AUC = 1.0 σημαίνει **τέλεια πρόβλεψη**, ενώ **AUC = 0.5** σημαίνει ότι η πρόβλεψη είναι τυχαία.

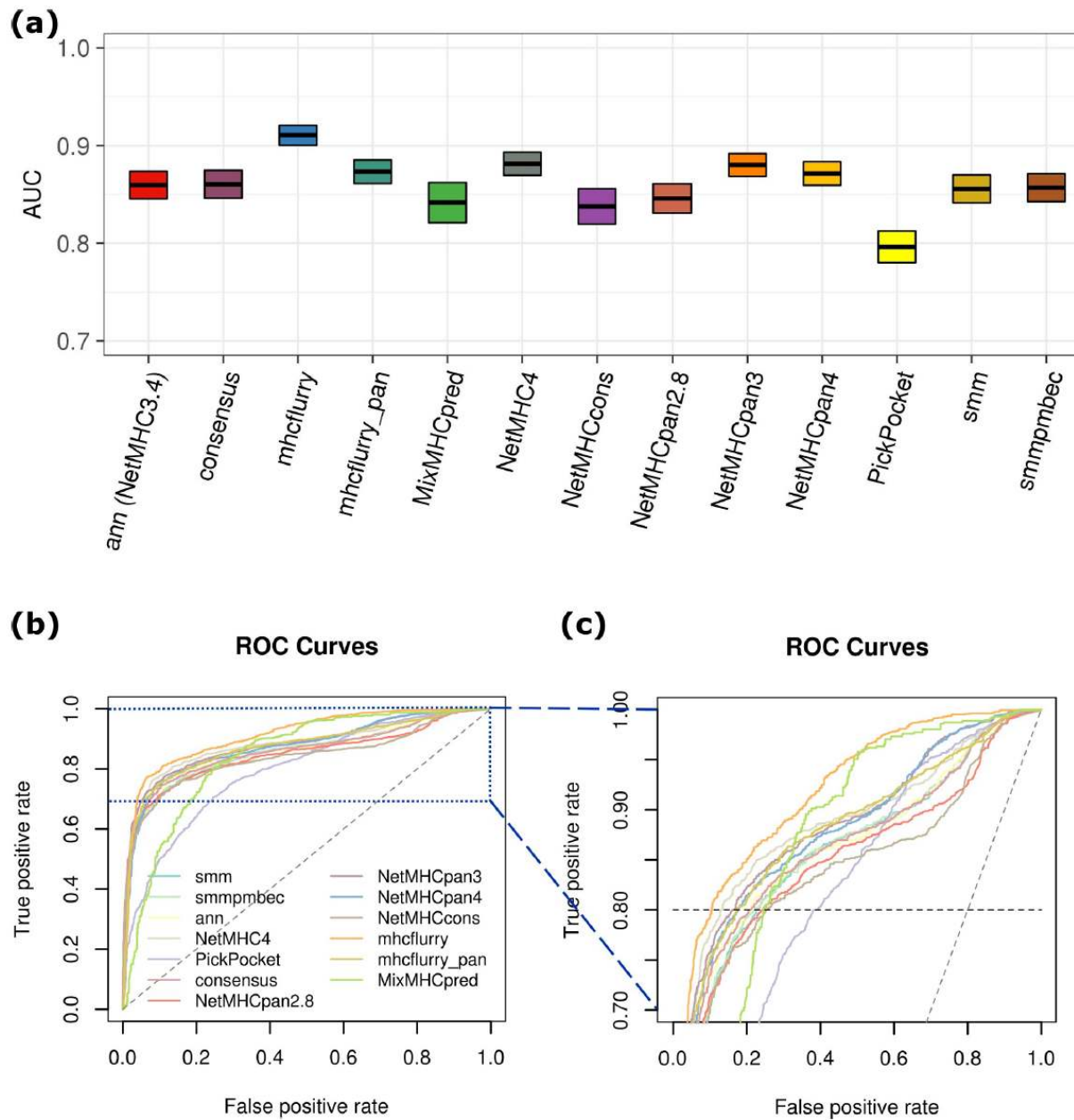


Fig 1. Binary classification (binder vs. non-binder) performance. (a) AUC of MHC-I binding epitope prediction tools. (b) ROC curves. IC₅₀ = 500 nM was used as the cutoff for classifying experimentally measured epitopes. AUC was shown by box plot with upper and lower boundaries covering confidence level of 95%. (c) ROC curves enlarged for TPR between 0.7 and 1.0.

<https://doi.org/10.1371/journal.pcbi.1006457.g001>

Table 2. Methods Developed for T Cell Epitope Prediction (MHC II)^a

Tools	Link	MHC	ROC-AUC	Method
SYFPEITHI	http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm	Both	AUC: 0.90	Published motif
ProPred	http://crdd.osdd.net/raghava/propred/	Class II	AUC: 0.73	Quantitative matrix
IMTECH	http://crdd.osdd.net/raghava/mhc/	Class II	N/A	Matrix Optimization technique
RANKPEP	http://imed.med.ucm.es/Tools/rankpep.html	Both	AUC: 0.78	Position Specific Scoring Matrices (PSSMs)
MHCPred	http://www.ddg-pharmfac.net/mhcpred/MHCPred/	Both	AUC: 0.67	Additive method
SVMHC	https://www.hsls.pitt.edu/obrc/index.php?page=URL1153429637	Both	N/A	Machine Learning (Support Vector Machine)
MHC2Pred	http://crdd.osdd.net/raghava/mhc2pred/	Class II	AUC: 0.36	Machine Learning (Support Vector Machine)
NNAlign 2.0	https://services.healthtech.dtu.dk/services/NNAlign-2.0/	Both	N/A	Artificial Neural Network
NetMHCpan 4.0	https://services.healthtech.dtu.dk/services/NetMHCpan-4.0/	Both	AUC: 0.90	Artificial Neural Network
NetMHCII 2.3	https://services.healthtech.dtu.dk/services/NetMHCII-2.3/	Class II	AUC: 0.85	Artificial Neural Network
TCRMatch	http://tools.iedb.org/tcrmatch/	Both	AUC: 0.68	Machine Learning
Deepitope	Unavailable	Class II	AUC: 0.85	Artificial Neural Network (Convolutional Neural Network)
CAPTAn	https://gitlab.com/xavier-lab-computation/public/captan	Class II	AUC: 0.83	Machine Learning (Deep Learning)
TRAP	https://github.com/ChloeHJ/TRAP	Class II	AUC: 0.71	Machine Learning (Deep Learning)
TCRDock	https://github.com/phbradley/TCRdock	Both	AUC: 0.85	Machine Learning (Deep Learning)
MixMHC2pred 2.0	https://github.com/GfellerLab/MixMHC2pred	Class II	AUC: 0.85	Machine Learning
Prime 2.0	http://ec2-18-188-210-66.us-east-2.compute.amazonaws.com:3000/	Class II	AUC: 0.85	Machine Learning
TROLLOPE	https://pmlabqsar.pythonanywhere.com/TROLLOPE	Both	AUC: 0.76	Machine Learning
GTE	https://github.com/uta-smile/GTE	Both	AUC: 0.75	Artificial Neural Network
HLAIIMaster	https://github.com/TomasYang001/HLAIImaster	Class II	AUC: 0.93	Machine Learning (Deep Learning)

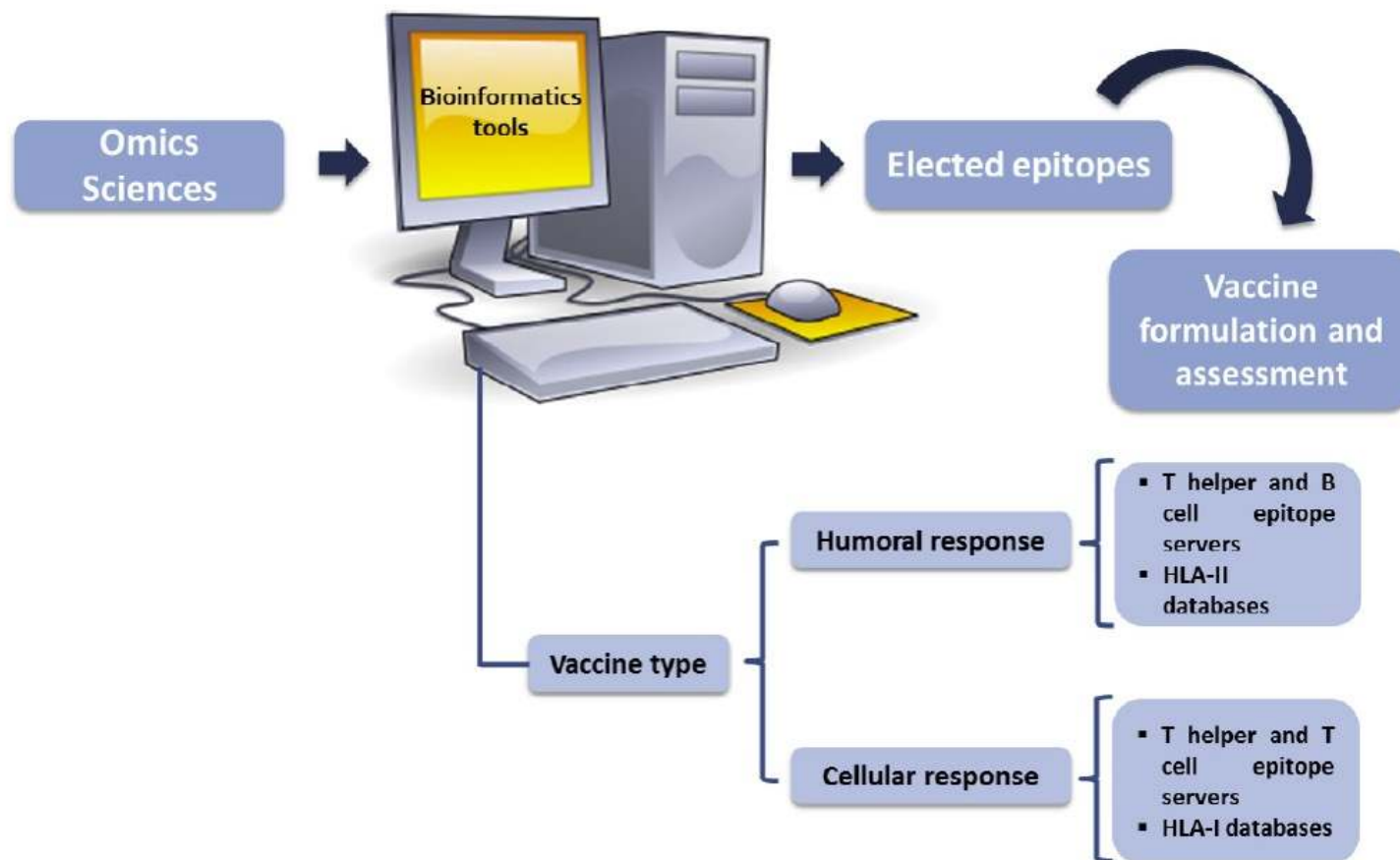


Fig. 1. Schematic representation of the workflow to identify epitopes for vaccine development.

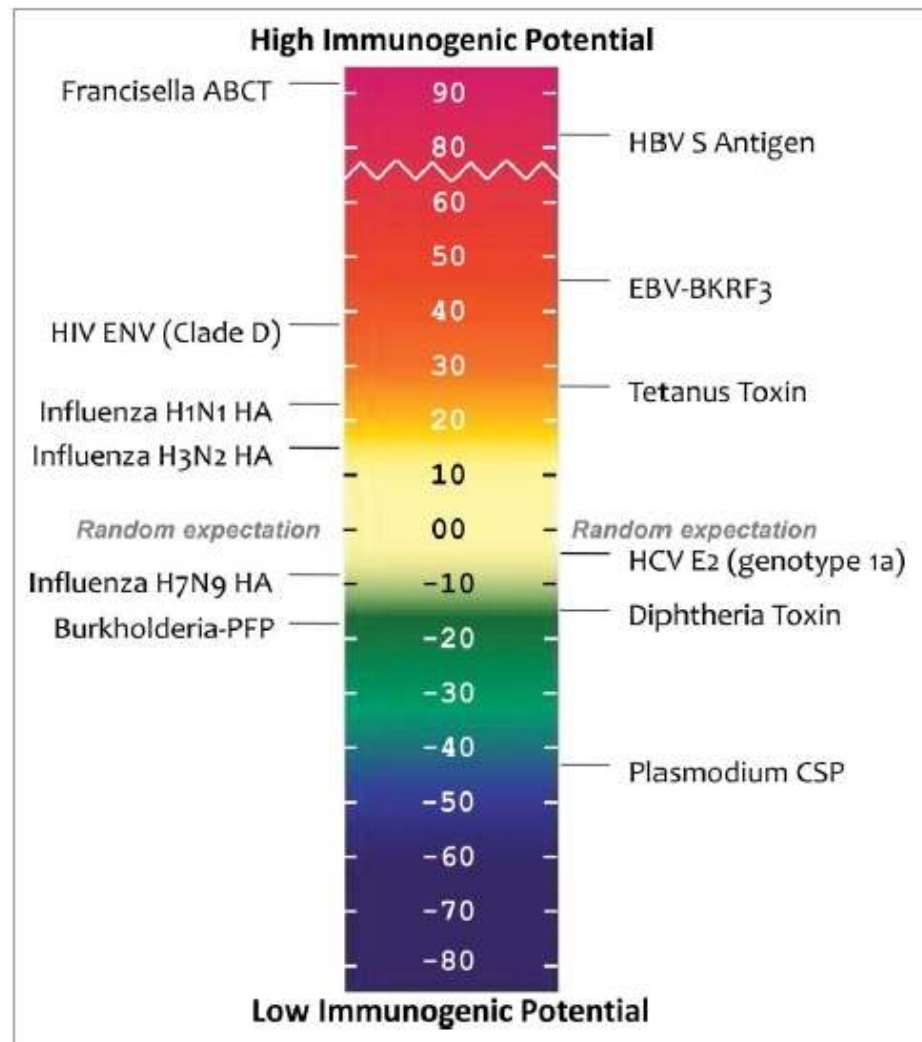


Figure 6. EpiMatrix protein immunogenicity scale. EpiMatrix protein immunogenicity scores higher than +20 are considered to be potentially immunogenic. On the left of the scale are well-known proteins for comparison. Low-scoring proteins near the bottom of the scale are known to engender little to no immunogenicity while higher scoring proteins near the top of the scale are known immunogens.

