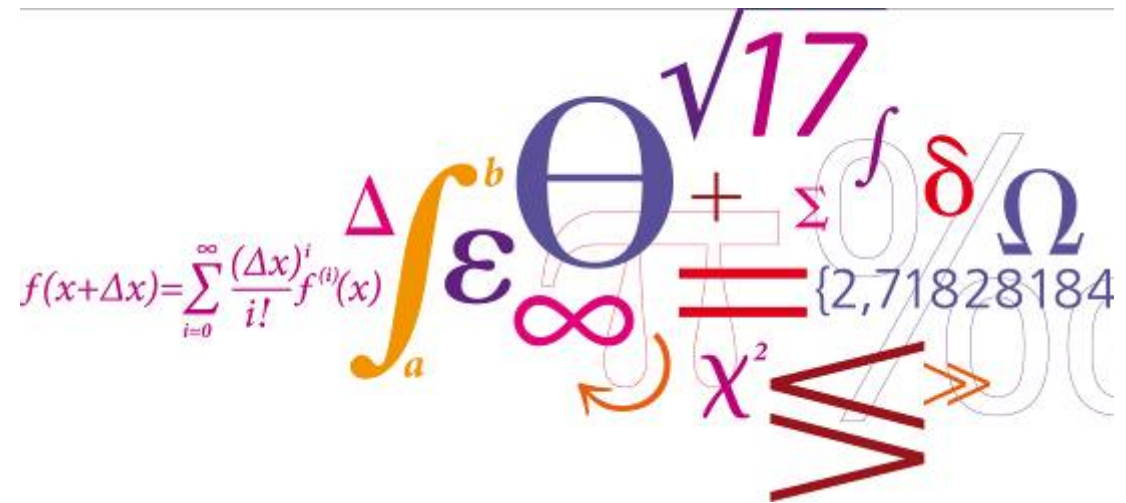


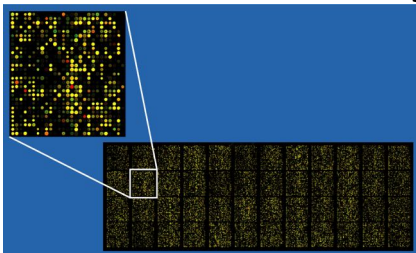
Prediction of immunogenicity of peptide drugs using bioinformatics





Informatics in biology

- 70' Little computing power
 - Analytical solutions
- 80' Computers, few data
 - Simulations of molecular/cellular dynamics
- 90' More data – Sequence and structure
 - Searching biological databases
 - Prediction of features by data driven methods
 - Analysis of gene expression data

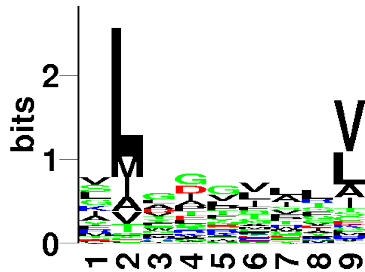


```
>polymerase"  
MERIKELRDLMSQSRTRILLTKTTVDHMAIIKKYTSGRQEKNPALRMKMMAMKYPITAD  
KRIMEMIPERNEQGQTLWSKTNDAGSDRVMVSP LAVTWNNRNGPTTSTVHYPKVYKTYFE  
KVERLKHGTFGPHFRNOVKIRRRVDINPGHADLSAKEAQDVIMEVVFPNEVGARILTSE  
SQLTTI TKEKKEELQDCKIAPLMVAYMLERELVRKTRFLPVAGGTSSVYIEVLHLTQGTGW  
EQMYTPEGVEVRNDDVDQSLIIAARNIVRRATVSADPLASLLEMCHSTQIGGIRMDILRQ  
NP'TEEQAVDICKAAMGLRISSSPFGGPTFKRTNGSSVKKEEVLGTGNLQTLKIKVHEGY  
EEFTMVGRRATAILRKATRRLIQLIVSGRDEQSIABAIIVAMVFSQEDCMIKAVRGDLNF  
...
```

