MSc course BIOINFORMATICS

Predicting Peptide Binding to

Major Histocompatibility Complex (MHC) molecules

ImmunoInformatics – Computational Immunology

Professor Katerina Chlichlia

Email: achlichl@mbg.duth.gr

Τηλ. 25510-30630

26.11.2023

- Bioinformatics has broad applicability to Immunology → IMMUNOINFORMATICS

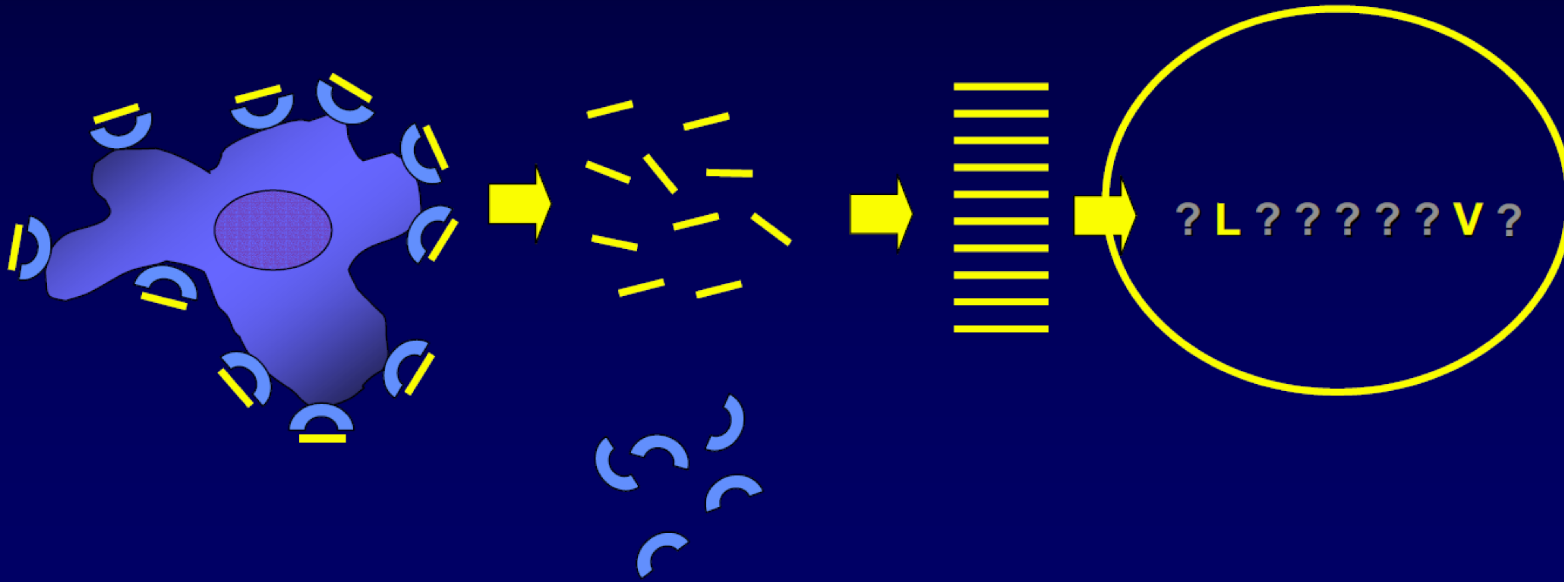
- Development of in-silico models of entire systems – towards a virtual immune system

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

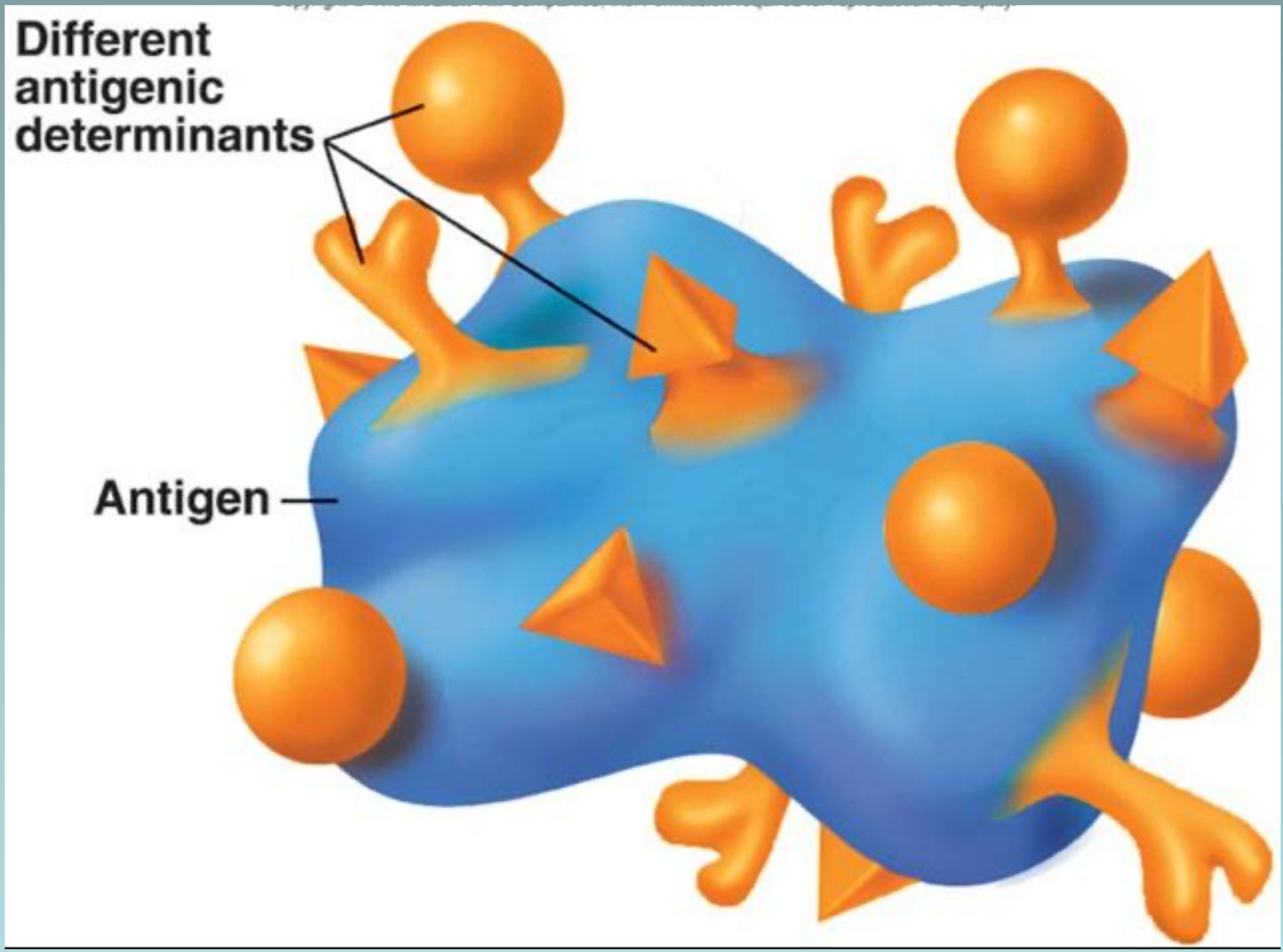
Overview of the lecture

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

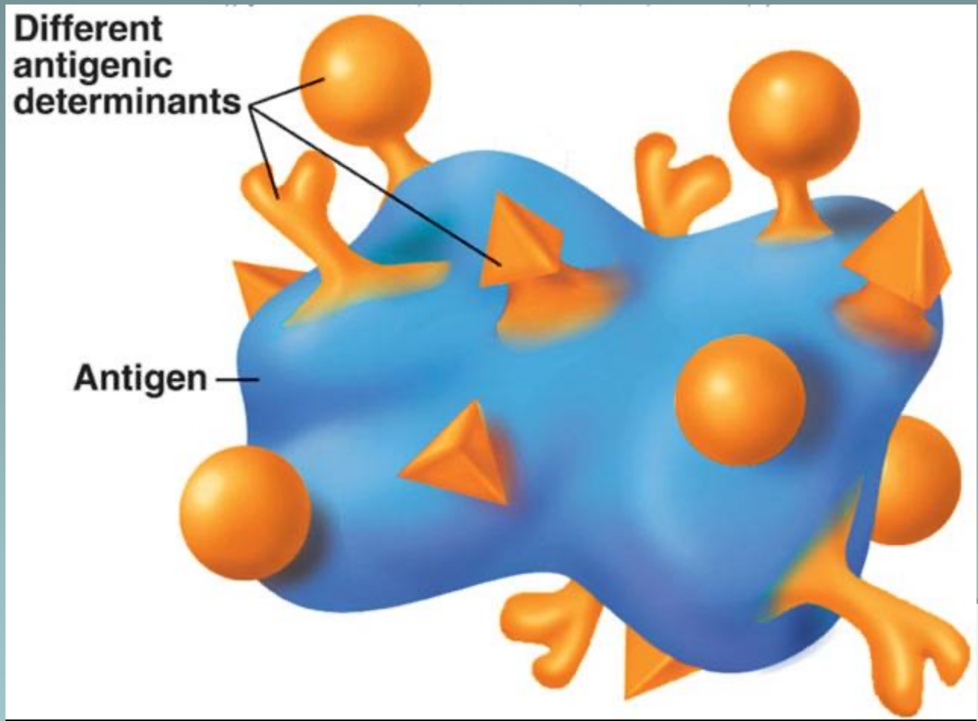
search for epitopes



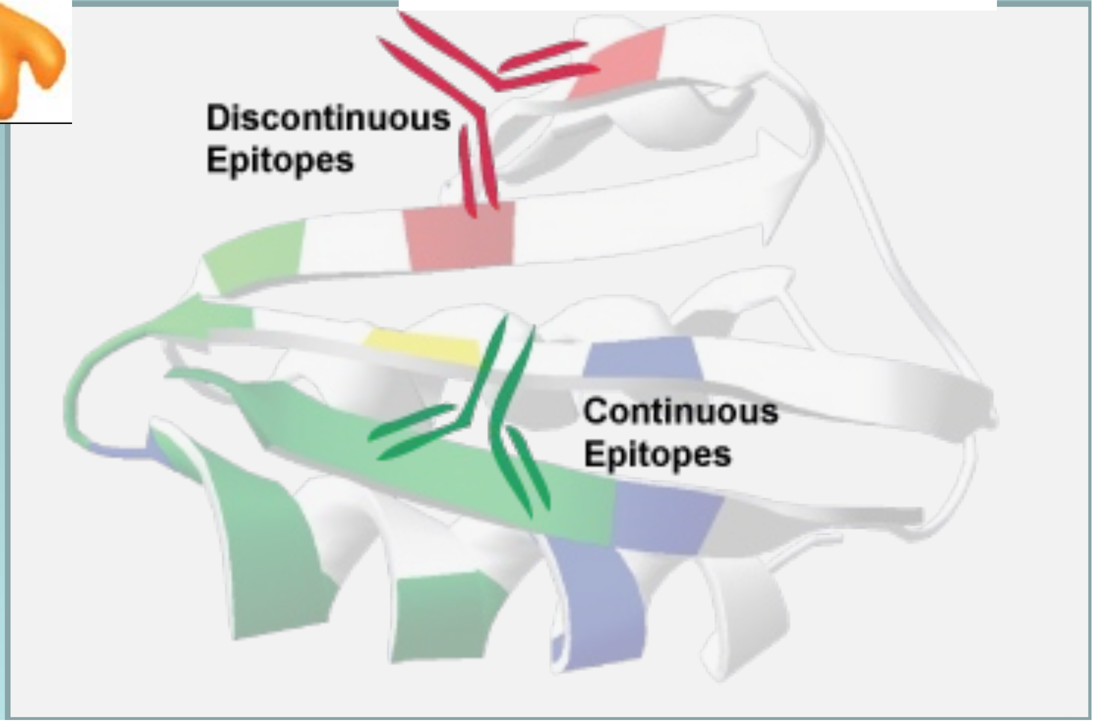
EPITOPE = antigenic determinant



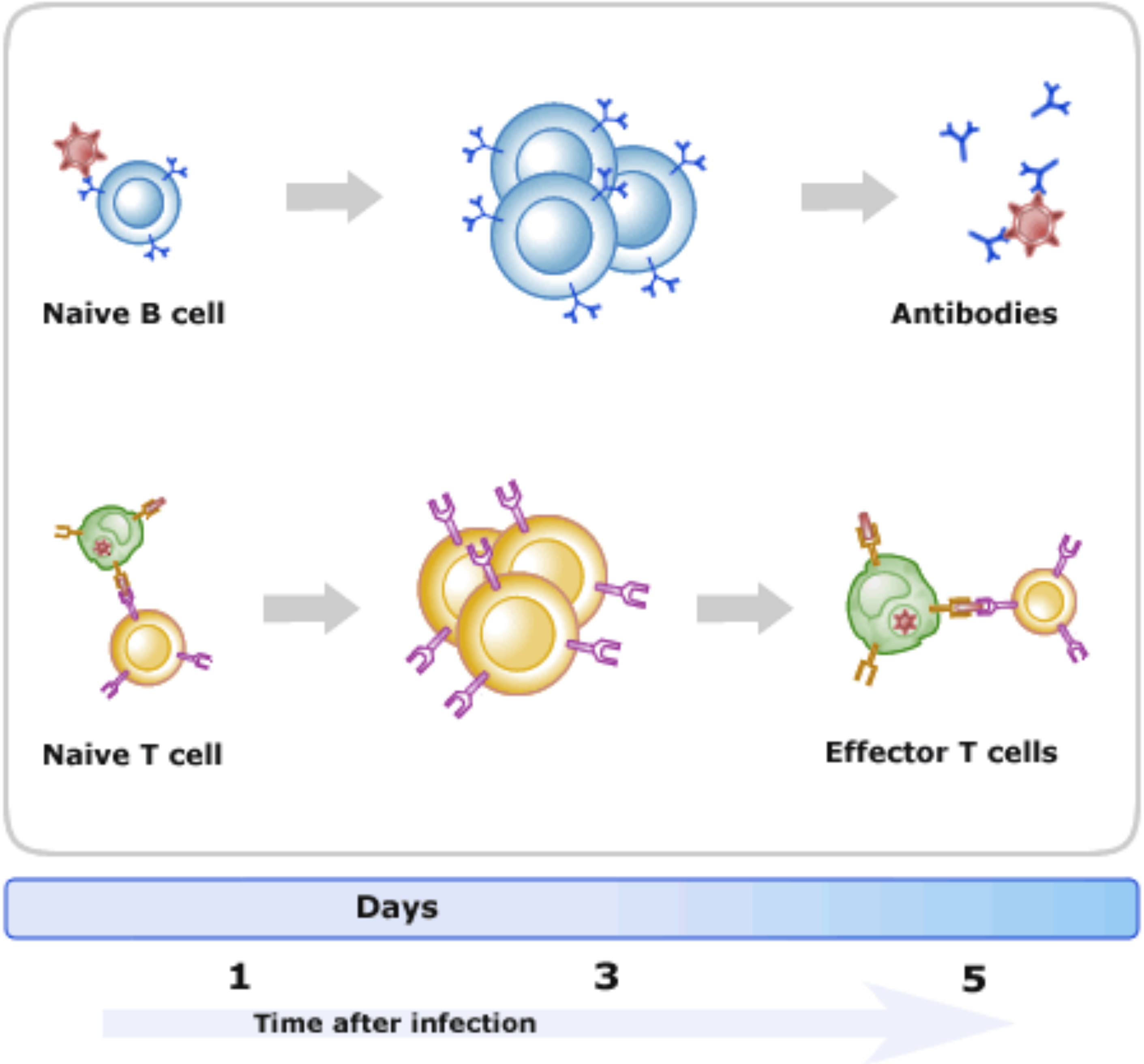
EPITOPE = antigenic determinant



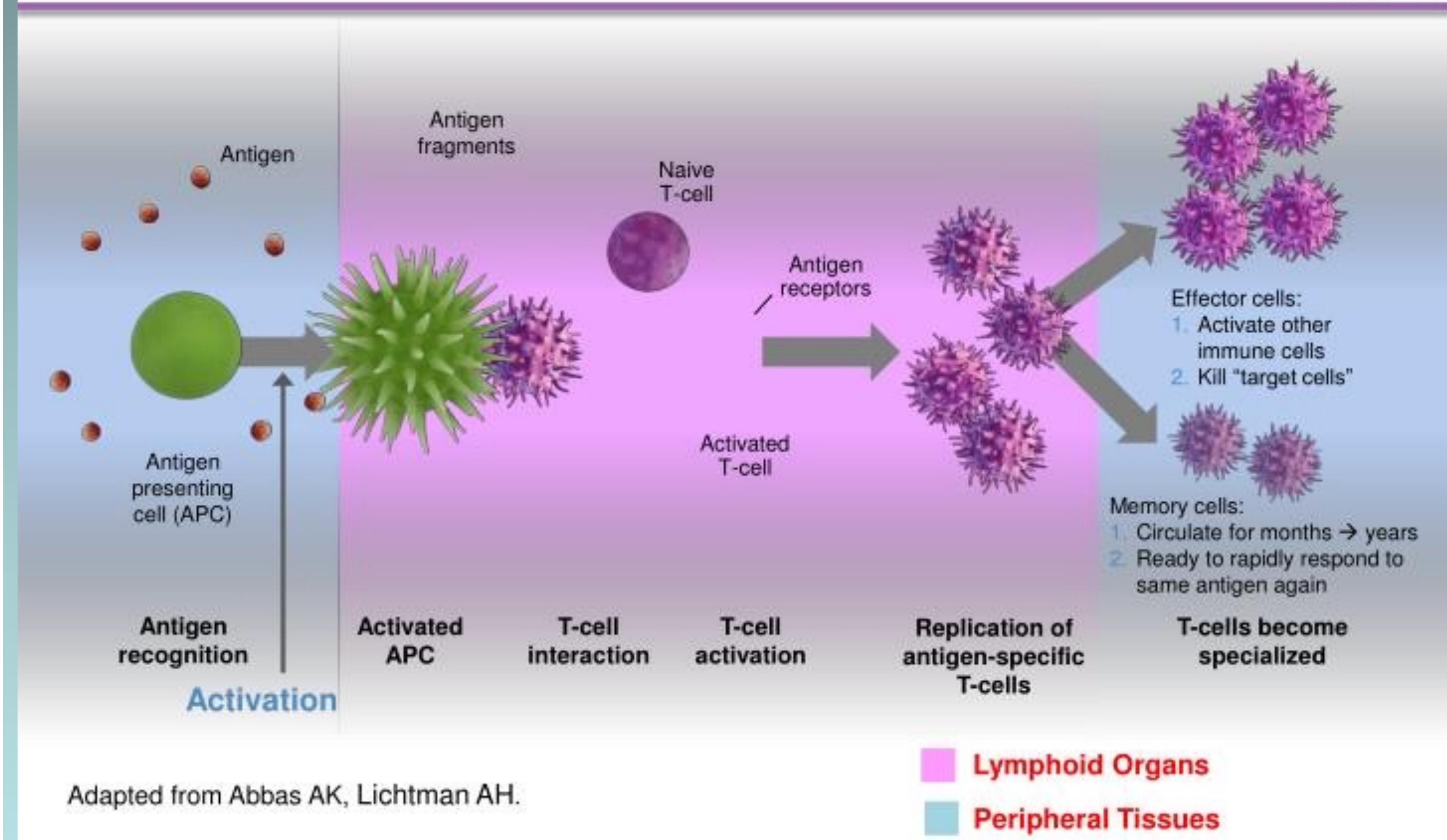
2 types of epitopes



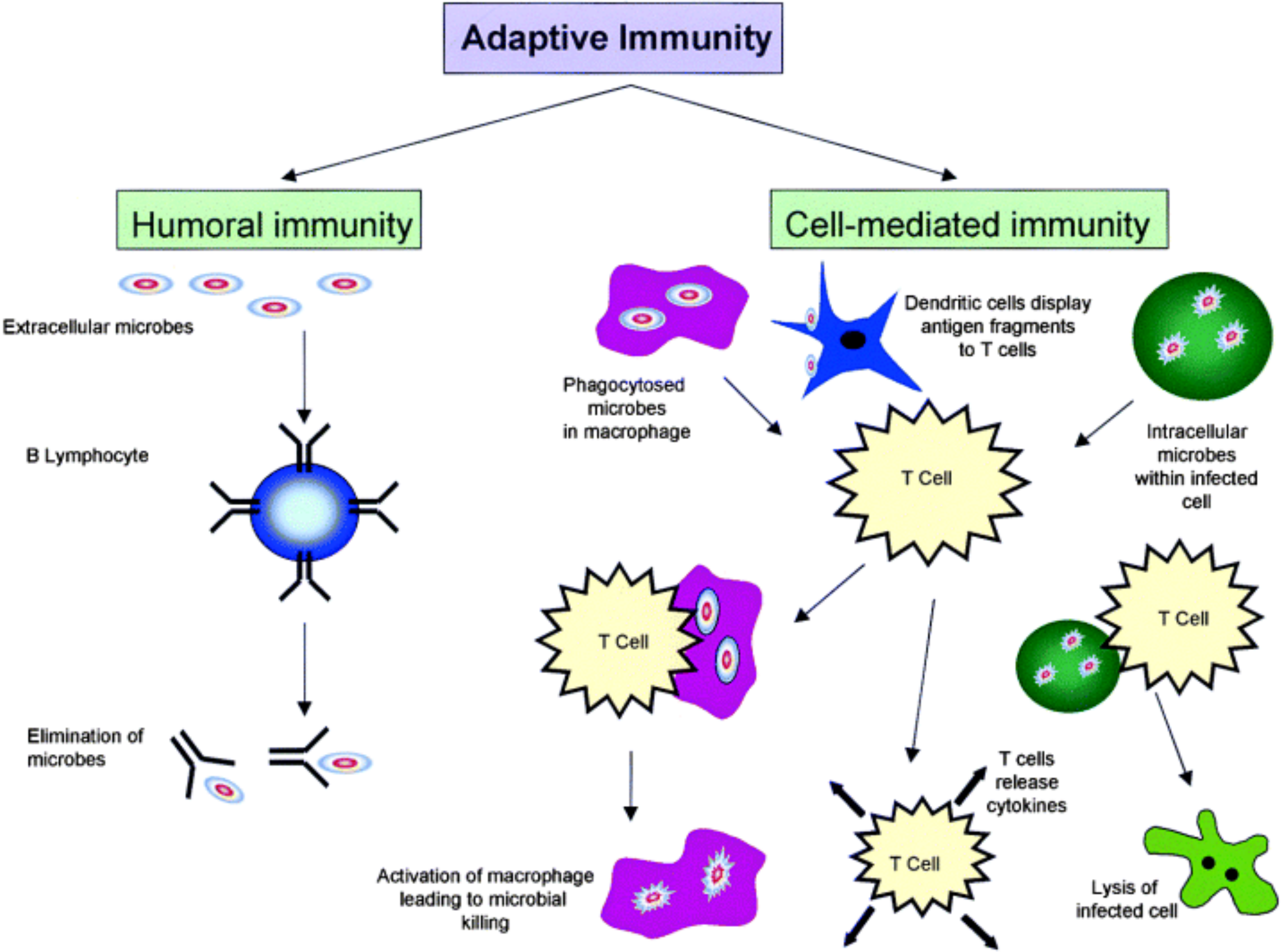
Adaptive Immunity



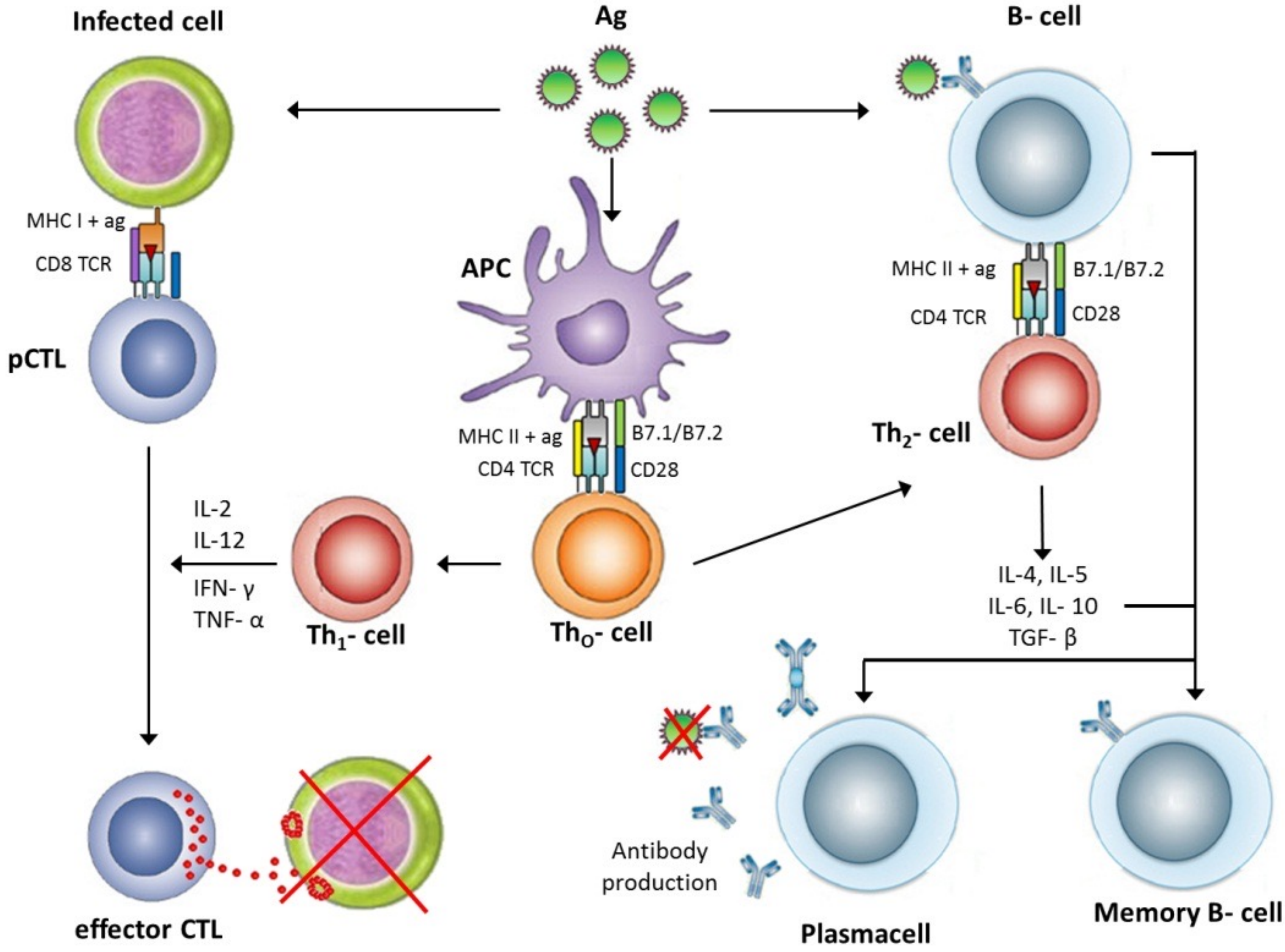
Initiation of Immune Response: Key Components



Adapted from Abbas AK, Lichtman AH.

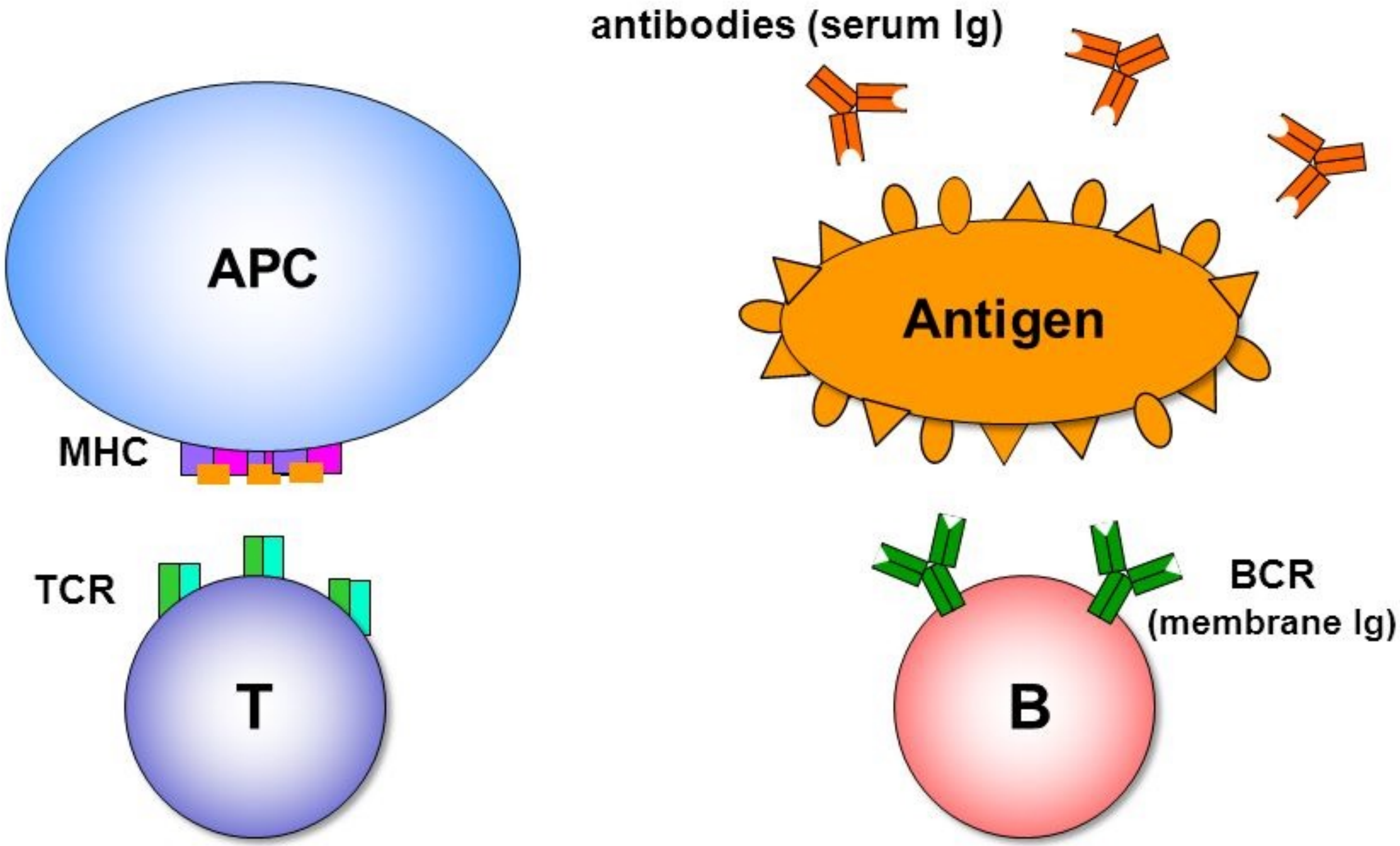


Adaptive Immunity
Humoral immunity
Cell-mediated immunity
Extracellular microbes
B Lymphocyte
Elimination of microbes
Phagocytosed microbes in macrophage
Dendritic cells display antigen fragments to T cells
Intracellular microbes within infected cell
T Cell
T Cell
T Cell
T Cell
T cells release cytokines
Activation of macrophage leading to microbial killing
Lysis of infected cell



Infected cell
Ag
B- cell
MHC I + ag
CD8 TCR
pCTL
APC
MHC II + ag
B7.1/B7.2
CD4 TCR
CD28
MHC II + ag
B7.1/B7.2
CD4 TCR
CD28
Th2- cell
IL-2
IL-12
IFN- γ
TNF- α
Th1- cell
Tho- cell
IL-4, IL-5
IL-6, IL- 10
TGF- β
Antibody production
effector CTL
Plasmacell
Memory B- cell

ANTIGEN RECOGNITION BY LYMPHOCYTES



B cells recognise native antigens
T cells recognise processed antigens

T cell recognition of antigen peptide

T-cell recognition of antigen

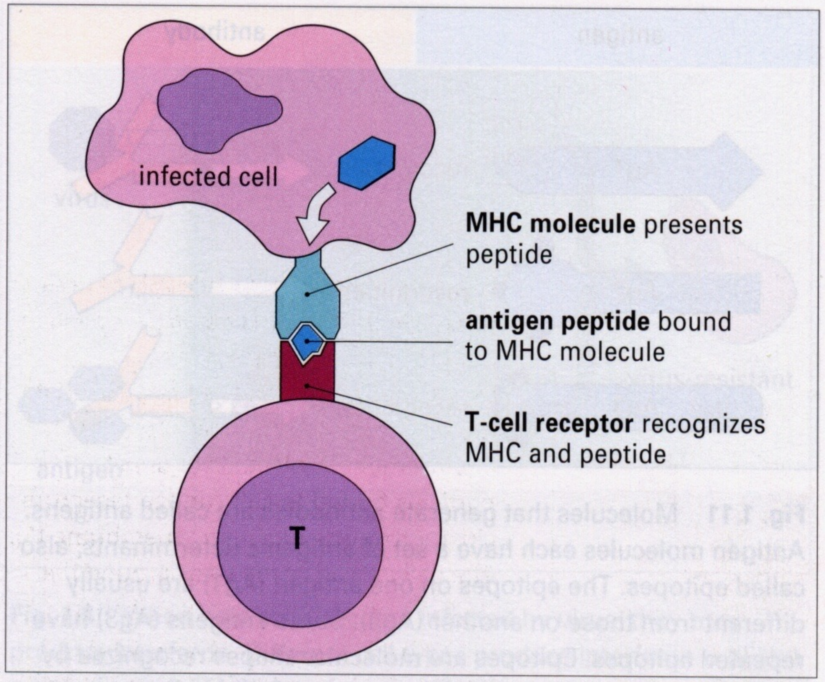
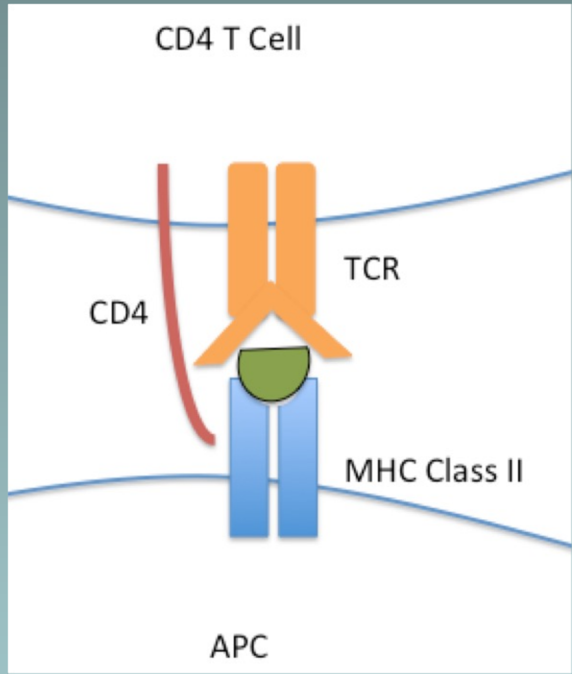
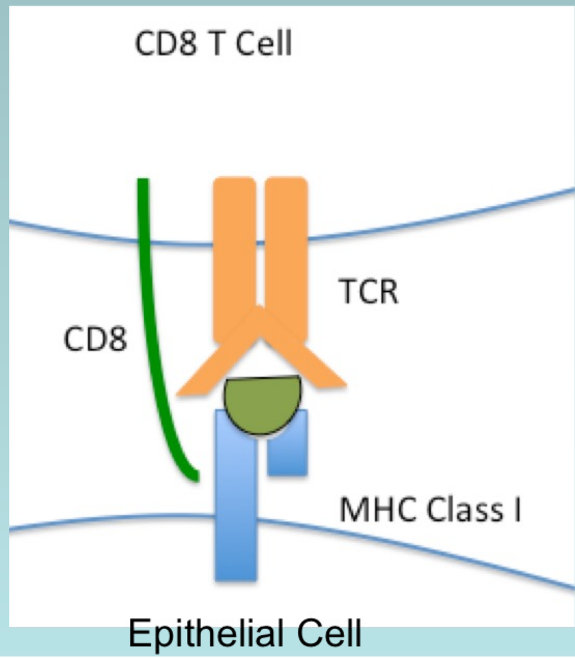
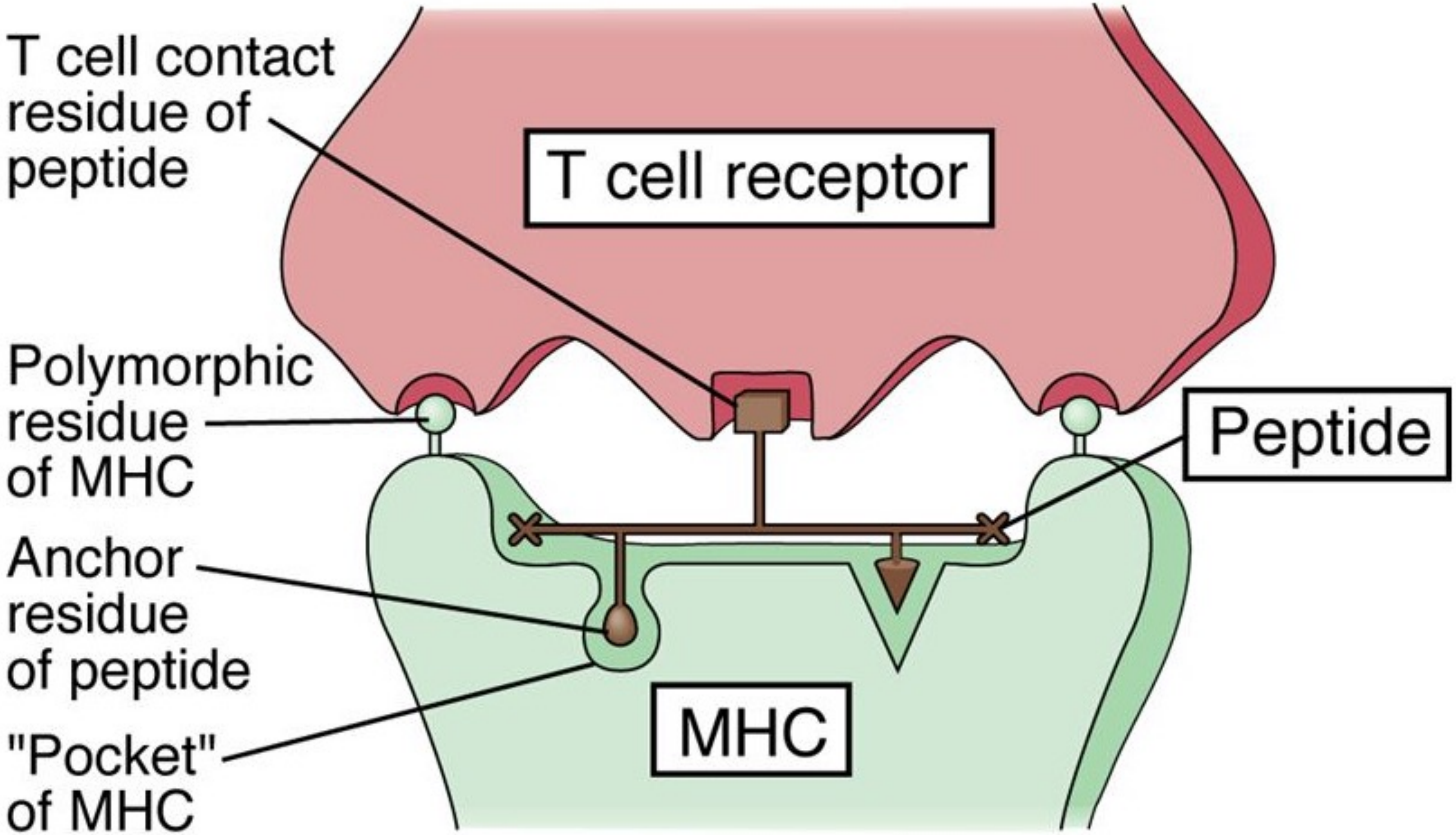