How these MHC predictions work

- Breaks the sequence into all possible peptides (of chosen length).
- Predicts the binding affinity for each peptide based on the method.
- Compares the predicted affinity to that of a large set of randomly selected peptides.
- Assigns a percentile rank depending on individual predicted affinity.
- Consensus picks the median rank of the methods used.

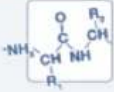
START YOUR SEARCH HERE ?

Epitope ?

☒ Any Epitopes
 ☐ Linear Epitope

Exact M Ex: SIINFEKL

☐ Discontinuous Epitopes
 ☐ Non-peptidic Epitopes



Assay ?

☒ Positive Assays Only
 ☒ T Cell Assays
 ☒ B Cell Assays
 ☒ MHC Ligand Assays

Ex: neutralization Find



Antigen ?

Organism

Ex: influenza, peanut

Antigen Name

Ex: core, capsid, myosin



Host ?

☒ Any Host
 ☐ Humans
 ☐ Mice
 ☐ Non-human Primates

Ex: dog, camel Find



MHC Restriction ?

☒ Any MHC Restriction
 ☐ MHC Class I
 ☐ MHC Class II
 ☐ MHC Nonclassical

Ex: HLA-A*02:01 Find



Disease ?

☒ Any Disease
 ☐ Infectious Disease
 ☐ Allergic Disease
 ☐ Autoimmune Disease

Ex: asthma, diabetes Find



Reset Search

Epitope Analysis Resource

T Cell Epitope Prediction ?

Scan an antigen sequence for amino acid patterns indicative of:

- MHC I Binding
- MHC II Binding
- MHC I Processing (Proteasome, TAP)
- MHC I Immunogenicity

B Cell Epitope Prediction ?

Predict linear B cell epitopes using:

- Antigen Sequence Properties

Predict discontinuous B cell epitopes using antigen structure via:

- Discotope
- ElliPro

Epitope Analysis Tools ?

Analyze epitope sets of:

- Population Coverage
- Conservation Across Antigens
- Clusters with Similar Sequences

Fig. 2. Metadata associated with epitope identification capture by the IEDB. The IEDB homepage (www.iedb.org), and initial search fields are shown. The IEDB is an NIH-NIAID funded publicly available database of T and B cell epitopes curated from the published literature or by direct submission from NIH-NIAID funded large scale epitope discovery contracts. From the homepage, epitopes can be search using selected criteria, and subsequent results can be further filtered with additional criteria, to include specific assays or receptor(s).

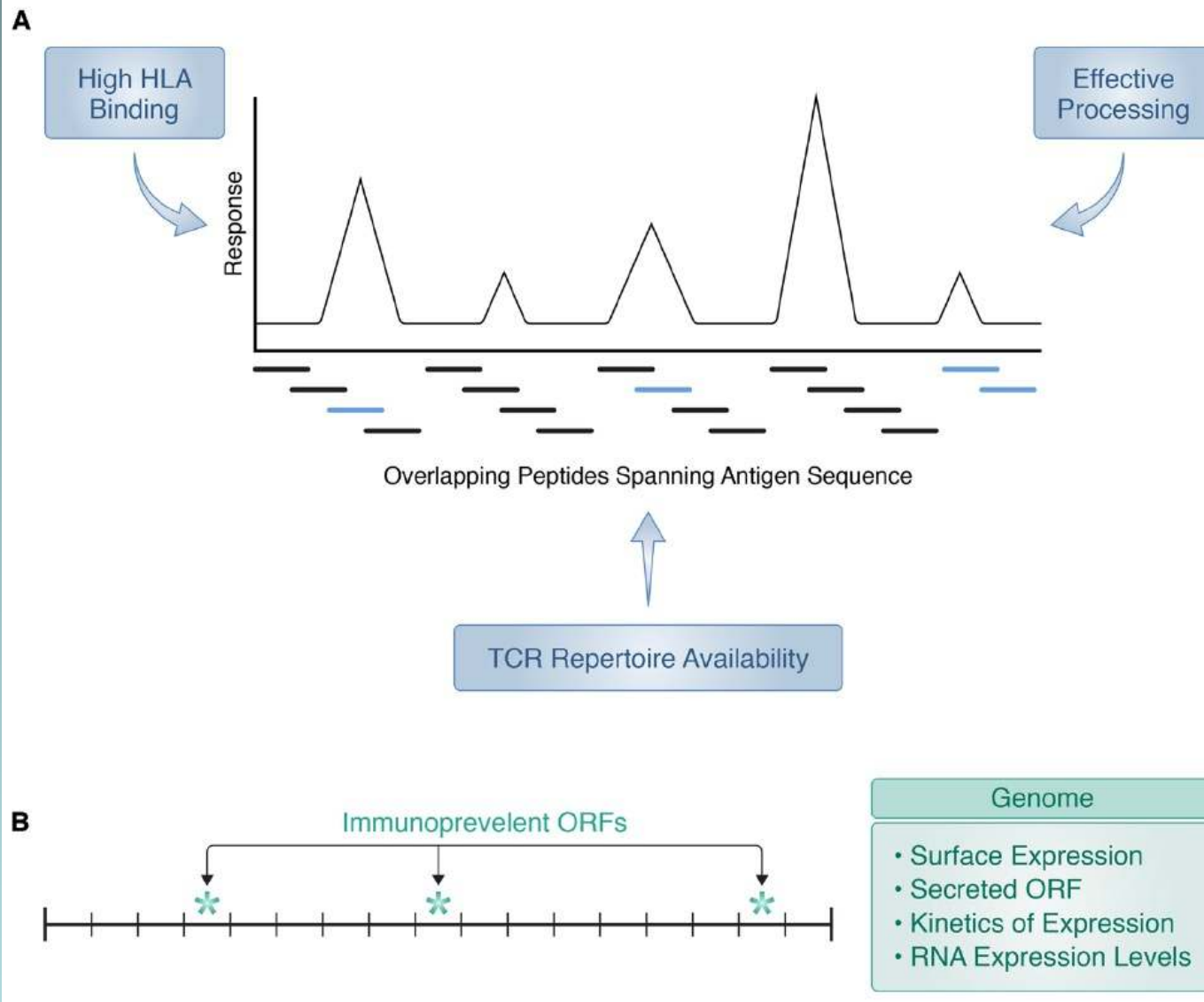


Fig. 4. Immunodominance and immunoprevalence. A) Epitopes effectively generated by processing and for which a TCR repertoire is available (light blue bars), and capable of eliciting the (relatively) strongest T cell responses, are termed immunodominant. B) Epitopes, antigens or ORFs that elicit responses with high frequency in an out-bred population are termed immunoprevalent. Typically, only a few epitopes/ORFs/antigens are found to be immunoprevalent, and may be associated with high levels of surface or RNA expression or secretion (class II); kinetics (time) of expression is often associated with dominance of class I or class II responses.

SYFPEITHI

www.syfpeithi.de

Immunogenetics (1999) 50:213–219

REVIEW

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

www.syfpeithi.de

SYFPEITHI: database for MHC ligands and peptide motifs

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.

REVIEW

www.syfpeithi.de

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

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)

	Position	Source	Accession No. EMBL database
	1 2 3 4 5 6 7 8 9		
Anchor residues	H		
Preferred	I E P A V V V F		
residues	Y A D G I R R M		
	T S E P M P E		
	G G N K T		
	E K R A		
	Q		
	V		
Examples for ligands	G H D P R A Q G T L	HLA-DP α chain (220–229)	X00457
	D H C V A H K L	Cytochrome C reductase (66–73)	M36647
	I H E D S T N R R R L	Heat shock protein 90 b (440–450)	M16660
	E H A H N M R V M	Elongation factor 2 (489–497)	M19997
	G H L E N N P A L	60 S acidic rib. protein PQ (67–75)	M17885
	H H S G A K V V L	Chaperonin cont. TCP-1 η (282–290)	AF026292
	I H D P G R G A P L	60 S ribosomal protein L8 (49–58)	Z28407
	T H T Q P G V Q L	Septin 2 homologue (70–78)	D50918
	T H Y V A P R R L	Transcription activator SNF2L4 (899–907)	U29175
	Y H G H G V S A F	Human EST	T96718
	Y Q E K G V R V L	Actin-related protein Arp2 (402–410)	AF006082
	I H E P E P H I L	Cyclin-dep. kin. reg. subunit 1 (59–67)	X54941
	A H S T I M P R L	DNA repl. lic. factor MCM4 (694–702)	X74794
	E H A G V I S V L	HBV X interacting protein (40–48)	AF029890

Table 5 Motif prediction of HLA-B*1510 self peptides and epitopes

Actin-related protein 2 (ARP2; human)

AA pos.	Score	Sequence
378	19	YQEKGVRL
227	15	IEQEQKLAL
29	14	EHIFPALVG
60	14	EASELRSML
164	14	THICPVYEG

Septin 2 homologue (SEP2; human)

AA pos.	Score	Sequence
70	25	THTQPGVQL
156	22	GHSLKSLDL
371	22	LHQDEKKKL
129	17	EELKIRRVL
84	15	DLQESNVRL

60 S acidic ribosomal protein RQ (RLA0; human)

AA pos.	Score	Sequence
67	25	GHLENNPAL
80	19	PHIRGNVGF
241	18	IINGYKRVL
46	15	SLRGKAVVL
196	15	GSINPEVL

Elongation factor 2 (EF2; human)

AA pos.	Score	Sequence
489	24	EHAHNMRVM
356	18	IHLPSVTA
147	17	IAERIKPVL
491	17	AHNMRVMKF
843	16	GLKEGIPAL

Table 2 Peptide motif and natural ligands of HLA-DRB1*0301

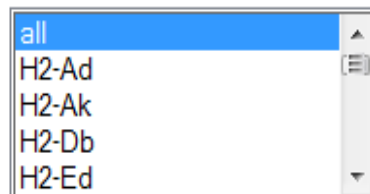
	Position									Source protein	Reference													
	1	2	3	4	5	6	7	8	9															
Anchors	L			D				K	Y		Malcherek et al. 1993; Geluk et al. 1992, 1994													
	I							R	L															
	F							E	F															
	M							Q																
	V							N																
Examples for ligands																								
	I	S	N	Q	L	T	L	D	S	N	T	K	Y	F	H	K	L	N	Apolipoprotein B-100 (2877–2894)	Malcherek et al. 1993				
	I	S	N	Q	L	T	L	D	S	N	T	K	Y	F	H	K	L	N	Apolipoprotein B-100 (2877–2893)	Malcherek et al. 1993				
	I	S	N	Q	L	T	L	D	S	N	T	K	Y	F	H	K			Apolipoprotein B-100 (2877–2892)	Malcherek et al. 1993				
	V	D	T	F	L	E	D	V	K	N	L	Y	H	S	E	A			α 1-Antitrypsin (149–164)	Malcherek et al. 1993				
	K	P	R	A	I	V	V	D	P	V	H	G	F	M	Y				LDL receptor (518–532)	Malcherek et al. 1993				
	K	Q	T	I	S	P	D	Y	R	N	M	I							IG2a (384–395)	Malcherek et al. 1993				
	Y	P	D	F	I	M	D	P	K	E	K	D	K	V					Unknown	Malcherek et al. 1993				
	N	I	Q	L	I	N	D	Q	E	V	A	R	F	D					Unknown	Malcherek et al. 1993				
	L	L	S	F	V	R	D	L	N	Q	Y	R	A	D	I				Transferrin receptor (618–632)	Malcherek et al. 1993				
L	P	K	P	P	K	P	V	S	K	M	R	M	A	T	P	L			Invariant chain (97–111)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992				
L	P	K	P	P	K	P	V	S	K	M	R	M	A	T	P	L	L	M	Q	A	L	P	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992	
L	P	K	P	P	K	P	V	S	K	M	R	M	A	T	P	L	L	M	Q	A	L	P	M	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
P	K	P	P	K	P	V	S	K	M	R	M	A	T	P	L				Invariant chain (98–113)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992				
P	K	P	P	K	P	V	S	K	M	R	M	A	T	P	L	L	M	Q	A				Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992	
K	P	P	K	P	V	S	K	M	R	M	A	T	P	L	L	M	Q		Invariant chain (99–116)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992				
K	P	P	K	P	V	S	K	M	R	M	A	T	P	L	L	M	Q	A	L	P	M		Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992	
V	D	D	T	Q	F	V	R	F	D	S	D	A	A	S	Q				HLA-A30 (52–67)	Chicz et al. 1993				
A	T	K	Y	G	N	M	T	E	D	H	V	M	H	L	L	Q	N	A		Invariant chain (131–149)	Chicz et al. 1993			
V	F	L	L	L	L	A	D	K	V	P	E	T	S	L	S				ACh receptor (289–304)	Chicz et al. 1993				
L	N	K	I	L	L	D	E	Q	A	Q	W	K							ICAM-2 (64–76)	Chicz et al. 1993				
G	P	P	K	L	D	I	R	K	E	E	K	Q	I	M	I	D	I	F	H				Chicz et al. 1993	
G	P	P	K	L	D	I	R	K	E	E	K	Q	I	M	I	D	I	F	H	P			Chicz et al. 1993	
G	F	K	A	I	R	P	D	K	K	S	N	P	I	I	R	T	V			Cyt-B5 (155–172)	Chicz et al. 1993			
Y	A	N	I	L	L	D	R	R	V	P	Q	T	D	M	T	F				Apolipoprotein B-100 (1207–1224)	Chicz et al. 1993			
N	L	F	L	K	S	D	G	R	I	K	Y	T	L	N	K	N	S	L	K				Chicz et al. 1993	
I	P	D	N	L	F	L	K	S	D	G	R	I	K	Y	T	L	N	K	N				Chicz et al. 1993	
I	P	D	N	L	F	L	K	S	D	G	R	I	K	Y	T	L	N	K					Chicz et al. 1993	
I	P	D	N	L	F	L	K	S	D	G	R	I	K	Y	T	L	N						Malcherek et al. 1993; Chicz et al. 1993	
I	P	D	N	L	F	L	K	S	D	G	R	I	K	Y	T	L							Chicz et al. 1993	
N	L	F	L	K	S	D	G	R	I	K	Y	T	L	N	K								Chicz et al. 1993	
N	L	F	L	K	S	D	G	R	I	K	Y	T	L	N									Chicz et al. 1993	
V	T	T	L	N	S	D	L	K	Y	N	A	L	D	L	T	N							Chicz et al. 1993	
V	G	S	D	W	R	F	L	R	G	Y	H	Q	Y	A						HLA-A2 (103–117)	Chicz et al. 1993			
T-cell epitopes																								
G	D	V	V	A	V	V	D	I	K	E	K	G	K	D	K	W	I	E	L	K			Lol pol. P1 (171–190)	Geluk et al. 1994
K	T	I	A	Y	D	E	E	A	R	R													HSP65 (cattle) (3–13)	Hawes et al. 1995
M	G	R	S	I	K	V	Q	L	Q														M. tuberculosis 30/31 kD protein (56–65)	Geluk et al. 1997
S	D	K	N	P	L	F	L	D	E	Q	L	I											M. tuberculosis HSP70 (257–269)	Geluk et al. 1997

Epitope prediction

This page allows you to find out the ligation strength to a defined HLA type for a sequence of aminoacids. The algorithmus 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.