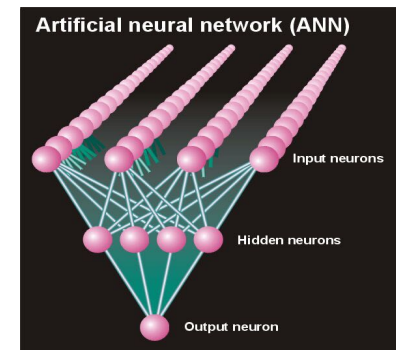
Data driven predictions

List of peptides that have a given biological feature

Y**M**NGTMSQ**V**
G**I**LGFV**F**TL
A**L**WGFFP**V**
I**L**KEPV**H**GV
I**L**GFV**F**TLT
L**L**FGYP**V**YV
G**L**SPT**V**WLS
W**L**SLL**V**PFV
F**L**PSDF**F**PS
C**V**GGL**L**TM**V**
F**I**AGNS**A**YE



Mathematical model (neural network, hidden Markov model)



Search databases for other biological sequences with the same feature/property

>polymerase"
MERIKELRDLMSQSRTRILLTKTTVDHMAIIKKYTSGRQEKNPALRMKMMAMKYPITAD
KRIMEMIPERNEQGQTLWSKTNDAAGSDRVMVSPLAVTWNNRNGPTTSTVHYPKVYKTYFE
KVERLKHGTFPGVHFRNOVKIRRRVDINPGHADLSAKEAQDVIMEVVFPNEVGARLLTSE
SQLTITKEKKKEELQDCKIAPLMVAYMLERELVRKTRFLPVAGGTSSVYIEVLHLTQGTGW
EQMYTPGGEVRNDDVDQSLIIAARNIVRRATVSADPLASLLEMCHSTQIGGIRMDVILRQ
NPTEQAVDICKAAMGLRISSSPFGGPTFKRTNGSSVKKEEVLGTGLKIKVHEGY
EEFTMVGRRATAILRKATRRLIQLIVSGRDEQSIABAIIVAMVFSQEDCMIKAVRGDLNF
...

Immune system

- Innate – fast...
- Addaptive – remembers...
 - Cellular
 - Cytotoxic T lymphocytes (CTL)
 - Helper T lymphocytes (HTL)
 - Humoral
 - B lymphocytes