Fig. 1.12 T cells recognize antigens that originate within other cells, such as viral peptides from infected cells. They do this by binding specifically to antigenic peptides presented on the surface of the infected cells by molecules encoded by the major histocompatibility complex (MHC molecules). The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

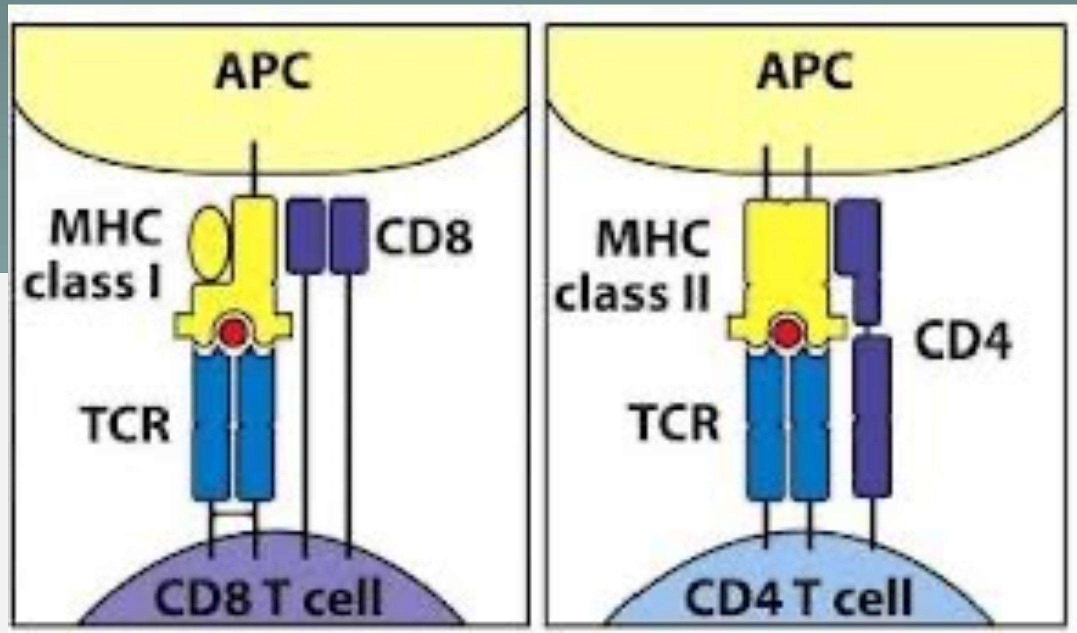
These **peptides** bind to MHC molecules and the MHC-peptide complex is then transported to the cell membrane from where it elicits an immune response via **T-cell binding**.

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 / Tc

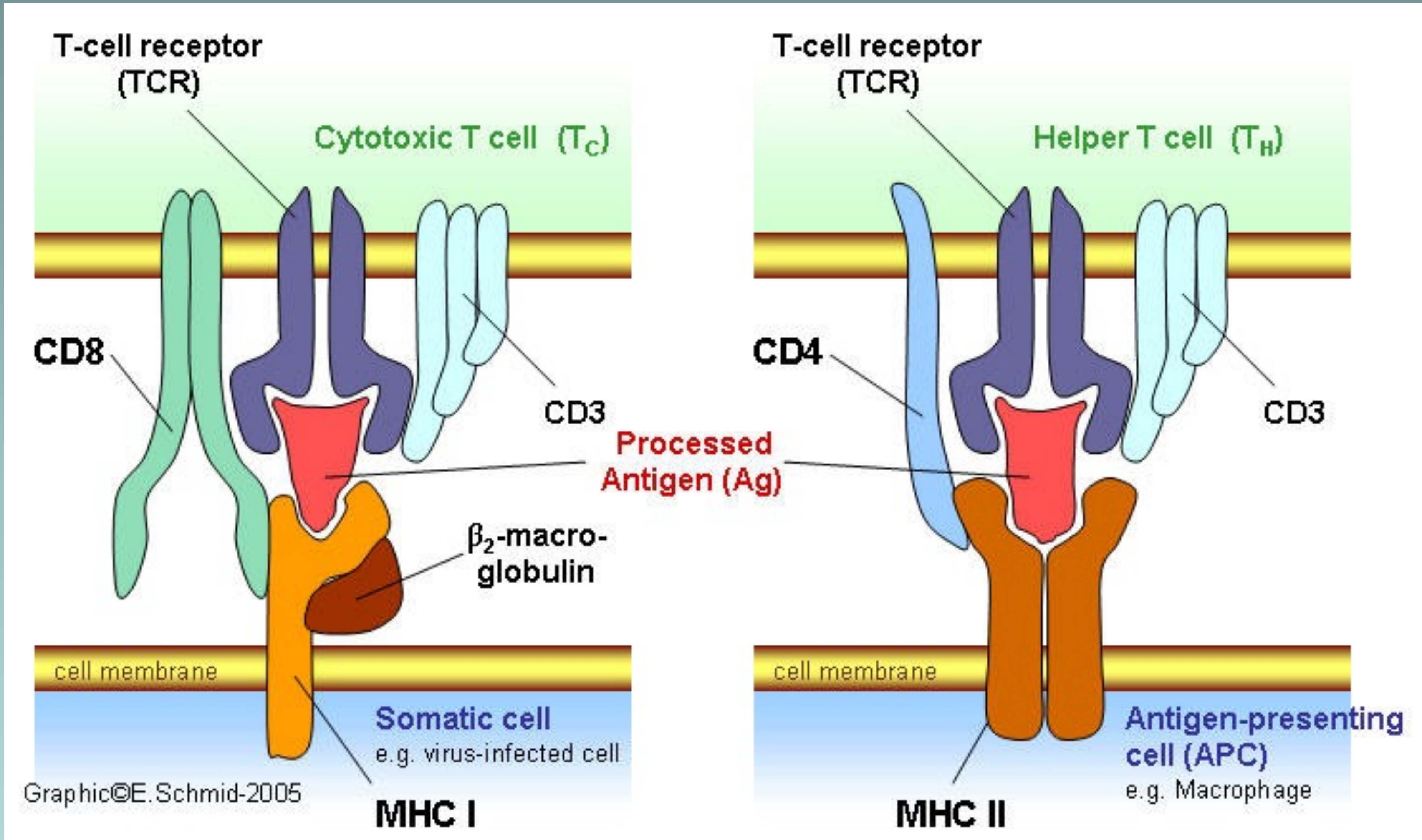
peptides in the cytoplasm

endogenous proteins

MHC τάξης II → CD4 / T_H

peptides in vesicles

exogenous proteins



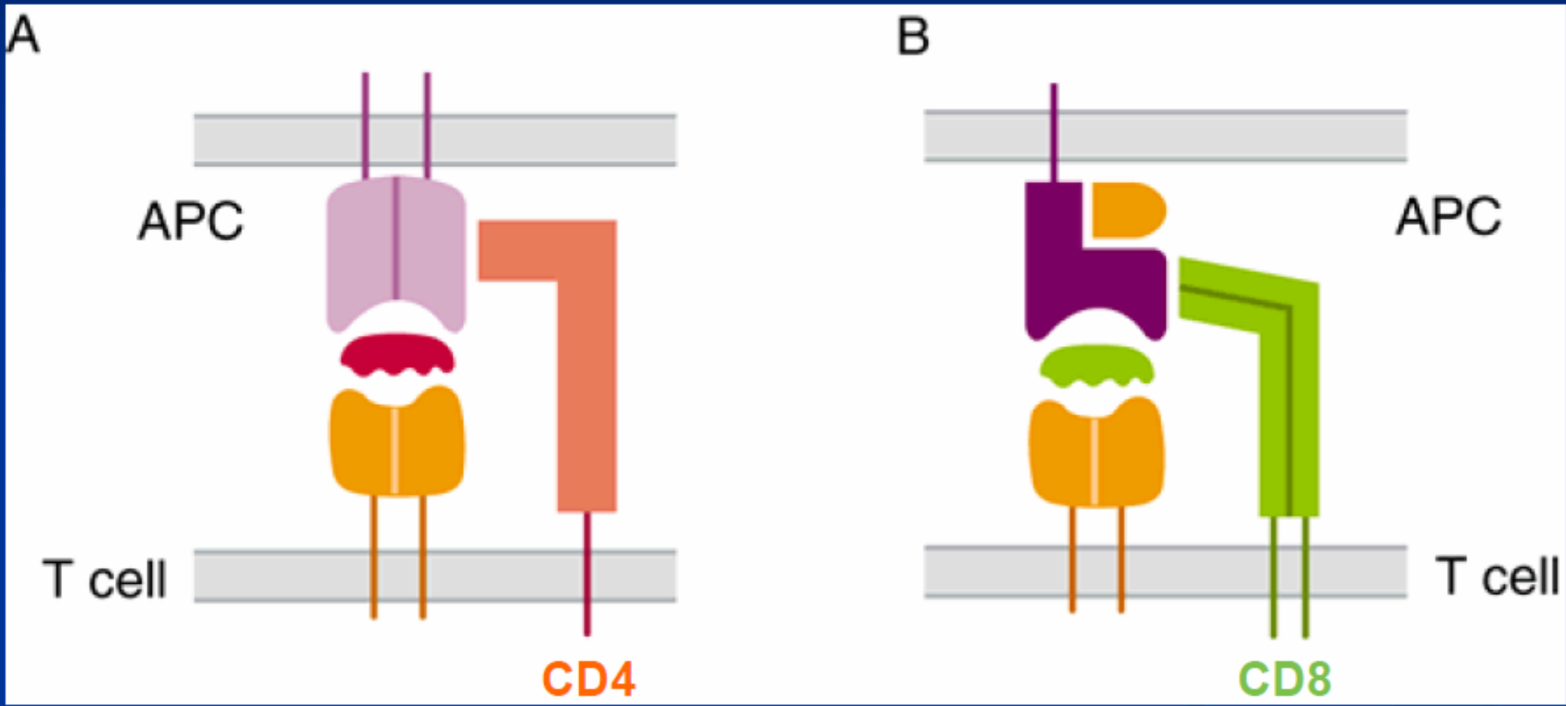
T-cell receptor (TCR)
Cytotoxic T cell (T_C)
CD8
CD3
Processed Antigen (Ag)
β_2-macro-globulin
cell membrane
Somatic cell
e.g. virus-infected cell
Graphic©E.Schmid-2005
MHC I
T-cell receptor (TCR)
Helper T cell (T_H)
CD4
CD3
cell membrane
Antigen-presenting cell (APC)
e.g. Macrophage
MHC II

εξωγενή Ag

ενδογενή Ag

MHC τάξης II

MHC τάξης I



HELP

KILL

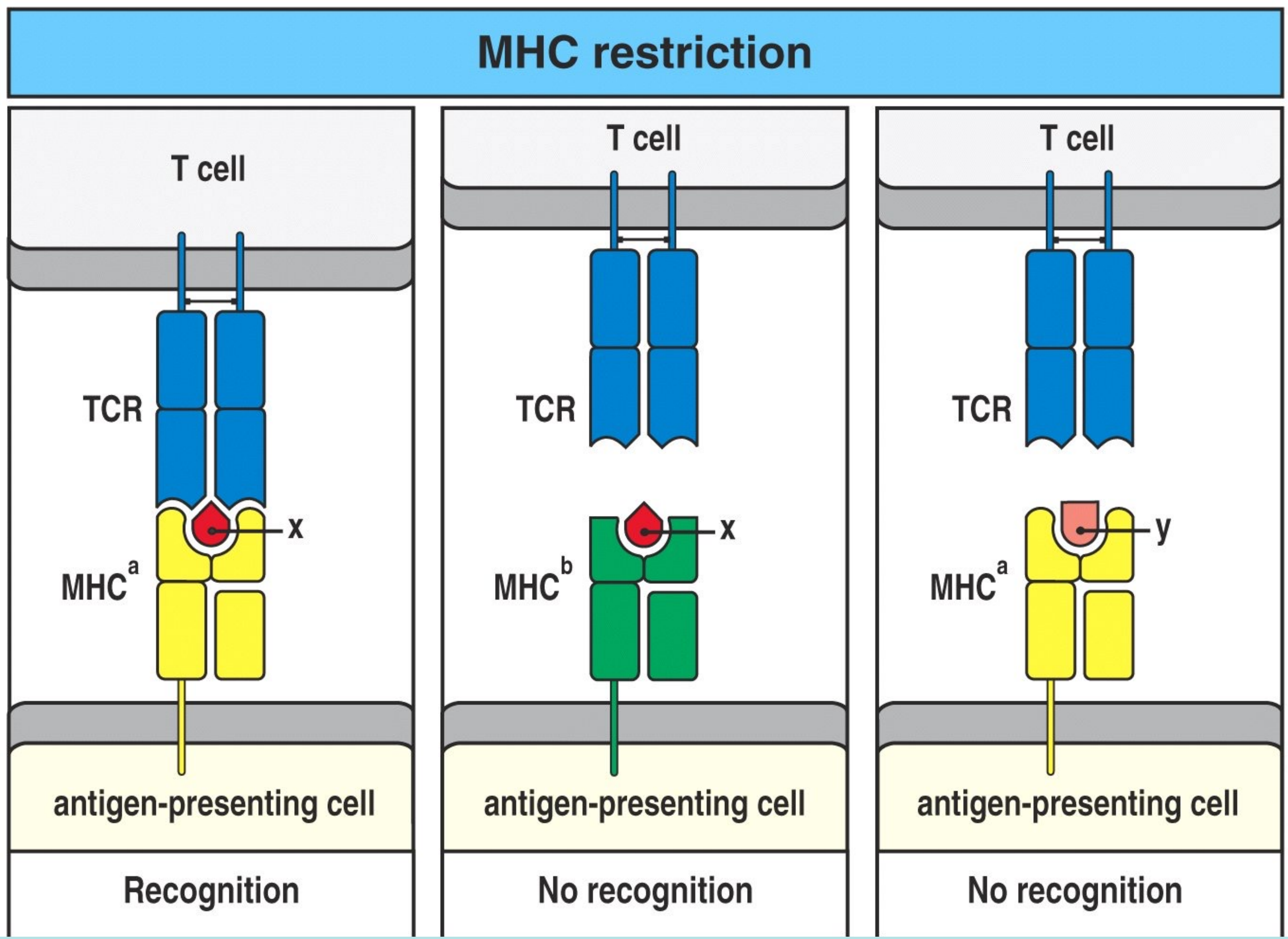
Antigen processing and presentation to T lymphocytes

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

MHC restriction of cytotoxic T cells

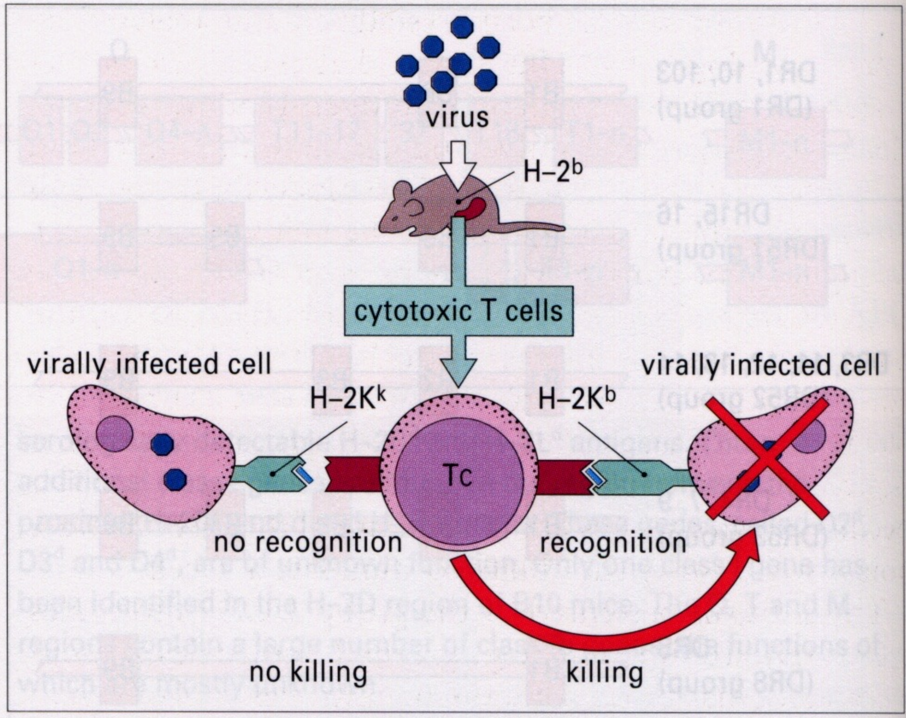


Fig. 5.19 A mouse of the H-2^b haplotype is primed with virus and the Tc cells thus generated are isolated and tested for their ability to kill H-2^b and H-2^k cells infected with the same virus. The Tc cells kill H-2^b, but not H-2^k cells. In this instance, it is the H-2K class I gene product which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

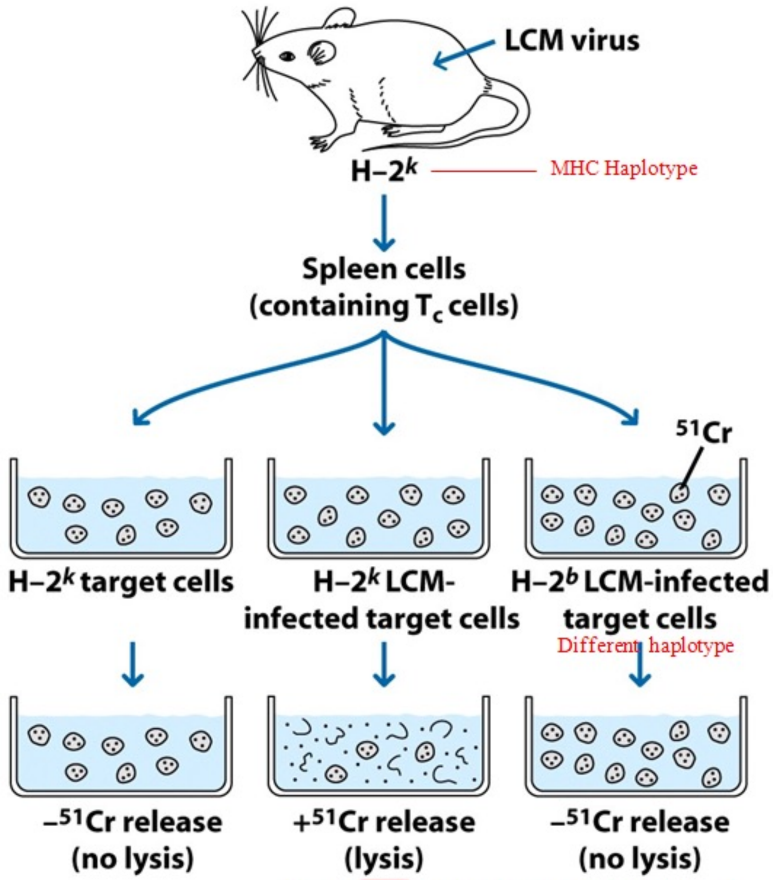


Figure 8-15
Kuby IMMUNOLOGY, Sixth Edition

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

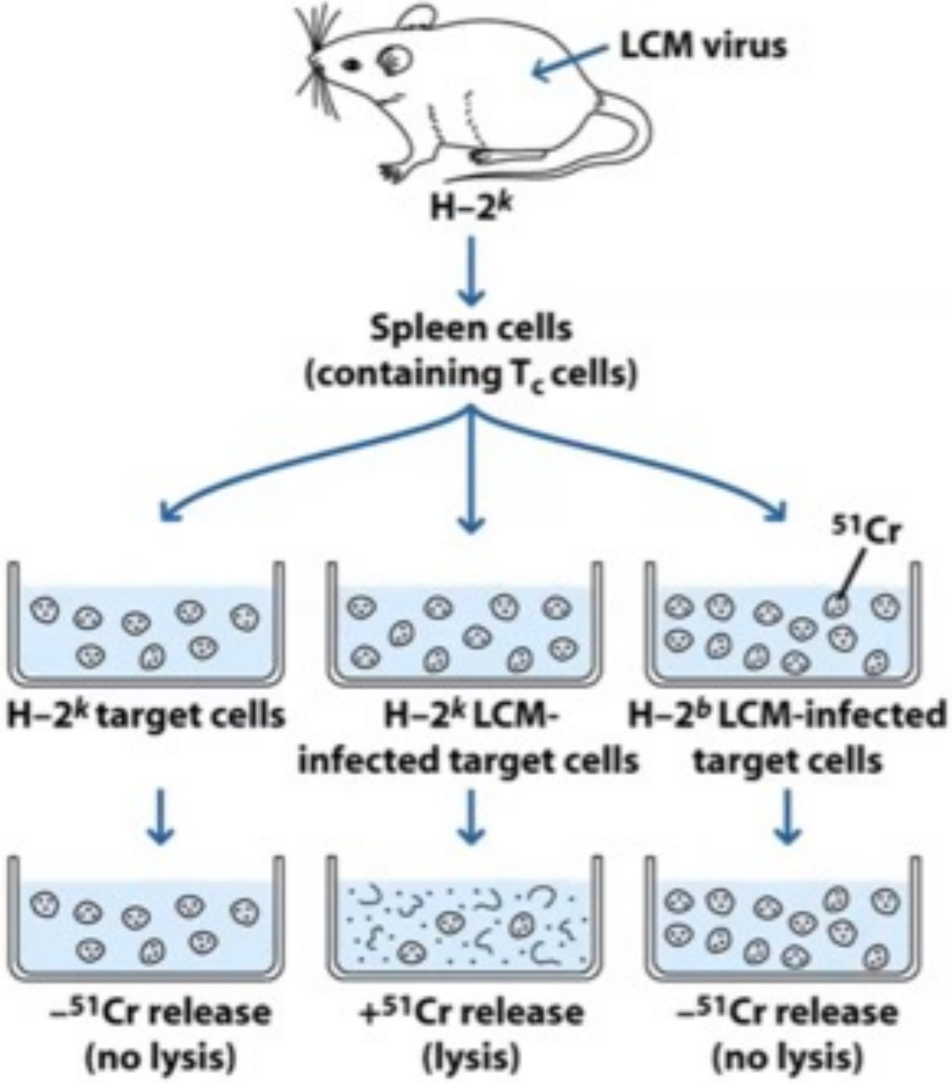


Figure 8-15
Kuby IMMUNOLOGY, Sixth Edition

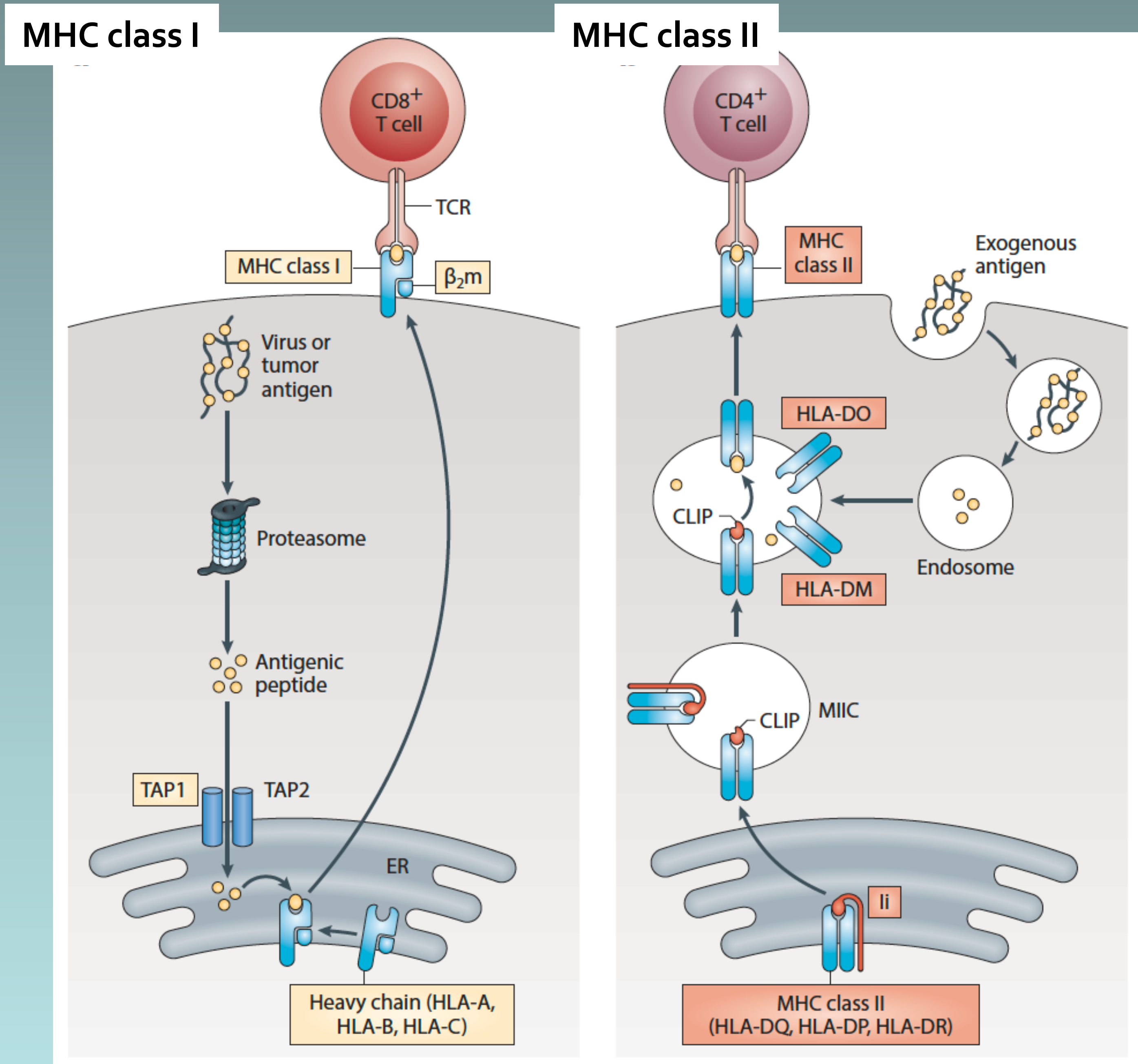
The concept of MHC restriction- Experimental evidence that antigen recognition by T cells depends on the presence of specific SELF MHC molecules in APCs is known as a phenomenon of MHC restriction.

$CD8^+$ T cells are MHC class I restricted.

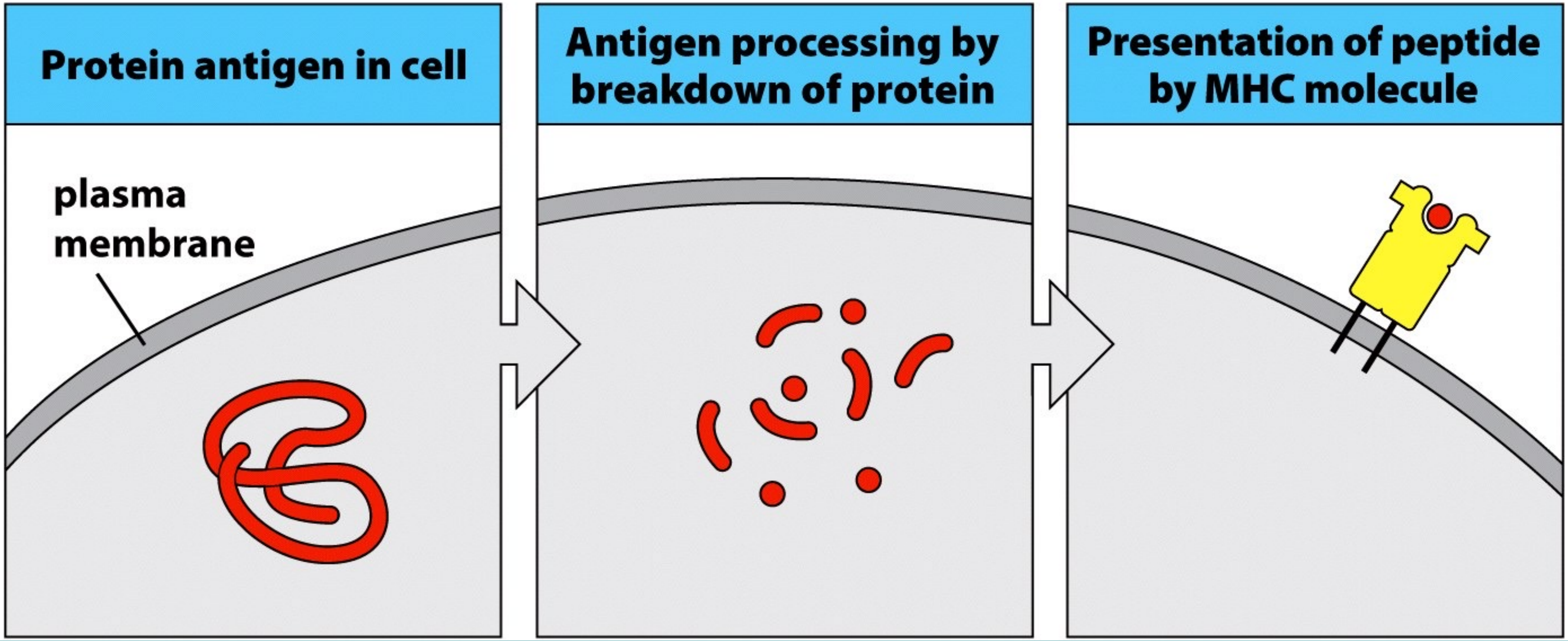
Doherty and Zinkernagel (1975, JEM) Nobel prize in 1996

- 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)

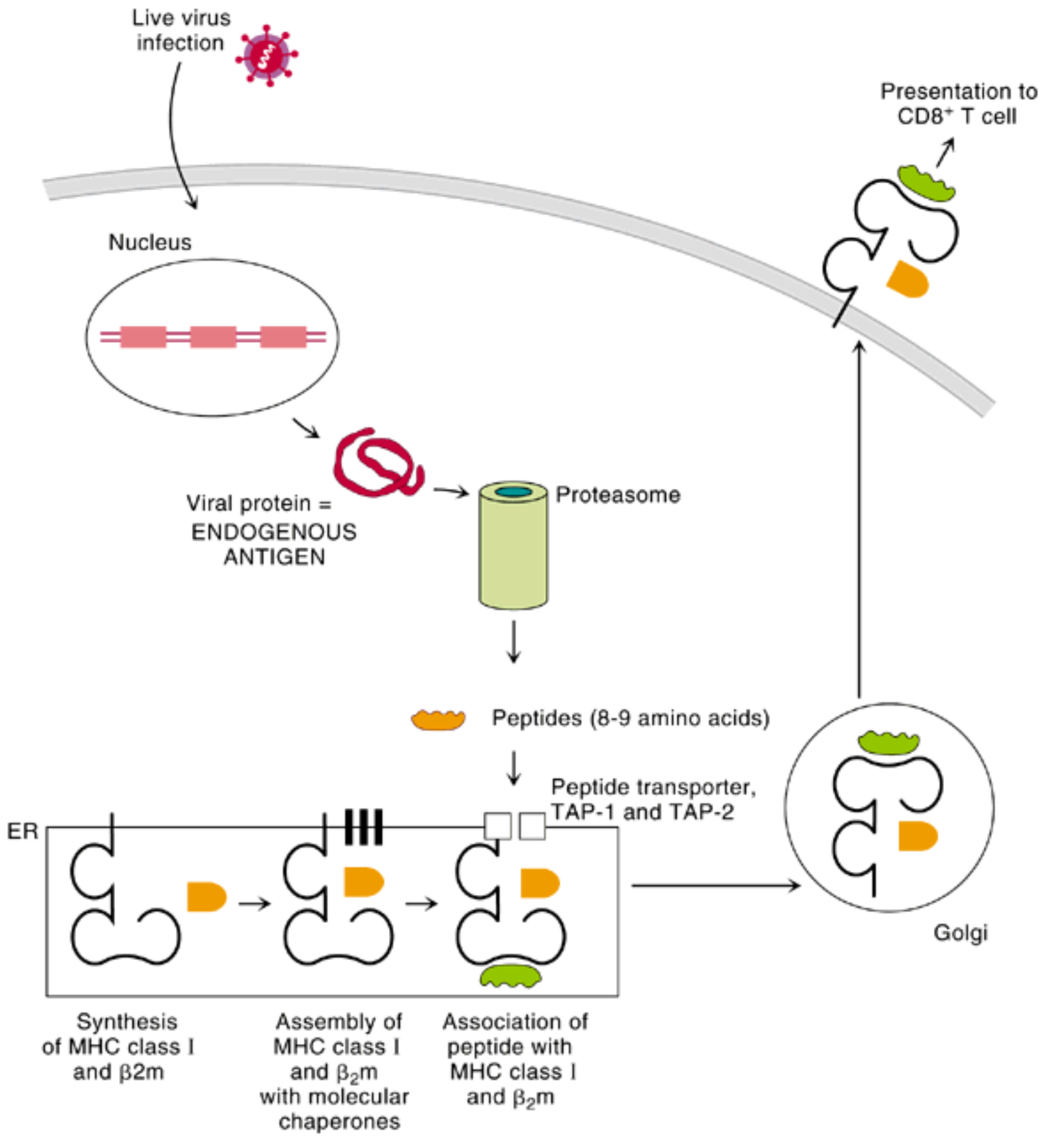


Antigen Processing (επεξεργασία) and Presentation (παρουσίαση)