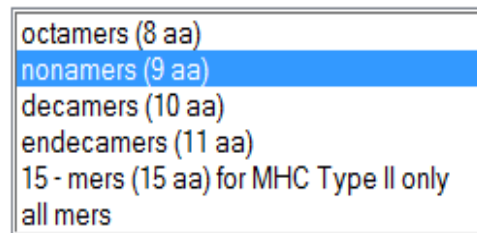
1. Select MHC type

If you chose "all", max. sequence length is 100 aminoacids (letters)!

A dropdown menu with a blue header bar containing the word "all". Below the header, the following options are listed: "H2-Ad", "H2-Ak", "H2-Db", and "H2-Ed". A small icon of a document with a checkmark is visible to the right of the list.

Hold down ctrl key when clicking
to select multiple items

2. Choose a mer

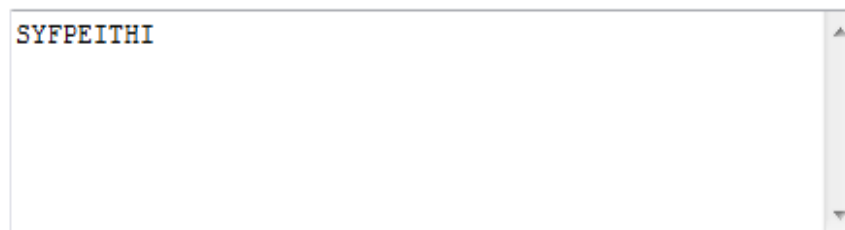
A dropdown menu with a blue header bar containing the text "nonamers (9 aa)". Below the header, the following options are listed: "octamers (8 aa)", "decamers (10 aa)", "endecamers (11 aa)", "15 - mers (15 aa) for MHC Type II only", and "all mers".

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.

A text input field with a light gray border and a vertical scrollbar on the right. The text "SYFPEITHI" is entered at the top of the field.

4. Choose Run to start analysis

Run

Reset

Home

IEDB - AR

www.iedb.org

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 Peters^{1,*}

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.

 IMMUNE EPITOPE DATABASE
AND ANALYSIS RESOURCE

[Home](#) [Browse](#) [Search](#) [Tools](#) [Support](#) [More IEDB](#)

Search

Epitope Structure

☒ Any

☐ Linear Peptide:

☐ Discontinuous Peptide

☐ Non-Peptidic:

Epitope Source

Source Organism:

Source Antigen:

Immune Mediated Disease Association

Disease Name:

Immune Recognition Context

☒ B Cell Response

☒ T Cell Response

☒ MHC Binding

☒ MHC Ligand Elution

Host Organism:

MHC Restriction:

MHC Class:

[❖ Help With Common Queries?](#)

Welcome!

 The IEDB contains data related to antibody and T cell epitopes for humans, non-human primates, rodents, and other animal species. Curation of peptidic and non-peptidic epitope data relating to all infectious diseases (including [NIAID Category A, B, and C priority pathogens](#) and [NIAID Emerging and Re-emerging infectious diseases](#)), allergens, autoimmune diseases, and transplant/alloantigens is current and constantly being updated. [More...](#)

Summary Metric	Count
Peptidic Epitopes	94560
Non-Peptidic Epitopes	1860
T Cell Assays	196437
B Cell Assays	152822
MHC Ligand Elution Assays	6687
MHC Binding Assays	221745
Epitope Source Organisms	2976
Restricting MHC Alleles	632
References	14386

[See all Metrics](#)

Resources

We have provided a variety of resources to analyze our data and enhance your IEDB experience:

- [T Cell Epitope Prediction](#)
- [B Cell Epitope Prediction](#)
- [Epitope Analysis Tools](#)
- [Database Export](#)
- [IEDB Ontology](#)
- [Data Field Descriptions](#)
- [Video Tutorials](#)



News

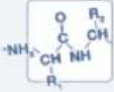
START YOUR SEARCH HERE ?

Epitope ?

☒ Any Epitopes
 ☐ Linear Epitope

Exact M

☐ Discontinuous Epitopes
 ☐ Non-peptidic Epitopes



Assay ?

☒ Positive Assays Only
 ☒ T Cell Assays
 ☒ B Cell Assays
 ☒ MHC Ligand Assays



Antigen ?

Organism
Antigen Name



Host ?

☒ Any Host
 ☐ Humans
 ☐ Mice
 ☐ Non-human Primates



MHC Restriction ?

☒ Any MHC Restriction
 ☐ MHC Class I
 ☐ MHC Class II
 ☐ MHC Nonclassical



Disease ?

☒ Any Disease
 ☐ Infectious Disease
 ☐ Allergic Disease
 ☐ Autoimmune Disease



Epitope Analysis Resource

T Cell Epitope Prediction ?

Scan an antigen sequence for amino acid patterns indicative of:

- MHC I Binding
- MHC II Binding
- MHC I Processing (Proteasome, TAP)
- MHC I Immunogenicity

B Cell Epitope Prediction ?

Predict linear B cell epitopes using:

- Antigen Sequence Properties

Predict discontinuous B cell epitopes using antigen structure via:

- Discotope
- ElliPro

Epitope Analysis Tools ?

Analyze epitope sets of:

- Population Coverage
- Conservation Across Antigens
- Clusters with Similar Sequences

Fig. 2. Metadata associated with epitope identification capture by the IEDB. The IEDB homepage (www.iedb.org), and initial search fields are shown. The IEDB is an NIH-NIAID funded publicly available database of T and B cell epitopes curated from the published literature or by direct submission from NIH-NIAID funded large scale epitope discovery contracts. From the homepage, epitopes can be search using selected criteria, and subsequent results can be further filtered with additional criteria, to include specific assays or receptor(s).



The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design

Ward Fleri*, Sinu Paul, Sandeep Kumar Dhanda, Swapnil Mahajan, Xiaojun Xu, Bjoern Peters and Alessandro Sette

Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA

The task of epitope discovery and vaccine design is increasingly reliant on bioinformatics analytic tools and access to depositories of curated data relevant to immune reactions and specific pathogens. The Immune Epitope Database and Analysis Resource (IEDB) was indeed created to assist biomedical researchers in the development of new vaccines, diagnostics, and therapeutics. The Analysis Resource is freely available to all researchers and provides access to a variety of epitope analysis and prediction tools. The tools include validated and benchmarked methods to predict MHC class I and class II binding. The predictions from these tools can be combined with tools predicting antigen processing, TCR recognition, and B cell epitope prediction. In addition, the resource contains a variety of secondary analysis tools that allow the researcher to calculate epitope conservation, population coverage, and other relevant analytic variables. The researcher involved in vaccine design and epitope discovery will also be interested in accessing experimental published data, relevant to the specific indication of interest. The database component of the IEDB contains a vast amount of experimentally derived epitope data that can be queried through a flexible user interface. The IEDB is linked to other pathogen-specific and immunological database resources.

Immune Epitope Database (IEDB) hosts a series of Machine Learning-based tools, each trained on specific dataset of experimental peptide-MHC binding affinity matrix.

These different tools encompass two common approaches of machine learning, namely, **linear regression (LR)** and **artificial neural network (ANN)**.

Η Βάση Δεδομένων Ανοσολογικών Επιτόπων IEDB φιλοξενεί μια σειρά από **εργαλεία βασισμένα στη Μηχανική Μάθηση**, καθένα από τα οποία έχει εκπαιδευτεί σε συγκεκριμένα σύνολα πειραματικών δεδομένων που αφορούν τη **συγγένεια δέσμευσης πεπτιδίων–MHC**.

Τα διάφορα αυτά εργαλεία ενσωματώνουν δύο βασικές προσεγγίσεις της μηχανικής μάθησης: την **γραμμική παλινδρόμηση (Linear Regression – LR)** και τα **τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks – ANN)**.

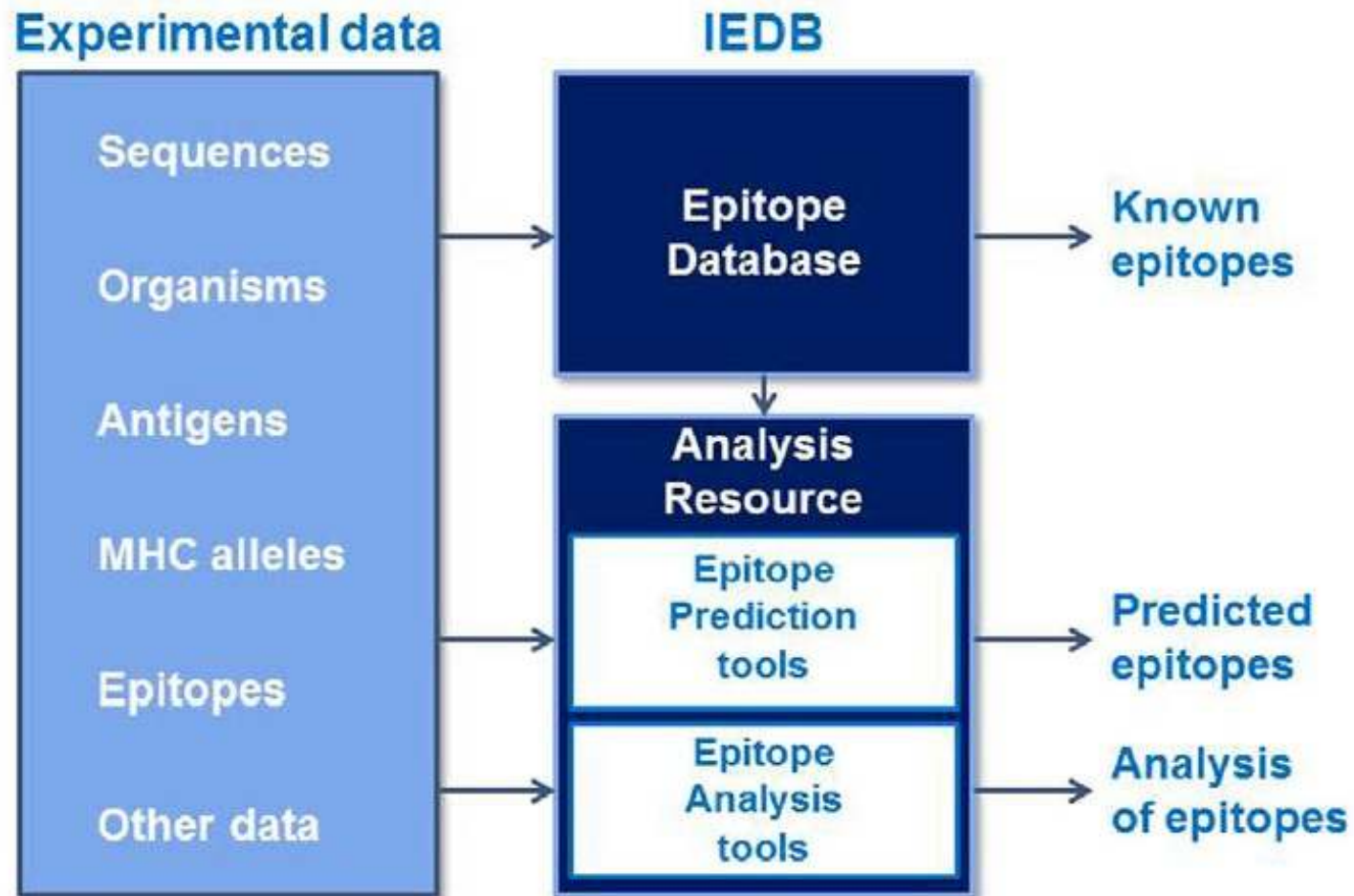


FIGURE 1 | The Immune Epitope Database and Analysis Resource captures experimental epitope data in a database and makes known epitopes freely available to the research community. These data are used to train epitope prediction tools in the Analysis Resource, which also contains tools to analyze sets of epitopes.