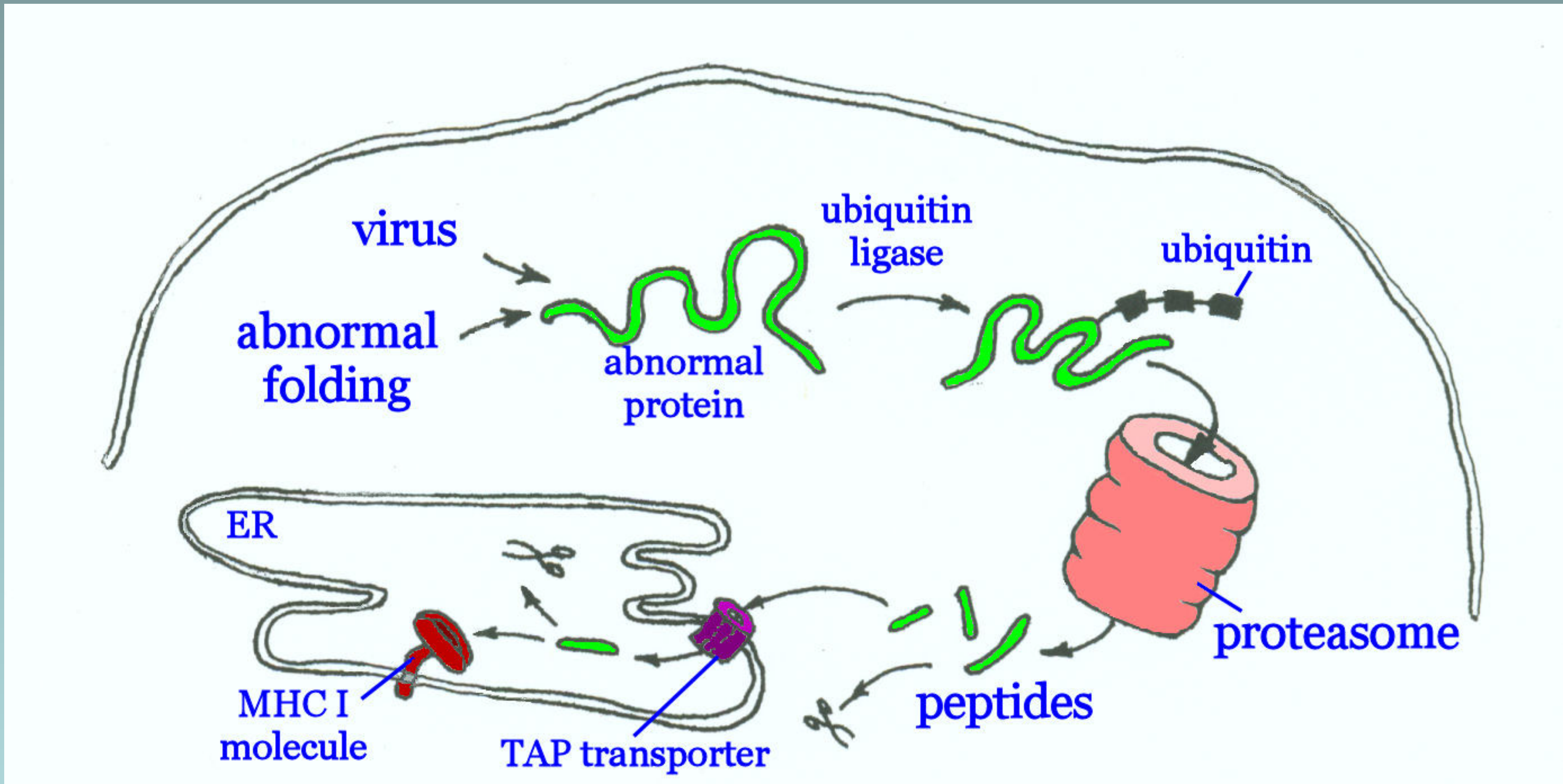


Antigen processing and Peptide binding to MHC τάξης I Presentation to CD8^{+} T cells

Endogenous Antigens

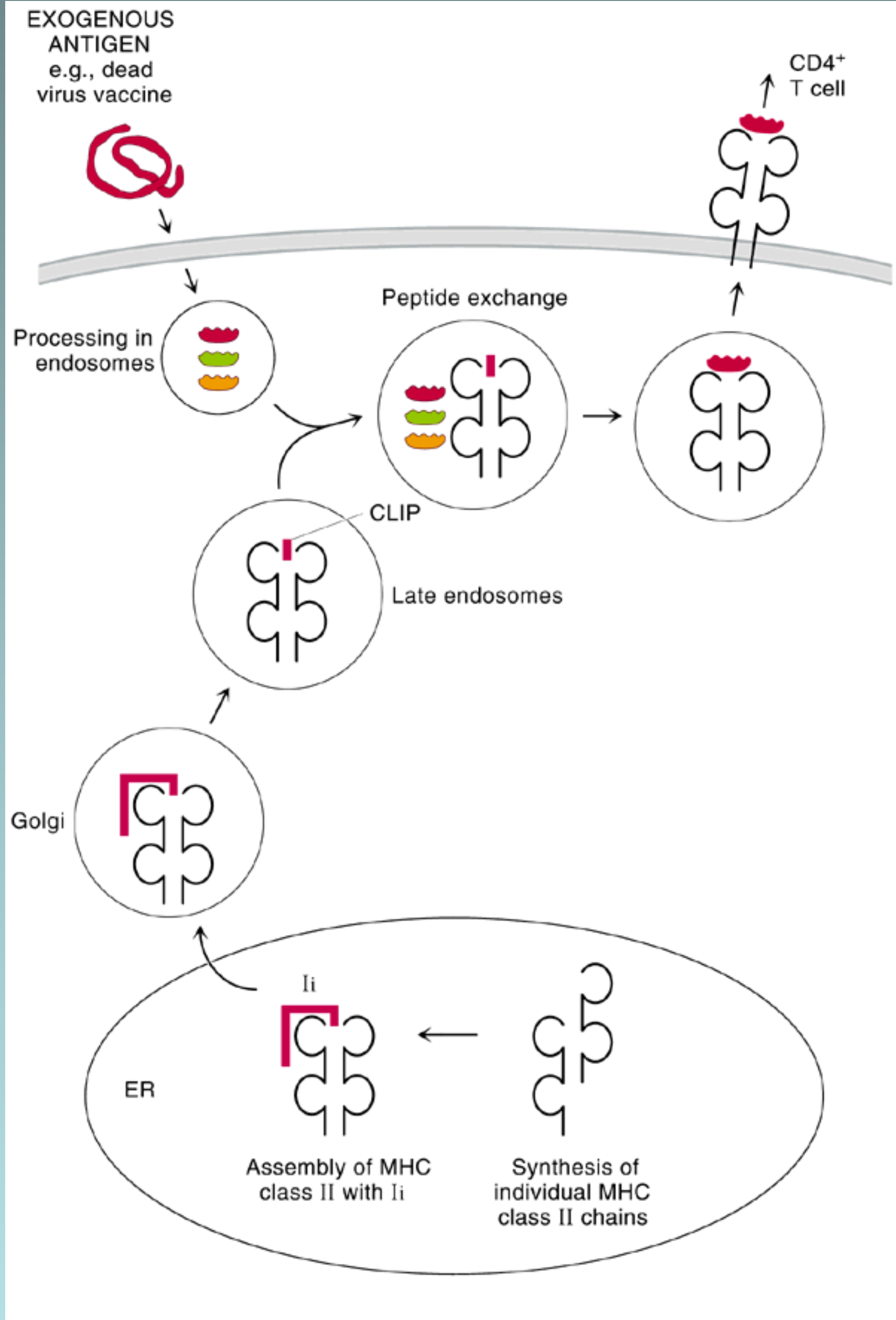


Antigen Processing for MHC class I



Antigen processing and Peptide binding to MHC τάξης II Presentation to CD4^{+} T cells

Exogenous Antigens



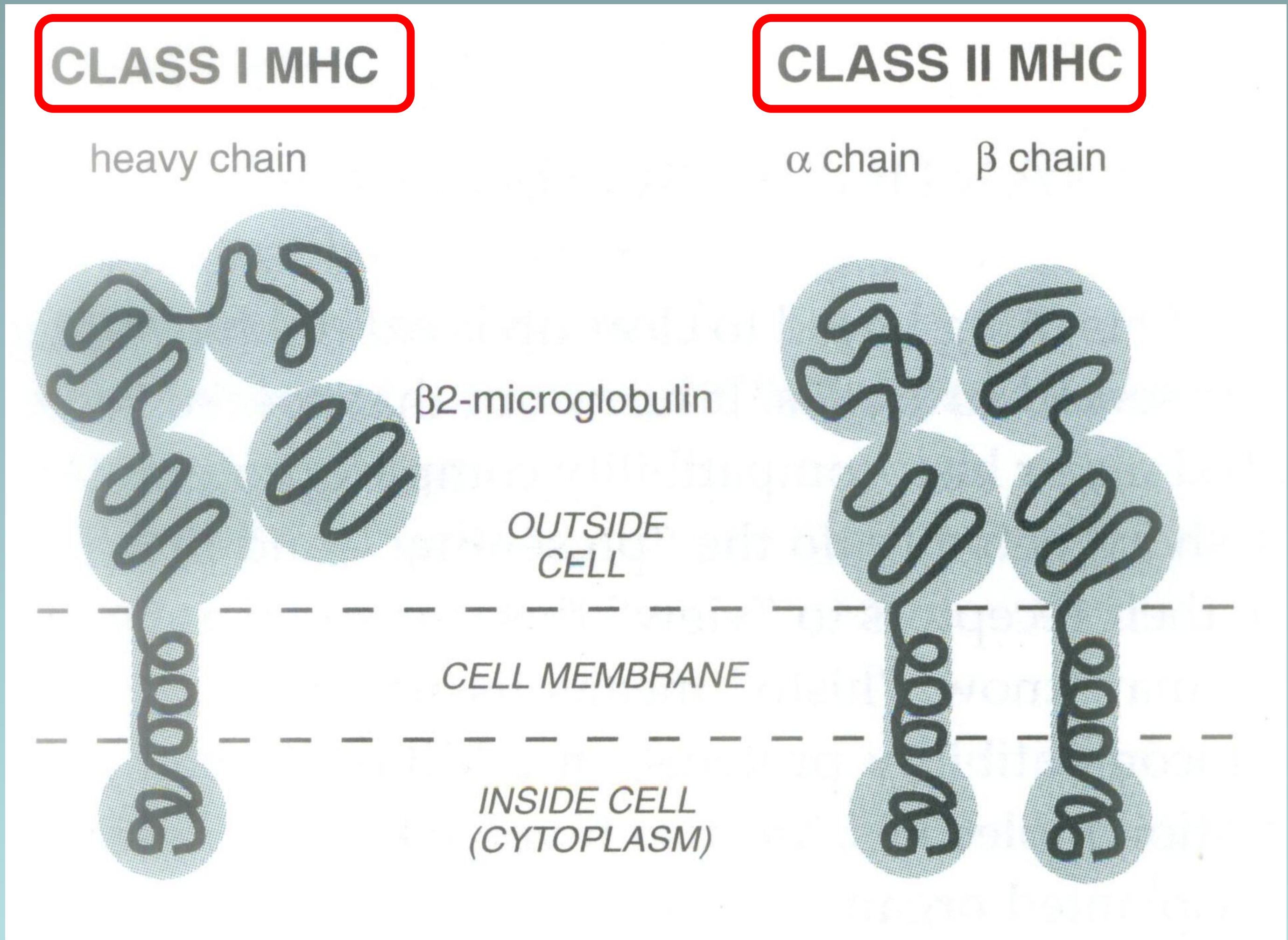
Antigen Processing Pathways

- 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 <u>polygenic</u>. Several different MHC class I and II genes.

MHC is highly <u>polymorphic</u>. Multiple variants of the gene within the population.

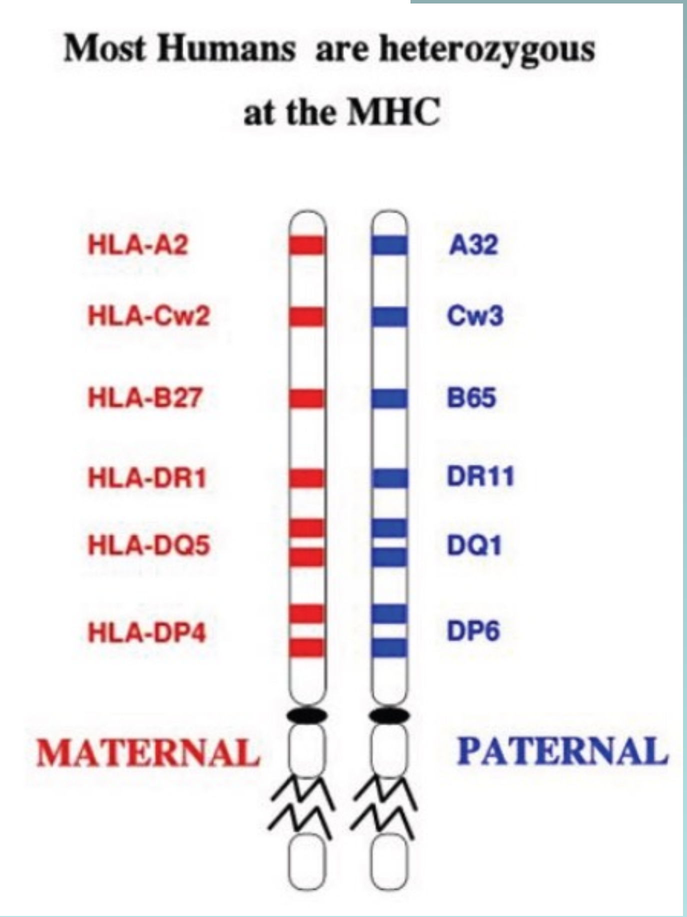
MHC highly polymorphic. More than 200 alleles that occur in high frequency.

Most individuals heterozygous at MHC locus.

MHC haplotype.

Expression co-dominant.

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



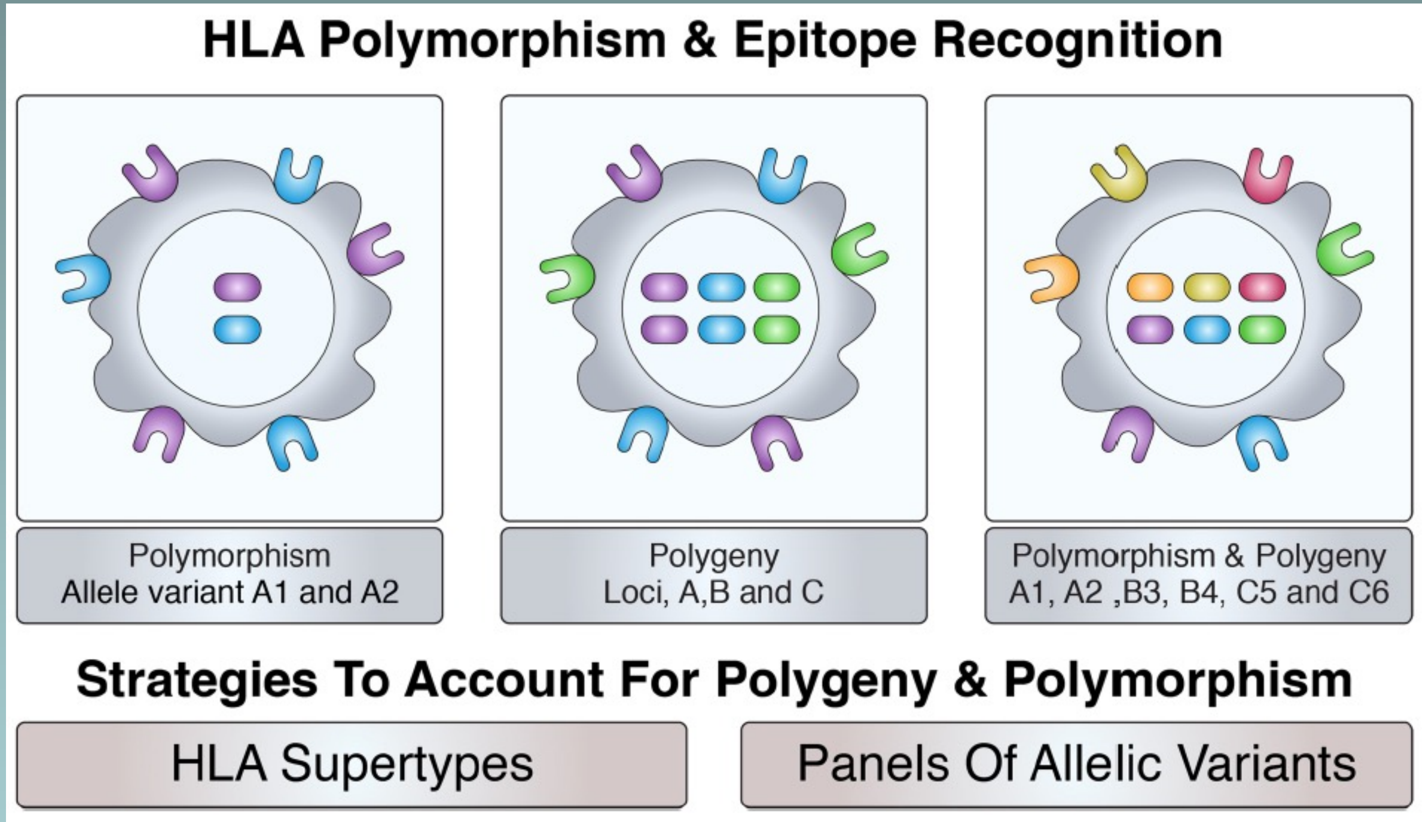
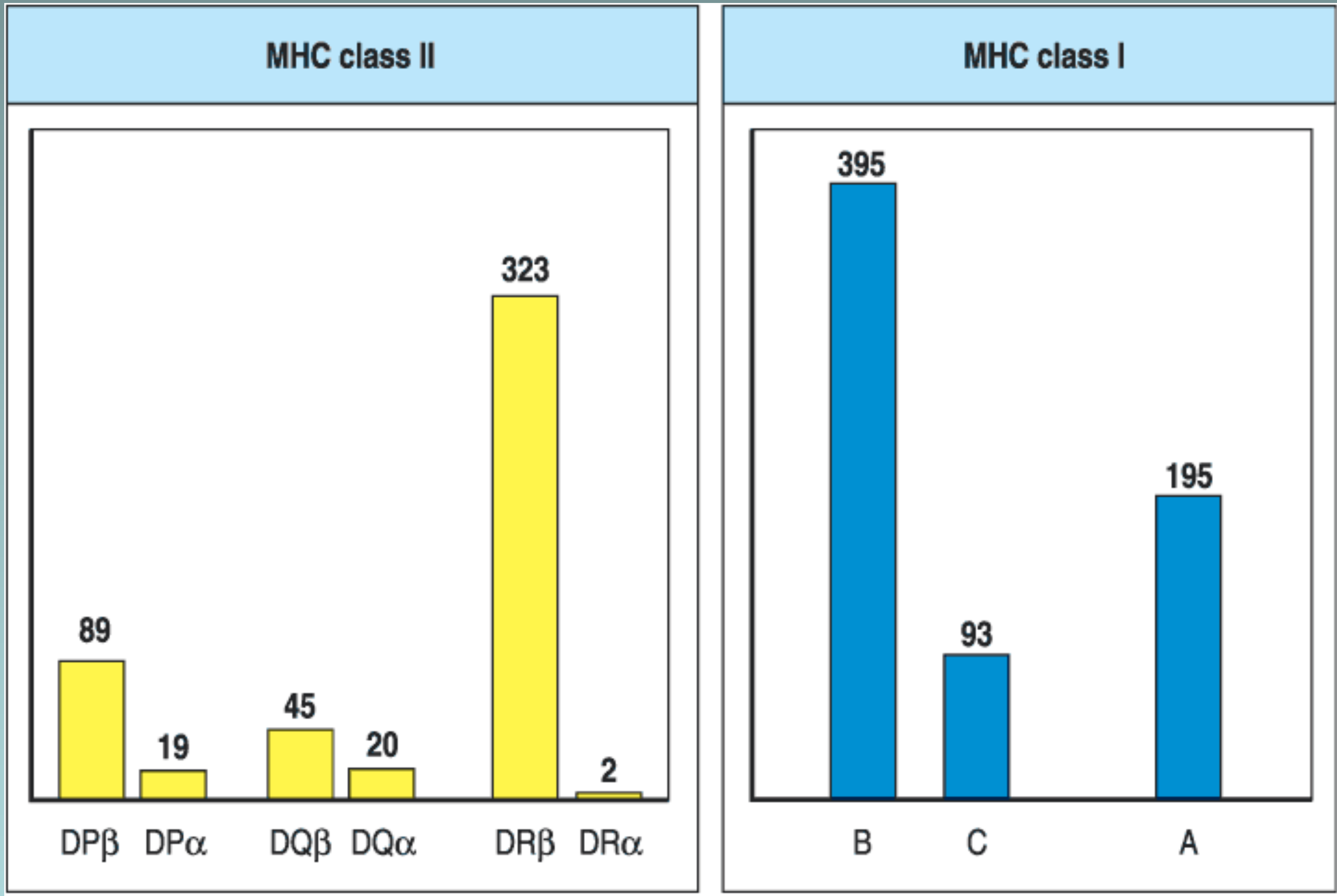


Fig. 5. HLA polymorphism and polygeny. The HLA region is highly polymorphic, with thousands of different allelic variants expressed at each locus. With heterozygosity, each individual will express up to two different alleles at each locus. Further, with class I and II presentation across several loci (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.

Sidney et al. *Semin Immunol* 2020


MHC class II
89
19
45
20
323
2
DPβ DPα DQβ DQα DRβ DRα
MHC class I
395
93
195
B C A

Mouse H-2 complex

Complex	H-2						
MHC class	I	II		III		I	
Region	K	IA	IE	S		D	
Gene products	H–2K	IA αβ	IE αβ	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 αβ	DQ αβ	DR αβ	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 α chain): **HLA-A, HLA-B, και HLA-C**

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

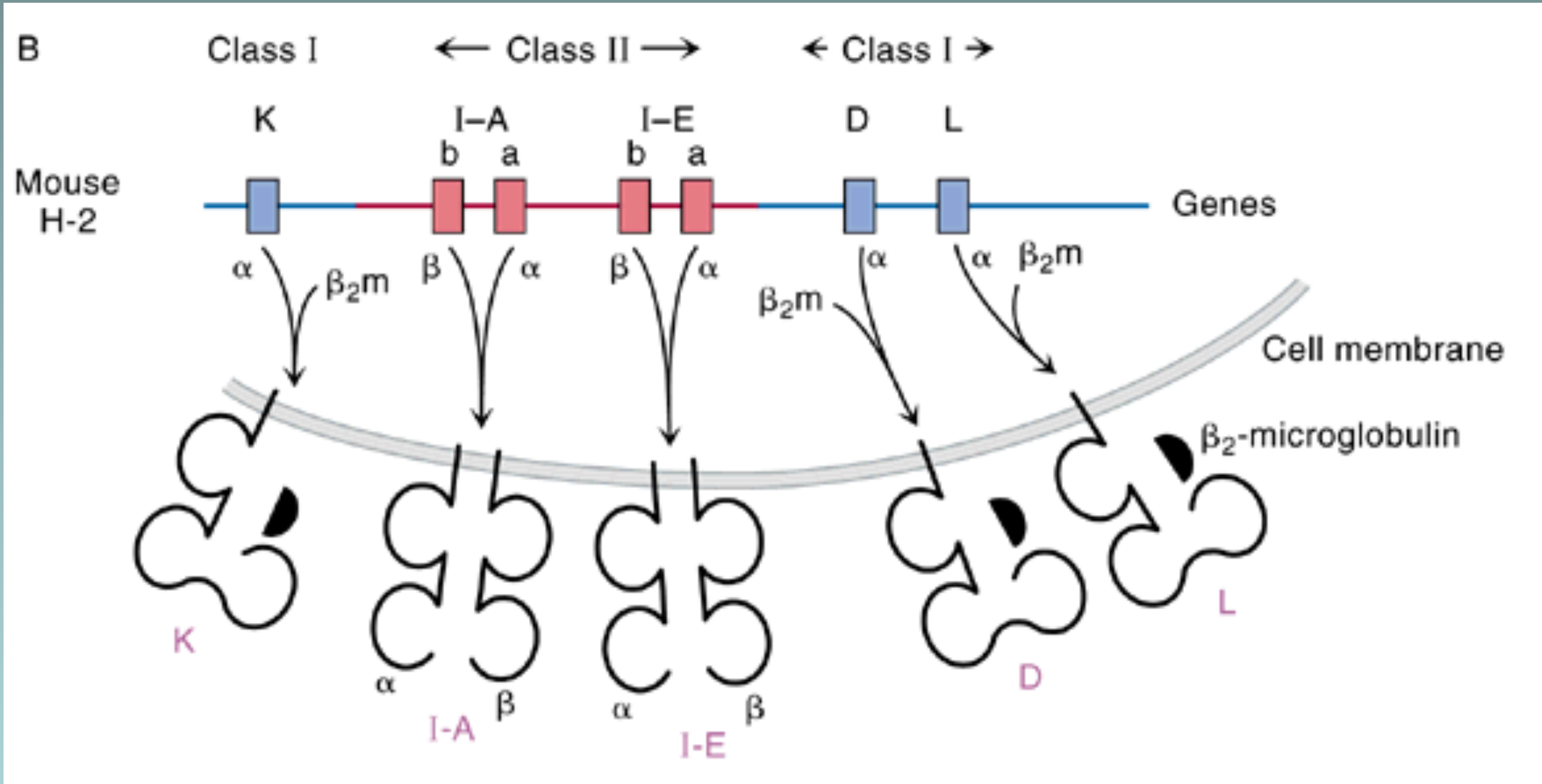
H-2 (mouse)

In mice, MHC is called H-2

In mice:

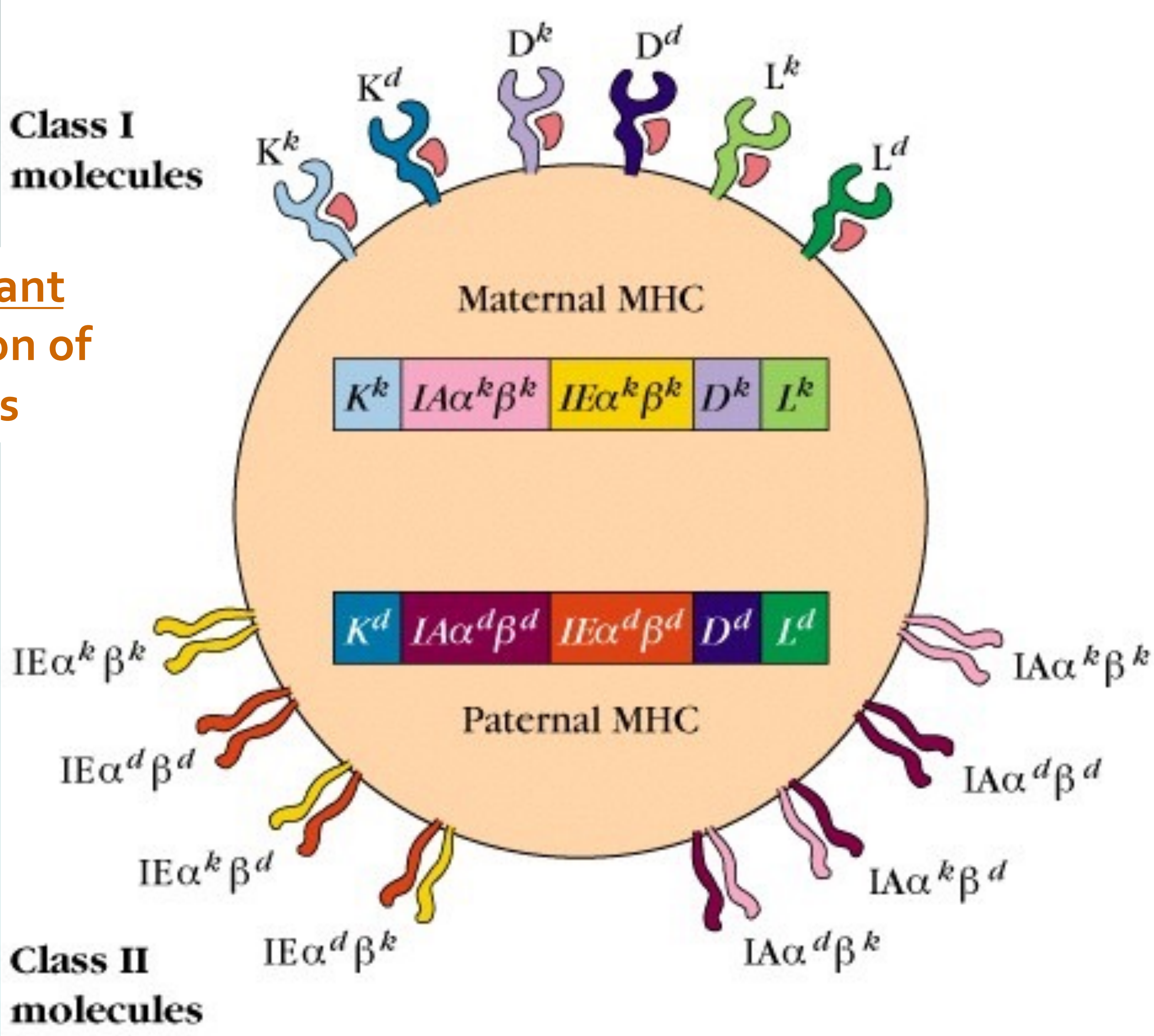
- **There are three MHC class I genes (coding for the α chain): H-2K, H-2D, and H-2L**

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


B
Class I
← Class II →
← Class I →
K
I–A
I–E
D
L
b
a
b
a
Mouse
H-2
Genes
α
β_2m
β
α
β
α
α
β_2m
α
β_2m
Cell membrane
β_2-microglobulin
K
α
I-A
β
α
I-E
β
D
L

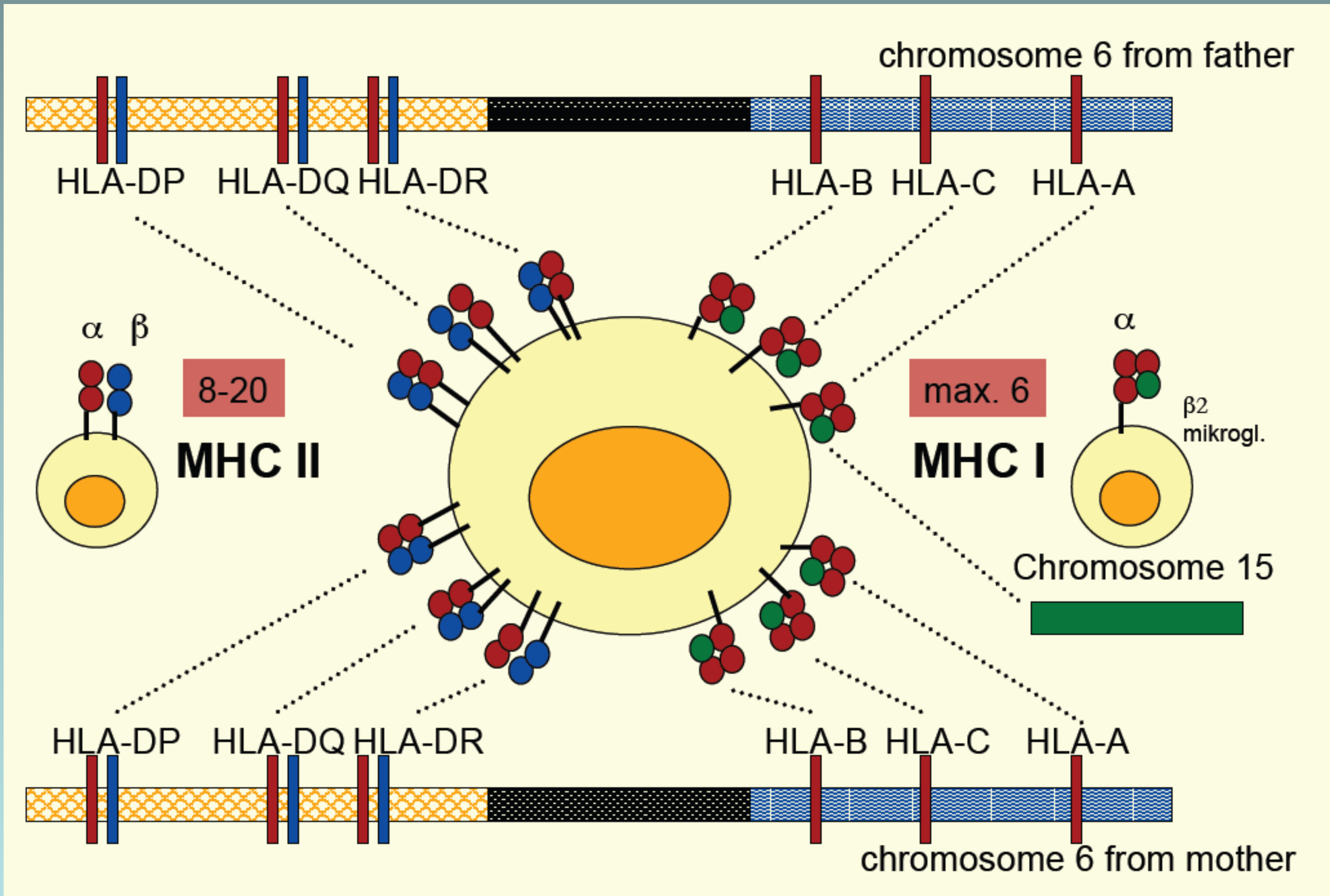
Mice

codominant expression of H-2 genes

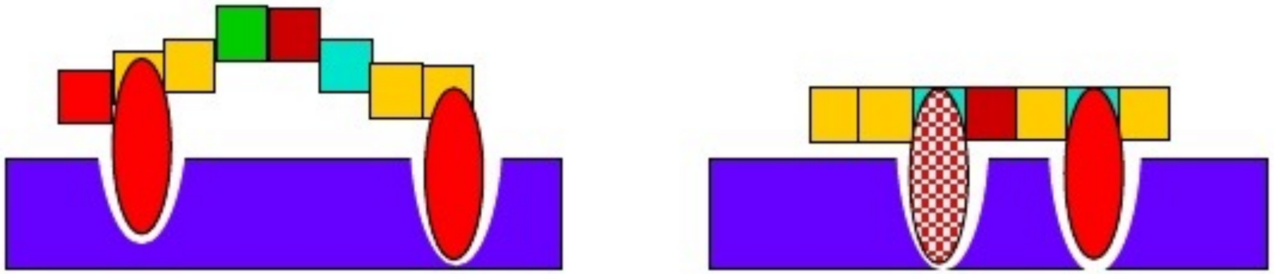


Humans

Codominant expression of HLA genes



Polymorphism in the MHC affects peptide antigen binding



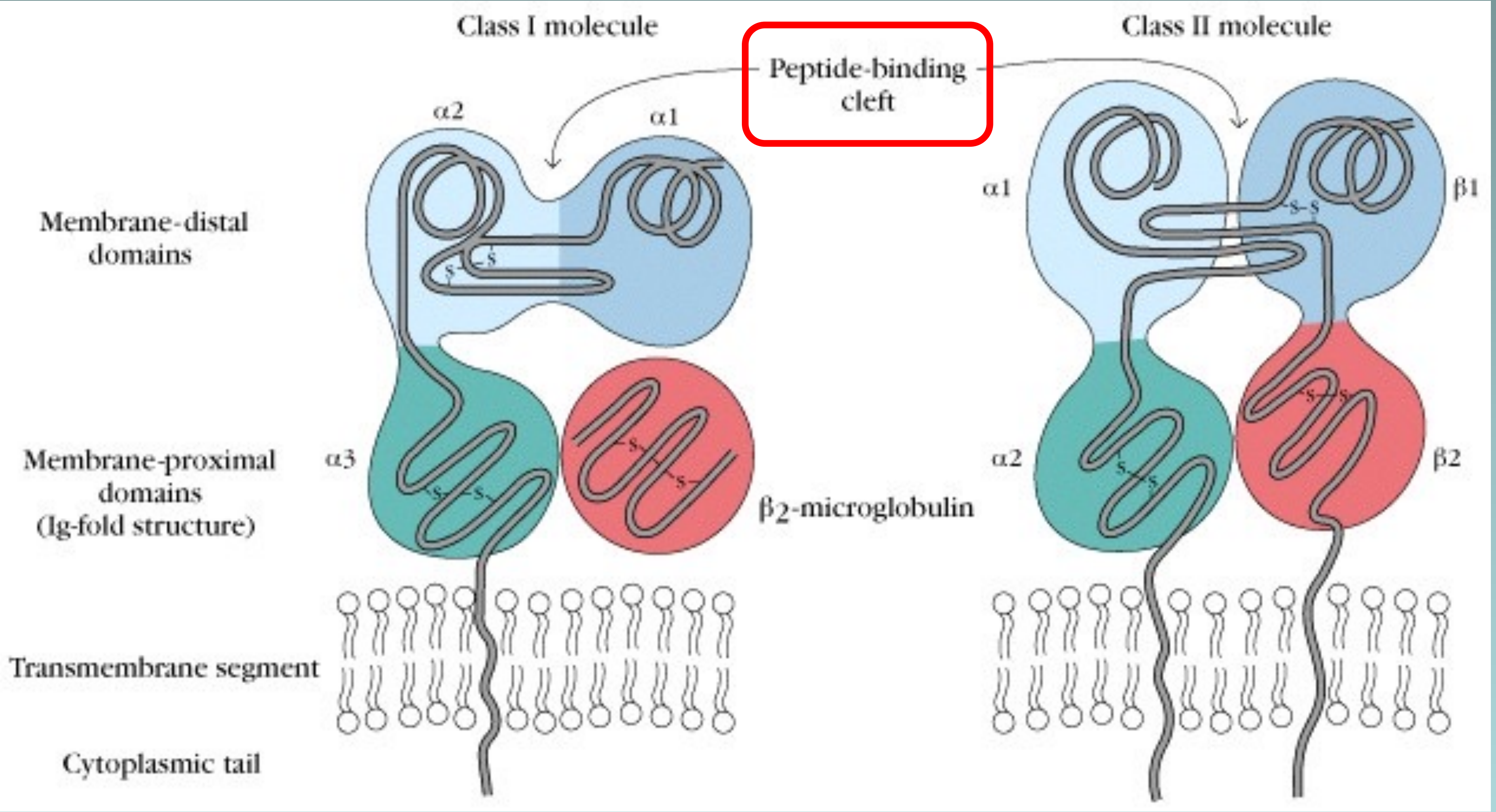
MHC allele A MHC allele B

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

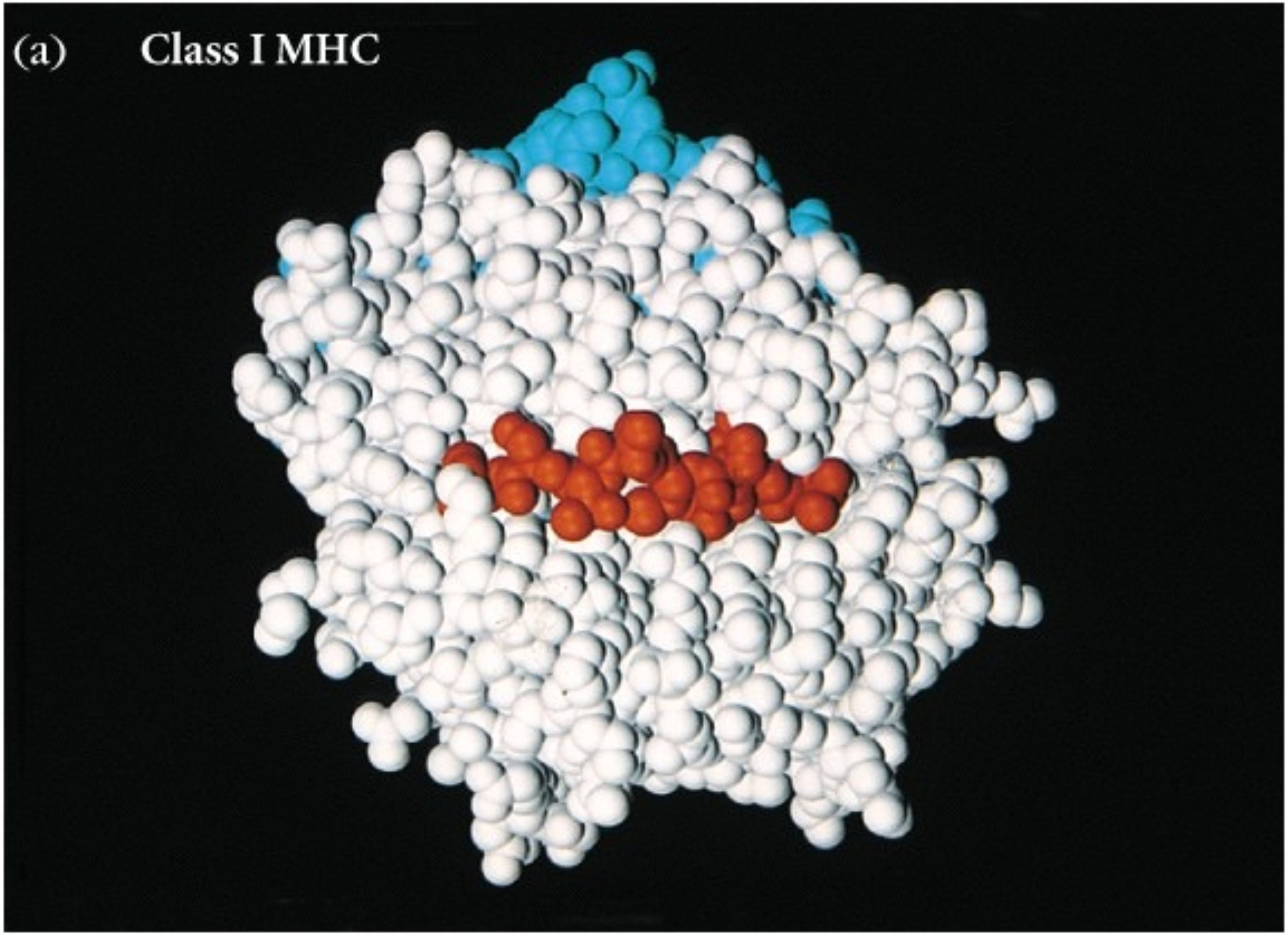
S Y I P S A K I MHC allele A

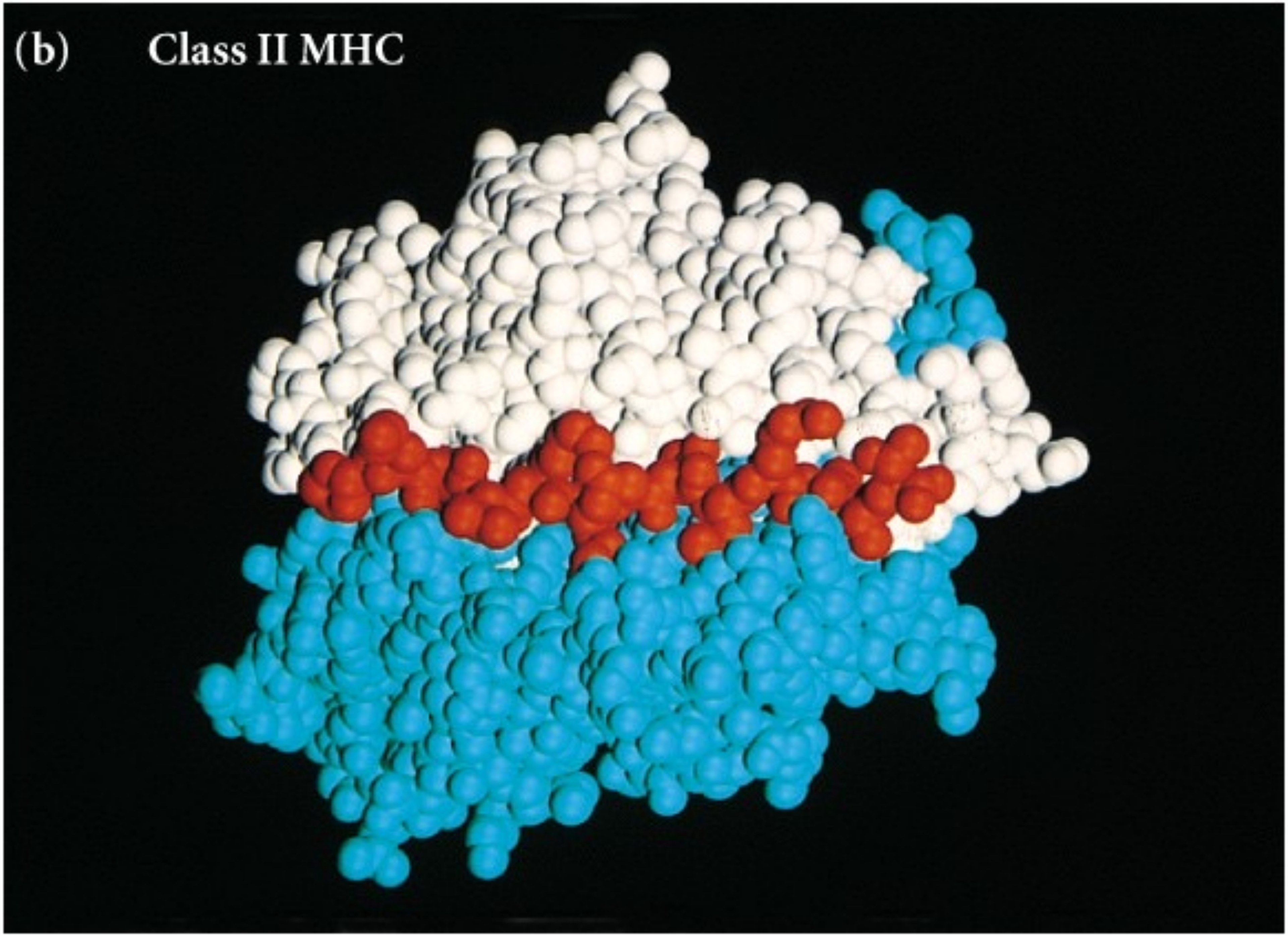
R G Y V Y Q Q L MHC allele B

Products of different MHC alleles bind a different repertoire of peptides



Class I molecule
Peptide-binding
cleft
Class II molecule
α2
α1
α1
β1
Membrane-distal
domains
S
S
S–S
Membrane-proximal
domains
(Ig-fold structure)
α3
S
S
S
S
β2-microglobulin
α2
S
S
S–S
β2
Transmembrane segment
Cytoplasmic tail

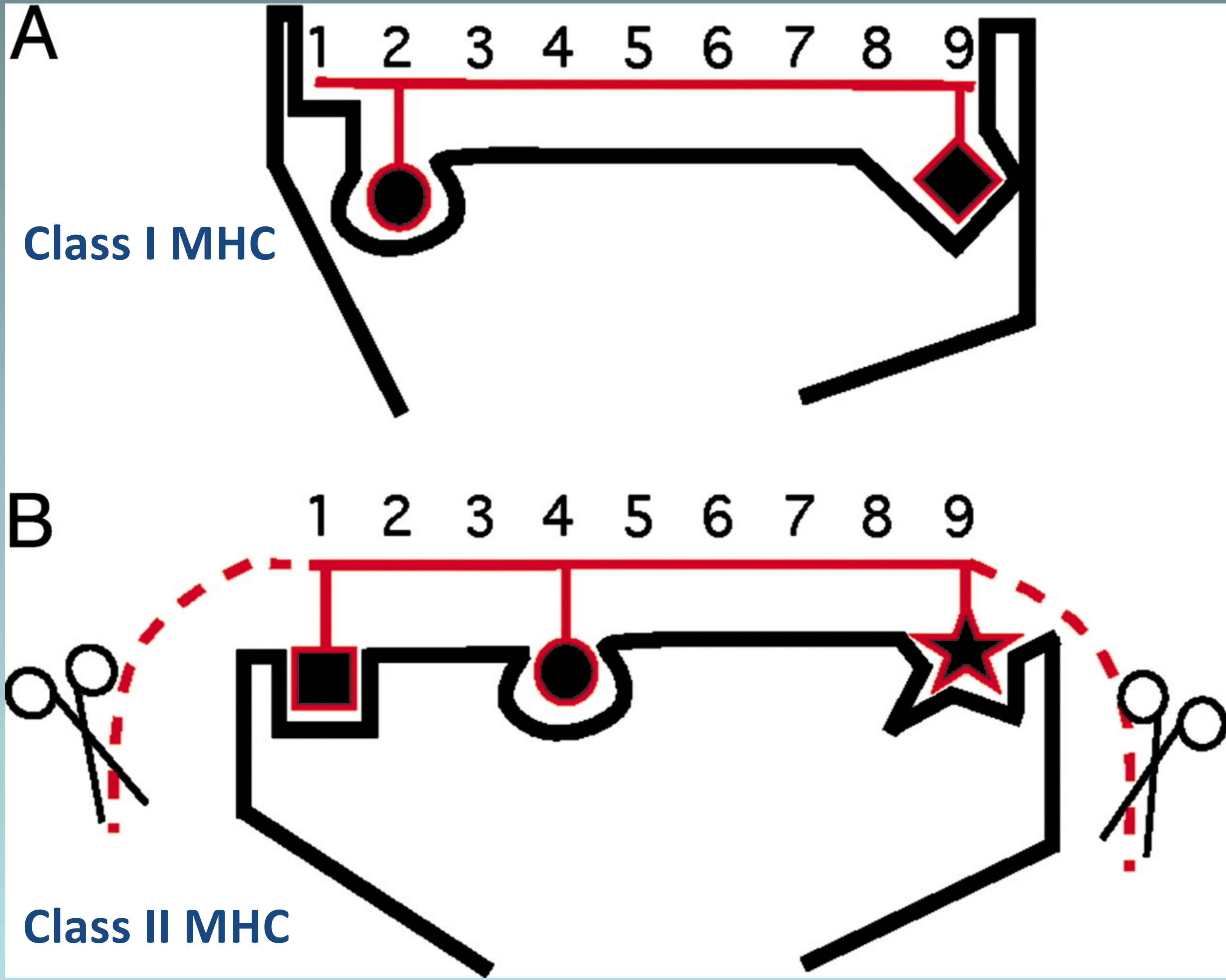

(a) Class I MHC

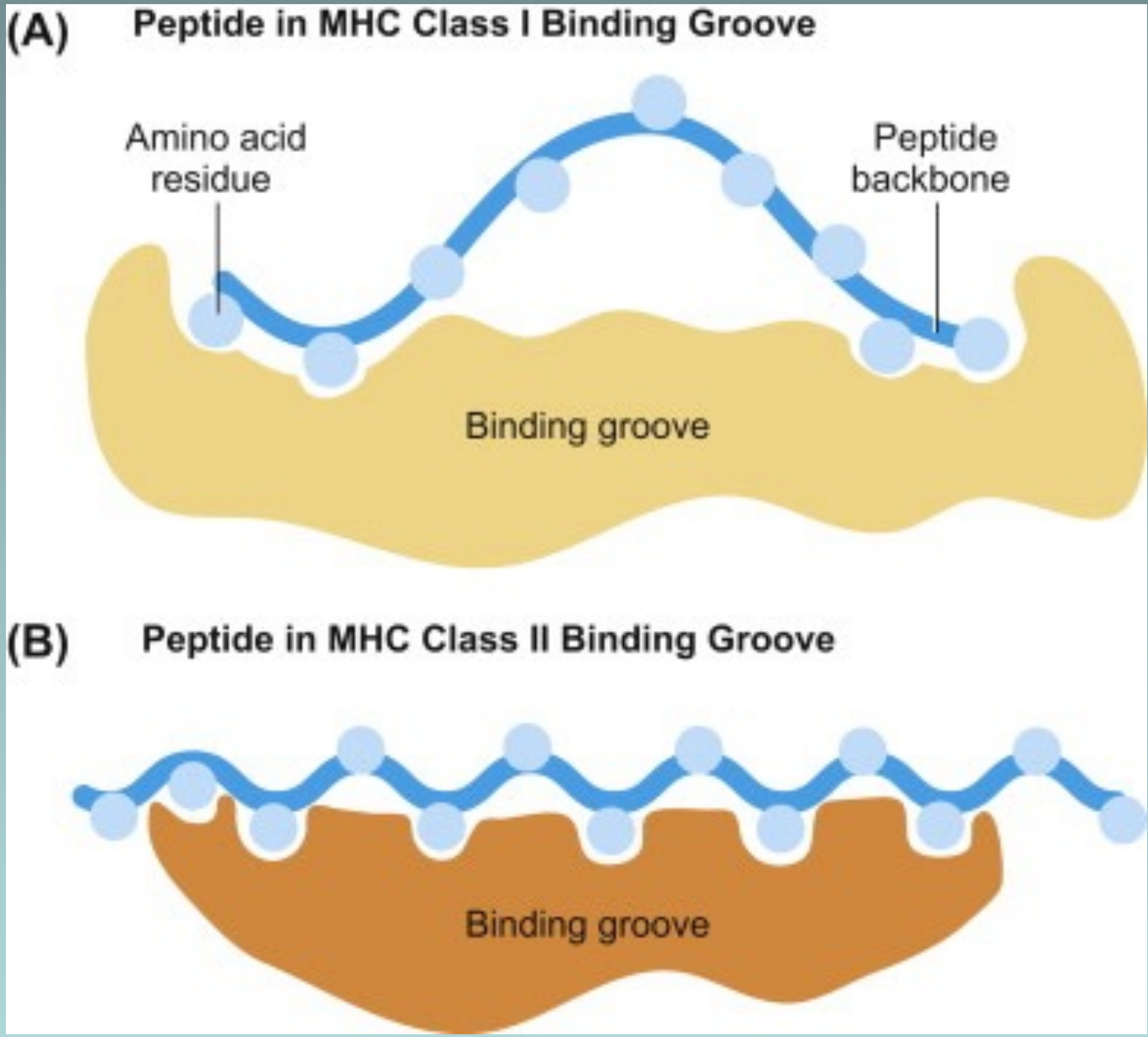

(b) Class II MHC

Anchor residues

(προσδετικά κατάλοιπα ή κατάλοιπα αγκυροβόλησης)

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


A
1 2 3 4 5 6 7 8 9
Class I MHC
B
1 2 3 4 5 6 7 8 9
Class II MHC



(A) Peptide in MHC Class I Binding Groove
Amino acid residue
Peptide backbone
Binding groove
(B) Peptide in MHC Class II Binding Groove
Binding groove

Eluted peptides from MHC molecules have different sequences but contain motifs

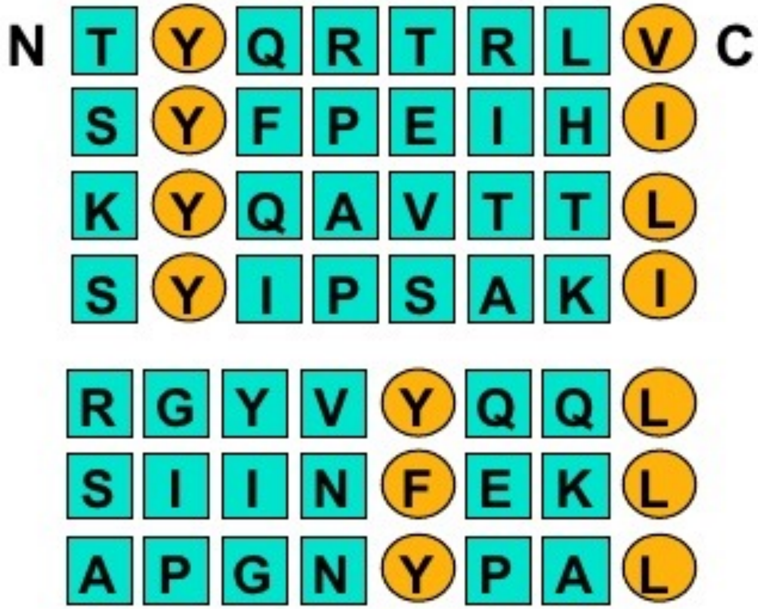
Peptides bound to a particular type of MHC class I molecule have conserved patterns of amino acids

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

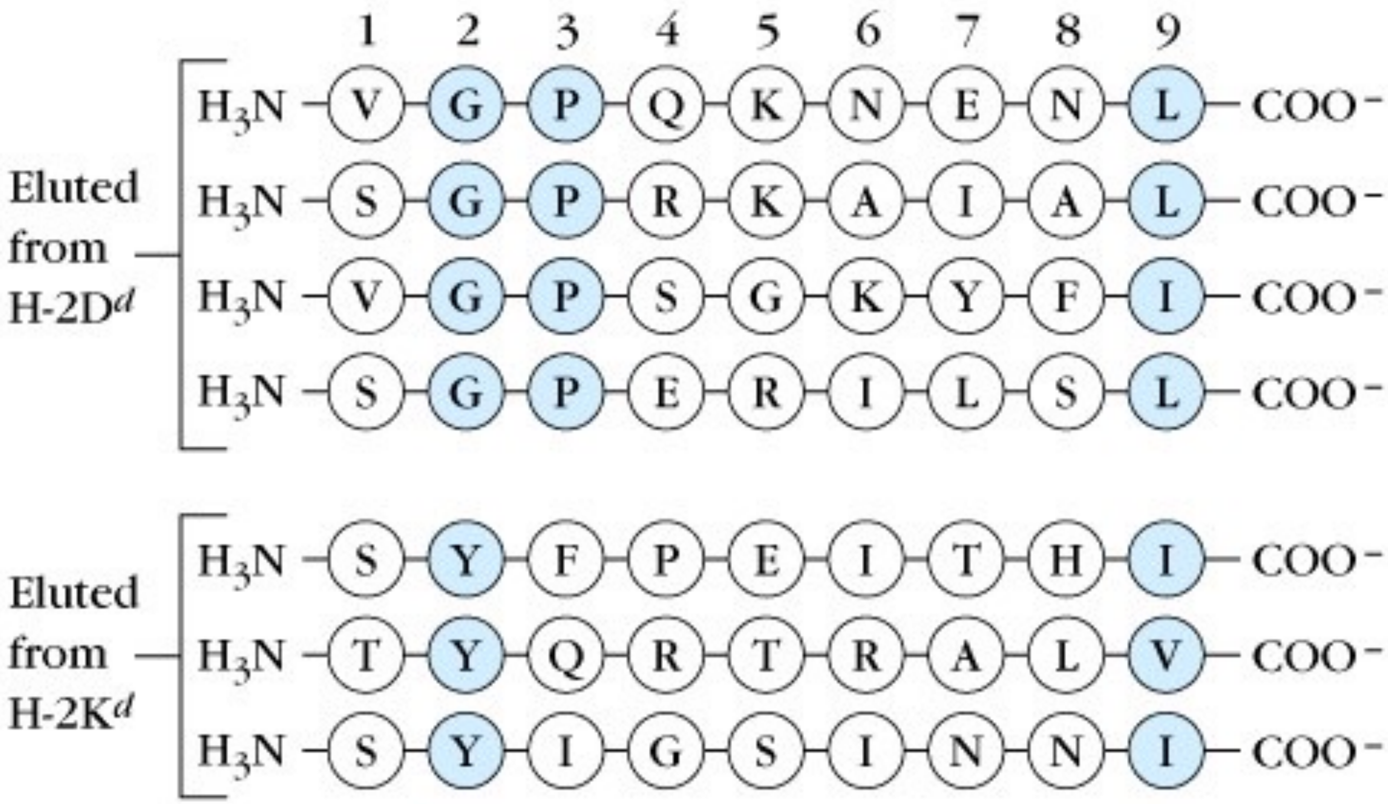
Amino acids common to many peptides tether the peptide to structural features of the MHC molecule
ANCHOR RESIDUES

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	K = lysine	R = arginine
E = glutamic acid	L = leucine	S = serine
F = phenylalanine	N = asparagine	T = threonine
G = glycine	P = proline	V = valine
H = histidine	Q = glutamine	Y = tyrosine
I = isoleucine		

MHC τάξης I 8-11 αα

MHC-I peptide binding

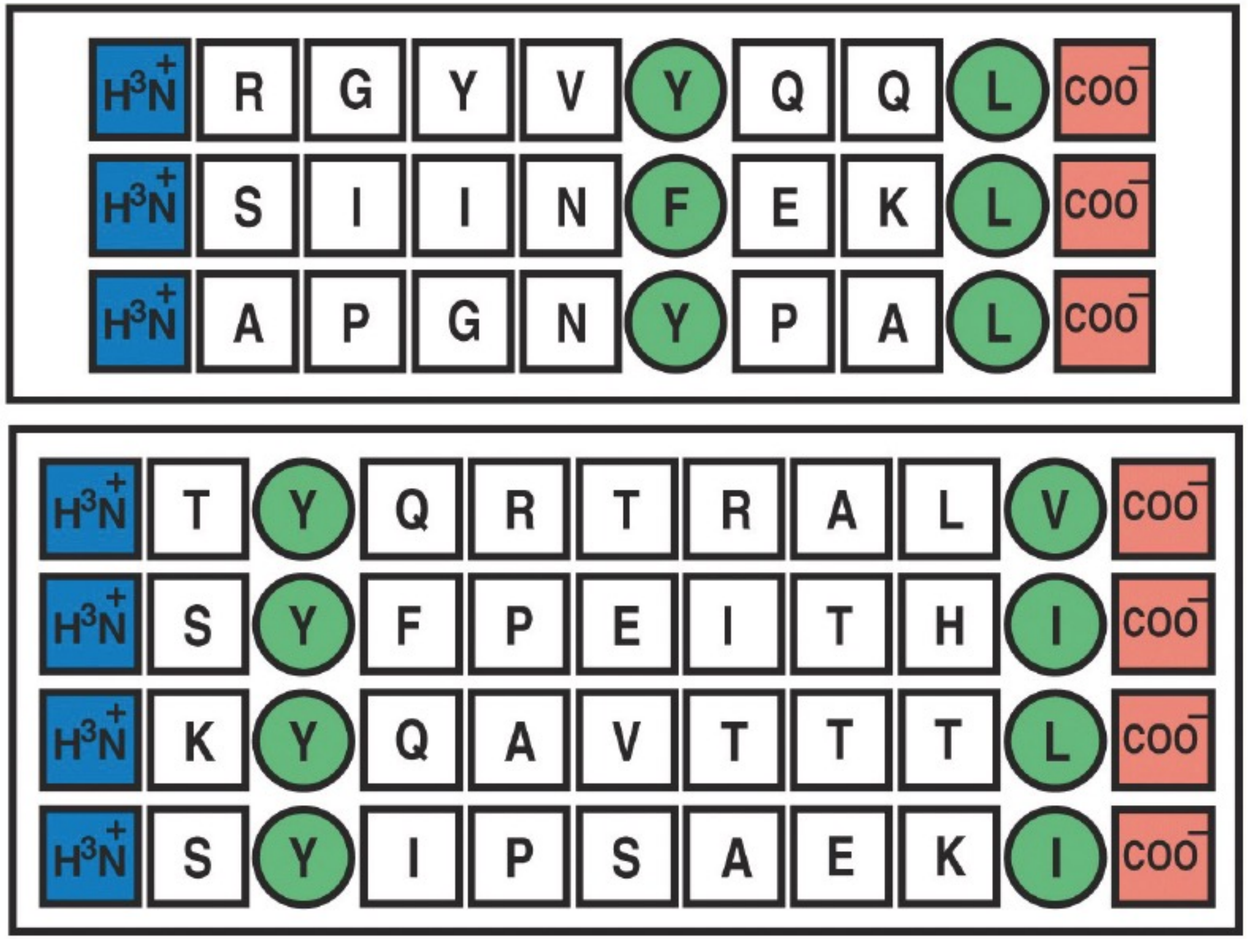
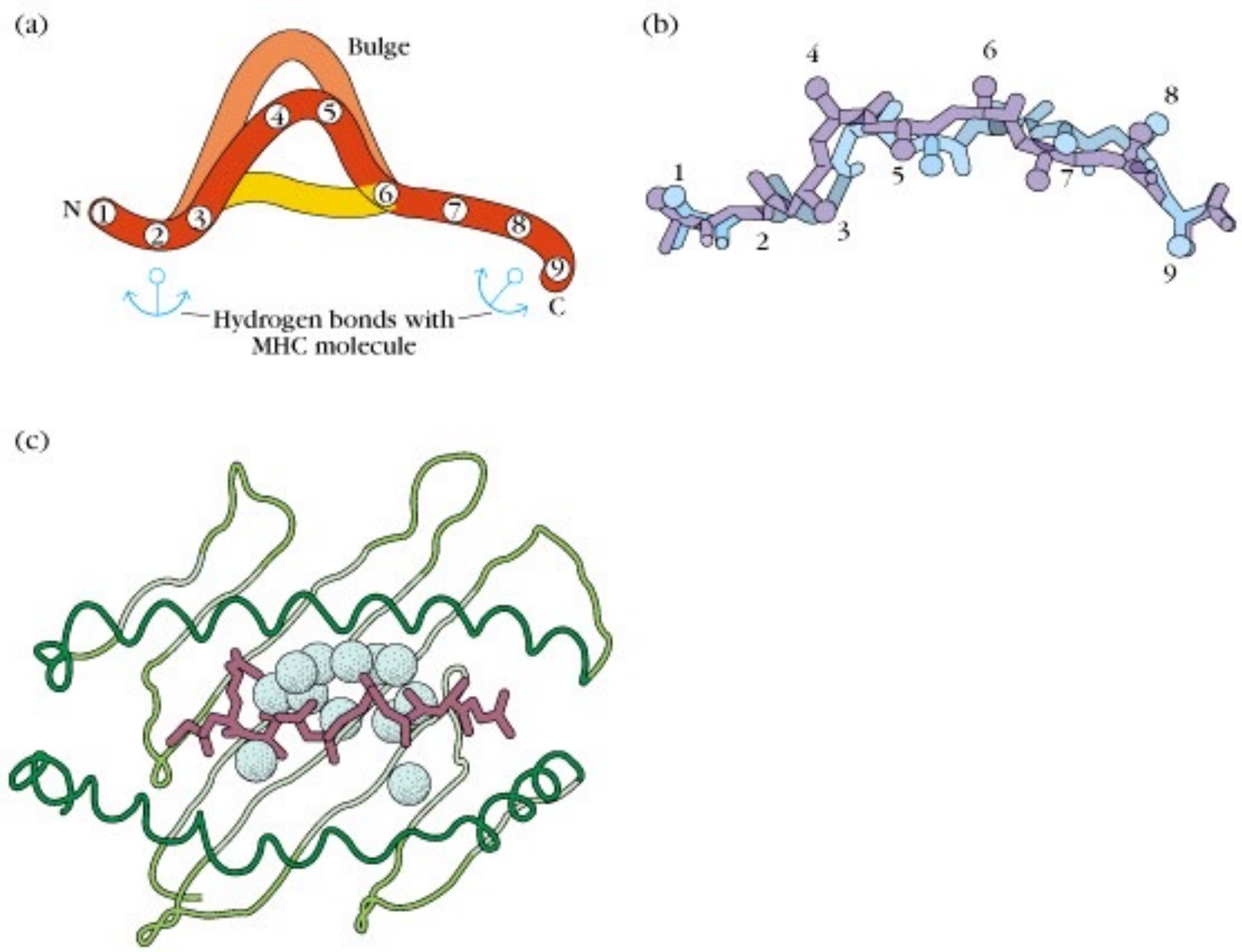
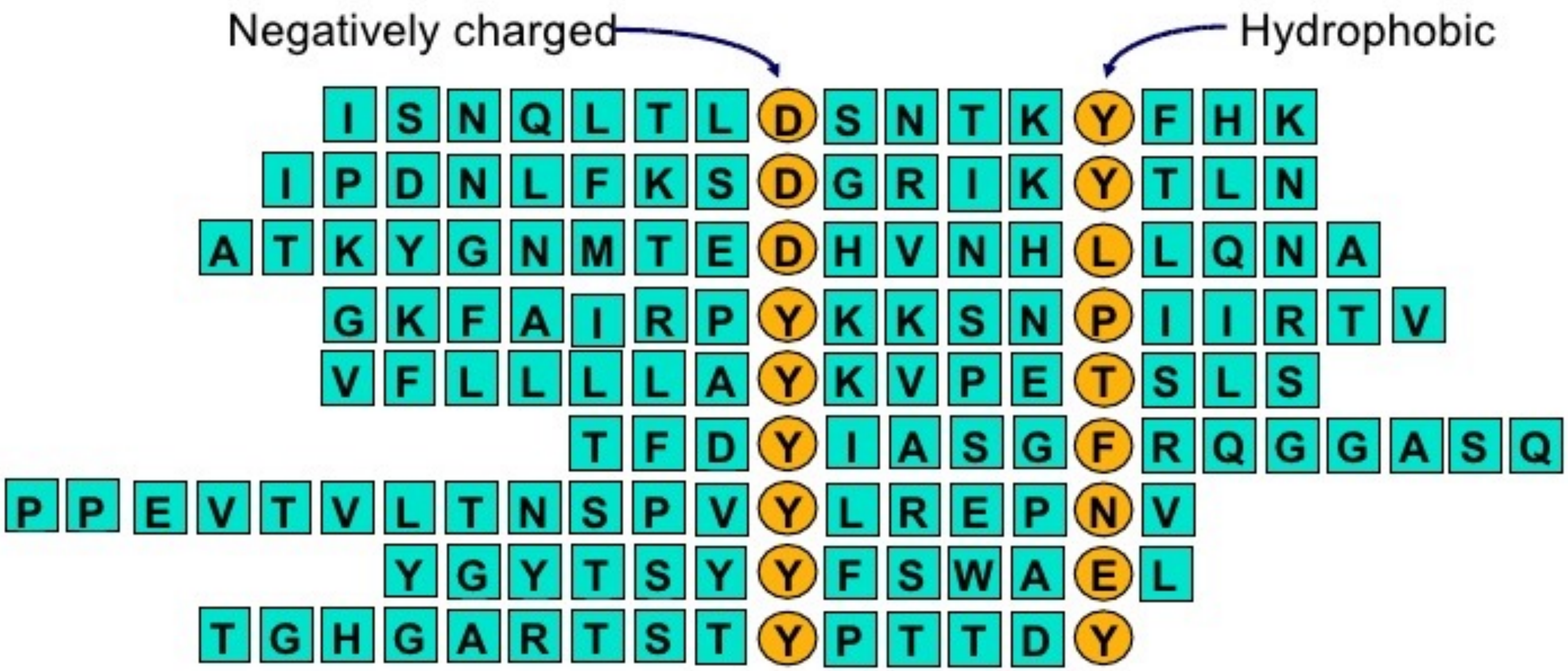


Figure 3-24 Immunobiology, 6/e. (© Garland Science 2005)

Closed peptide-binding cleft:
Conserved AA residues bind the terminal -NH2 and –COOH groups


(a)
Bulge
N
C
Hydrogen bonds with MHC molecule
(b)
(c)