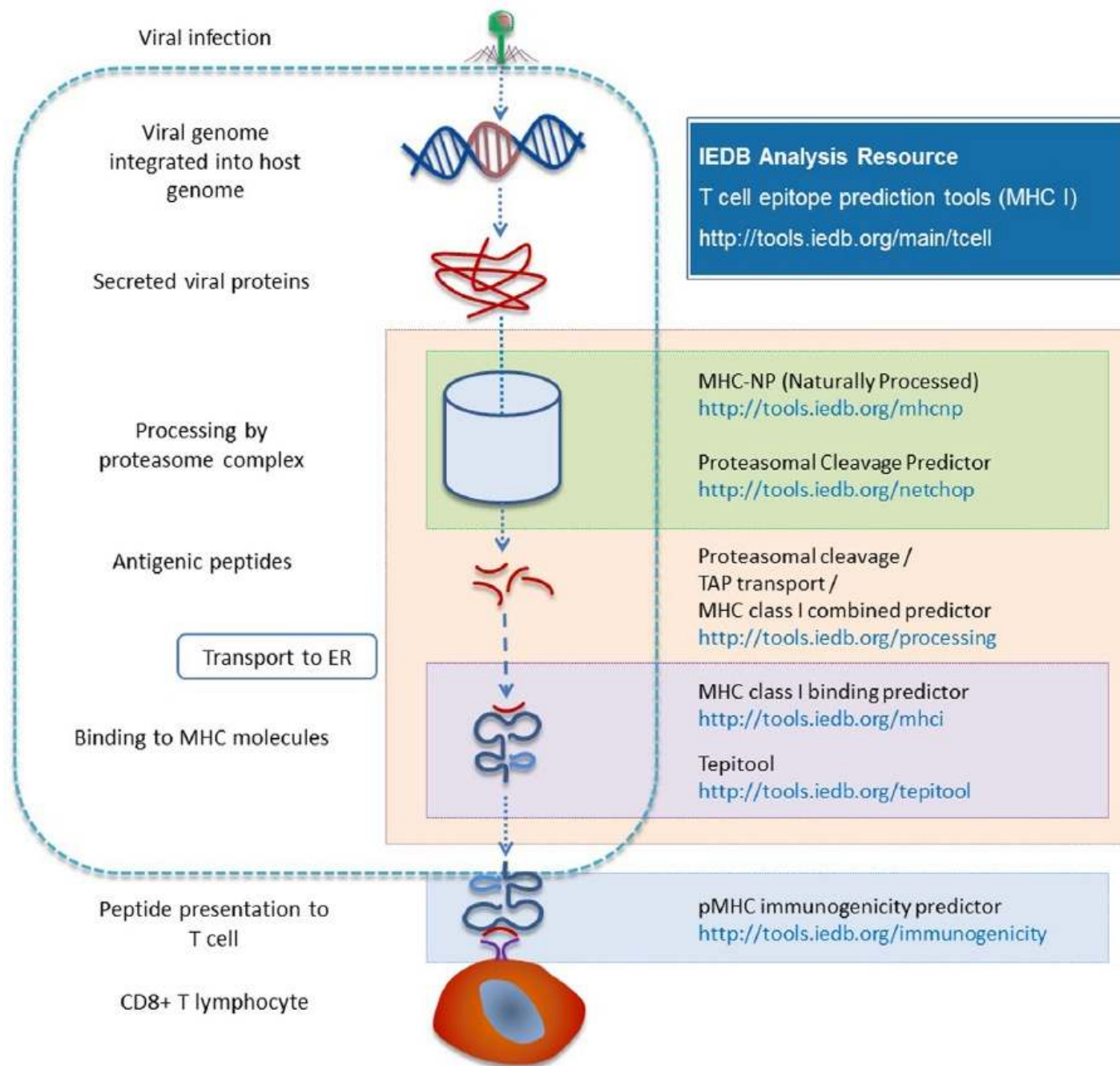


FIGURE 3 | Different prediction tools are available in the Analysis Resource with respect to different stages of MHC I antigen processing.

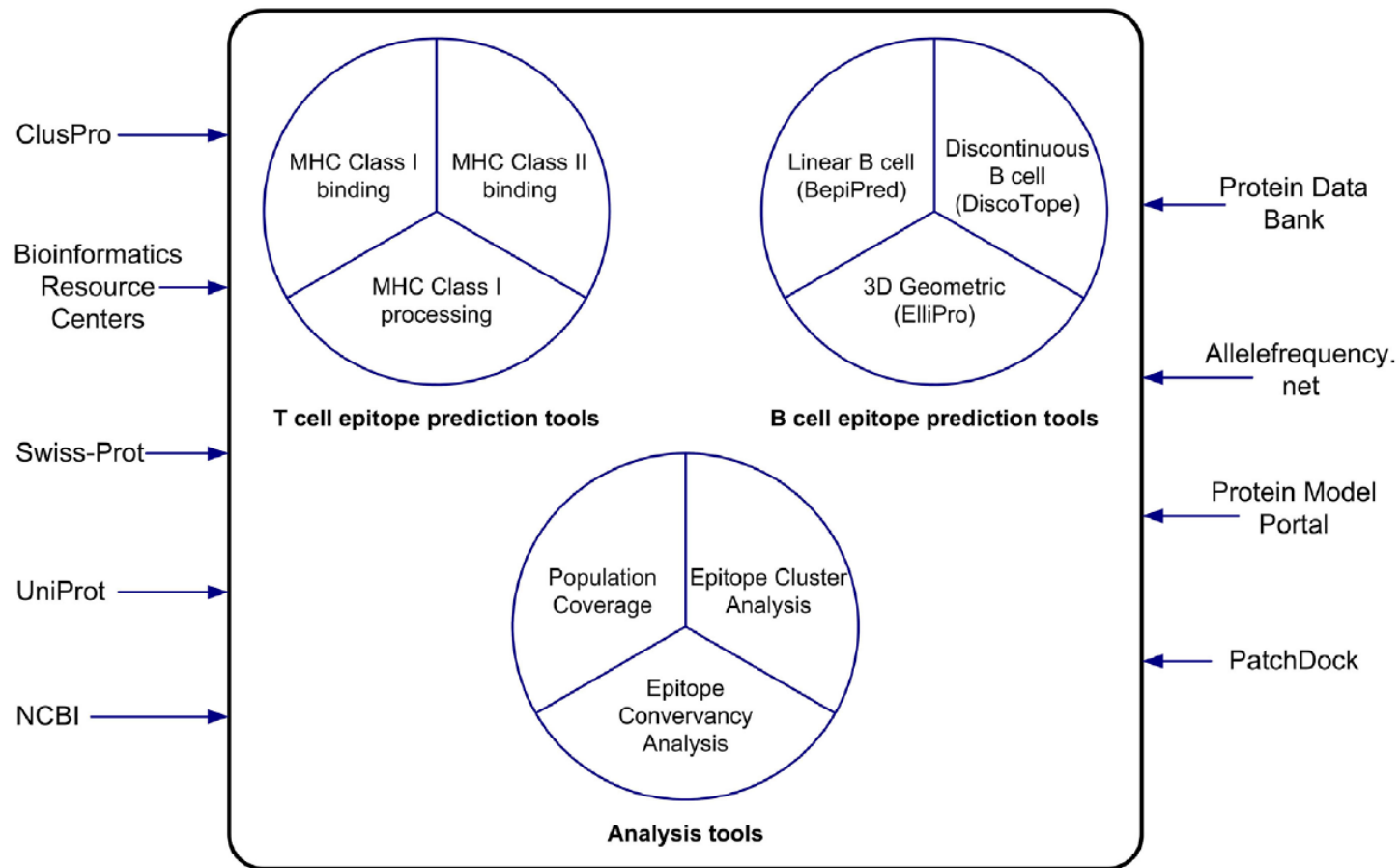


FIGURE 6 | The tools of the Analysis Resource can be used to predict T cell and B cell epitopes and to analyze sets of epitopes. The Analysis Resource interacts with a range of bioinformatics resources.

NetMHCpan-4.0

www.cbs.dtu.dk/services/NetMHCpan/

NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data

Vanessa Jurtz,^{*} Sinu Paul,[†] Massimo Andreatta,[‡] Paolo Marcatili,^{*} Bjoern Peters,[†] and Morten Nielsen^{*,‡}

Cytotoxic T cells are of central importance in the immune system's response to disease. They recognize defective cells by binding to peptides presented on the cell surface by MHC class I molecules. Peptide binding to MHC molecules is the single most selective step in the Ag-presentation pathway. Therefore, in the quest for T cell epitopes, the prediction of peptide binding to MHC molecules has attracted widespread attention. In the past, predictors of peptide–MHC interactions have primarily been trained on binding affinity data. Recently, an increasing number of MHC-presented peptides identified by mass spectrometry have been reported containing information about peptide-processing steps in the presentation pathway and the length distribution of naturally presented peptides. In this article, we present NetMHCpan-4.0, a method trained on binding affinity and eluted ligand data leveraging the information from both data types. Large-scale benchmarking of the method demonstrates an increase in predictive performance compared with state-of-the-art methods when it comes to identification of naturally processed ligands, cancer neoantigens, and T cell epitopes. *The Journal of Immunology*, 2017, 199: 3360–3368.

We trained the NetMHCpan method version 4.0 for the prediction of the interaction of peptides with MHC class I molecules integrating BA and MS EL data. Combined training was achieved by

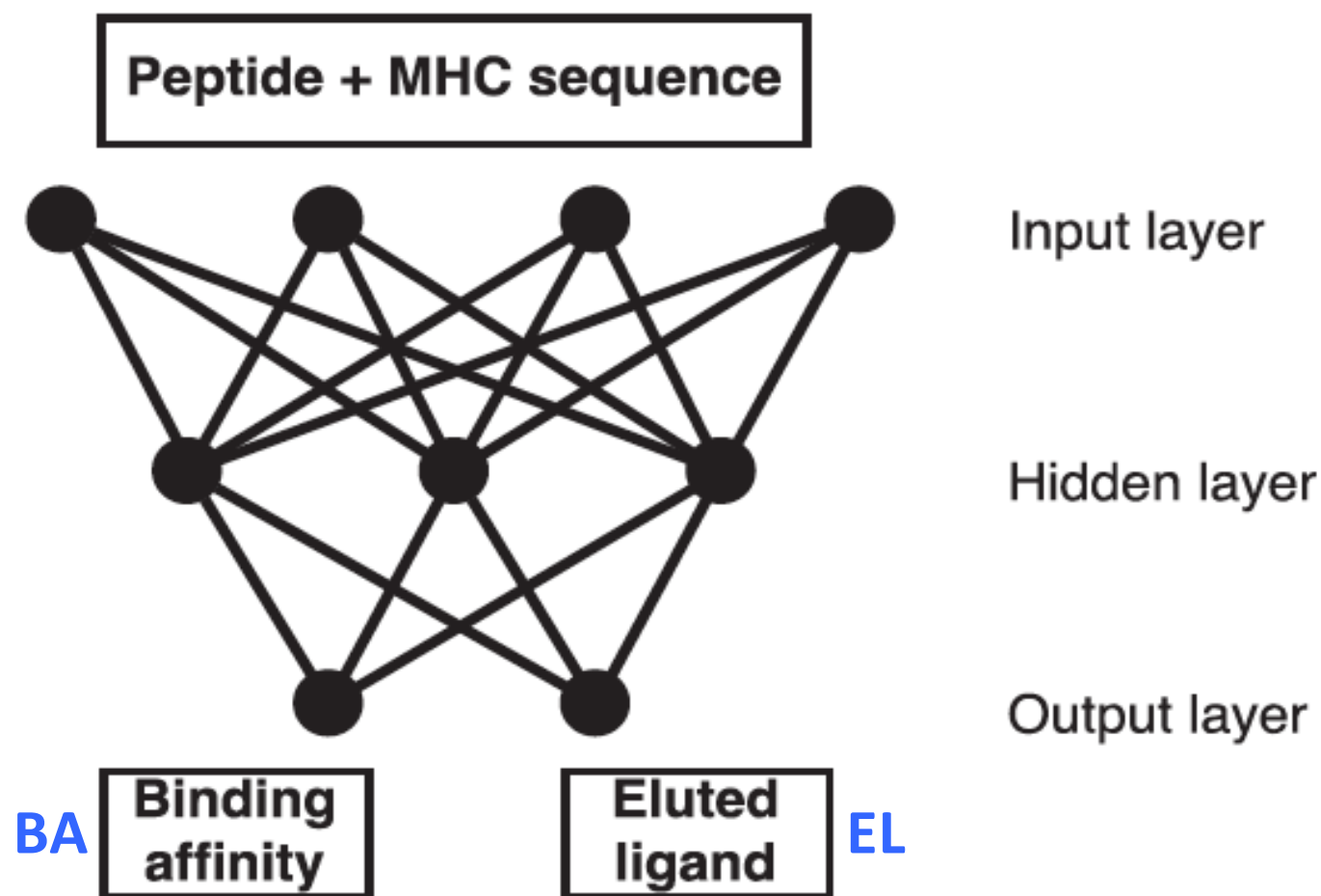


FIGURE 1. Visualization of the neural networks with two output neurons used for combined training on BA and EL data.

Cell Systems

MHCflurry: Open-Source Class I MHC Binding Affinity Prediction

Authors

Timothy J. O'Donnell, [Alex Rubinsteyn](#),
[Maria Bonsack](#), [Angelika B. Riemer](#),
[Uri Laserson](#), [Jeff Hammerbacher](#)

Correspondence

timothy.odonnell@icahn.mssm.edu

In Brief

Accurate prediction servers for MHC I ligands have been in wide use for some time, but these tools are typically closed source, may be trained only by their developers, and can be challenging to integrate into high-throughput workflows required for tumor neoantigen discovery. We introduce a prediction package that exposes a programmatic interface, may be modified and re-trained, and is much faster than existing tools.