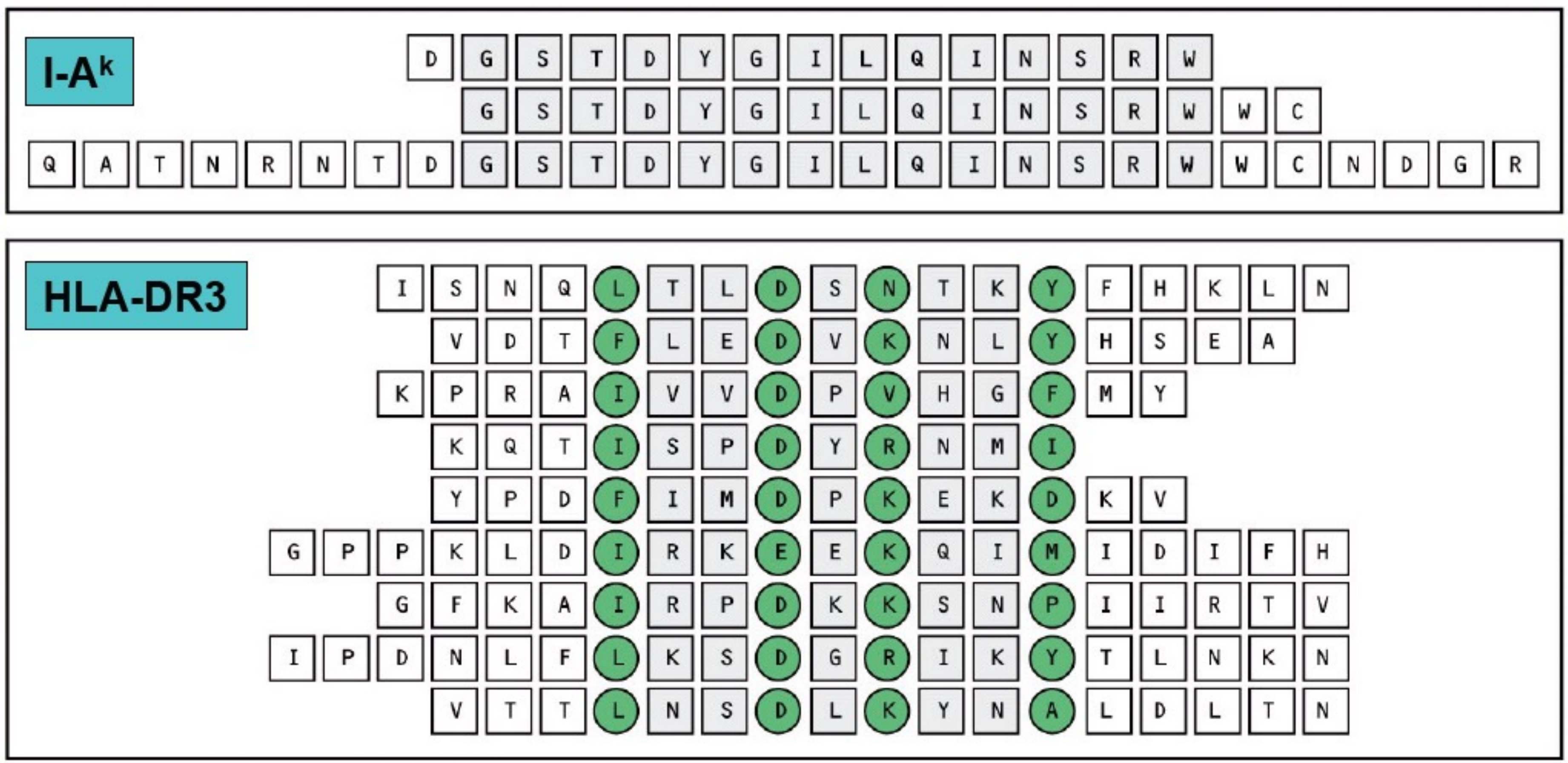
Peptide antigen binding to MHC class II molecules



- 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

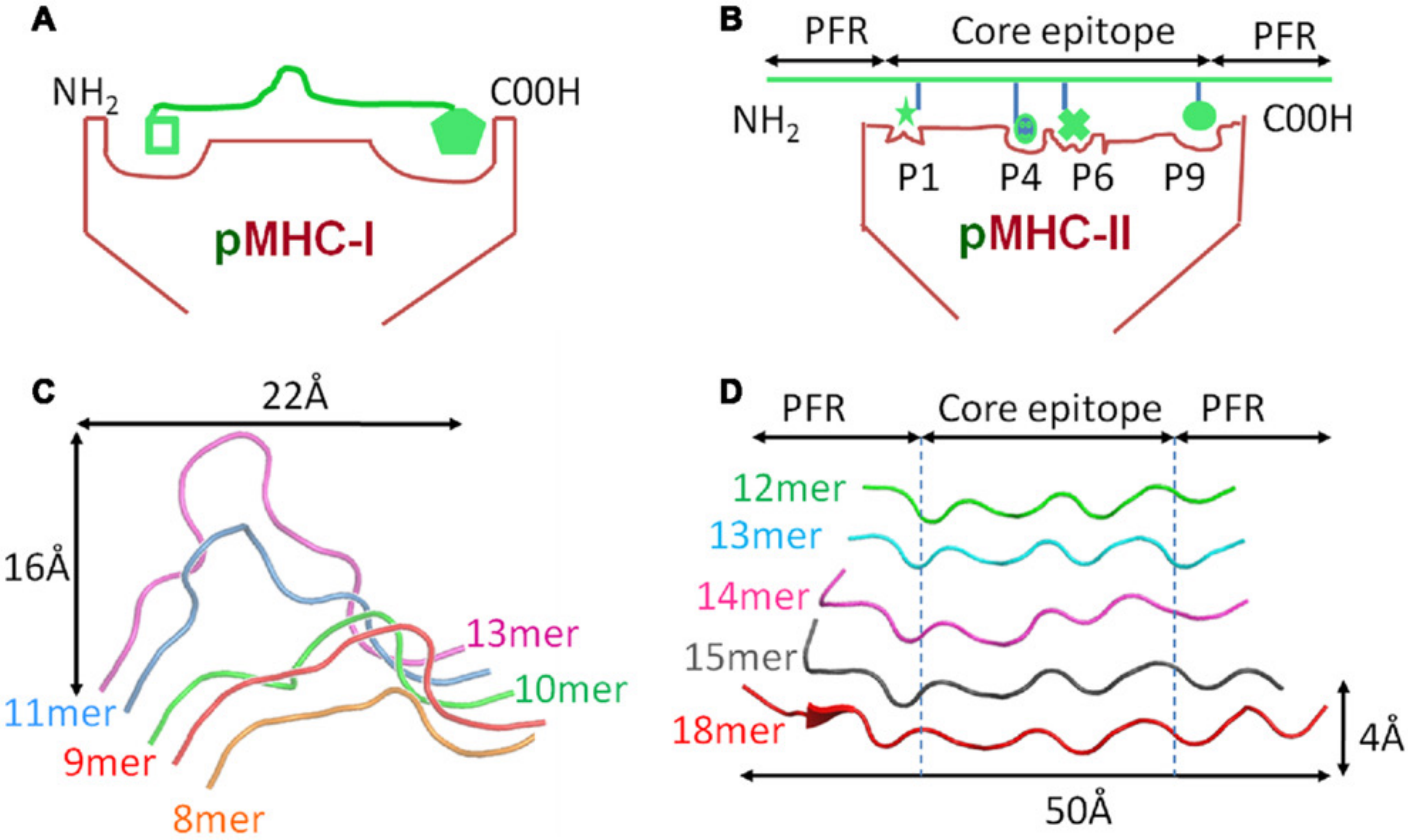
MHC τάξης II   > 13 αα



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
NH2
COOH
pMHC-I
B
PFR
Core epitope
PFR
NH2
COOH
P1
P4
P6
P9
pMHC-II
C
22Å
16Å
13mer
10mer
11mer
9mer
8mer
D
PFR
Core epitope
PFR
12mer
13mer
14mer
15mer
18mer
4Å
50Å

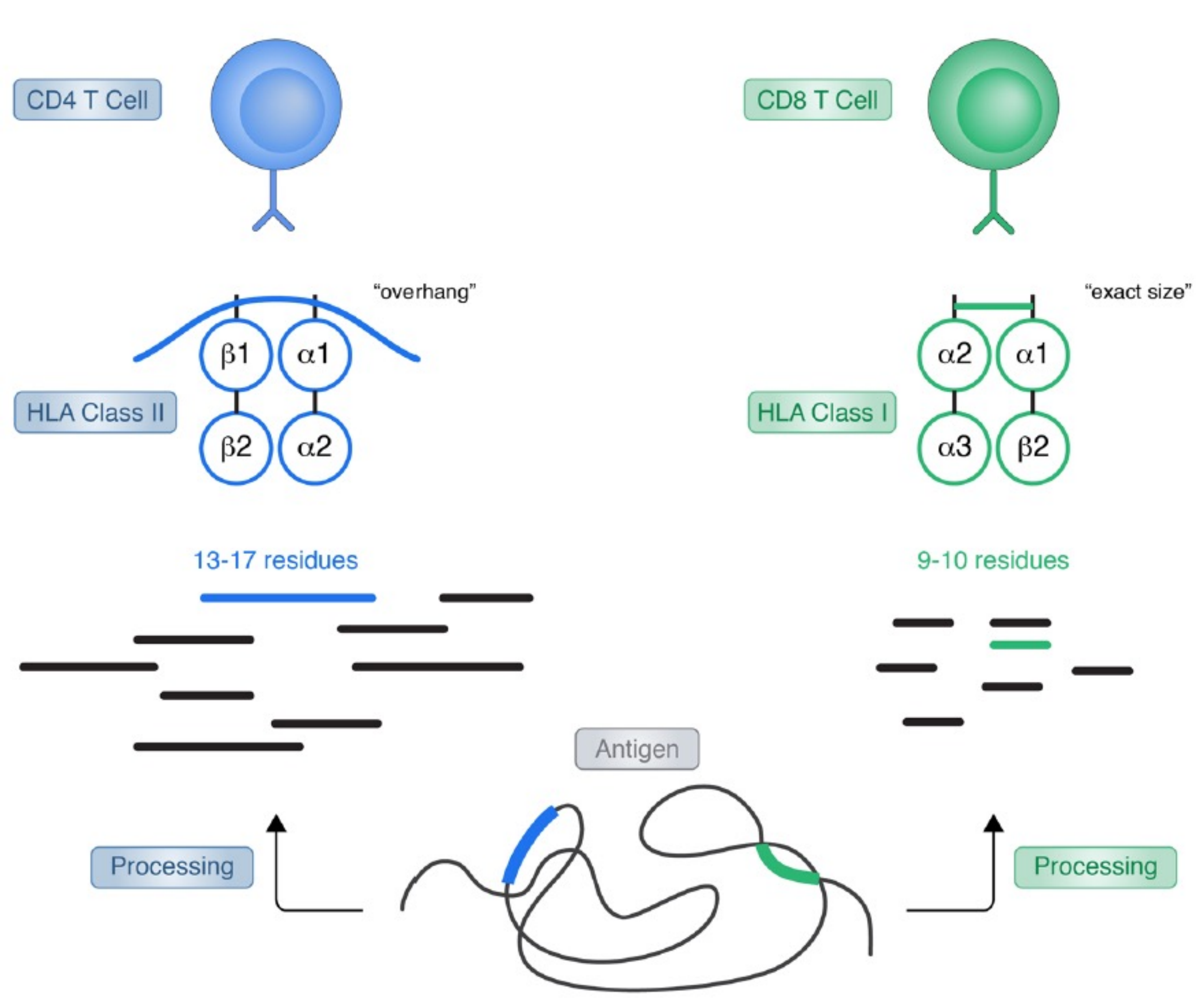


Fig. 1. Epitope definition. Presentation of peptide ligands by HLA class I and class II molecules is depicted. A) HLA class I and II ligands are generated by proteolytic processing of endogenously expressed proteins (antigen) (class I) or from proteins degraded in endocytic compartments (class II). Longer 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.

Sidney et al. *Semin Immunol* 2020

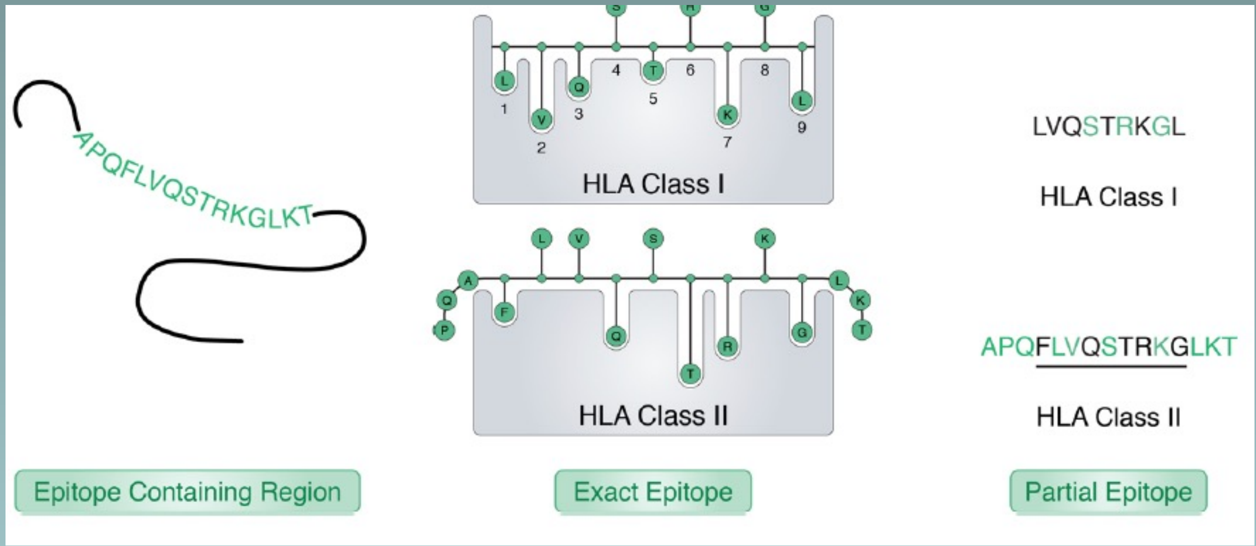


Fig. 1. Epitope definition. Presentation of peptide ligands by HLA class I and class II molecules is depicted. A) HLA class I and II ligands are generated by proteolytic processing of endogenously expressed proteins (antigen) (class I) or from proteins degraded in endocytic compartments (class II). Longer 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.

Sidney et al. *Semin Immunol* 2020

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

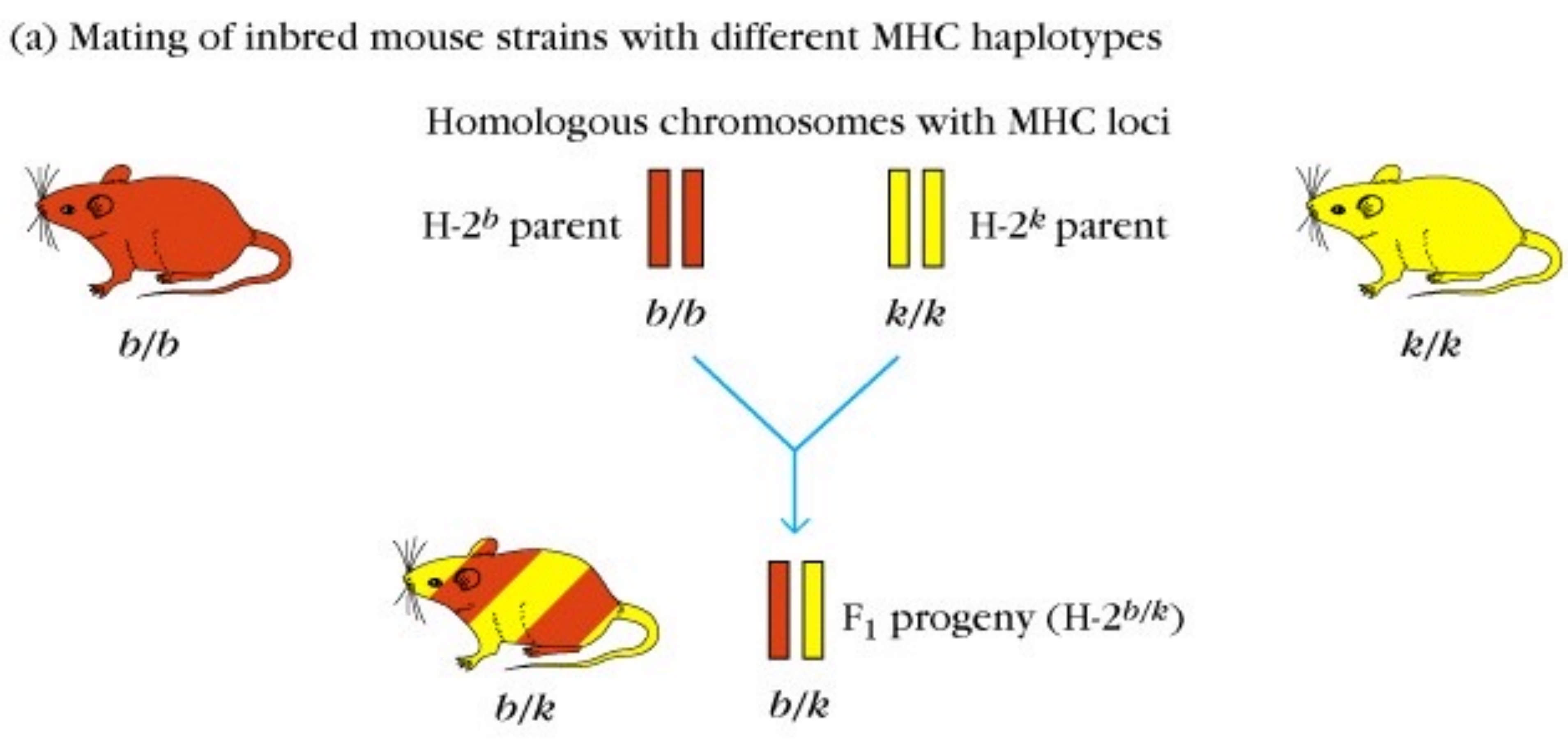
TABLE 7-2 PEPTIDE BINDING BY CLASS I AND CLASS II MHC MOLECULES

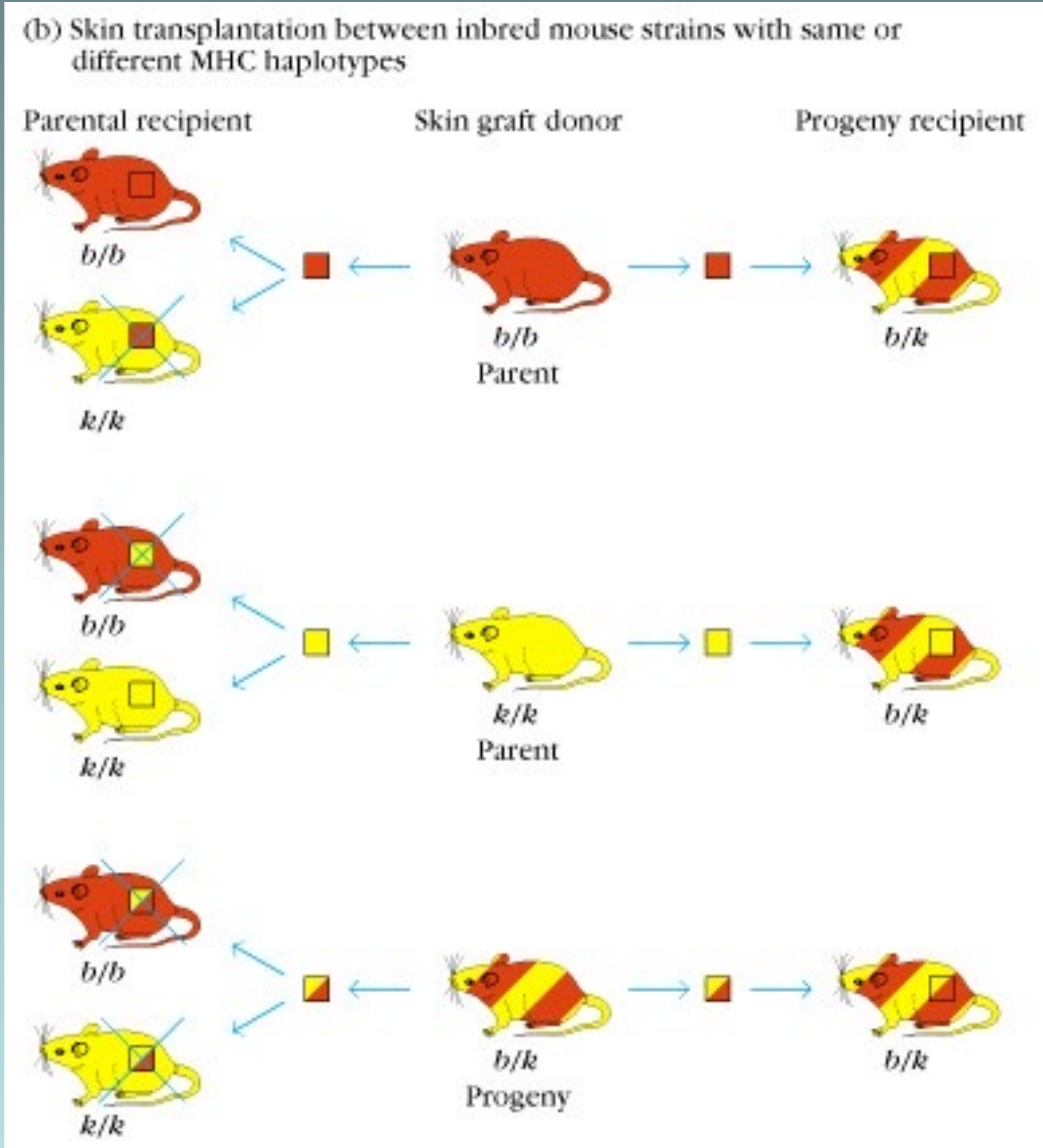
	Class I molecules	Class II molecules
Peptide-binding domain	α1/α2	α1/β1
Nature of peptide-binding cleft	Closed at both ends	Open at both ends
General size of bound peptides	8–10 amino acids	13–18 amino acids
Peptide motifs involved in binding to MHC molecule	Anchor residues at both ends of peptide; generally 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	$\alpha1/\alpha2$	$\alpha1/\beta2$
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

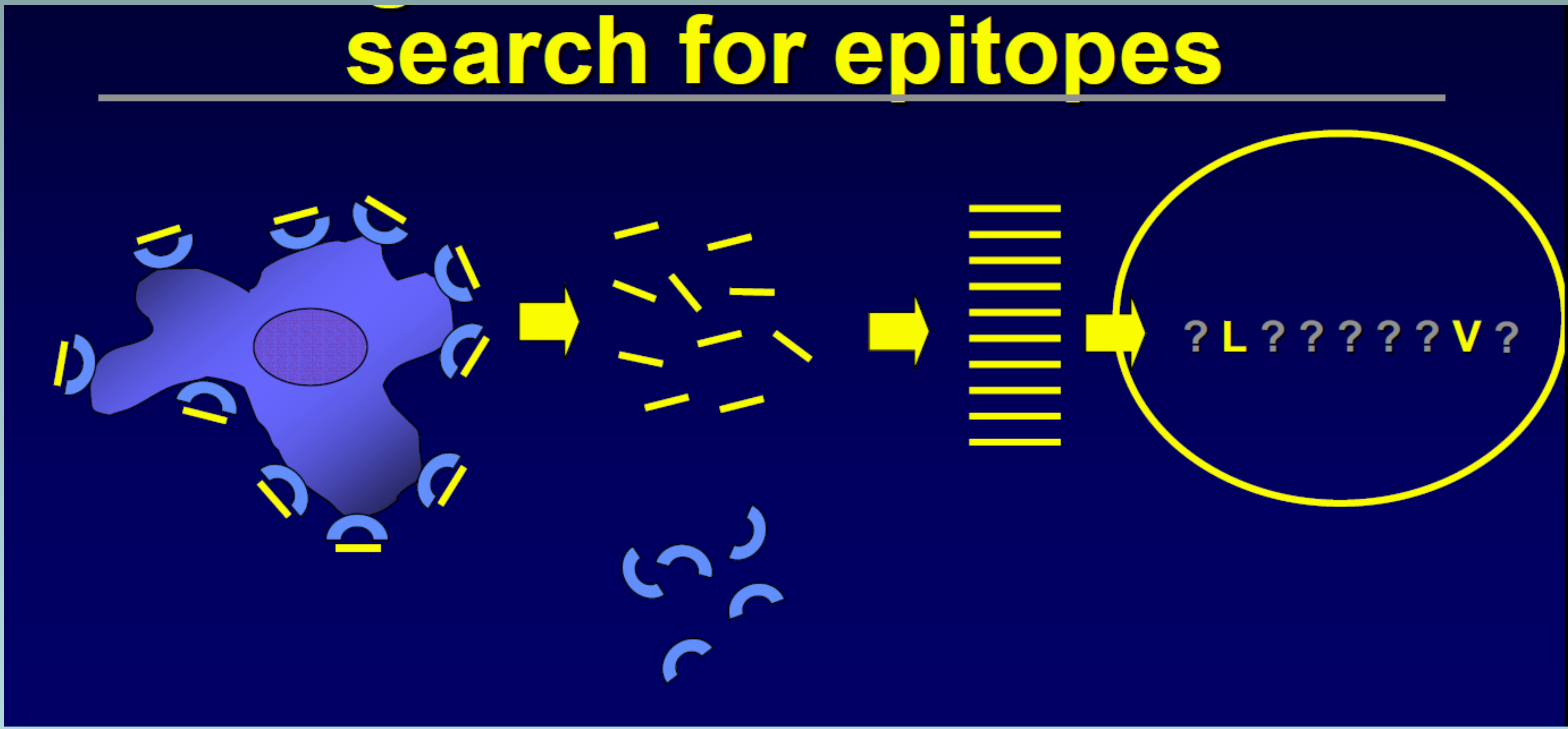
TABLE 7-1 H-2 HAPLOTYPES OF SOME MOUSE STRAINS

			H-2 alleles				
Prototype strain	Other strains with the same haplotype	Haplotype	*K*	*IA*	*IE*	*S*	*D*
CBA	AKR, C3H, B10.BR, C57BR	*k*	*k*	*k*	*k*	*k*	*k*
DBA/2	BALB/c, NZB, SEA, YBR	*d*	*d*	*d*	*d*	*d*	*d*
C57BL/10 (B10)	C57BL/6, C57L, C3H.SW, LP, 129	*b*	*b*	*b*	*b*	*b*	*b*
A	A/He, A/Sn, A/Wy, B10.A	*a*	*k*	*k*	*k*	*d*	*d*
A.SW	B10.S, SJL	*s*	*s*	*s*	*s*	*s*	*s*
A.TL		*t1*	*s*	*k*	*k*	*k*	*d*
DBA/1	STOLI, B10.Q, BDP	*q*	*q*	*q*	*q*	*q*	*q*


(a) Mating of inbred mouse strains with different MHC haplotypes
Homologous chromosomes with MHC loci
H-2^{b} parent
H-2^{k} parent
b/b
k/k
b/b
k/k
b/k
b/k
F$_{1}$ progeny (H-2$^{b/k}$)


(b) Skin transplantation between inbred mouse strains with same or different MHC haplotypes
Parental recipient
Skin graft donor
Progeny recipient
b/b
k/k
b/b
Parent
b/k
b/b
k/k
k/k
Parent
b/k
b/b
k/k
b/k
Progeny
b/k

How to search for T cell epitopes?



Understanding the specificity and sensitivity of the binding process of peptides to the respective MHC is challenging.

Since there are over 512 billion potential binding peptides for each MHC molecule, an empirical approach is not feasible.

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

Immunology Databases

- Management and analysis of immunological data
- Improve the efficiency of immunological research

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

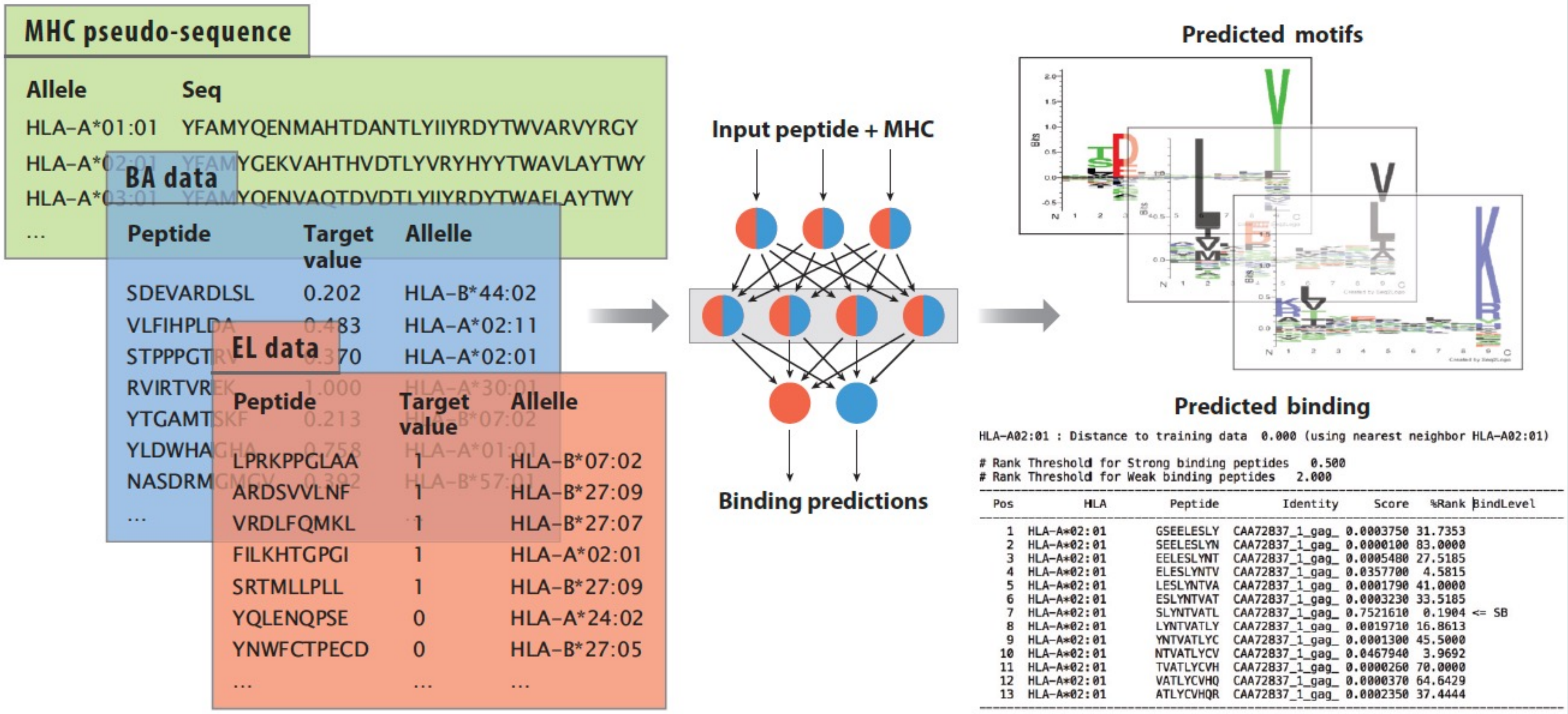


Figure 3

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.

Nielsen M et al. Annu Rev Biomed Data Sci 2020

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

Backert & Kohlbacher. *Genome Med* 2015

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]

Backert & Kohlbacher. *Genome Med* 2015

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]

Backert & Kohlbacher. *Genome Med* 2015

Annual Review of Biomedical Data Science

Immunoinformatics: Predicting Peptide–MHC Binding

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.