<http://openvax.github.io/mhcflurry/>

Cell Systems

MHCflurry: Open-Source Class I MHC Binding Affinity Prediction

Highlights

- Open-source software package for peptide/MHC class I binding prediction
- Easily installed Python package with command line and library interfaces
- Trained on affinity measurements and MHC ligands identified by mass spectrometry

<http://openvax.github.io/mhcflurry/>

Methods

MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing

Timothy J. O'Donnell,^{1,5,*} Alex Rubinsteyn,^{2,3} and Uri Laserson^{1,4}

¹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

²Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

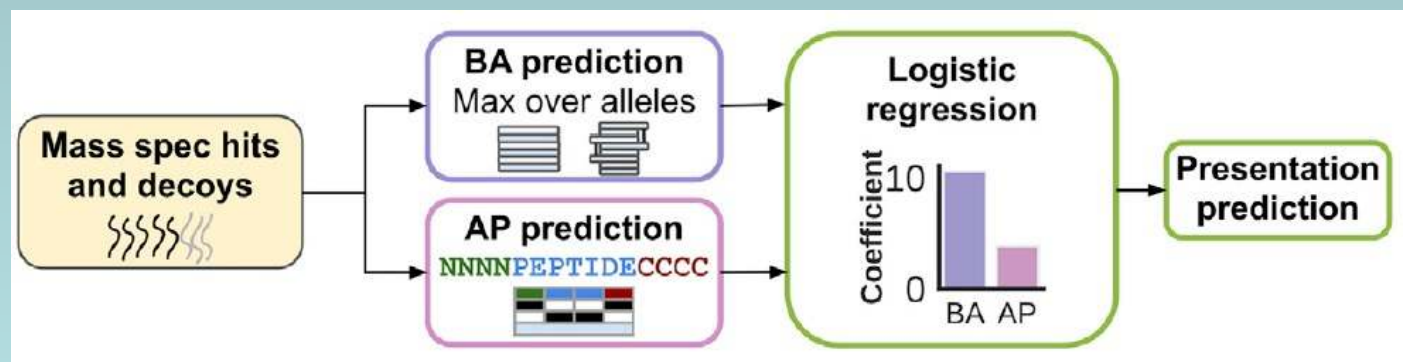
³Department of Genetics, UNC School of Medicine, Chapel Hill, NC 27599, USA

⁴Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

⁵Lead Contact

*Correspondence: tim@openvax.org

<https://doi.org/10.1016/j.cels.2020.06.010>



BA: binding affinity and AP: antigen processing

RESEARCH ARTICLE

Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes

Weilong Zhao , Xinwei Sher *

Global Research IT, Merck & Co., Inc., Boston, MA, United States of America

* xinwei_sher@merck.com

Computationally predicting antigen peptide sequences that elicit T-cell immune response has broad and significant impact on vaccine design. The most widely accepted approach is to rely on machine learning classifier, trained on large-scale major-histocompatibility complex (MHC)-binding peptide dataset. Because of the constant development of machine learning algorithms and expanding training data, providing comprehensive benchmarking of existing algorithms on blind testing dataset is important for recognizing the pros and cons of different algorithms and providing guidelines on specific applications. Here we present a study of such benchmarking by characterizing on a wide array of accuracy metrics, highlighting the best-in-class algorithms as well as their limitations. The rising concept that “naturally presented” antigen epitopes are more likely to generate effective T-cell immune response has led us to also consider the accuracy of these machine learning algorithms on predicting naturally presented peptides. We demonstrate that recent advance in incorporating high-quality naturally presented peptide data from mass spectrometry experiments has improved the accuracy. Our benchmarking of machine learning predictors for MHC-binding and MHC-naturally presented antigen peptides contributes to establishing best practice of computational T-cell epitope analysis, which also has implication in tumor neoantigen-based cancer vaccine discovery.

in silico

bacterial antigen
sequence

predict all possible spliced
and non-spliced peptides

reduce peptide database
with selection criteria

final candidate list
of epitopes

in vivo / ex vivo

Infection



day 0

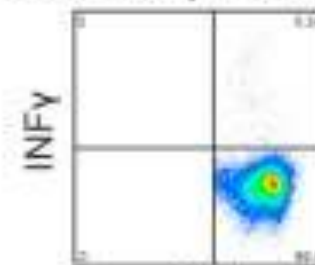


day 7

incubation with
candidate epitopes



detection of
CD8⁺ INFγ⁺ T cells



Seq2Logo

Nucleic Acids Research Advance Access published May 25, 2012

Nucleic Acids Research, 2012, 1–7
doi:10.1093/nar/gks469

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/biotools/Seq2Logo/>



<https://services.healthtech.dtu.dk/services/Seq2Logo-2.0/>

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/biotools/Seq2Logo> (14 May 2012, date last accessed).

Seq2Logo - 2.0

Sequence logo generator

Seq2Logo is a web-based sequence logo generation 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.

Note that Seq2Logo as default includes a pseudo count correction for lowcounts. This means that the amino acid frequencies displayed in the sequence logos are corrected for low number of observations using a Blosum amino acid similarity matrix. To turn this feature off, the Weight on prior must be set to zero.

Submission

Instructions

Output format

Abstract

Downloads

Submission

Provide Input (*MSA(Fasta and ClustalW), peptide, PSSM)*

[Switch to file upload](#)

Select Logo type: Kullback-Leibler

Clustering method: Clustering (Hobohm1)

Specify threshold for clustering (Hobohm1) 0.63 Threshold (Hobohm1)

Weight on prior (pseudo counts): 200

Select information content units: Bits Text on y-axis: Bits (the text on the y-axis can be edited at will)

Note: The PSSM of non-weight-matrix inputs will always be calculated in Bits*! *or Halfbits if chosen.

Available Output Formats. (multi)

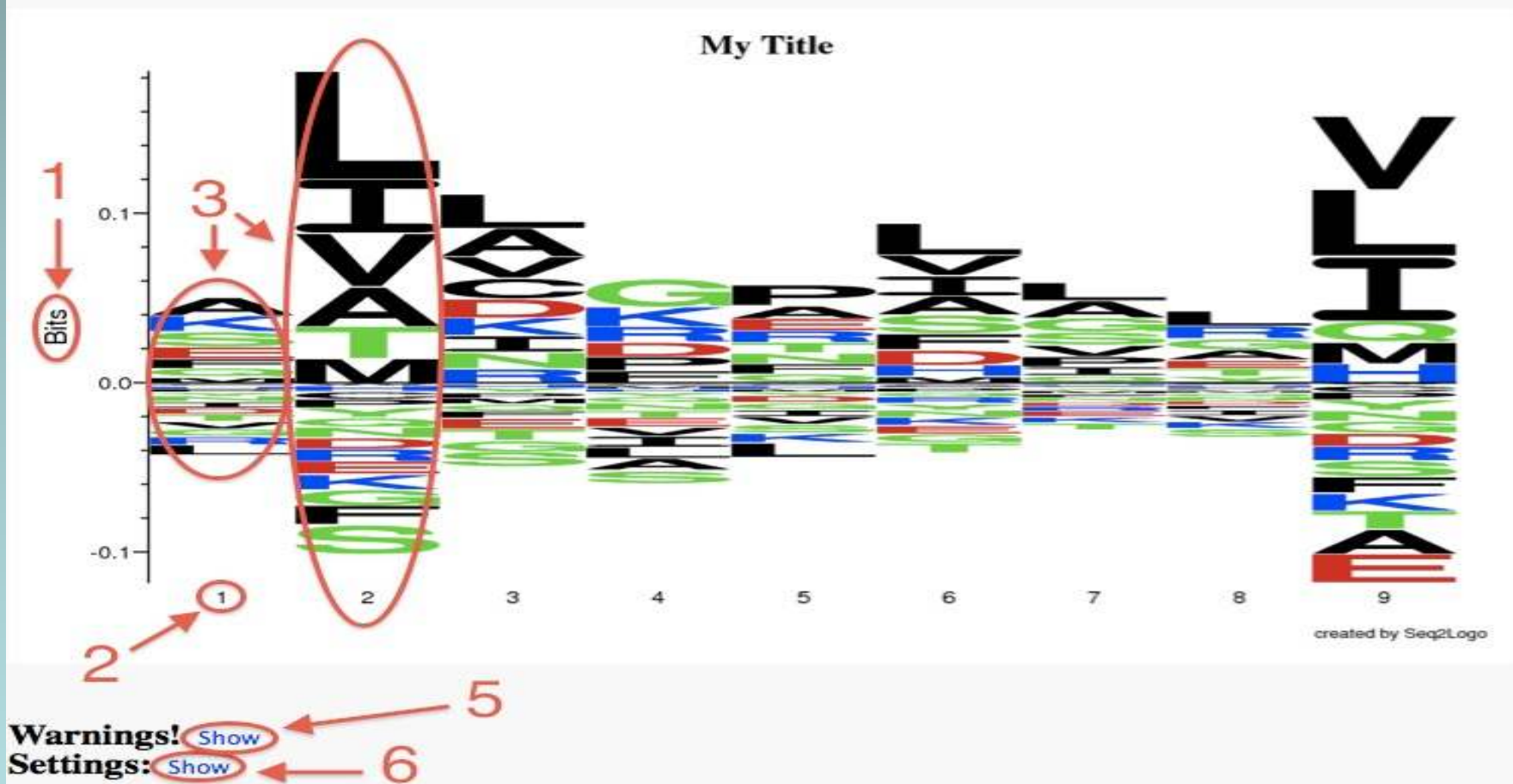
IPFG

Seq2Logo 1.0 results

Technical University of Denmark

Download logo as: [EPS](#) [PNG\(1\)](#) [Weightmatrix](#)

← 4



Prediction of B cell epitopes in proteins using a novel sequence similarity-based method

Alvaro Ras-Carmona¹, Alexander A. Lehmann^{1,2}, Paul V. Lehmann² & Pedro A. Reche^{1✉}

Prediction of B cell epitopes that can replace the antigen for antibody production and detection is of great interest for research and the biotech industry. Here, we developed a novel BLAST-based method to predict linear B cell epitopes. To that end, we generated a BLAST-formatted database upon a dataset of 62,730 known linear B cell epitope sequences and considered as a B cell epitope any peptide sequence producing ungapped BLAST hits to this database with identity $\geq 80\%$ and length ≥ 8 . We examined B cell epitope predictions by this method in tenfold cross-validations in which we considered various types of non-B cell epitopes, including 62,730 peptide sequences with verified negative B cell assays. As a result, we obtained values of accuracy, specificity and sensitivity of $72.54 \pm 0.27\%$, $81.59 \pm 0.37\%$ and $63.49 \pm 0.43\%$, respectively. In an independent dataset incorporating 503 B cell epitopes, this method reached accuracy, specificity and sensitivity of 74.85%, 99.20% and 50.50%, respectively, outperforming state-of-the-art methods to predict linear B cell epitopes. We implemented this BLAST-based approach to predict B cell epitopes at <http://imath.med.ucm.es/bepiblast>.

BepiBlast

a

No file selected.

Peptide length:

Min. Identity ≥ %

Gaps:

☐ Only neutralizing

☒ Map in sequence

b

Start	End	Predicted Epitope	Scores	Accsb	Flexb
39	52	GTLVKITDDQIEV	68	0.22	0.34
79	93	DCTLIDALLGDPHCD	65	0.19	-0.50
101	108	DLFVERSK	43	0.20	-0.07
110	126	FSNCYPYDVPDYASLRS	96	0.19	0.36
143	150	WTGVTQNG	46	0.27	0.34
161	170	SGFFSRLNWL	57	0.21	-0.41
197	205	GIHHPSTNQ	56	0.23	-0.56
230	237	IPNIGSRP	48	0.15	0.13
240	248	RGLSSRISI	36	0.25	-0.28
251	258	TIVKPGDV	38	0.17	-0.54
288	295	APIDTCIS	35	0.26	0.33
321	328	CPKYVKQN	47	0.18	0.28
337	344	RNVPEKQT	39	0.21	0.07
347	360	LPGAIAAGPIENGWE	75	0.19	0.15
380	398	AADLKSTQAAIDQINGKLN	64	0.24	-0.82
401	417	IEKTNEKFHQIEKEFSE	87	0.17	-1.18
439	447	YNAELLVAL	40	0.21	-0.42
451	461	HTIDLTDSMN	80	0.18	0.41
506	513	VYRDEALN	44	0.15	0.82

Figure 2. BepiBlast web server. (a) BepiBlast interface. (b) Representative BepiBlast output obtained with default settings. The shown results were obtained for hemagglutinin from Influenza A virus (UniProt Id: P03437). BepiBlast main result consists of a table displaying the following information (from left to right): peptide starting position; peptide ending position; predicted B cell epitope; bit score; accessibility value and flexibility value.

Tool	Algorithm	Training dataset		Validation	URL	Reference
		B cell epitopes	Non-B cell epitopes			
BepiBlast	BLAST	62,730	–	X, I	http://imath.med.ucm.es/bepiblast/	–
Bceps	Support vector machine	555	555 (a)	X, I, E	http://imath.med.ucm.es/bceps/	18
BepiPred 2.0 ^a	Random forest	3542	36,785	X, I, E	https://services.healthtech.dtu.dk/service.php?BepiPred-2.0	20
LBtope ^b	Support vector machine	14,876	23,321 (b)	X, I	https://webs.iiitd.edu.in/raghava/lbtope/	17
IBCE-EL	Random tree with boosting	4440	5485 (b)	X, I	http://www.thegleelab.org/iBCE-EL/	28
DLBEpitope	Deep neural network	22,012	201,563 (b)	X, I	http://ccb1.bmi.ac.cn:81/dlbepitope/index.php?	15
ILBE	Random Forest	4440	5485 (b)	X, I	http://kurata14.bio.kyutech.ac.jp/iLBE/	41
ABCPred	Neural network	700	700 (a)	X, I	https://webs.iiitd.edu.in/raghava/abcpred/	14
BCPREDS	Support vector machine	701	701 (a)	X, I, E	http://ailab.ist.psu.edu/bcpred/	32
SVMtrip	Support vector machine	4925	4925 (b)	X	http://sysbio.unl.edu/SVMTriP/prediction.php	16



MHC-I binding prediction - example

tools.iedb.org/mhci/

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

MHC-I Binding Predictions

Prediction Method Version: v2.24 [\[Older versions\]](#)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences.

Specify Sequence(s)

```
>LCMV Armstrong, Protein GP
MGQIVTMFEALPHIIDVINIVIVLIVITGIKAVYNFATCGIFALISFLLAGRSCGM
YGLKGPDIYKGVYQFKSVEFDMSHLNLTM PNACSANN SHHYISMGTSGLELTFTNDSII
SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA
QSAQSQCRTRFRGRVLD MFRTAFGGKYMRS GWGWTGSDGKTTWCSQTSYQYLIQNRTWE
NHCTYAGPFGMSRILLSQEKT KFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE
LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSLISDQ
LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVSI FLHLVKIPTHRH IKGGSCKPKP
HRLTNKGICSCGAFKVPGVKTVWKRR
```

Or select file containing sequence(s)

No file chosen

**Epitope
sequence**
(copy or upload)

MHC-I binding prediction - example

tools.iedb.org/mhci/

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

MHC-I Binding Predictions

Prediction Method Version: v2.24 [\[Older versions\]](#)

Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences.

>LCMV Armstrong, Protein GP
MGQIVTMFEALPHIIDEVINIIVLIVITGKAVYNFATCGIFALISFLLAGRSCGM
YGLKGPDIYKGVYQFSVEFDMSHLNTMPNACSAANSHHYISMGTSGLELTFTNDSII
SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA
QSAQSQCRTRFRGVLD MFRTAFGGKYMRSWGWTGSDGKTTWCSQTSYQYLIQNRTWE
NHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE
LKCFGNTAVAKCNVNHDAEFCMDMLRLIDYNKAALSFKEDVESALHLFKTTVNSLISDQ
LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTAYLVSIFLHLVKIPTHRHIKGGSCPKP
HRLTNKGICSCGAFKVPGVKTVWKRR

Or select file containing sequence(s)

No file chosen

Choose a Prediction Method

Prediction Method [?]

Show all the method versions: ☒

MHC source species

Show only frequently occurring alleles: ☒ [?]

Select MHC allele(s)

Select HLA allele reference set: ☐ [?]
(Specify MHC allele sequence)

Sort peptides by

Show

All predictions

Output format

XHTML table

Email address (optional)

IEDB recommended 2020.09 (NetMHCpan EL 4.1) [?]

IEDB recommended 2020.09 (NetMHCpan EL 4.1)

Consensus

NetMHCpan BA 4.1

IEDB recommended 2020.04 (NetMHCpan EL 4.0)

NetMHCpan BA 4.0

ANN 4.0

SMMPMBEC

SMM

CombLib_Sidney2008

PickPocket

netMHCcons

netMHCstabpan

[Help on prediction method selections](#)

Prediction
method

MHC-I binding prediction – example

tools.iedb.org/mhci/

Home Help Example Reference Download Contact

MHC-I Binding Predictions

Prediction Method Version: v2.24 [\[Older versions\]](#)

Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences.

```
>LCMV Armstrong, Protein GP
MGQIVTMFEALPHIIDEVINIVIVLIVITGIKAVYNFATCGIFALISFLLAGRSCGM
YGLKGPDIYKGVYQFQKSVFEDMSHLNLTMPNACSANNSHHYISMGTSGLELFTNDSII
SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIYNLTFSDA
QSAQSQCRTERGRVLDMFRTAFGGKYMRSWGWTGSDGKTTWCQSQSYQLIINRTWE
NHCTYAGPFGMSRILLSQEKTKFFTRRLAGFTWTLSDSSGVENPGGYCLTKWMILAAE
LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSLISDQ
LLMRNHLRDLMGVPYCNSKFWYLEHAKTGETSVKPCWLVTNGSYLNETHFSDQIEQEA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTAYLVSIFLHLVKIPTRHIKGGSCPKP
HRLTNKGICSCGAFKVPGVKTVVKRR
```

Or select file containing sequence(s): No file chosen

Choose a Prediction Method

Prediction Method [?] [Help on prediction method selections](#)

Show all the method versions: ☐

Specify what to make binding predictions for

MHC source species: [?]

Show only frequently occurring alleles: ☒ [?]

Select MHC allele(s): [?]

Select HLA allele reference set: ☐ [?] [\(Specify MHC allele sequence\)](#)

Length: [?] [Upload allele file](#) [?]

Specify Output

Sort peptides by: [?]

Show: [?]

Output format: [?]

Email address (optional): [?]

Choose species

MHC-I binding prediction – example

tools.iedb.org/mhci/

Home Help Example Reference Download Contact

MHC-I Binding Predictions

Prediction Method Version v2.24 [\[Older versions\]](#)

Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences.

```
>LCMV Armstrong, Protein GP
MGQIVTMFEALPHIIDEVINIVIVLIVITGIKAVYNFATCGIFALISFLLLAGRSCGM
YGLKGPDIYKGVYQFKSVEFDMSHLNLTPNACSAANSHHYISMGTSGLELFTNDSII
SHNFCNLTSAFNKKTDFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA
QSAQSQCRTRFRGRVLDMFRTAFGGKYMRSWGWTGSDGKTTWCSQTSYQYLIQNRTWE
NHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE
LKCFGNTAVAKCNVNHDAEFCMDLRLLIDYNKAALSFKEDVESALHLFKTTVNSLISDQ
LLMRNHLRDLMGVPYCNYSKEWYLEHAKTGETSVPKCWLVTNGSYLNETHESDQIEQEA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTAYLVISIFLHLVKIPTRHIKGGSCPKP
HRLTNKGICSCGAFKVPGVKTVWKR
```

Or select file containing sequence(s) No file chosen

Choose a Prediction Method

Prediction Method (?) IEDB recommended 2020.09 (NetMHCpan EL 4.1) [Help on prediction method selections](#)

Show all the method versions: ☐

Specify what to make binding predictions for

MHC source species human

Show only frequently occurring alleles ☒ (?)

Select MHC allele(s)

Allele	Length	
HLA-A*01:01	9	☐
HLA-B*07:02	10	☐

[Select HLA allele reference set: \(Specify MHC allele sequence\)](#) ☐ (?)

[Upload allele file \(?\)](#)

Specify Output

Sort peptides by Predicted IC50

Show All predictions

Output format XHTML table

Email address (optional) (?)

Complete set

Reference alleles

Specify allele(s) & peptide length
(select or upload)

Upload format:
HLA-A*01:01,9
HLA-B*07:02,10

Prediction method dependent allele selection

tools.iedb.org/mhci/

NetMHCpan prediction methods allow **FASTA sequence input**

The screenshot shows the IEDB MHC prediction interface. A red box highlights the 'Specify MHC allele sequence' section, which includes a text area for pasting a full-length MHC protein sequence in FASTA format. A red arrow points from the 'FASTA sequence input' text box to the 'Specify MHC allele sequence' section. Another red arrow points from the 'Specify MHC allele sequence' section to the 'Select MHC allele(s)' section, which includes a checkbox for 'Select HLA allele reference set' and a link to 'Input FASTA sequence (Select MHC allele(s))'.

Choose a Prediction Method

Prediction Method [?]
Show all the method versions: ☐ IEDB recommended 2020.09 (NetMHCpan EL 4.1) ^v [Help on prediction method selections](#)

Specify what to make binding predictions for

MHC source species

Select MHC allele(s)
[Select HLA allele reference set:](#) ☐ [?]
[Input FASTA sequence \(Select MHC allele\(s\)\)](#)

Paste a single full length MHC protein sequence in [FASTA](#) format:

Select "Specify MHC allele sequence"

MHC-I binding prediction – example

tools.iedb.org/mhci/

Home Help Example Reference Download Contact

MHC-I Binding Predictions

Prediction Method Version v2.24 [\[Older versions\]](#)

Specify Sequence(s)

Enter protein sequence(s) in FASTA format or as whitespace-separated sequences.

```
>LCMV Armstrong, Protein GP
MGQIVTMFEALPHIIDEVINIVIVLIVITGIKAVYNFATCGIFALISFLLLAGRSCGM
YGLKGPDIYKVYQFKSVFDMSHLNLTMNPACANNHHYISMGTSGLELFTNDSII
SHNFCNLTSFANKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQNLTFSDA
QSAQSQCRTRFRVLDMFRTAFGGKYMRSWGWTGSDGKTTWCSQTSYQYLIQNRWE
NHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE
LKCFGNTAVAKCNVNHDAEFCMDMLRLIDYNKAALSFKEDVESALHLFKTTVNSLISDQ
LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVLTNGSYLNETHFSDQIEQEA
DNMITEMLRKDYIKRQGSTPLALMDLLMFSTAYLVSIFLHLVKIPTRHIKGGSCPKP
HRLTNKGICSCGAFKVPGVKTVWKRR
```

Or select file containing sequence(s) No file chosen

Choose a Prediction Method

Prediction Method [?](#) IEDB recommended 2020.09 (NetMHCpan EL 4.1) [Help on prediction method selections](#)

Show all the method versions: ☐

Specify what to make binding predictions for

MHC source species human

Allele	Length	
HLA-A*01:01	9	☐
HLA-B*07:02	10	☐

Show only frequently occurring alleles: ☒ [?](#)

Select MHC allele(s)

Select HLA allele reference set: ☐ [?](#)
(Specify MHC allele sequence)

[Upload allele file](#) [?](#)

Specify Output

Sort peptides by Predicted IC50

Show Position in sequence

Output format All predictions

IC50 below [cutoff] nM

Percent rank below [cutoff]

Email address (optional) [?](#)

Input

Output

How the tool works

- Breaks the sequence into all possible peptides (of chosen length).
- Predicts the binding affinity for each peptide based on the method.
- Compares the predicted affinity to that of a large set of randomly selected peptides.
- Assigns a percentile rank depending on individual predicted affinity.
- Consensus picks the median rank of the methods used.

MHC-I binding prediction – example

tools.iedb.org/mhci/

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

MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	LCMV Armstrong, Protein GP	MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFL LLAGRSCGMYGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNHHY ISMGTSGLELTFTNDSIISHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIR GNSNYKAVSCDFNNGITIQYNLTFSDAQSAQSQCRTRGRVLDMFRTAFG GKYMRSWGWGTGSDGKTTWCSQTSYQYLIIQNRTWENHCTYAGPFGMSRI LLSQEKTKFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMLAAELKCFG NTAVAKCNVNHDAEFCMLRLIDYNKAALSKFKEDVESALHLFKTTVNSL ISDQLLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYL NETHFSQIEQEAADNMITEMLRKDYIKRQGSTPLALMDLLMFSTAYLVS IFLHLVKIPTHRHKGGSCPKPHRLTNKGICSCGAFKVPGVKTVWKRR

Prediction method: **IEDB recommended 2.22** | **Low Percentile Rank = good binders**
[Download result](#)

Citations
Check to expand the result: ☐

Allele	#	Start	End	Length	Peptide	Method used	Percentile Rank
HLA-A*01:01	1	92	100	9	CSANNSHHY	Consensus (ann/smm)	0.2
HLA-A*01:01	1	405	413	9	FSDQIEQEA	Consensus (ann/smm)	0.34
HLA-B*07:02	1	471	480	10	KPHRLTNKGI	Consensus (ann/smm)	0.35
HLA-A*01:01	1	417	425	9	ITEMLRKDY	Consensus (ann/smm)	0.6
HLA-A*01:01	1	361	369	9	LRDLMGVPY	Consensus (ann/smm)	0.68
HLA-A*01:01	1	233	241	9	RTWENHCTY	Consensus (ann/smm)	0.69
HLA-A*01:01	1	217	225	9	TTWCSQTSY	Consensus (ann/smm)	0.71
HLA-A*01:01	1	439	447	9	LLMFSTAY	Consensus (ann/smm)	0.75
HLA-A*01:01	1	147	155	9	LSIRGNSNY	Consensus (ann/smm)	1.25
HLA-B*07:02	1	425	434	10	YIKRQGSTPL	Consensus (ann/smm)	1.27
HLA-B*07:02	1	243	252	10	GPFGMSRILL	Consensus (ann/smm)	1.35
HLA-A*01:01	1	191	199	9	VLDMFRTAF	Consensus (ann/smm)	1.6
HLA-A*01:01	1	474	482	9	SSDQSSHHY	Consensus (ann/smm)	1.75

Input sequence

Output

(sorted low-to-high by
percentile rank)

A percentile rank for a peptide is the percentage of randomly sampled peptides scoring better than the peptide.