Nielsen M et al. Annu Rev Biomed Data Sci 2020

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
mhcflurry	ANN	32 or 64 hidden neurons; trained on 9-mer peptides	IEDB—2014
mhcflurry-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

Zhao W & Sher X. *PLOS Comput Biol* 2018

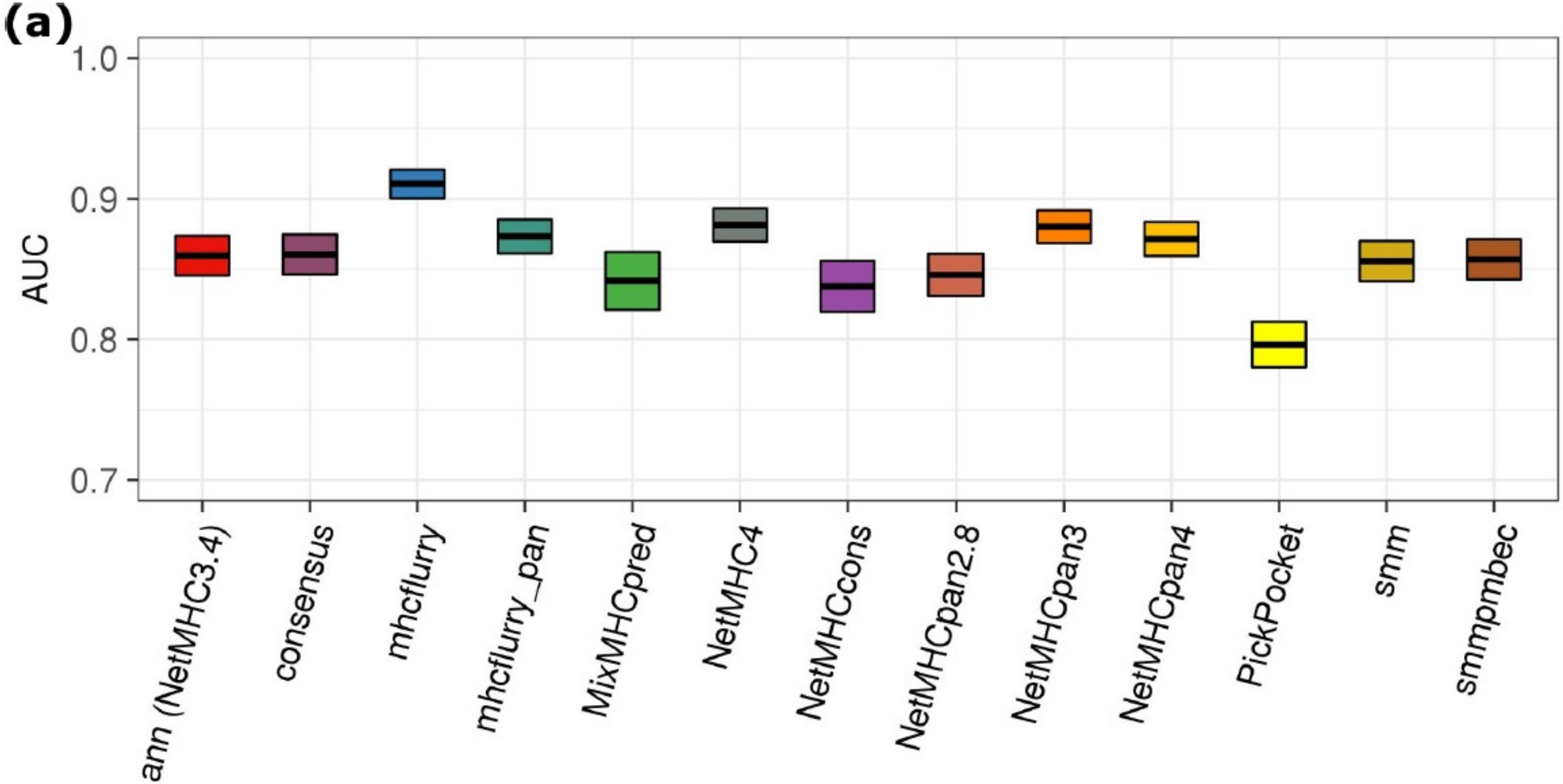
MHC Class II Binding Predictor

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

Zhao W & Sher X. *PLOS Comput Biol* 2018



Zhao W & Sher X. *PLOS Comput Biol* 2018

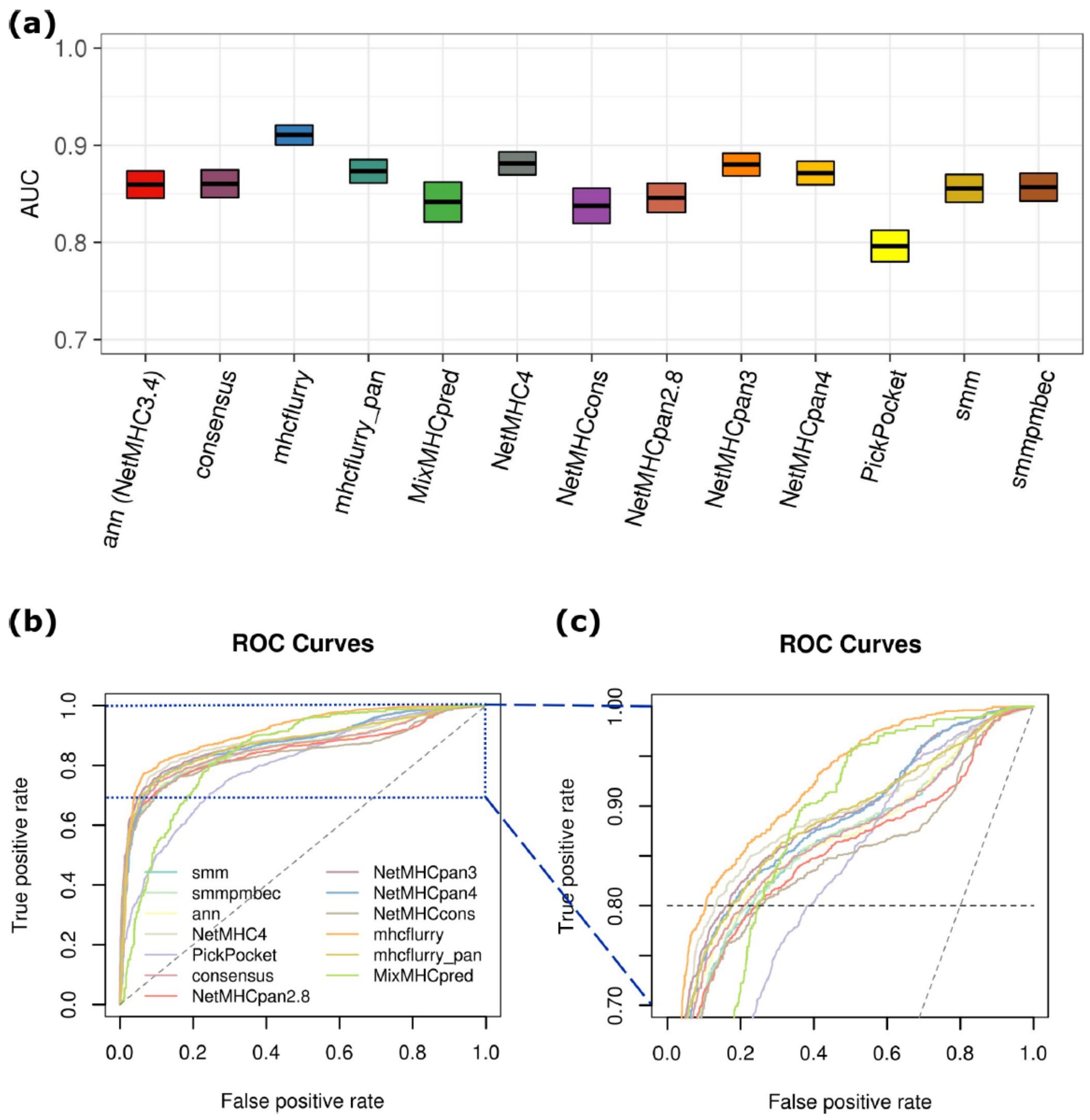


Fig 1. Binary classification (binder vs. non-binder) performance. (a) AUC of MHC-I binding epitope prediction tools. (b) ROC curves. IC50 = 500 nM was used as the cutoff for classifying experimentally measured epitopes. AUC was shown by box plot with upper and lower boundaries covering confidence level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

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

Zhao W & Sher X. *PLOS Comput Biol* 2018

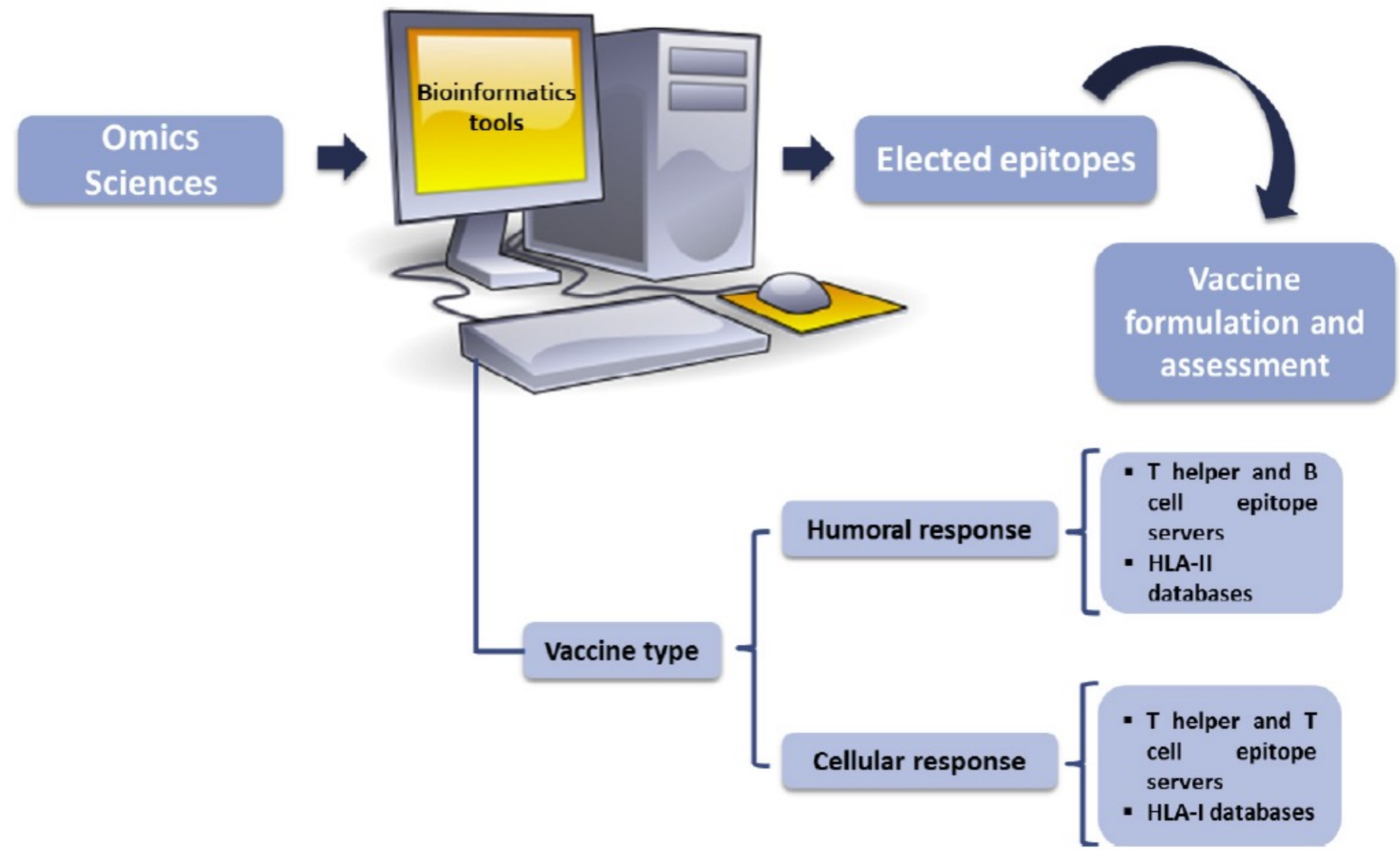


Fig. 1. Schematic representation of the workflow to identify epitopes for vaccine development.

Soria-Guerra RE et al. *J Biomed Inf* 2015

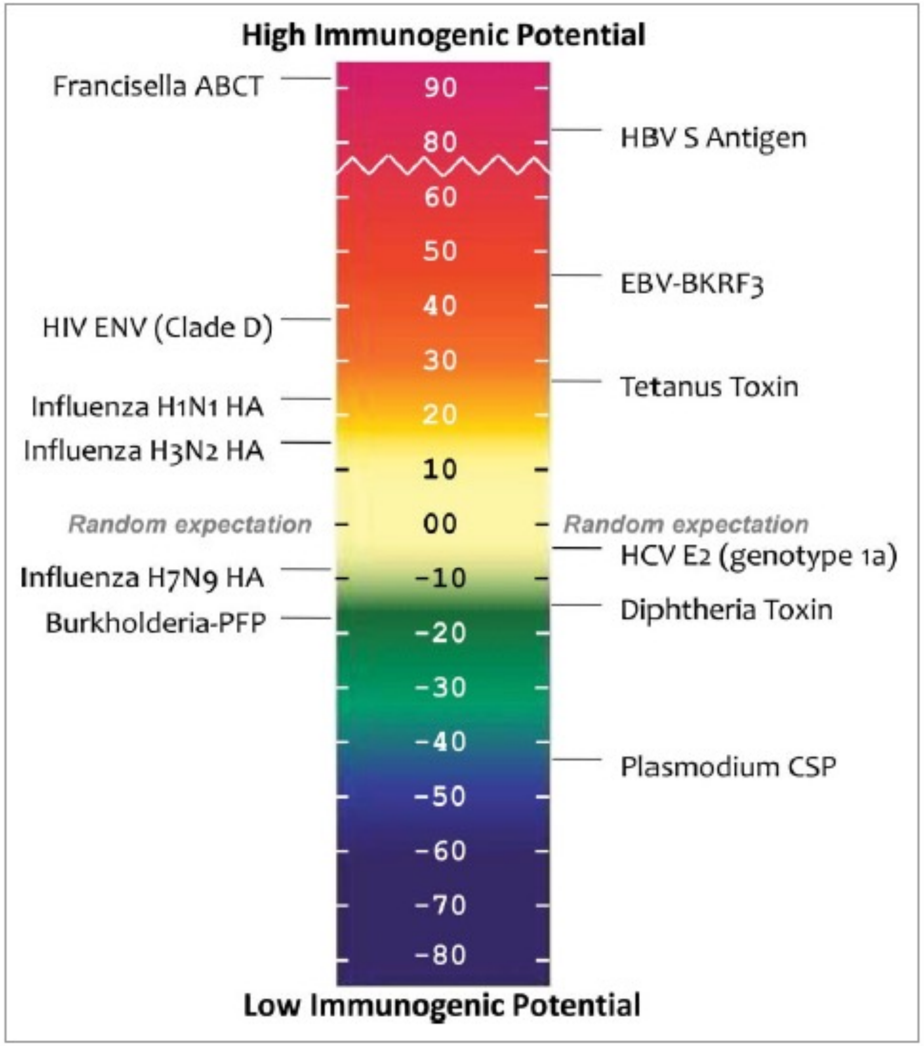


Figure 6. EpiMatrix protein immunogenicity scale. EpiMatrix protein immunogenicity scores higher than +20 are considered to be potentially immunogenic. On the left of the scale are well-known proteins for comparison. Low-scoring proteins near the bottom of the scale are known to engender little to no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

Epitope prediction and identification- adaptive T cell responses in humans

John Sidney[a], Bjoern Peters[a,b], Alessandro Sette[a,b,*]

[a] Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology, 9420 Athena Circle, La Jolla, CA, 92037, USA
[b] Department of Medicine, Division of Infectious Diseases and Global Public Health, University of California, San Diego, La Jolla, CA, 92037, USA

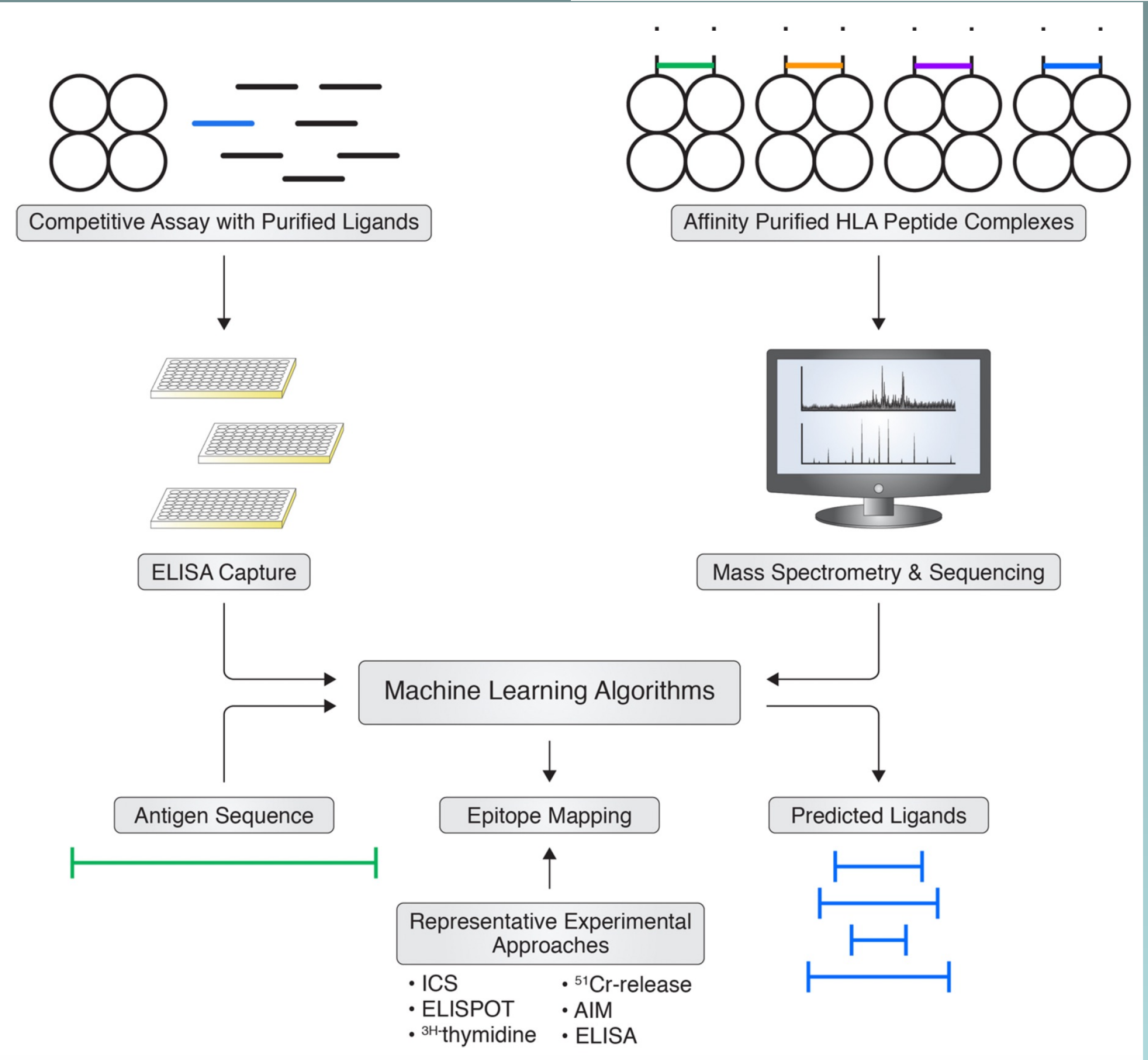


Fig. 3. Epitope predictions based on HLA interactions. The development of bioinformatic tools to aid in identification of T cell epitopes is based on data generated from, for example, HLA-ligand assays or by mass spectrometry analysis of eluted ligands. This data is then utilized to develop machine learning tools to predict potential HLA binding peptides, and in turn candidate T cell epitopes.

Sidney et al. *Semin Immunol* 2020

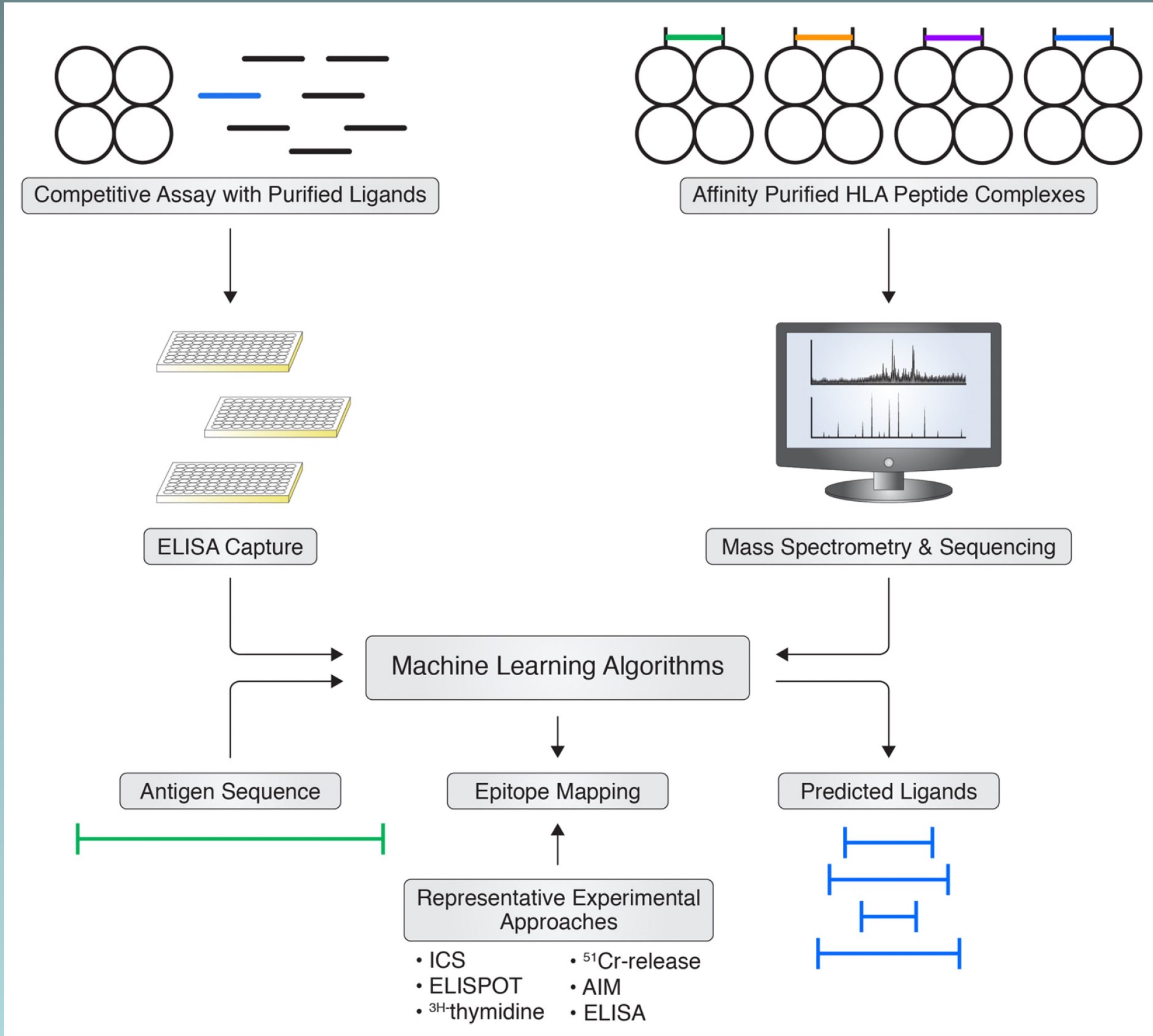


Fig. 3. Epitope predictions based on HLA interactions. The development of bioinformatic tools to aid in identification of T cell epitopes is based on data generated from, for example, HLA-ligand assays or by mass spectrometry analysis of eluted ligands. This data is then utilized to develop machine learning tools to predict potential HLA binding peptides, and in turn candidate T cell epitopes.

Sidney et al. *Semin Immunol* 2020

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).

Sidney et al. *Semin Immunol* 2020

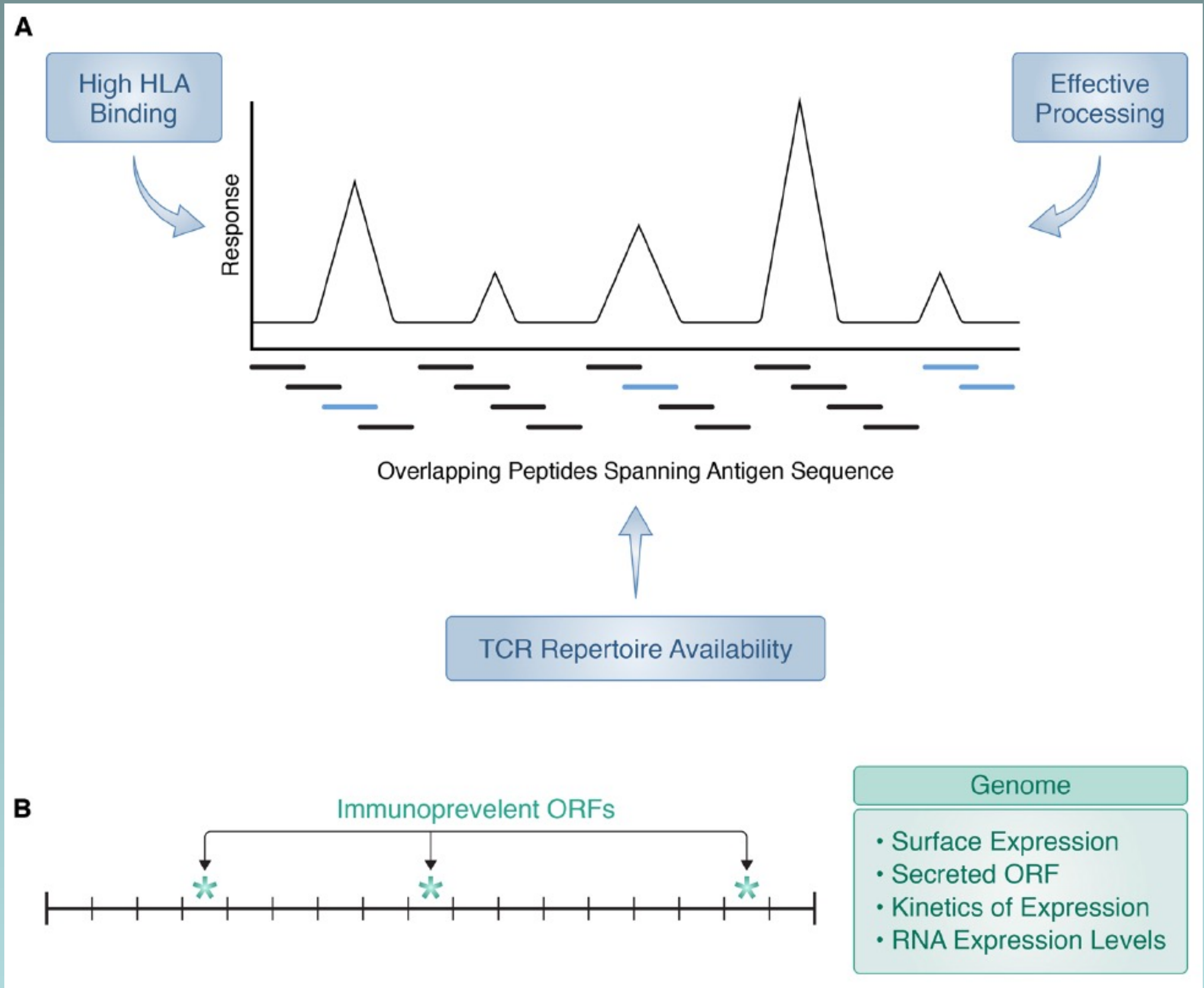


Fig. 4. Immunodominance and immunoprevalence. A) Epitopes effectively generated by processing and for which a TCR repertoire is available (light blue bars), and capable of eliciting the (relatively) strongest T cell responses, are termed immunodominant. B) Epitopes, antigens or ORFs that elicit responses 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.

Sidney et al. *Semin Immunol* 2020

SYFPEITHI

www.syfpeithi.de

Immunogenetics (1999) 50:213–219

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

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.

Immunogenetics (1999) 50:213–219

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)

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

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

Actin-related protein 2 (ARP2; human)

AA pos.	Score	Sequence
378	19	YQEKGVRVL
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	GSIYNPEVL

Elongation factor 2 (EF2; human)

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

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

	Position 1	2	3	4	5	6	7	8	9	Reference
Anchors	**L**			**D**		**K**			**Y**	Malcherek et al. 1993; Geluk et al. 1992, 1994
	I					**R**			**L**	
	F					**E**			**F**	
	M					**Q**				
	V					**N**				

Peptide	Source protein	Reference
Examples for ligands		
ISNQLTLDSNTKYFHKLN	Apolipoprotein B-100 (2877–2894)	Malcherek et al. 1993
ISNQLTLDSNTKYFHKL	Apolipoprotein B-100 (2877–2893)	Malcherek et al. 1993
ISNQLTLDSNTKYFHK	Apolipoprotein B-100 (2877–2892)	Malcherek et al. 1993
VDTFLEDVKNLYHSEA	α1-Antitrypsion (149–164)	Malcherek et al. 1993
KPRAIVVDPVHGFMY	LDL receptor (518–532)	Malcherek et al. 1993
KQTISPDYRNMI	IG2a (384–395)	Malcherek et al. 1993
YPDFIMDPKEKDKV	Unknown	Malcherek et al. 1993
NIQLINDQEVARFD	Unknown	Malcherek et al. 1993
LLSFVRDLNQYRADI	Transferrin receptor (618–632)	Malcherek et al. 1993
LPKPPKPVSKMRMATPL	Invariant chain (97–111)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
LPKPPKPVSKMRMATPLLMQALP	Invariant chain (97–119)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
LPKPPKPVSKMRMATPLLMQALPM	Invariant chain (97–120)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
PKPPKPVSKMRMATPL	Invariant chain (98–113)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
PKPPKPVSKMRMATPLLMQA	Invariant chain (98–117)	Riberdy et al. 1992; Chicz et al. 1993:Sette et al. 1992
KPPKPVSKMRMATPLLMQ	Invariant chain (99–116)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
KPPKPVSKMRMATPLLMQALPM	Invariant chain (99–119)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
VDDTQFVRFDSDAASQ	HLA-A30 (52–67)	Chicz et al. 1993
ATKYGNMTEDHVMHLLQNA	Invariant chain (131–149)	Chicz et al. 1993
VFLLLLADKVPETSLS	ACh receptor (289–304)	Chicz et al. 1993
LNKILLDEQAQWK	ICAM-2 (64–76)	Chicz et al. 1993
GPPKLDIRKEEKQIMIDIFH	IFNγ receptor (128–147)	Chicz et al. 1993
GPPKLDIRKEEKQIMIDIFHP	IFNγ receptor (128–148)	Chicz et al. 1993
GFKAIRPDKKSNPIIRTV	Cyt-B5 (155–172)	Chicz et al. 1993
YANILLDRRVPQTDMTF	Apolipoprotein B-100 (1207–1224)	Chicz et al. 1993
NLFLKSDGRIKYTLNKNSLK	Apolipoprotein B-100 (1276–1295)	Chicz et al. 1993
IPDNLFLKSDGRIKYTLNKN	Apolipoprotein B-100 (1273–1292)	Chicz et al. 1993
IPDNLFLKSDGRIKYTLNK	Apolipoprotein B-100 (1273–1291)	Chicz et al. 1993
IPDNLFLKSDGRIKYTLN	Apolipoprotein B-100 (1273–1290)	Malcherek et al. 1993; Chicz et al. 1993
IPDNLFLKSDGRIKYTL	Apolipoprotein B-100 (1273–1289)	Chicz et al. 1993
NLFLKSDGRIKYTLNK	Apolipoprotein B-100 (1276–1291)	Chicz et al. 1993
NLFLKSDGRIKYTLN	Apolipoprotein B-100 (1276–1290)	Chicz et al. 1993
VTTLNSDLKYNALDLTN	Apolipoprotein B-100 (1294–1310)	Chicz et al. 1993
VGSDWRFLRGYHQYA	HLA-A2 (103–117)	Chicz et al. 1993
T-cell epitopes		
GDVVAVVDIKEKGKDKWIELK	Lol pol. P1 (171–190)	Geluk et al. 1994
KTIAYDEEARR	HSP65 (cattle) (3–13)	Hawes et al. 1995
MGRSIKVQLQ	M. tuberculosis 30/31 kD protein (56–65)	Geluk et al. 1997
SDKNPLFLDEQLI	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)!

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

Hold down ctrl key when clicking
to select multiple items

2. Choose a mer

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

3. Paste your sequence here:

Max. input 2048 aminoacids (letters)!
Letters only, no numbers or non-ASCII-symbols please.
You may use 'SYFPEITHI' with H2-Kd to see an example.

```
SYFPEITHI
```

4. Choose Run to start analysis

Run Reset Home

IEDB - AR

www.iedb.org

Published online 31 May 2008 *Nucleic Acids Research, 2008, Vol. 36, Web Server issue* ***W513–W518***
doi:10.1093/nar/gkn254

Immune epitope database analysis resource (IEDB-AR)

Qing Zhang[1], Peng Wang[1], Yohan Kim[1], Pernille Haste-Andersen[2], John Beaver[3], Philip E. Bourne[3], Huynh-Hoa Bui[1], Soren Buus[2], Sune Frankild[2], Jason Greenbaum[1], Ole Lund[2], Claus Lundegaard[2], Morten Nielsen[2], Julia Ponomarenko[3], Alessandro Sette[1], Zhanyang Zhu[3] and Bjoern Peters[1,*]

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

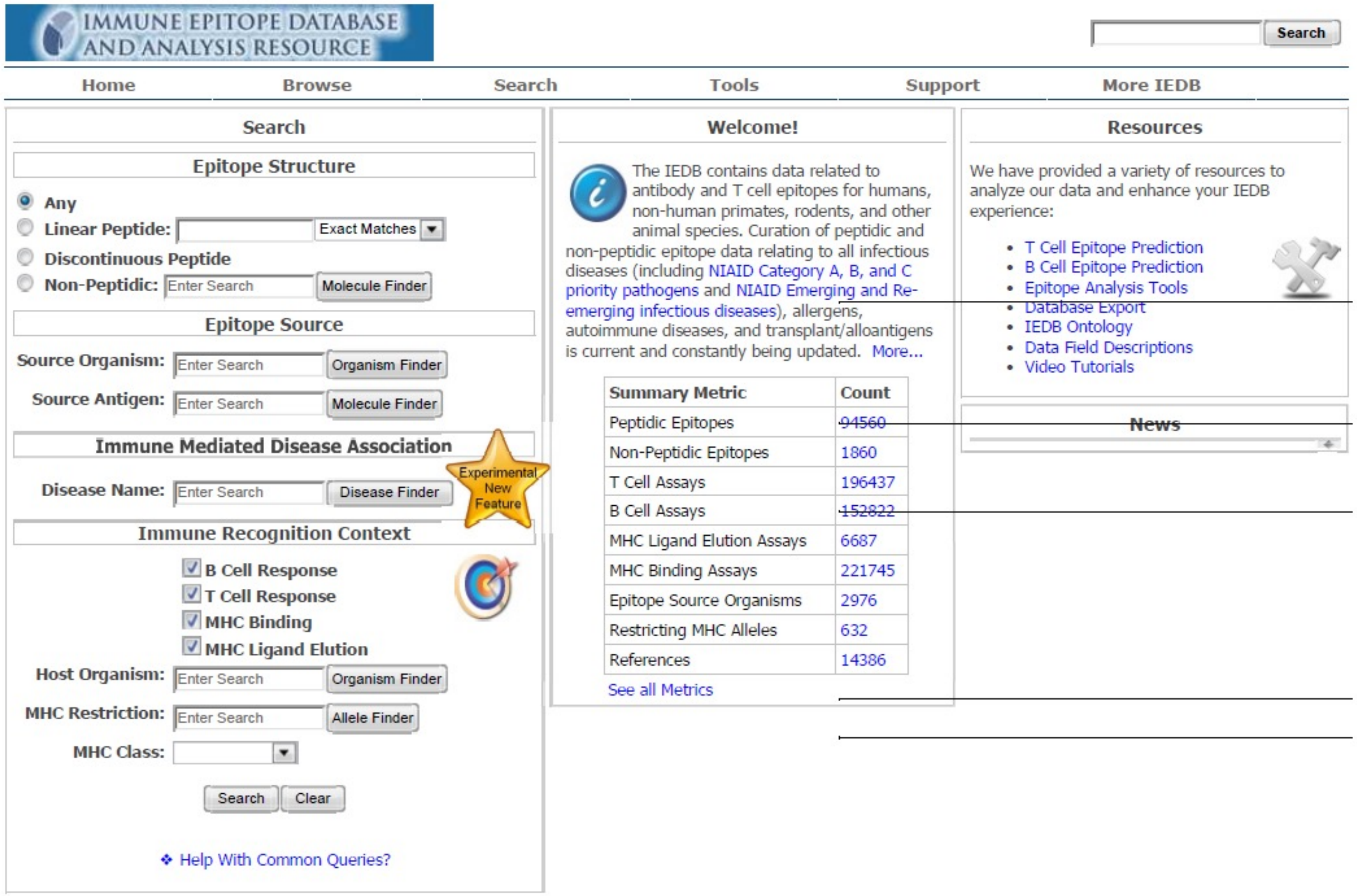
Received January 31, 2008; Revised April 14, 2008; Accepted April 20, 2008

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.

www.iedb.org


IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE
Search
Home
Browse
Search
Tools
Support
More IEDB
Search
Epitope Structure
Any
Linear Peptide:
Exact Matches
Discontinuous Peptide
Non-Peptidic:
Enter Search
Molecule Finder
Epitope Source
Source Organism:
Enter Search
Organism Finder
Source Antigen:
Enter Search
Molecule Finder
Immune Mediated Disease Association
Experimental New Feature
Disease Name:
Enter Search
Disease Finder
Immune Recognition Context
B Cell Response
T Cell Response
MHC Binding
MHC Ligand Elution
Host Organism:
Enter Search
Organism Finder
MHC Restriction:
Enter Search
Allele Finder
MHC Class:
Search
Clear
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

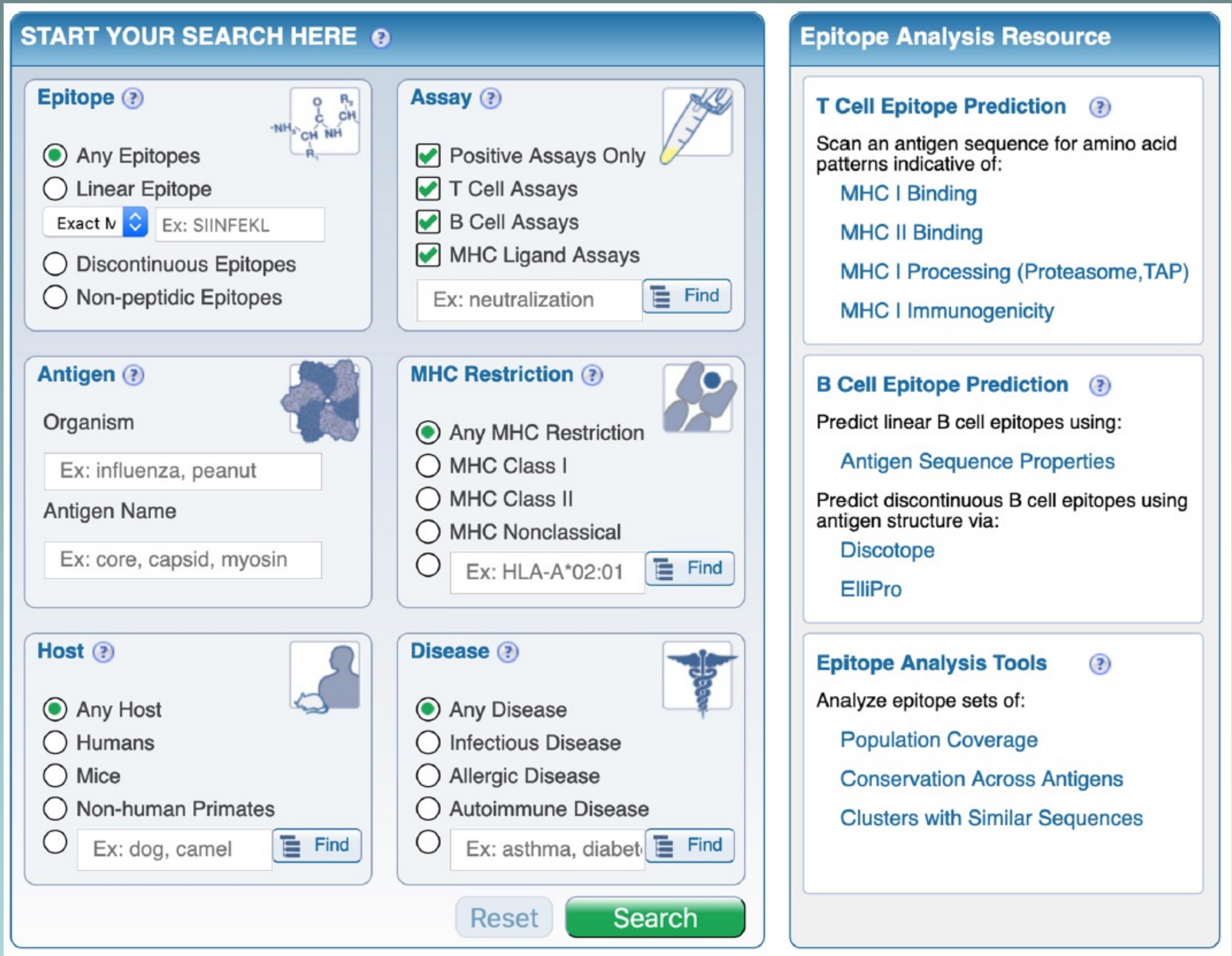


Fig. 2. Metadata associated with epitope identification capture by the IEDB. The IEDB homepage (www.iedb.org), and initial search fields are shown. The IEDB is an NIH-NIAID funded publicly available database of T and B cell epitopes curated from the published literature or by direct submission from NIH-NIAID 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).