MHC-I binding prediction – example

tools.iedb.org/mhci/

Individual scores for different methods

Prediction method: IEDB recommended 2.22 | Low Percentile Rank = good binders

[Download result](#) 

Citations

Check to expand the result ☒

Allele	#	Start	End	Length	Peptide	Method used	Percentile Rank	ANN IC50(nM)	ANN rank	SMM IC50(nM)	SMM rank
HLA-A*01:01	1	92	100	9	CSANNSHHY	Consensus (ann/smm)	0.2	25.62	0.09	173.60	0.3
HLA-A*01:01	1	405	413	9	FSDQIEQEA	Consensus (ann/smm)	0.34	121.15	0.27	360.21	0.4
HLA-B*07:02	1	471	480	10	KPHRLTNKGI	Consensus (ann/smm)	0.35	46.84	0.2	112.67	0.5
HLA-A*01:01	1	417	425	9	ITEMLRKDY	Consensus (ann/smm)	0.6	591.06	0.71	426.14	0.5
HLA-A*01:01	1	361	369	9	LRDLMGVPY	Consensus (ann/smm)	0.68	799.14	0.85	421.26	0.5
HLA-A*01:01	1	233	241	9	RTWENHCTY	Consensus (ann/smm)	0.69	552.60	0.68	694.30	0.7
HLA-A*01:01	1	217	225	9	TTWCSQTSY	Consensus (ann/smm)	0.71	604.36	0.72	653.96	0.7
HLA-A*01:01	1	439	447	9	LLMFSTSAY	Consensus (ann/smm)	0.75	724.33	0.8	728.70	0.7
HLA-A*01:01	1	147	155	9	LSIRGNSNY	Consensus (ann/smm)	1.25	3116.42	2.0	448.28	0.5
HLA-B*07:02	1	425	434	10	YIKRQGSTPL	Consensus (ann/smm)	1.27	59.83	0.24	575.20	2.3
HLA-B*07:02	1	243	252	10	GPFGMSRILL	Consensus (ann/smm)	1.35	418.14	1.2	351.41	1.5
HLA-A*01:01	1	191	199	9	VLDMFRTAF	Consensus (ann/smm)	1.6	2586.86	1.8	1457.30	1.4
HLA-A*01:01	1	174	182	9	FSDAQAQAS	Consensus (ann/smm)	1.75	2437.12	1.7	1934.42	1.8

Downloaded prediction results

	A	B	C	D	E	F	G	H	I	J	K	L
1	allele	seq_num	start	end	length	peptide	method	Percentile Rank	ann_ic50	ann_rank	smm_ic50	smm_rank
2	HLA-A*01:01	1	92	100	9	CSANNSHHY	Consensus (ann/smm)	0.2	25.62	0.09	173.6	0.3
3	HLA-A*01:01	1	405	413	9	FSDQIEQEA	Consensus (ann/smm)	0.34	121.15	0.27	360.21	0.4
4	HLA-B*07:02	1	471	480	10	KPHRLTNKGI	Consensus (ann/smm)	0.35	46.84	0.2	112.67	0.5
5	HLA-A*01:01	1	417	425	9	ITEMLRKDY	Consensus (ann/smm)	0.6	591.06	0.71	426.14	0.5
6	HLA-A*01:01	1	361	369	9	LRDLMGVYPY	Consensus (ann/smm)	0.68	799.14	0.85	421.26	0.5
7	HLA-A*01:01	1	233	241	9	RTWENHCTY	Consensus (ann/smm)	0.69	552.6	0.68	694.3	0.7
8	HLA-A*01:01	1	217	225	9	TTWCSQTSY	Consensus (ann/smm)	0.71	604.36	0.72	653.96	0.7
9	HLA-A*01:01	1	439	447	9	LLMFSTSAY	Consensus (ann/smm)	0.75	724.33	0.8	728.7	0.7
10	HLA-A*01:01	1	147	155	9	LSIRGNSNY	Consensus (ann/smm)	1.25	3116.42	2	448.28	0.5
11	HLA-B*07:02	1	425	434	10	YIKRQGSTPL	Consensus (ann/smm)	1.27	59.83	0.24	575.2	2.3
12	HLA-B*07:02	1	243	252	10	GPFGMSRILL	Consensus (ann/smm)	1.35	418.14	1.2	351.41	1.5
13	HLA-A*01:01	1	191	199	9	VLDMFRTAF	Consensus (ann/smm)	1.6	2586.86	1.8	1457.3	1.4
14	HLA-A*01:01	1	174	182	9	FSDAQSAQS	Consensus (ann/smm)	1.75	2437.12	1.7	1934.42	1.8
15	HLA-A*01:01	1	52	60	9	LAGRSCGMY	Consensus (ann/smm)	2.05	4721.07	2.5	1692.58	1.6
16	HLA-A*01:01	1	220	228	9	CSQTSYQYL	Consensus (ann/smm)	2.15	5007.72	2.6	1826.21	1.7
17	HLA-A*01:01	1	219	227	9	WCSQTSYQY	Consensus (ann/smm)	2.2	2051.4	1.6	3009.89	2.8
18	HLA-A*01:01	1	86	94	9	LTMPNACSA	Consensus (ann/smm)	2.25	4423.31	2.4	2215.9	2.1
19	HLA-B*07:02	1	320	329	10	RLIDYNKAAL	Consensus (ann/smm)	2.25	1113.26	2.2	595.42	2.3
20	HLA-B*07:02	1	190	199	10	RVLDMFRTAF	Consensus (ann/smm)	2.4	567.7	1.5	816.24	3.3
21	HLA-A*01:01	1	272	280	9	LSDSSGVEN	Consensus (ann/smm)	2.45	8300.79	3.9	913.17	1
22	HLA-A*01:01	1	369	377	9	YCNYSKFWY	Consensus (ann/smm)	2.45	5677.63	2.9	2145.61	2
23	HLA-A*01:01	1	436	444	9	LMDLLMFST	Consensus (ann/smm)	2.5	3758.17	2.2	3037.74	2.8
24	HLA-B*07:02	1	432	441	10	TPLALMDLLM	Consensus (ann/smm)	2.6	767.22	1.8	854.71	3.4
25	HLA-A*01:01	1	166	174	9	ITIQYNLTF	Consensus (ann/smm)	2.75	8692.54	4	1583.25	1.5
26	HLA-A*01:01	1	364	372	9	LMGVPCYNY	Consensus (ann/smm)	2.75	5142.58	2.7	3009.89	2.8
27	HLA-A*01:01	1	104	112	9	GTSGLELTF	Consensus (ann/smm)	2.8	7192.3	3.4	2374.38	2.2
28	HLA-A*01:01	1	222	230	9	QTSYQYLII	Consensus (ann/smm)	2.9	8442.18	4	1873.05	1.8
29	HLA-A*01:01	1	448	456	9	LVSIFLHLV	Consensus (ann/smm)	2.95	5023.73	2.7	3424.13	3.2
30	HLA-B*07:02	1	262	270	10	PLAETEMAFI	Consensus (ann/smm)	2.95	1227.48	2.4	1062.7	4.4

Selection of “binders”

- Pick peptides below percentile rank 1.0
- Pick peptides below predicted binding affinity of 500 nM
 - $IC_{50} < 50$ nM - high affinity
 - $IC_{50} < 500$ nM - intermediate affinity
 - $IC_{50} < 5000$ nM - low affinity
- Pick top 1% of peptides for each allele/length combination to cover most of immune responses

TepiTool

tools.iedb.org/tepitool/

- New interface to prediction of class I and class II epitope candidates
- Motivation:
 - Make tools more user friendly
 - Provide recommendations as default
 - Provide a set of top peptides as concise results
- In the form of a step-by-step wizard (6 steps)
- Provides recommendations as default values
- Input parameters can be adjusted as desired
- New methods incorporated





Prediction of T cell epitopes

Use of bioinformatics tools to predict peptide binding of selected proteins to MHC class I

Exercise:

1. Identification of MHC class I binding motif and identify potential epitopes of a selected protein

Search the SYFPEITHI, IEDB & NetMHCpan4.0 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**

ΠΙΝΑΚΑΣ ΕΡΓΑΣΙΑΣ

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 τρεις βάσεις δεδομένων (SYFPEITHI, IEDB, NETMHCPan4.0) για τις ίδιες πρωτεΐνες.

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

cDNA	Species	MHC type	Group
EPCAM	HUMAN	HLA-A*01	A
BIRC5	HUMAN	HLA-B*08	B
MTA1	MOUSE	H2-Db	C
NPTN	MOUSE	H2-Kd	D
SPIKE	SARS-CoV-2	HLA-A22	E

1. SYFPEITHI: MHC Ligands and peptide motifs

- Κάντε «Αντιγραφή» της αλληλουχίας FASTA
- Στο δεύτερο παράθυρο ανοίξτε την βάση δεδομένων **SYFPEITHI**

www.syfpeithi.de

2. 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

3. NetMHCpan-4.0: Prediction of peptide MHC class I binding using artificial neural networks (ANN)

- Κάντε «Αντιγραφή» της αλληλουχίας FASTA
- Και προχωρήστε επιλέγοντας αλληλόμορφο, μήκος πεπτιδίου κλπ.

<http://www.cbs.dtu.dk/services/NetMHCpan/>

Αναζήτηση της πρωτεϊνικής αλληλουχίας στο Uniprot

<https://www.uniprot.org>

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB

UniProt Knowledgebase

Swiss-Prot
(558,681)



Manually
annotated and
reviewed.

TrEMBL
(133,507,323)



Automatically
annotated and not
reviewed.

UniRef

Sequence clusters

XXX

Literature citations

XXX

Cross-ref. databases



UniParc

Sequence archive



Supporting data

Taxonomy



Diseases

XXX

Proteomes



Subcellular locations



Keywords



News

Forthcoming changes

Planned changes for UniProt

UniProt release 2018_10

You're not coming in!

UniProt release 2018_09

Tubulin code: a long sought-after player identified

UniProt release 2018_08

Human brain development: slow and steady wins

 News archive

UniProtKB - Q9Y639 (NPTN_HUMAN)

Protein | **Neuroplastin**

Gene | **NPTN**

Organism | *Homo sapiens (Human)*

Status |  Reviewed - Annotation score:  - Experimental evidence at protein level

Function

Probable homophilic and heterophilic cell adhesion molecule involved in long term potentiation at hippocampal excitatory synapses through activation of p38MAPK. May also regulate neurite outgrowth by activating the FGFR1 signaling pathway. May play a role in synaptic plasticity (By similarity). Evidence: By similarity

GO - Molecular function

This isoform has been chosen as the 'canonical' sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.

[« Hide](#)

10	20	30	40	50
MSGSSLPSAL	ALSLLLVSGS	LLPGPGAQON	AGFVKSPMSE	TKLTGDAPFEL
60	70	80	90	100
YCDVVGSPPTP	EIQWWYAEVN	RAESFRQLWD	GARKRRVTVN	TAYGSNGVSV
110	120	130	140	150
LRITRLTLED	SGTYECRASN	DPKRNDLRQN	PSITWIRAQA	TISVLQKPRI
160	170	180	190	200
VTSEEVIIIRD	SPVLPVTLQC	NLTSSSHTLT	YSYWTKNGVE	LSATRKNASN
210	220	230	240	250
MEYRINKPRA	EDSGEYHCVY	HFVSAPKANA	TIEVKAAPDI	TGHKRSENKN
260	270	280	290	300
EGQDATMYCK	SVGYPHPDWI	WRKKENGMPM	DIVNTSGRFF	IINKENYTEL
310	320	330	340	350
NIVNLQITED	PGEYECNATN	AIGSASVVTV	LRVRSHLAPL	WPFLGILAEI
360	370	380	390	
IILVVIIVVY	EKRKRPEVP	DDDEPAGPMK	TNSTNNHKDK	NLRQRNTN

FASTA format

```
>sp|Q9Y639|NPTN_HUMAN Neuroplastin OS=Homo sapiens OX=9606 GN=NPTN PE=1 SV=2
MSGSSLPSALALSLLLVSGLLPGPAAQNAGFVKSPMSETKLTGDAFELYCDVVGSPPTP
EIQWWYAEVNRAESFRQLWDGARKRRVTVNTAYGSNGVSVLRITRLTLEDSTYECRASN
DPKRNDLRQNPSITWIRAQATISVLQKPRIVTSEEVIIRDSPVLPVTLQCNLTSSSHTLT
YSYWTKNGVELSATRKNASNMEYRINKPRAEDSGEYHCVYHFVSAPKANATIEVKAAPDI
TGHKRSENKNEGQDATMYCKSVGYPHPDWIWRKKENGMPMDIVNTSGRFFIINKENYTEL
NIVNLQITEDPGEYECNATNAIGSASVTVLRVRSHLAPLWPFLGILAEIIILVVIIVVY
EKRRKPDEVPPDDEPAGPMKTNSTNNHKDKNLRQRNTN
```

Copy the sequence in FASTA format and insert it in the database (SYFPEITHI, IEDB, netMHCpan-4.0)

www.syfpeithi.de

01/12/2016, 16:03

supported by
[DFG-Sonderforschungsbereich 685](#) and the European Union:
EU BIOMED CT95-1627, BIOTECH CT95-0263,
and EU QLQ-CT-1999-00713



**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.

Last update: August 27th 2012

Institute for Cell Biology
Department
of
Immunology



ADVERTISEMENT

SYFPEITHI Offline Version
Now available. Click here!



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

Detect antigen-specific T Cells EVEN the low affinity ones!
Order your MHC Dextramers at www.immudex.com

Immudex

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.Steinkamp. The probability of being processed and presented is given in order to predict T-cell epitopes.

1. Select MHC type

If you chose "all", 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 as an example.

SYFPEITHI

4. Choose Run to start analysis

Run

Reset

Home



Αποτελέσματα

☒ Advertising on SYFPEITHI is now available. Click here to find out more!

Detect antigen-specific T Cells EVEN the low affinity ones!
Order your MHC Dextramers at www.immudex.com

Immudex



Your search Results

[Return to search conditions](#)

[H2-Kd nonamers](#)

H2-Kd nonamers

Pos

92
341
4
128
14
214
13
149
273
291
322
328
35
42
95
201

1	2	3	4	5	6	7	8	9
A	Y	G	S	N	G	V	S	V
P	F	L	G	I	L	A	E	I
S	S	L	P	G	A	L	A	L
R	Q	N	P	S	I	T	W	I
L	L	L	V	S	G	S	L	L
E	Y	H	C	V	Y	H	F	V
S	L	L	L	V	S	G	S	L
R	I	V	T	S	E	E	V	I
K	E	N	G	V	F	E	E	I
T	N	K	E	N	Y	T	E	L
G	S	A	S	V	S	T	V	L
T	V	L	R	V	R	S	H	L
K	S	P	M	S	E	T	K	L
K	L	T	G	D	A	F	E	L
S	N	G	V	S	V	L	R	I
E	Y	R	I	N	K	P	R	A

score

24
22
21
21
20
20
18
18
18
18
18
17
17
17
17

[go to top](#)



Welcome

The IEDB is a free resource, funded by a contract from the [National Institute of Allergy and Infectious Diseases](#). It offers easy searching of experimental data characterizing antibody and T cell epitopes studied in humans, non-human primates, and other animal species. Epitopes involved in infectious disease, allergy, autoimmunity, and transplant are included.