Sidney et al. *Semin Immunol* 2020

frontiers in Immunology

REVIEW
published: 14 March 2017
doi: 10.3389/fimmu.2017.00278

Check for updates

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).

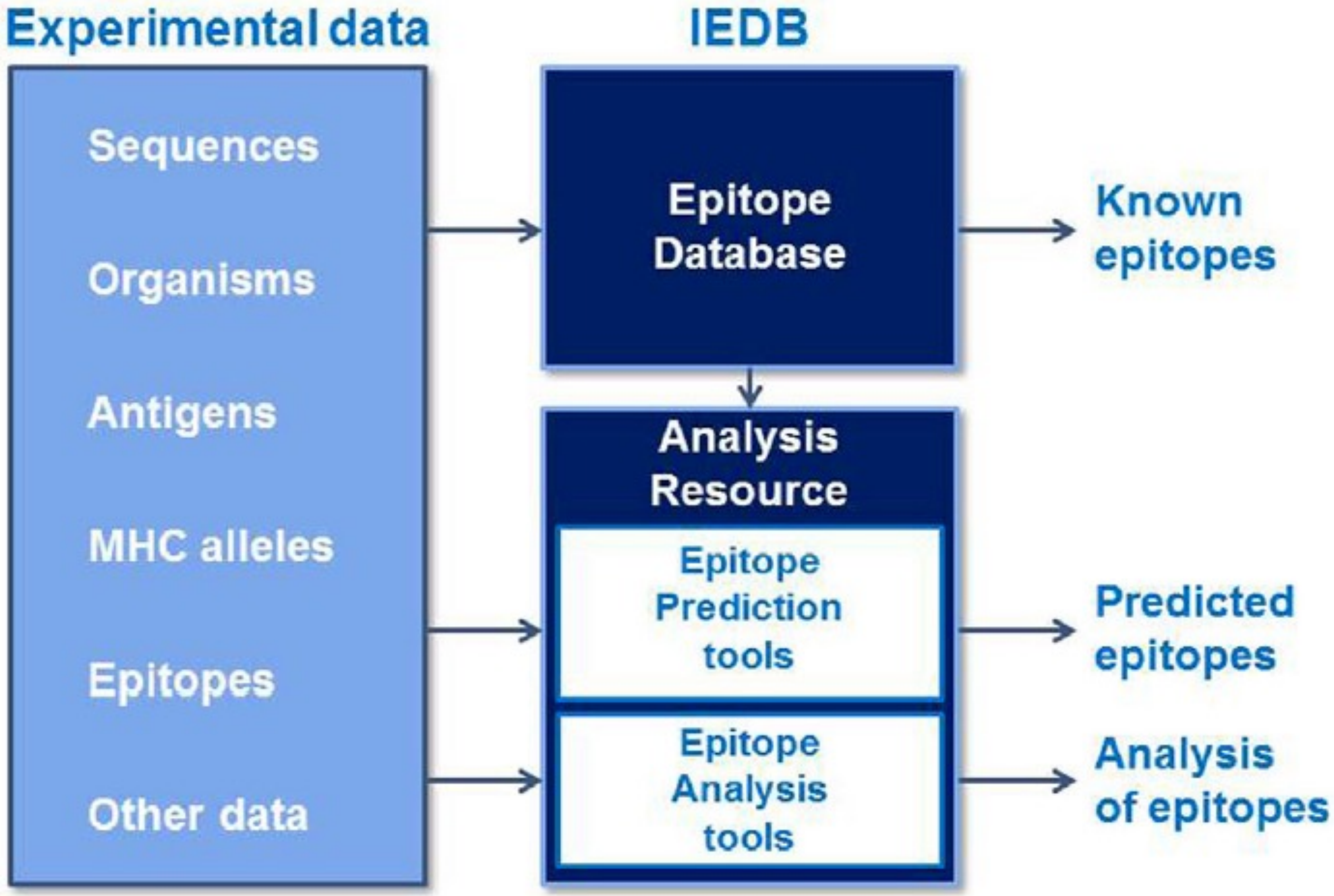


FIGURE 1 | The Immune Epitope Database and Analysis Resource captures experimental epitope data in a database and makes known epitopes freely available to the research community. These data are used to train epitope prediction tools in the Analysis Resource, which also contains tools to analyze sets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

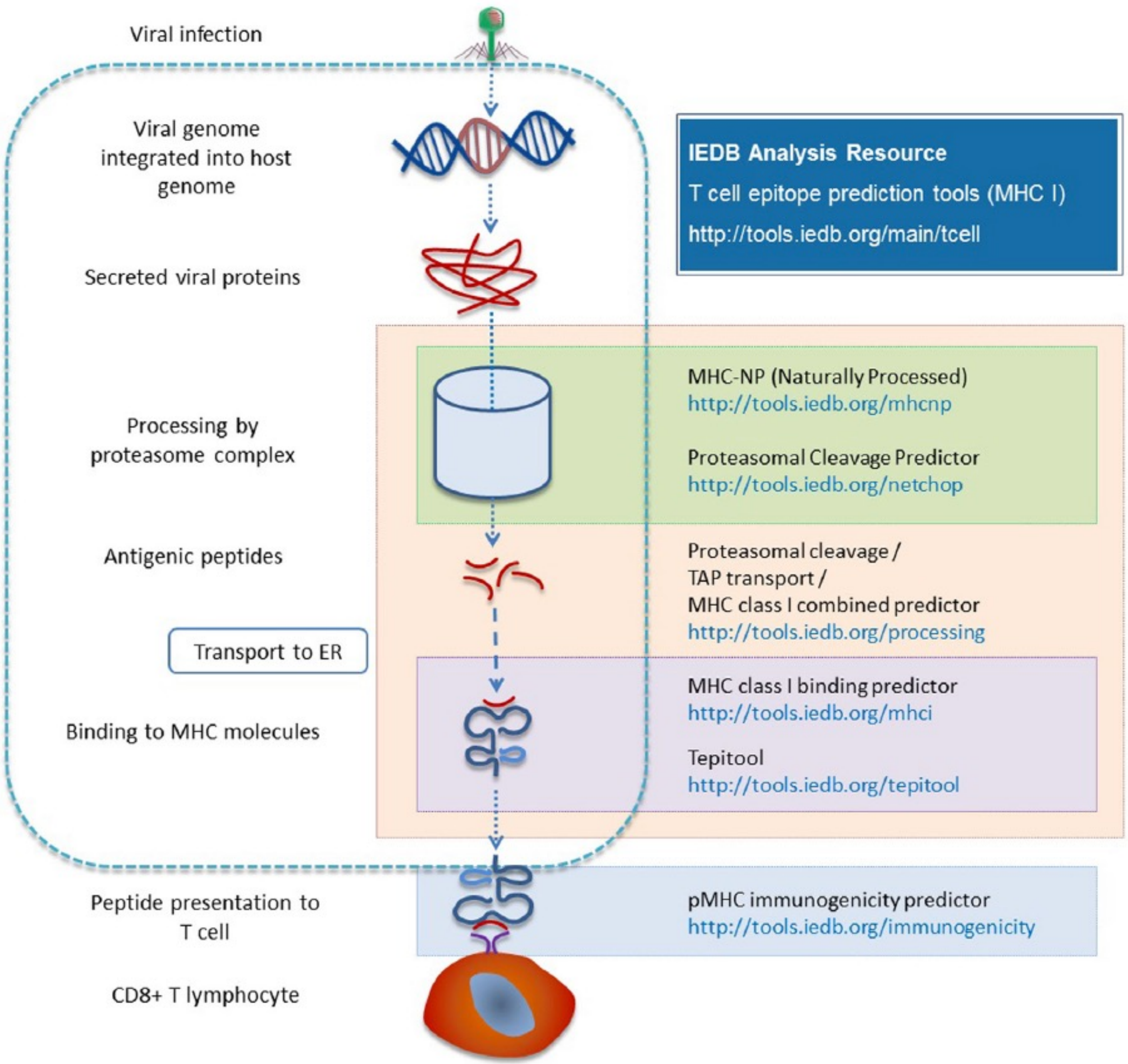


FIGURE 3 | Different prediction tools are available in the Analysis Resource with respect to different stages of MHC I antigen processing.

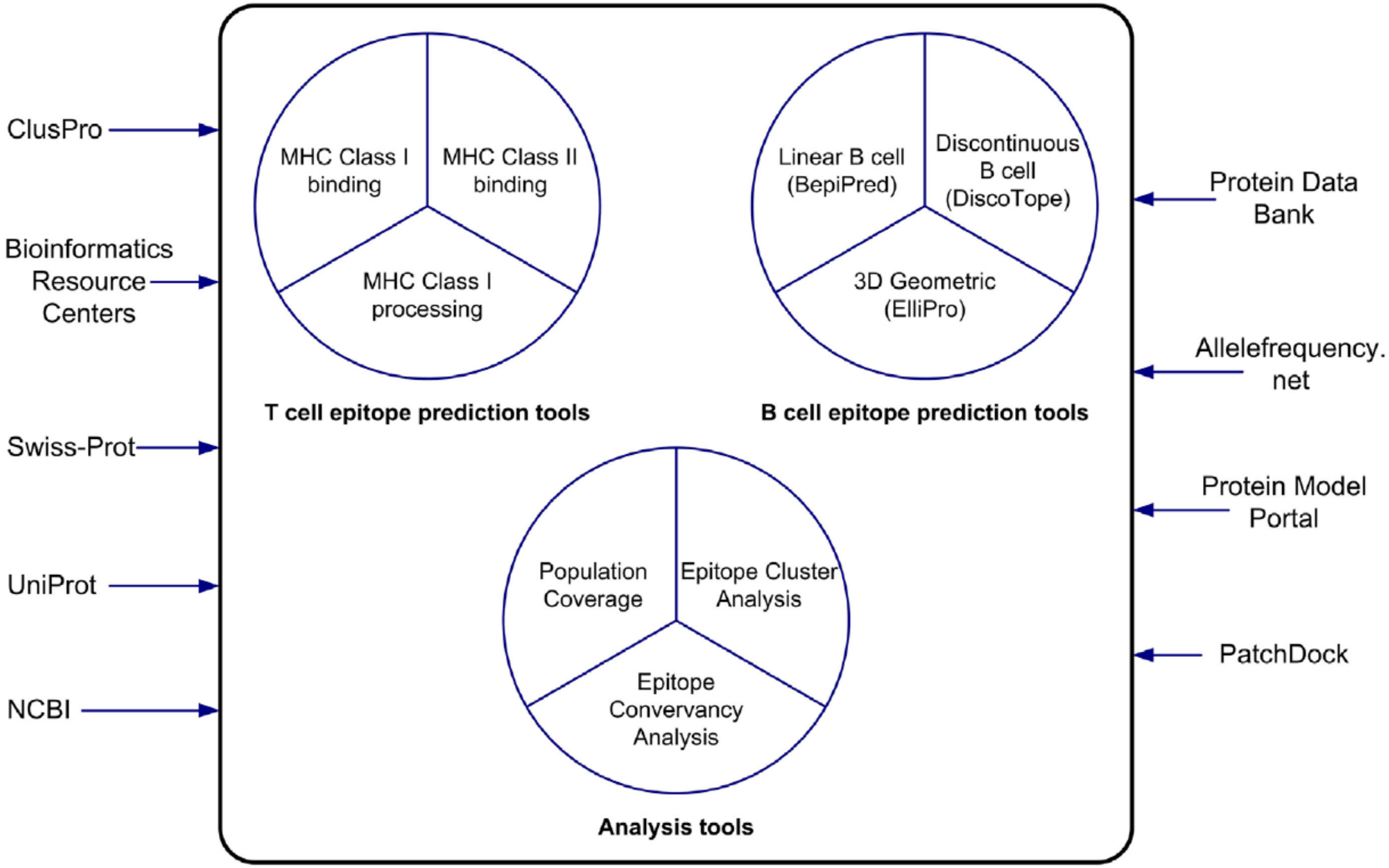


FIGURE 6 | The tools of the Analysis Resource can be used to predict T cell and B cell epitopes and to analyze sets of epitopes. The Analysis Resource interacts with a range of bioinformatics resources.

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

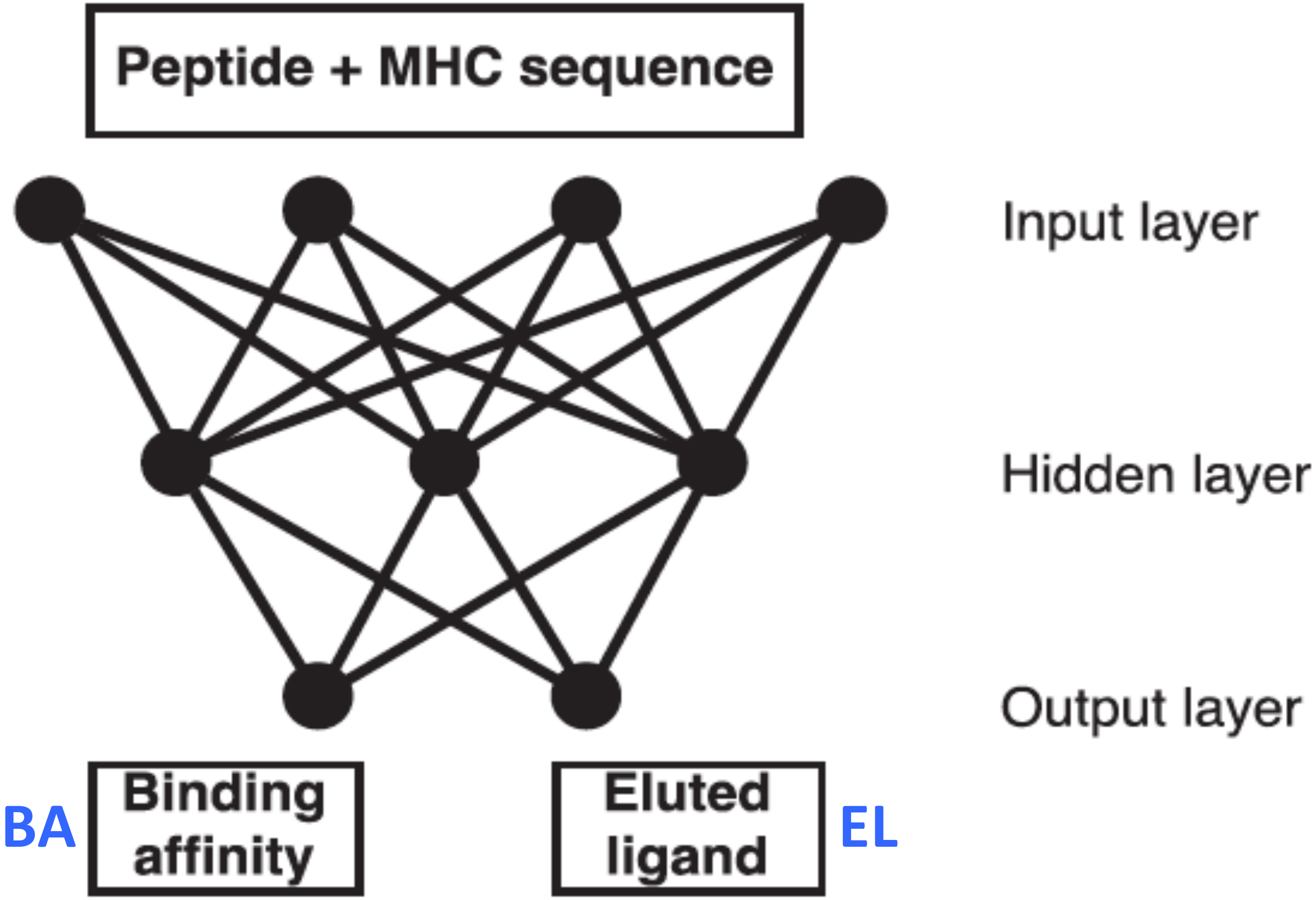


FIGURE 1. Visualization of the neural networks with two output neurons used for combined training on BA and EL data.

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-retrained, 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/

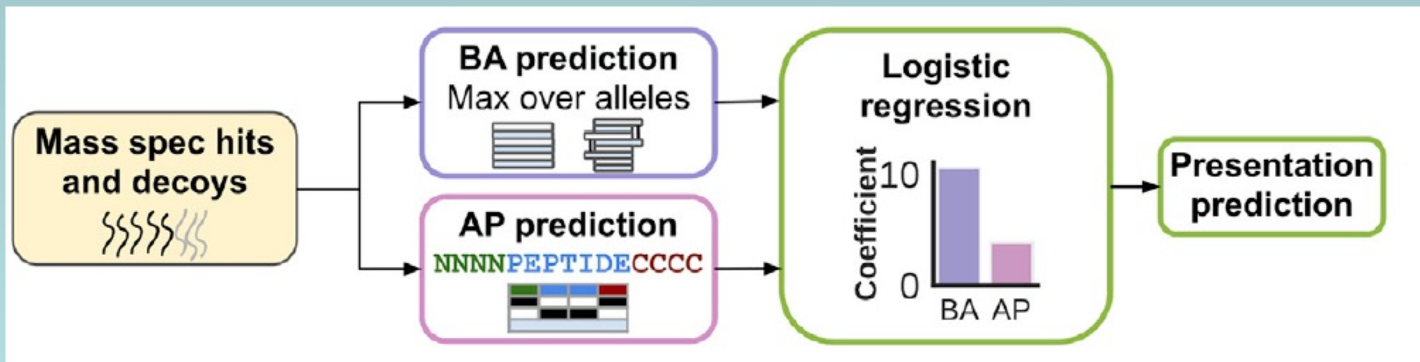
Cell Systems

CellPress
OPEN ACCESS

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]
[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[3]Department of Genetics, UNC School of Medicine, Chapel Hill, NC 27599, USA
[4]Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[5]Lead Contact
*Correspondence: tim@openvax.org
https://doi.org/10.1016/j.cels.2020.06.010



BA: binding affinity and AP: antigen processing

PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

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

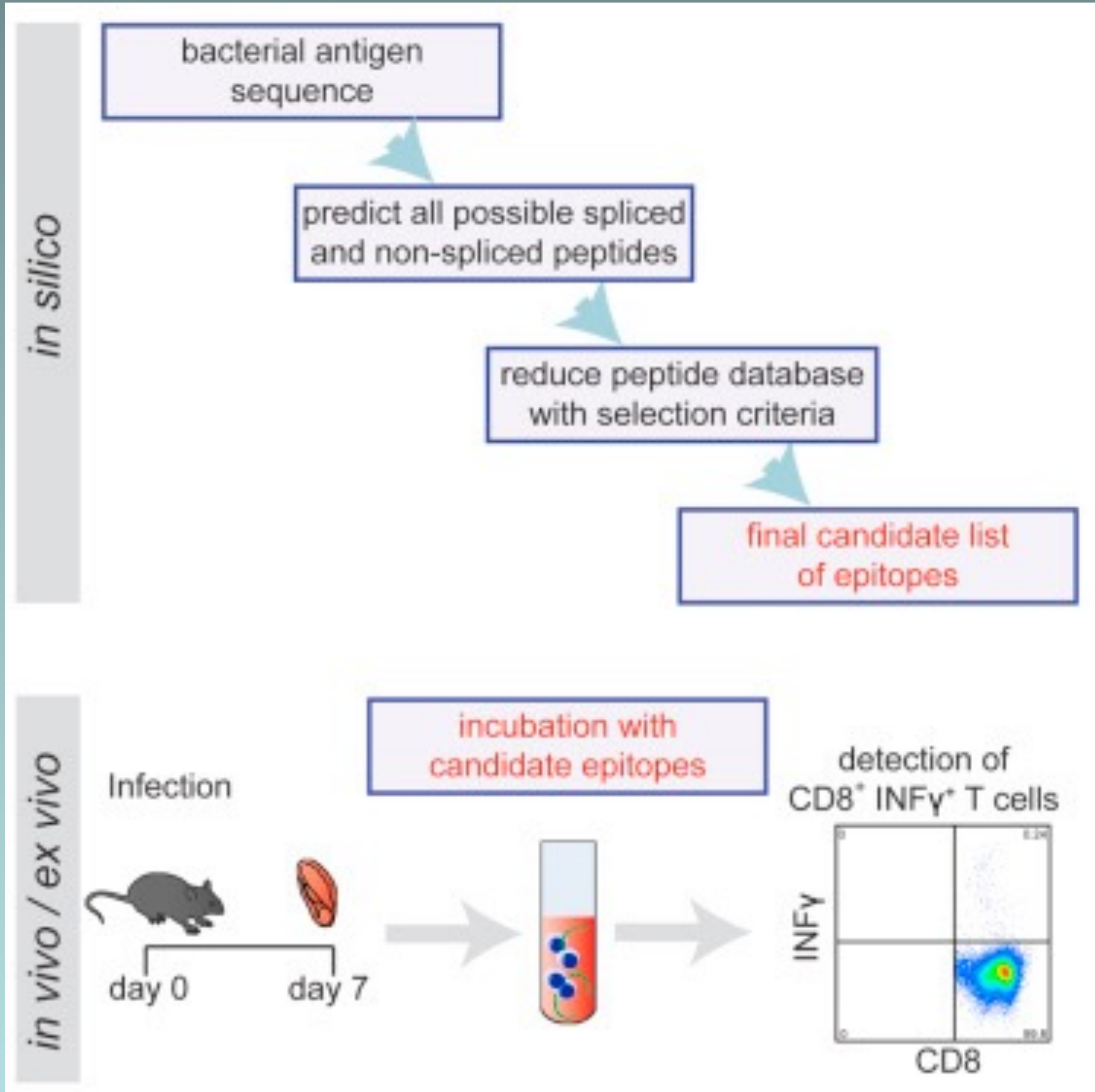
Weilong Zhao, Xinwei Sher *

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

* xinwei_sher@merck.com

PLOS Computational Biology | https://doi.org/10.1371/journal.pcbi.1006457 November 8, 2018

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
INFγ
CD8

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).

DTU Health Tech

Department of Health Technology

Contact

Research Education Collaboration Services and Products News About

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)

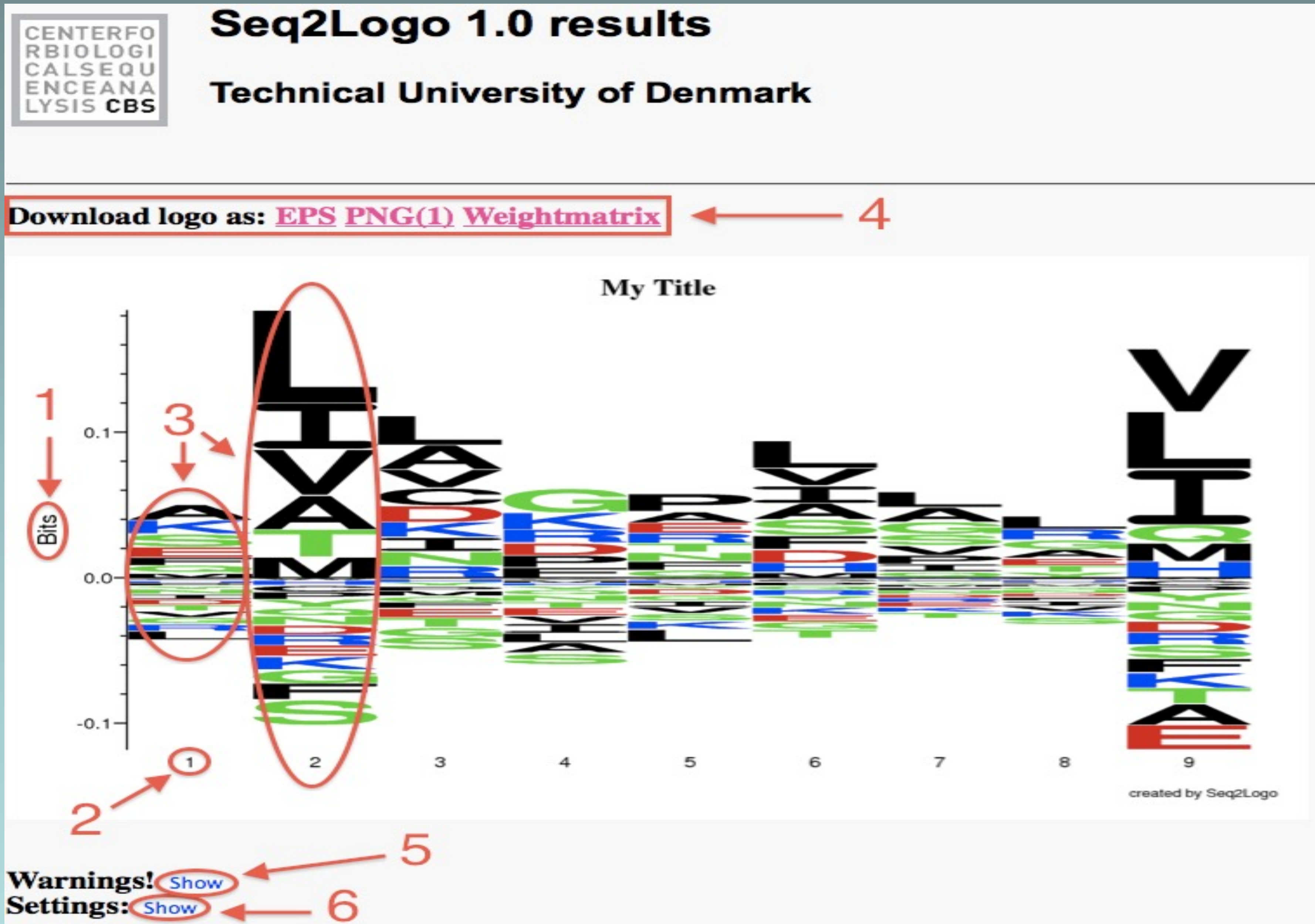
Weight on prior (pseudo counts): 200

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)

JPEG


CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS
Seq2Logo 1.0 results
Technical University of Denmark
Download logo as: EPS PNG(1) Weightmatrix
4
My Title
1
Bits
3
0.1
0.0
-0.1
1
2
3
4
5
6
7
8
9
2
created by Seq2Logo
5
Warnings! Show
Settings: Show
6

BepiBlast http://imath.med.ucm.es/bepiblast/

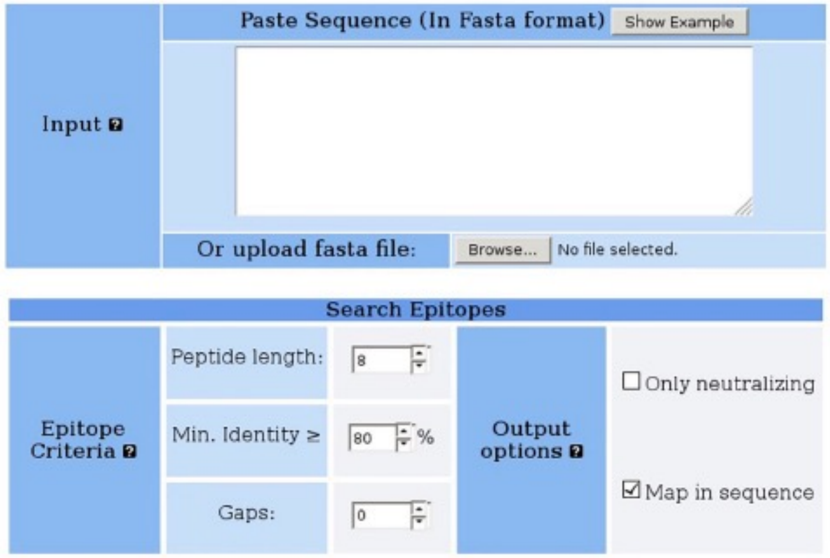
Prediction of B cell epitopes in proteins using a novel sequence similarity-based method

Alvaro Ras-Carmona[1], Alexander A. Lehmann[1,2], Paul V. Lehmann[2] & 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 ≥ 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 ± 0.27%, 81.59 ± 0.37% and 63.49 ± 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



b

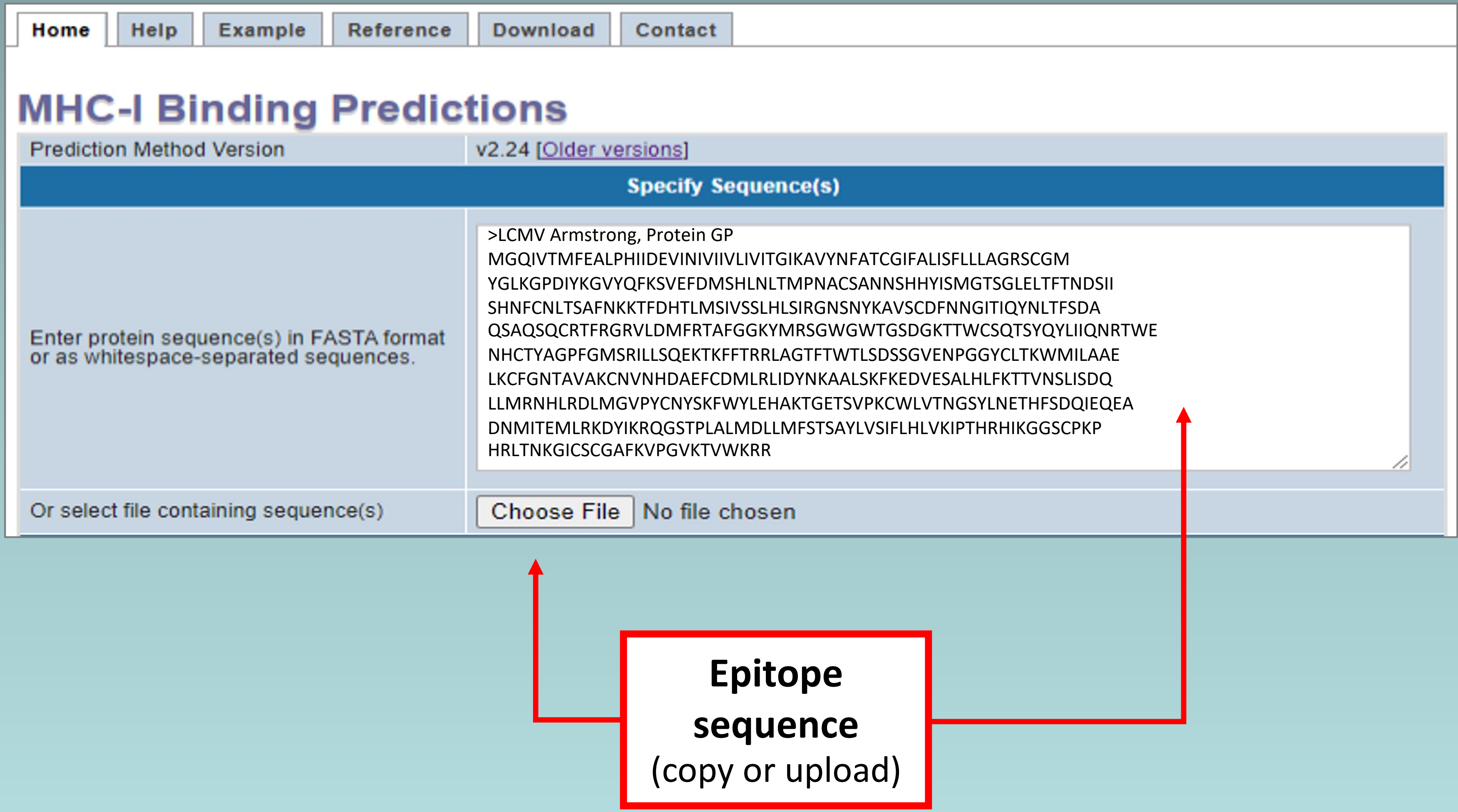
Start	End	Predicted Epitope	Scores	Accsb	Flexb
39	52	GTLVKTITDDQIEV	68	0.22	0.34
79	93	DCTLIDALLGDPHCD	65	0.19	-0.50
101	108	DLFVERSK	43	0.20	-0.07
110	126	FSNCYPYDVPDYASLRS	96	0.19	0.36
143	150	WTGVTQNG	46	0.27	0.34
161	170	SGFFSRLNWL	57	0.21	-0.41
197	205	GIHHPSTNQ	56	0.23	-0.56
230	237	IPNIGSRP	48	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	LFGAIAGFIENGWE	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	HTIDLTDSEMN	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