The IEDB also hosts tools to assist in the prediction and analysis of B cell and T cell epitopes.

[Learn More](#)

Summary Metrics

Peptidic Epitopes	533,957
Non-Peptidic Epitopes	2,720
T Cell Assays	343,098
B Cell Assays	471,916
MHC Ligand Assays	1,072,460

START YOUR SEARCH HERE

Epitope

- ☐ Any Epitopes
☐ Linear Epitope
 Ex: SIINFELK
☐ Discontinuous Epitopes
☐ Non-peptidic Epitopes

Antigen

Organism

Ex: influenza, peanut

Antigen Name

Ex: core, capsid, myosin

Host

- ☐ Any Host
☐ Humans
☐ Mice

Assay

- ☐ Positive Assays Only
☐ T Cell Assays
☐ B Cell Assays
☐ MHC Ligand Assays

Ex: neutralization

[Find](#)

MHC Restriction

- ☐ Any MHC Restriction
☐ MHC Class I
☐ MHC Class II
☐ MHC Nonclassical
☐ Ex: HLA-A*02:01

[Find](#)

Disease

- ☐ Any Disease
☐ Infectious Disease
☐ Allergic Disease

Epitope Analysis Resources

T Cell Epitope Prediction

Scan an antigen sequence for patterns indicative of:

[MHC I Binding](#)

[MHC II Binding](#)

[MHC I Processing \(Pro](#)

[MHC I Immunogenicity](#)

B Cell Epitope Prediction

Predict linear B cell epitope:

[Antigen Sequence Prop](#)

Predict discontinuous B cell antigen structure via:

[Discotope](#)

[EliPro](#)

Epitope Analysis Tools

Analyze epitope sets of:

[Population Coverage](#)

[Conservation Across A](#)

IEDB Analysis Resource

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

MHC-I Binding Predictions

Prediction Method Version

2013-02-22 [Order versions](#)

Specify Sequence(s)

Enter protein sequence(s) in FASTA format
or as whitespace-separated sequences.
([Browse for sequences in NCBI](#))

Or select file containing sequence(s)

no file selected

Choose sequence format

Choose a Prediction Method

Prediction Method

[Help on prediction method selections](#)

Show all the method versions: ☐ [?](#)

Specify what to make binding predictions for

MHC source species

Show only frequently occurring alleles: ☒ [?](#)

Select MHC allele(s)

[Upload allele file](#) [?](#)

Specify Output

Sort peptides by

Show

Output format

MHC-I Binding Predictions

Prediction
Method
Version

2013-02-22 [\[Older versions\]](#)

Specify Sequence(s)

Enter
protein
sequence
(s) in
FASTA
format
[\[Browse
for
sequences
in NCBI\]](#)

```
MSGSLLPGALALSLLLVSGSLLPGPGAAQNAAGFVKSPMSETKLTGDAFELYCDVVGSPTP  
EIQWVYAEVNRASFRQLWDGARKRRTVNTAYGSGNGSVLRITRLTLEDSTGYECRASN  
DPKRNDLRQNPSTIWIRAQATISVLQKPRIVTSEEVIRESLLPVTLCNLTSSTHTLMY  
SYWTRNGVELTATRKNASINMEYRINKPRAEDSGEYHCVYHFVSAPKANATIEVKAAPDIT  
GHRSENKNEGQDAMMYCKSVGYPHPEWIWRKKENGVFEEISNSSGRFFITNKENYTELS  
IVNLQITEDPGEYECNATNSIGSASVSTVLRVRSHLAPLWFFLGILAEIIILVVIIVVYE  
KRRRDEVPDDDEPAGPMKTNSTNNHKDKNLRQRNTN
```

Or select
file
containing
sequence
(s)

Αναζήτηση...

Choose
sequence
format

auto detect format

Choose a Prediction Method

Prediction
Method

IEDB recommended

[Help on prediction method selections](#)

Specify what to make binding predictions for

MHC
source
species

human

Show only
frequently
occurring
alleles:

☒ [?](#)

Select
MHC
allele(s)

[Select
HLA allele
reference
set](#) ☐

Allele	Length
H-2-Kd	9 ?
H-2-Kd	10 ?
H-2-Kd	8 ?

[Upload allele file](#) [?](#)

Specify Output

Sort
peptides
by

Percentile Rank

Show

All predictions

Output
format

XHTML table

Submit

Reset

MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	sequence 1	MSGSSSLPGALALSLLLVSGSLLPGPGAAQNAGFVKSPMSETKLTGDAFEL YCDVVGSPTEIQWWYAEVNRAESFRQLWDGARKRRVTVNTAYGSNGVSV LRITRLTLEDSTGYECRASNDPKRNDLRQNPSITWIRAQATISVLQKPRI VTSEEVIIRESLLPVTLQCNLTSSSHTLMYSYWTRNGVELTATRKNASNM EYRINKPRAEDSGEYHCVYHFVSAPKANATIEVKAAPDITGHKRSENKNE GQDAMMYCKSVGYPHPEWIWRKKENGVFEEISNSSGRFFITNKENYTELS IVNLQITEDPGEYECNATNSIGSASVSTVLRVRSHLAPLWPFILGILAEII ILVVIIVVYEKRKRPEVDPDDDEPAGPMKTNSTNNHKDKNLRQRNTN

Prediction method: IEDB recommended | Low percentile_rank = good binders

[Download result](#) 

Citations

Check to expanded the result: ☐

Allele	#	Start	End	Length	Peptide	Method used	Percentile rank
H-2-Kd	1	92	101	10	AYGSNGVSVL	Consensus (ann/smm)	0.65
H-2-Kd	1	312	321	10	EYECNATNSI	Consensus (ann/smm)	0.65
H-2-Kd	1	181	190	10	SYWTRNGVEL	Consensus (ann/smm)	0.85
H-2-Kd	1	214	221	8	EYHCVYHF	Consensus (ann/smm)	0.9
H-2-Kd	1	328	336	9	TVLRVRSHL	Consensus (ann/comblib_sidney2008/smm)	1.3
H-2-Kd	1	320	329	10	SIGSASVSTV	Consensus (ann/smm)	1.35



www.cbs.dtu.dk/services/NetMHCpan/

NetMHCpan 4.0 Server

01/12/2018, 10:02

 **DTU Bioinformatics**
Department of Bio and Health Informatics
[Home](#)

NetMHCpan-4.0

NetMHCpan 4.0 Server

Prediction of peptide-MHC class I binding using artificial neural networks (ANNs).

New In this version: the method is trained on naturally eluted ligands AND on binding affinity data. It returns two properties: either the likelihood of a peptide becoming a natural ligands, or the predicted binding affinity.

View the [version history](#) of this server. All previous versions are available online, for comparison and reference.

NetMHCpan server predicts binding of peptides to **any MHC molecule of known sequence** using artificial neural networks (ANNs). The method is trained on a combinatino of more than 180,000 quantitative binding data and MS derived MHC eluted ligands. The binding affinity data covers 172 MHC molecules from human (HLA-A, B, C, E), mouse (H-2), cattle (BoLA), primates (Patr, Mamu, Gogo) and swine (SLA). The MS eluted ligand data covers 55 HLA and mouse allelee. Furthermore, the user can obtain redictions to the any custom MHC class I molecule by uploading a full length MHC protein sequence.

Predictions can be made for peptides of any length.

The project is a collaboration between CBS, [ISIM](#), and [LIAI](#).

[Instructions](#)

[Output format](#)

[Motif viewer](#)

[Article abstract](#)

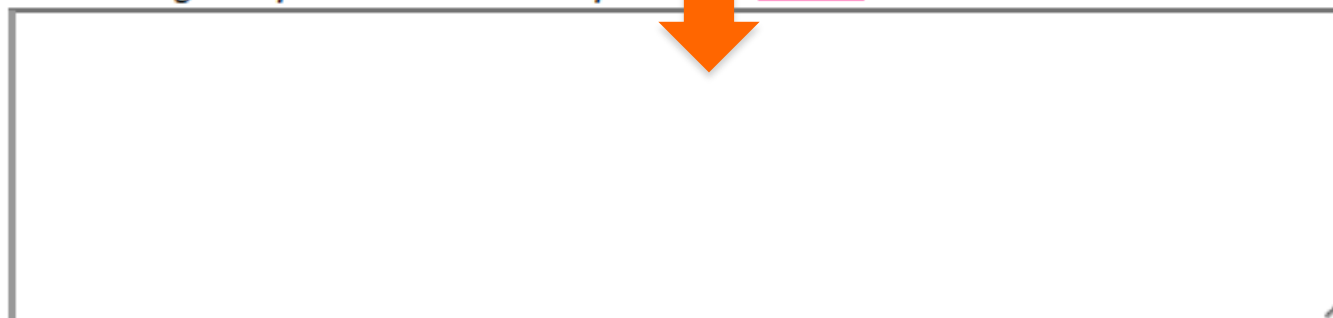
SUBMISSION

SUBMISSION

Hover the mouse cursor over the  symbol for a short description of the options

Type of Input 

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:

no file selected

Peptide length (you may select multiple lengths): 

8mer peptides
9mer peptides
10mer peptides
11mer peptides

Select species/loci 

Select Allele (max 20 per submission) 

NetMHCpan version 4.0

Tmpdir made /usr/opt/www/webface/tmp/server/netmhpcan/5C0241FB000031EB936C79F8/netMHCpanav4can

Input is in FSA format

Peptide length 9

Make Eluted ligand likelihood predictions

HLA-A02:01 : Distance to training data 0.000 (using nearest neighbor HLA-A02:01)

Rank Threshold for Strong binding peptides 0.500

Rank Threshold for Weak binding peptides 2.000

Pos	HLA	Peptide	Core	Of	Gp	Gl	Ip	Il	Icore	Identity	Score	%Rank	Binding Level
1	HLA-A*02:01	MSGSSLPSA	MSGSSLPSA	0	0	0	0	0	MSGSSLPSA	Sequence	0.0037910	13.0474	
2	HLA-A*02:01	SGSSLPSAL	SGSSLPSAL	0	0	0	0	0	SGSSLPSAL	Sequence	0.0016820	17.9111	
3	HLA-A*02:01	GSSLPSALA	GSSLPSALA	0	0	0	0	0	GSSLPSALA	Sequence	0.0023990	15.5985	
4	HLA-A*02:01	SSLPSALAL	SSLPSALAL	0	0	0	0	0	SSLPSALAL	Sequence	0.0310230	4.9407	
5	HLA-A*02:01	SLPSALALS	SLPSALALS	0	0	0	0	0	SLPSALALS	Sequence	0.0428680	4.1732	
6	HLA-A*02:01	LPSALALSL	LPSALALSL	0	0	0	0	0	LPSALALSL	Sequence	0.0014260	19.0494	
7	HLA-A*02:01	PSALALSLL	PSALALSLL	0	0	0	0	0	PSALALSLL	Sequence	0.0003600	32.2069	
8	HLA-A*02:01	SALALSLLL	SALALSLLL	0	0	0	0	0	SALALSLLL	Sequence	0.0368040	4.5033	
9	HLA-A*02:01	ALALSLLLV	ALALSLLLV	0	0	0	0	0	ALALSLLLV	Sequence	0.6613280	0.2777	<= SB
10	HLA-A*02:01	LALSLLLV	LALSLLLV	0	0	0	0	0	LALSLLLV	Sequence	0.0002880	34.8462	
11	HLA-A*02:01	ALSLLLVSG	ALSLLLVSG	0	0	0	0	0	ALSLLLVSG	Sequence	0.0106590	8.4146	
12	HLA-A*02:01	LSLLLVSGS	LSLLLVSGS	0	0	0	0	0	LSLLLVSGS	Sequence	0.0000650	56.0526	
13	HLA-A*02:01	SLLLVSGSL	SLLLVSGSL	0	0	0	0	0	SLLLVSGSL	Sequence	0.1791990	1.6363	<= WB
14	HLA-A*02:01	LLLVSGSLL	LLLVSGSLL	0	0	0	0	0	LLLVSGSLL	Sequence	0.1702140	1.7099	<= WB
15	HLA-A*02:01	LLVSGSLLP	LLVSGSLLP	0	0	0	0	0	LLVSGSLLP	Sequence	0.0149230	7.1649	
16	HLA-A*02:01	LVSGSLLPG	LVSGSLLPG	0	0	0	0	0	LVSGSLLPG	Sequence	0.0001890	40.2857	
17	HLA-A*02:01	VSGSLLPGP	VSGSLLPGP	0	0	0	0	0	VSGSLLPGP	Sequence	0.0000850	52.0370	
18	HLA-A*02:01	SGSLLPGPG	SGSLLPGPG	0	0	0	0	0	SGSLLPGPG	Sequence	0.0000060	88.3333	
19	HLA-A*02:01	GSLLPGPAA	GSLLPGPAA	0	0	0	0	0	GSLLPGPAA	Sequence	0.0060090	10.8169	
20	HLA-A*02:01	SLLPGPGAA	SLLPGPGAA	0	0	0	0	0	SLLPGPGAA	Sequence	0.5013020	0.5227	<= WB
21	HLA-A*02:01	LLPGPGAAQ	LLPGPGAAQ	0	0	0	0	0	LLPGPGAAQ	Sequence	0.0013720	19.3462	
22	HLA-A*02:01	LLPGPGAAQ	LLPGPGAAQ	0	0	0	0	0	LLPGPGAAQ	Sequence	0.0001370	76.0000	

ΠΙΝΑΚΑΣ ΕΡΓΑΣΙΑΣ

ΓΟΝΙΔΙΟ	ΠΡΟΕΛΕΥΣΗ	ΜΗC	ΟΜΑΔΑ
EPCAM	HUMAN	HLA-A*01	A
BIRC5	HUMAN	HLA-B*08	B
MTA1	MOUSE	H2-Db	Γ
NPTN	MOUSE	H2-Kd	Δ

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