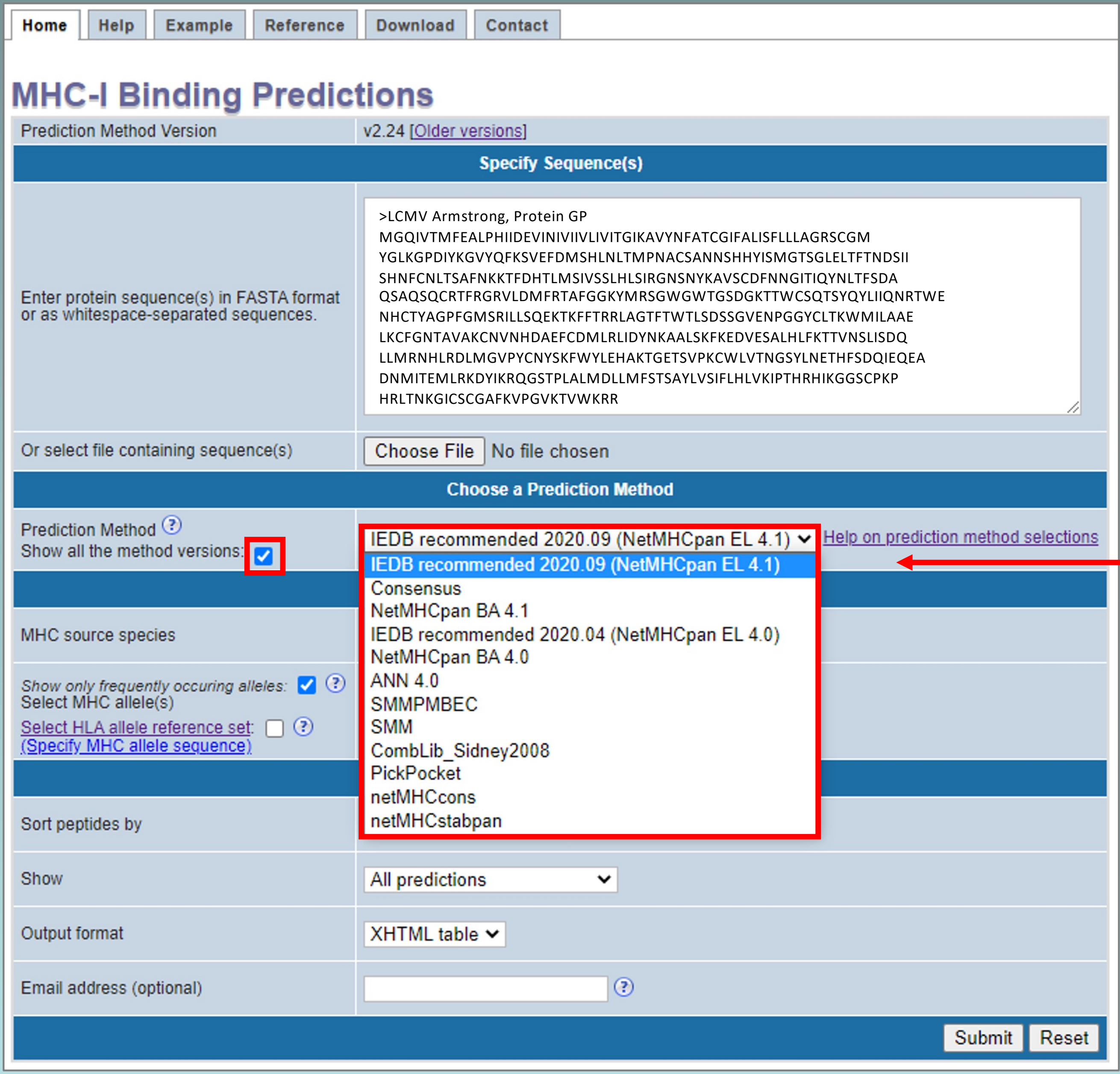
MHC-I binding prediction - example

tools.iedb.org/mhci/



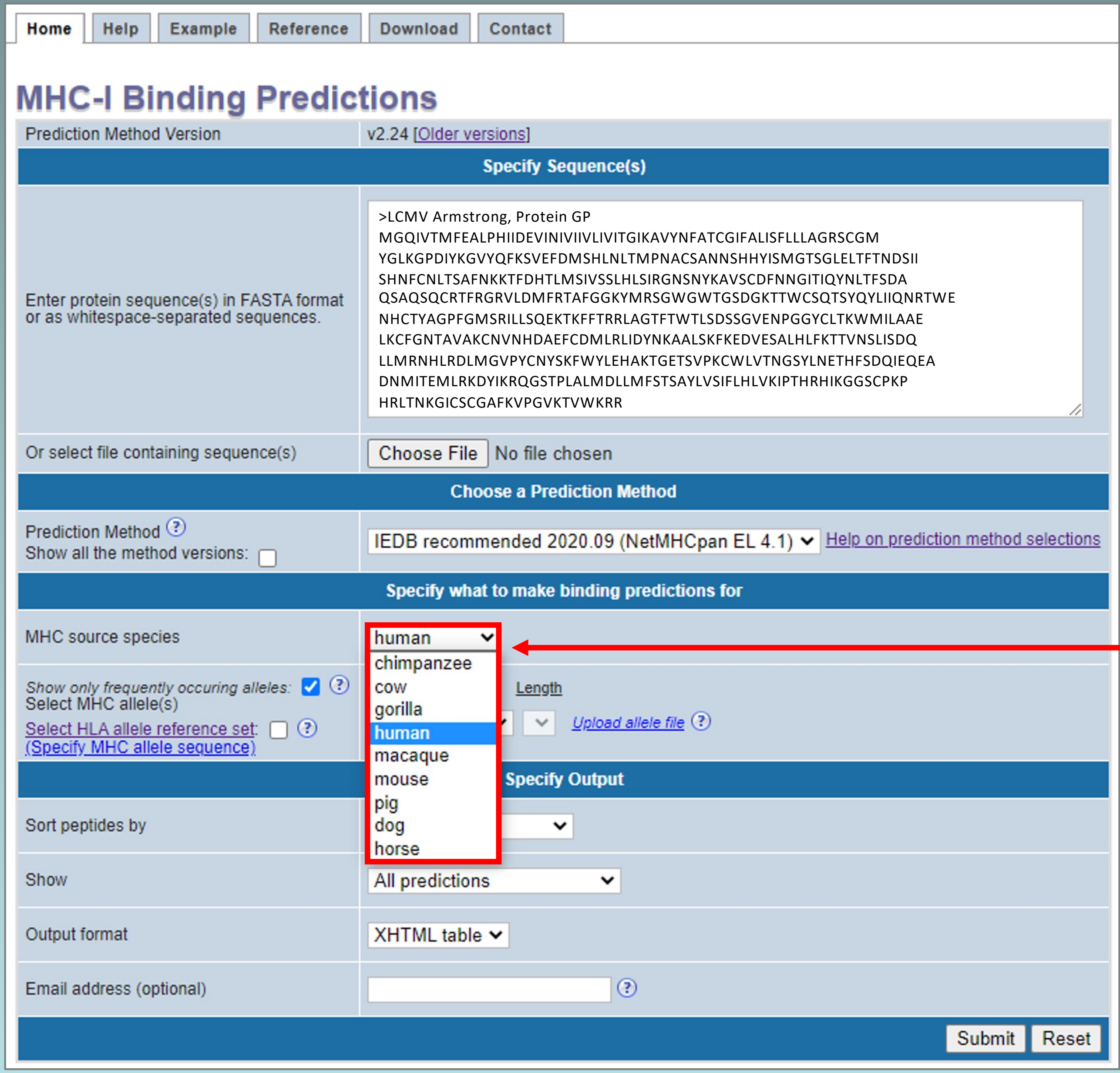
MHC-I binding prediction - example

tools.iedb.org/mhci/



Prediction method

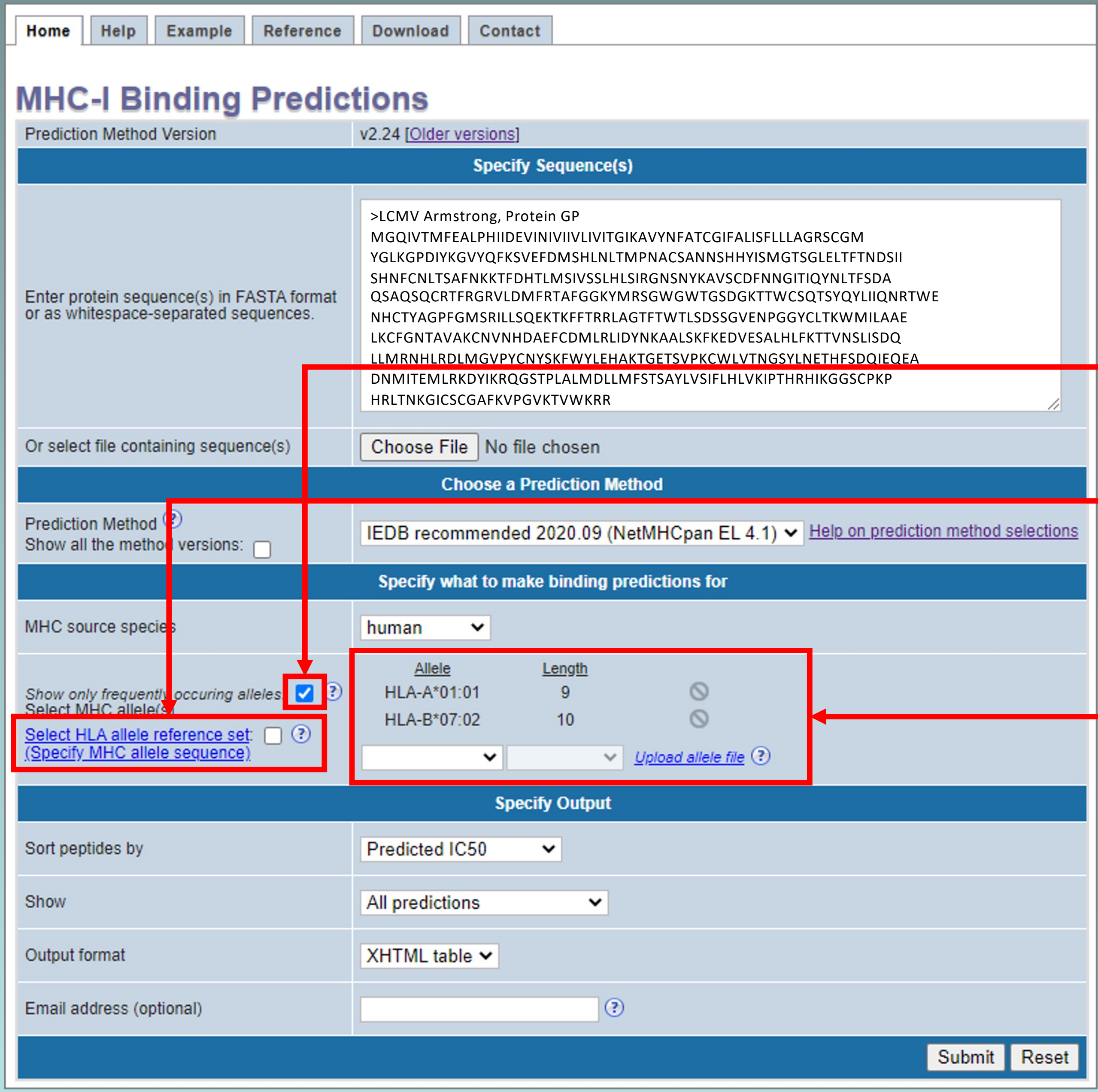
MHC-I binding prediction – example



tools.iedb.org/mhci/

Choose species

MHC-I binding prediction – example



tools.iedb.org/mhci/

Complete set

Reference alleles

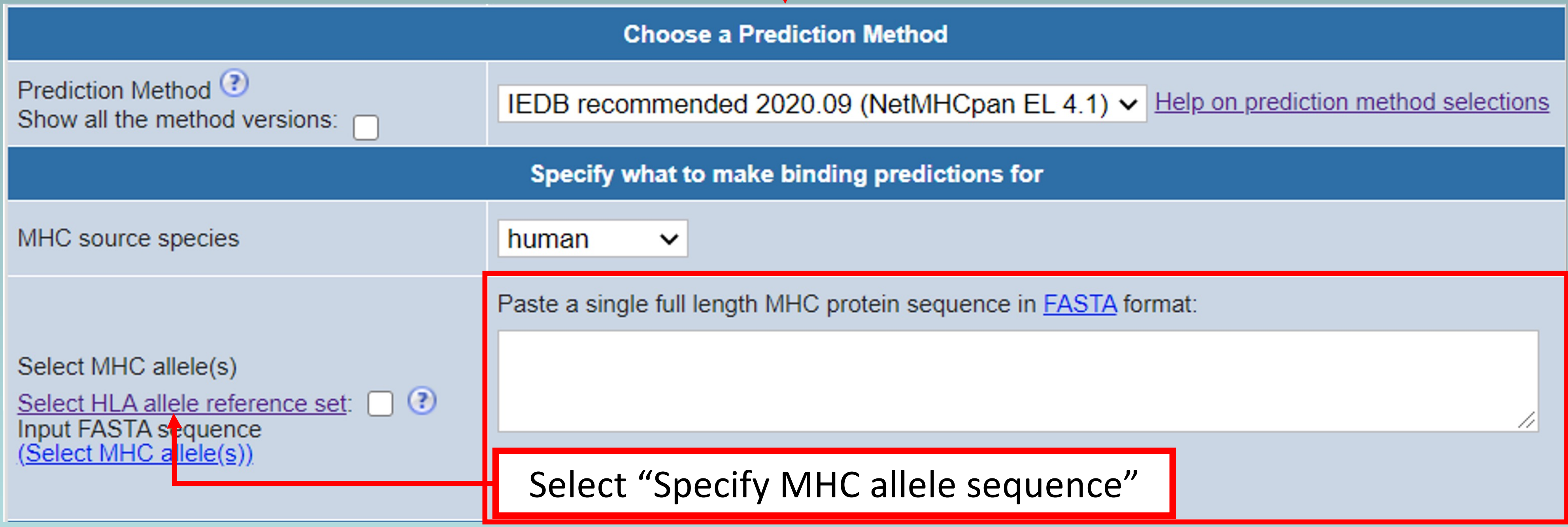
Specify allele(s) & peptide length
(select or upload)

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

Prediction method dependent allele selection

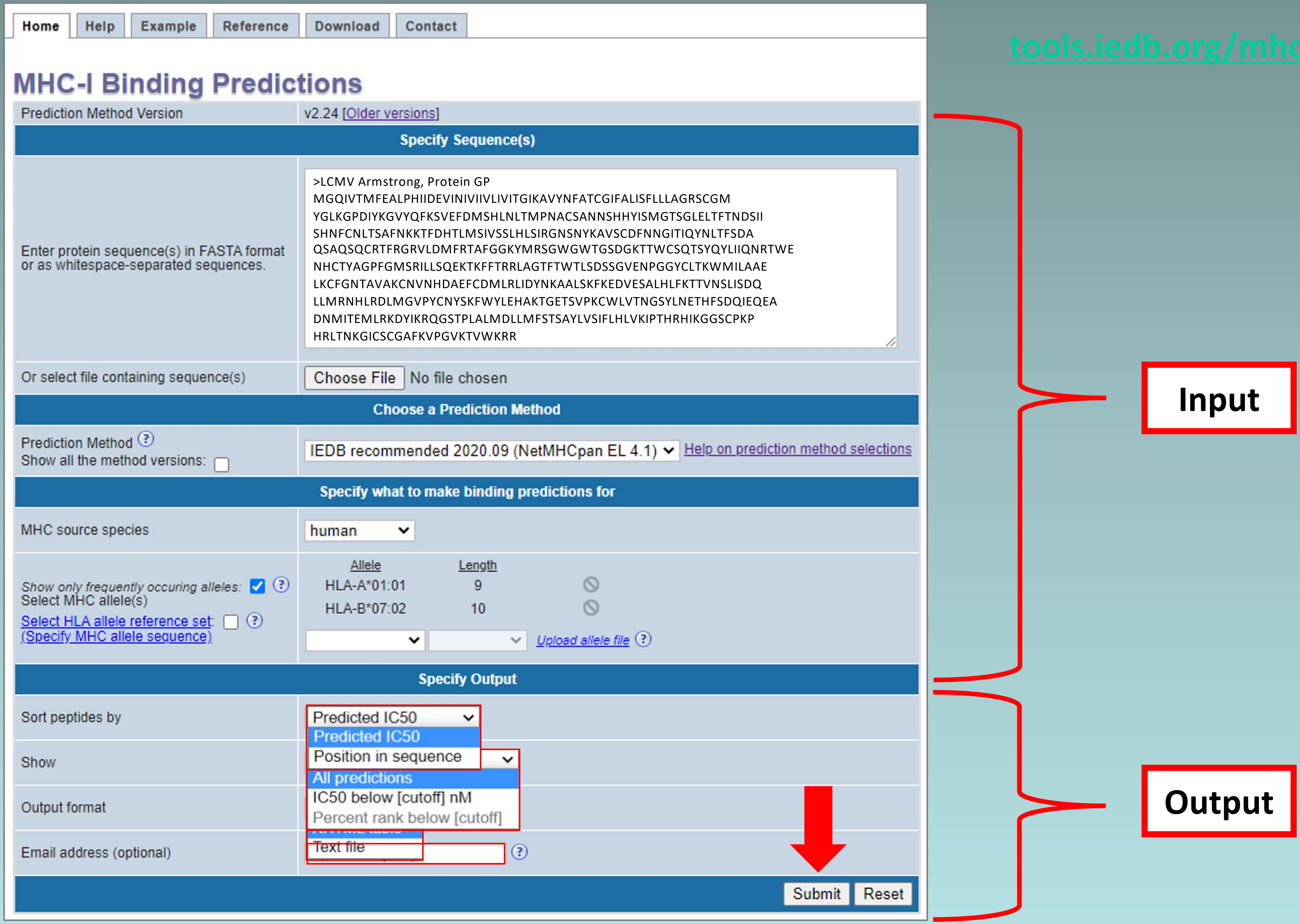
NetMHCpan prediction methods allow **FASTA sequence input**

tools.iedb.org/mhci/



MHC-I binding prediction – example

tools.iedb.org/mhci/



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

Home | Help | Example | Reference | Download | Contact

tools.iedb.org/mhci/

MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	LCMV Armstrong, Protein GP	MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFL LLAGRSCGMYGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHY ISMGTSGLELTFTNDSIISHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIR GNSNYKAVSCDFNNGITIQYNLTFSDAQSAQSQCRTFRGRVLDMFRTAFG GKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWENHCTYAGPFGMSRI LLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAELKCFG NTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSL ISDQLLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYL NETHFSDQIEQEADNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVS IFLHLVKIPTHRHIKGGSCPKPHRLTNKGICSCGAFKVPGVKTVWKRR

Input sequence

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	LLMFSTSAY	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

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	FSDAQSAQS	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	LRDLMGVPY	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	LMGVPYCNY	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	[illegible]	1	[illegible]	[illegible]	10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

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

UniParc

Sequence archive

Proteomes

Supporting data

Literature citations

Taxonomy

Subcellular locations

Cross-ref. databases

Diseases

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 tl sequence that appears in the downloadable versions of the entry.

« Hide

```
        10         20         30         40         50
MSGSSLPSAL ALSLLLVSGS LLPGPGAAQN AGFVKSPMSE TKLTGDAFEL
        60         70         80         90        100
YCDVVGSPTP EIQWWYAEVN RAESFRQLWD GARKRRVTVN TAYGSNGVSV
       110        120        130        140        150
LRITRLTLED SGTYECRASN DPKRNDLRQN PSITWIRAQA TISVLQKPRI
       160        170        180        190        200
VTSEEVIIRD 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 EKRKRPDEVP DDDEPAGPMK TNSTNNHKDK NLRQRNTN
```

01/12/2018, 10:20

FASTA format

```
>sp|Q9Y639|NPTN_HUMAN Neuroplastin OS=Homo sapiens OX=9606 GN=NPTN PE=1 SV=2
MSGSSLPSALALSLLLVSGSLLPGPGAAQNAGFVKSPMSETKLTGDAFELYCDVVGSPTP
EIQWWYAEVNRAESFRQLWDGARKRRVTVNTAYGSNGVSVLRITRLTLEDSGTYECRASN
DPKRNDLRQNPSITWIRAQATISVLQKPRIVTSEEVIIRDSPVLPVTLQCNLTSSSHTLT
YSYWTKNGVELSATRKNASNMEYRINKPRAEDSGEYHCVYHFVSAPKANATIEVKAAPDI
TGHKRSENKNEGQDATMYCKSVGYPHPDWIWRKKENGMPMDIVNTSGRFFIINKENYTEL
NIVNLQITEDPGEYECNATNAIGSASVVTVLRVRSHLAPLWPFLGILAEIIILVVIIVVY
EKRKRPDEVPDDDEPAGPMKTNSTNNHKDKNLRQRNTN
```

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

www.syfpeithi.de

01/12/2018, 16:03

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

SYFPEITHI

Click here
to continue

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 27 th 2012

Institute for Cell Biology
Department of Immunology

BMI - Heidelberg

UKT

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 e available online

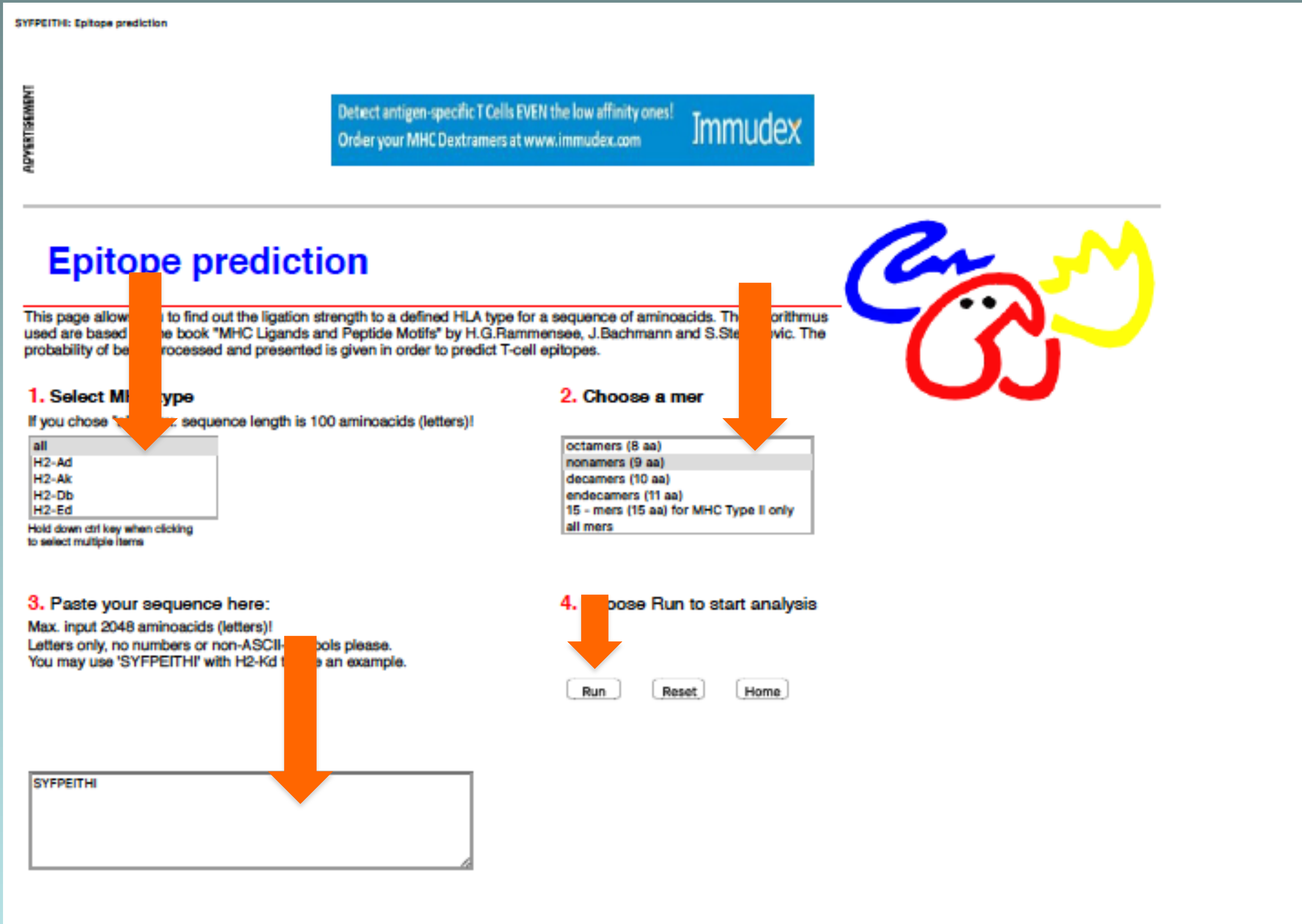
Epitope prediction and retrieval of sequences cording to their molecular mass is also possible

The following search options are available:

FIND YOUR MOTIF, LIGAND OR EPITOPE

EPITOPE PREDICTION

INFORMATION


SYFPEITHI: Epitope prediction
ADVERTISEMENT
Detect antigen-specific T Cells EVEN the low affinity ones!
Order your MHC Dextramers at www.immudex.com
Immudex
Epitope prediction
This page allow to find out the ligation strength to a defined HLA type for a sequence of aminoacids. Th orithmus used are based e book "MHC Ligands and Peptide Motifs" by H.G.Rammensee, J.Bachmann and S.Ste vic. The probability of be rocessed and presented is given in order to predict T-cell epitopes.
1. Select MI ype
If you chose : 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- bols please.
You may use 'SYFPEITHI' with H2-Kd e an example.
SYFPEITHI
4. oose Run to start analysis
Run
Reset
Home

Αποτελέσματα

Your search Results

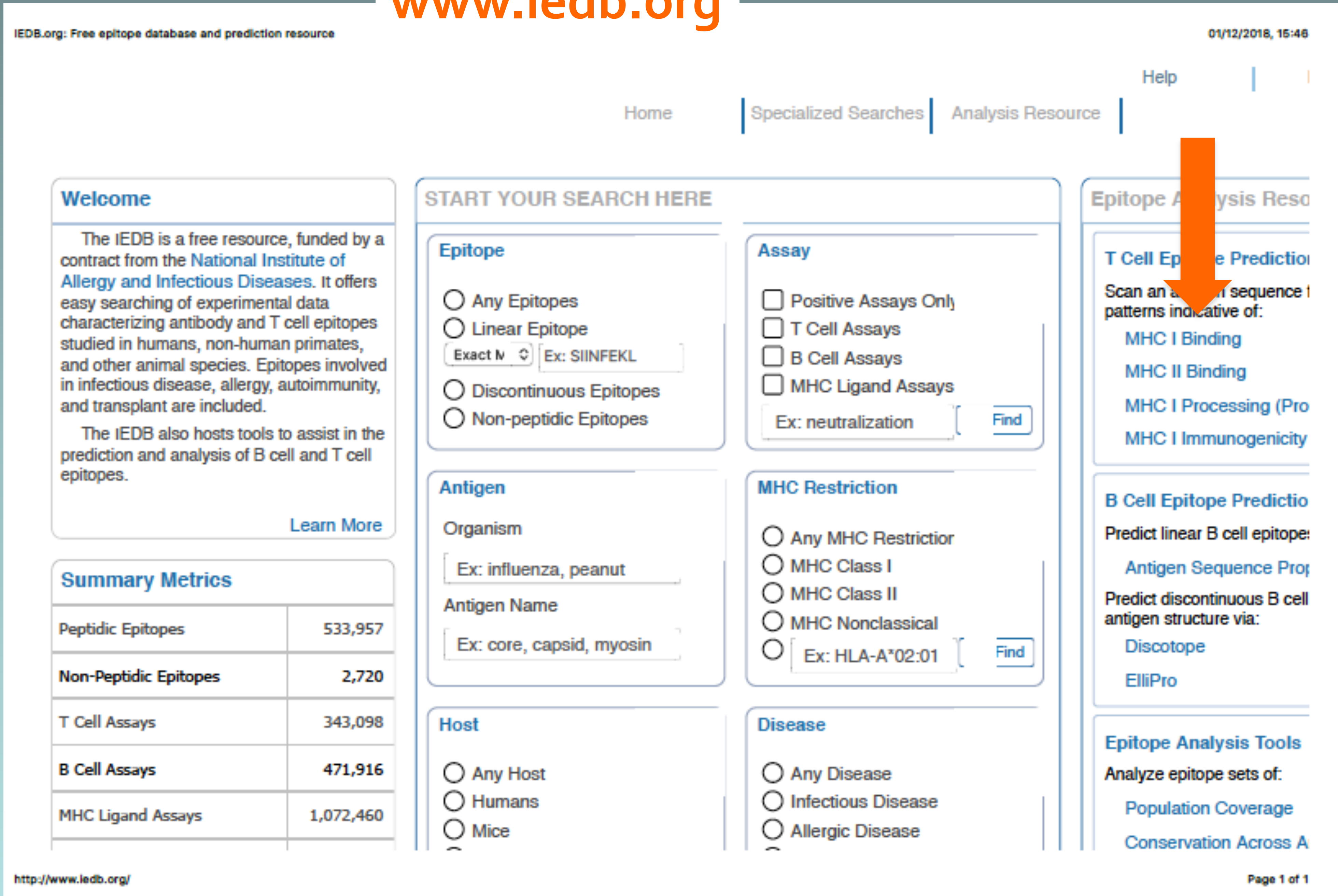
Return to search conditions
H2-Kd nonamers

H2-Kd nonamers

go to top

Pos	1 2 3 4 5 6 7 8 9	score
92	A Y G S N G V S V	24
341	P F L G I L A E I	22
4	S S L P G A L A L	21
128	R Q N P S I T W I	21
14	L L L V S G S L L	20
214	E Y H C V Y H F V	20
13	S L L L V S G S L	18
149	R I V T S E E V I	18
273	K E N G V F E E I	18
291	T N K E N Y T E L	18
322	G S A S V S T V L	18
328	T V L R V R S H L	18
35	K S P M S E T K L	17
42	K L T G D A F E L	17
95	S N G V S V L R I	17
201	E Y R I N K P R A	17

www.iedb.org



IEDB.org: Free epitope database and prediction resource
01/12/2018, 15:46
Help
Home
Specialized Searches
Analysis Resource
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
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
MHC Restriction
Any MHC Restriction
MHC Class I
MHC Class II
MHC Nonclassical
Ex: HLA-A*02:01
Find
Host
Any Host
Humans
Mice
Disease
Any Disease
Infectious Disease
Allergic Disease
Epitope A ysis Reso
T Cell Ep e Predictio
Scan an a n sequence f patterns indicative of:
MHC I Binding
MHC II Binding
MHC I Processing (Pro
MHC I Immunogenicity
B Cell Epitope Predictio
Predict linear B cell epitope
Antigen Sequence Pro
Predict discontinuous B cell antigen structure via:
Discotope
ElliPro
Epitope Analysis Tools
Analyze epitope sets of:
Population Coverage
Conservation Across A
http://www.iedb.org/
Page 1 of 1


MHC-I Binding
01/12/2018, 16:16
IEDB Analysis Resource
Home
Help
Example
Reference
Download
Contact
MHC-I Binding Predictions
Prediction Method Version
2013-02-22 [Older 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)
Choose File
no file selected
Choose sequence format
auto detect format
Choose a Prediction Method
Prediction Method
Show all the method versions:
IEDB recommended 2.18
Help on prediction method selections
Specify what to make binding predictions for
MHC source species
human
Allele
Length
Show only frequently occuring alleles:
Select MHC allele(s)
Upload allele file
Specify Output
Sort peptides by
Predicted IC50
Show
All predictions
Output format
XHTML table
http://tools.iedb.org/mhci/
Page 1 of 2

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)

```
MSGSSLPGALALSLLLVSGSLLPGPGAAQNAGFVKSPMSETKLTGDAFELYCDVVGSPTP
EIQWWYAEVNRAESFRQLWDGARKRRVTVNTAYGSNGVSVLRITRLTLEDSGTYECRASN
DPKRNDLRQNPSITWIRAQATISVLQKPRIVTSEEVIIRESLLPVTLQCNLTSSSHTLMY
SYWTRNGVELTATRKNASNMEYRINKPRAEDSGEYHCVYHFVSAPKANATIEVKAAPDIT
GHKRSENKNEGQDAMMYCKSVGYPHPEWIWRKKENGVFEEISNSSGRFFITNKENYTELS
IVNLQITEDPGEYECNATNSIGSASVSTVLRVRSHLAPLWPFLGILAEIIILVVIIVVYE
KRKRPDEVPDDDEPAGPMKTNSTNNHKDKNLRQRNTN
```

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 occuring alleles: ☑

Select MHC allele(s)

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

Upload allele file

Select HLA allele reference set: ☐

Specify Output

Sort peptides by: Percentile Rank

Show: All predictions

Output format: XHTML table

Submit | Reset

Αποτελέσματα

IEDB Analysis Resource

Home | Help | Example | Reference | Download | Contact

MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	sequence 1	MSGSSLPGALALSLLLVSGSLLPGPGAAQNAGFVKSPMSETKLTGDAFEL YCDVVGSPTPEIQWWYAEVNRAESFRQLWDGARKRRVTVNTAYGSNGVSV LRITRLTLEDSGTYECRASNDPKRNDLRQNPSITWIRAQATISVLQKPRI VTSEEVIIRESLLPVTLQCNLTSSSHTLMYSYWTRNGVELTATRKNASNM EYRINKPRAEDSGEYHCVYHFVSAPKANATIEVKAAPDITGHKRSENKNE GQDAMMYCKSVGYPHPEWIWRKKENGVFEEISNSSGRFFITNKENYTELS IVNLQITEDPGEYECNATNSIGSASVSTVLRVRSHLAPLWPFLGILAEII ILVVIIVVYEKRKRPDEVPDDDEPAGPMKTNSTNNHKDKNLRQRNTN

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/

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.

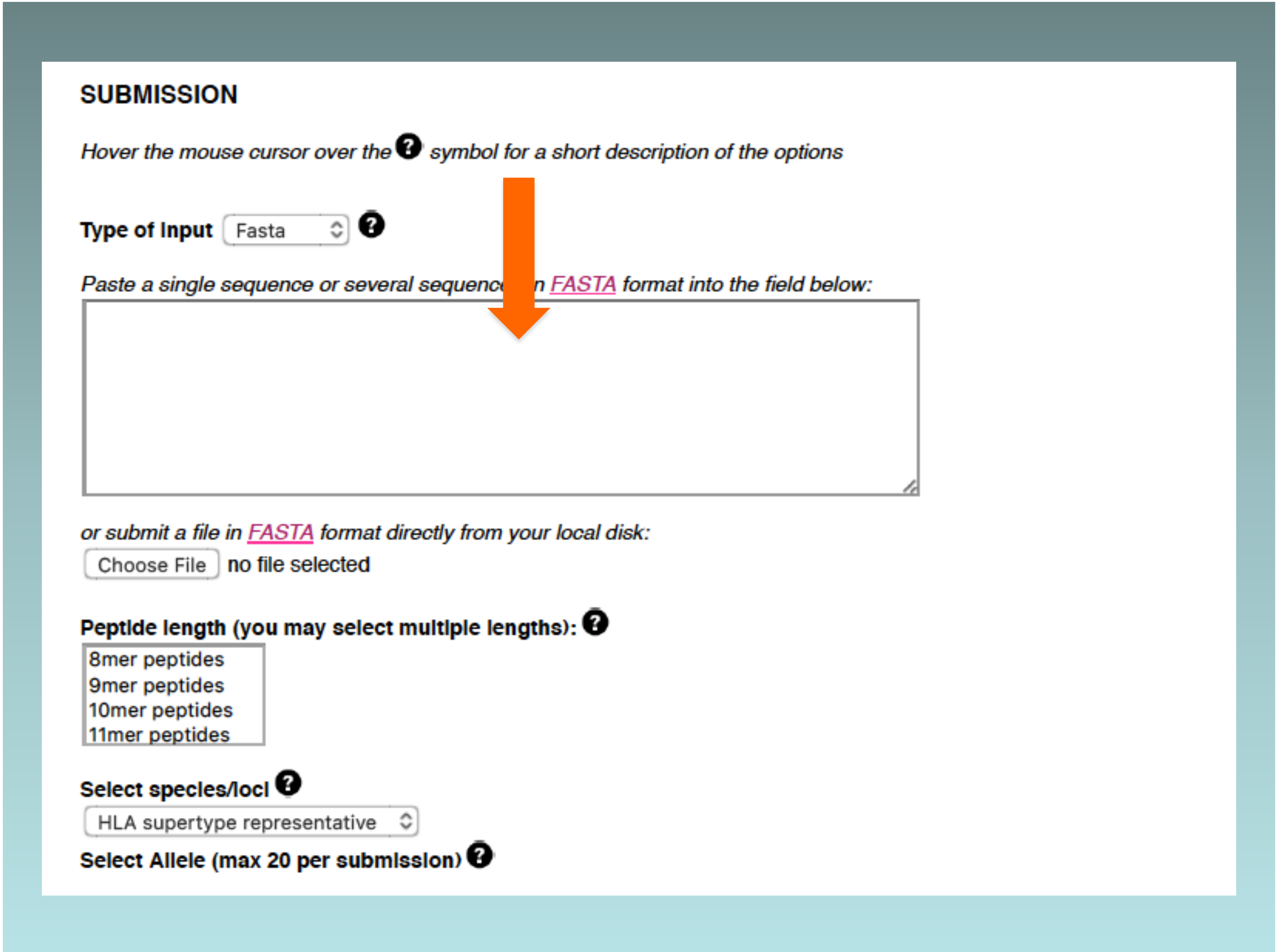
NetMHCpan server predicts binding of peptides to **any MHC molecule of known sequence** using artificial neural networks (ANNs). The method is trained on a combinatino of more than 180,000 quantitative binding data and MS derived MHC eluted ligands. The binding affinity data covers 172 MHC molecules 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
Fasta
Paste a single sequence or several sequenc n FASTA format into the field below:
or submit a file in FASTA format directly from your local disk:
Choose File
no file selected
Peptide length (you may select multiple lengths):
8mer peptides
9mer peptides
10mer peptides
11mer peptides
Select species/loci
HLA supertype representative
Select Allele (max 20 per submission)

NetMHCpan Server - prediction results

Αποτελέσματα

Technical University of Denmark

CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS

```
# NetMHCpan version 4.0

# Tmpdir made /usr/opt/www/webface/tmp/server/netmhcpan/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	B 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	LALSLLLVS	LALSLLLVS	0	0	0	0	0	LALSLLLVS	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	GSLLPGPGA	GSLLPGPGA	0	0	0	0	0	GSLLPGPGA	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	

ΠΙΝΑΚΑΣ ΕΡΓΑΣΙΑΣ

ΓΟΝΙΔΙΟ	ΠΡΟΕΛΕΥΣΗ	MHC	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.