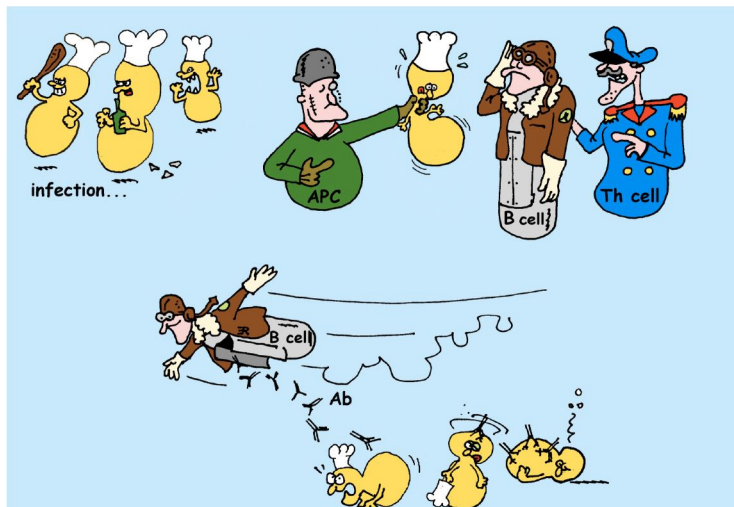
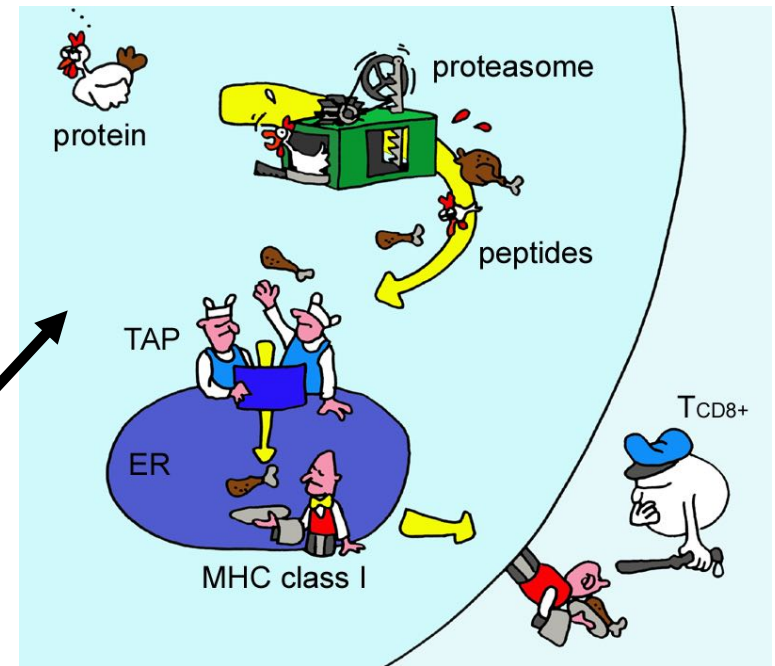
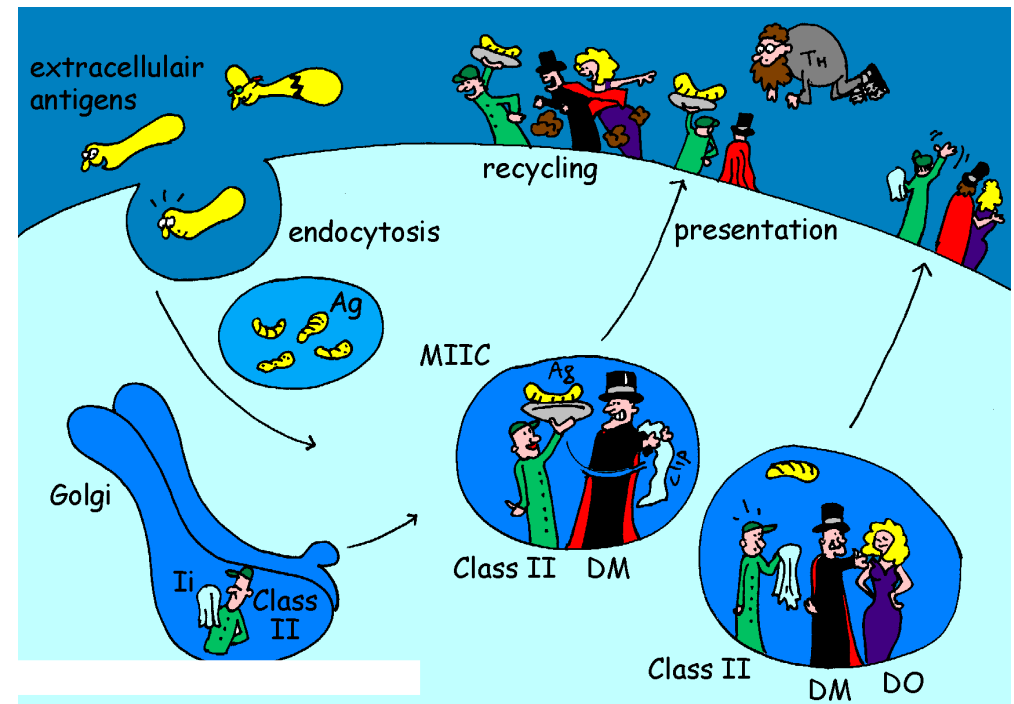


Figure by Eric A.J. Reits



Human MHC:
~1000 variants
distributed
over 12 types

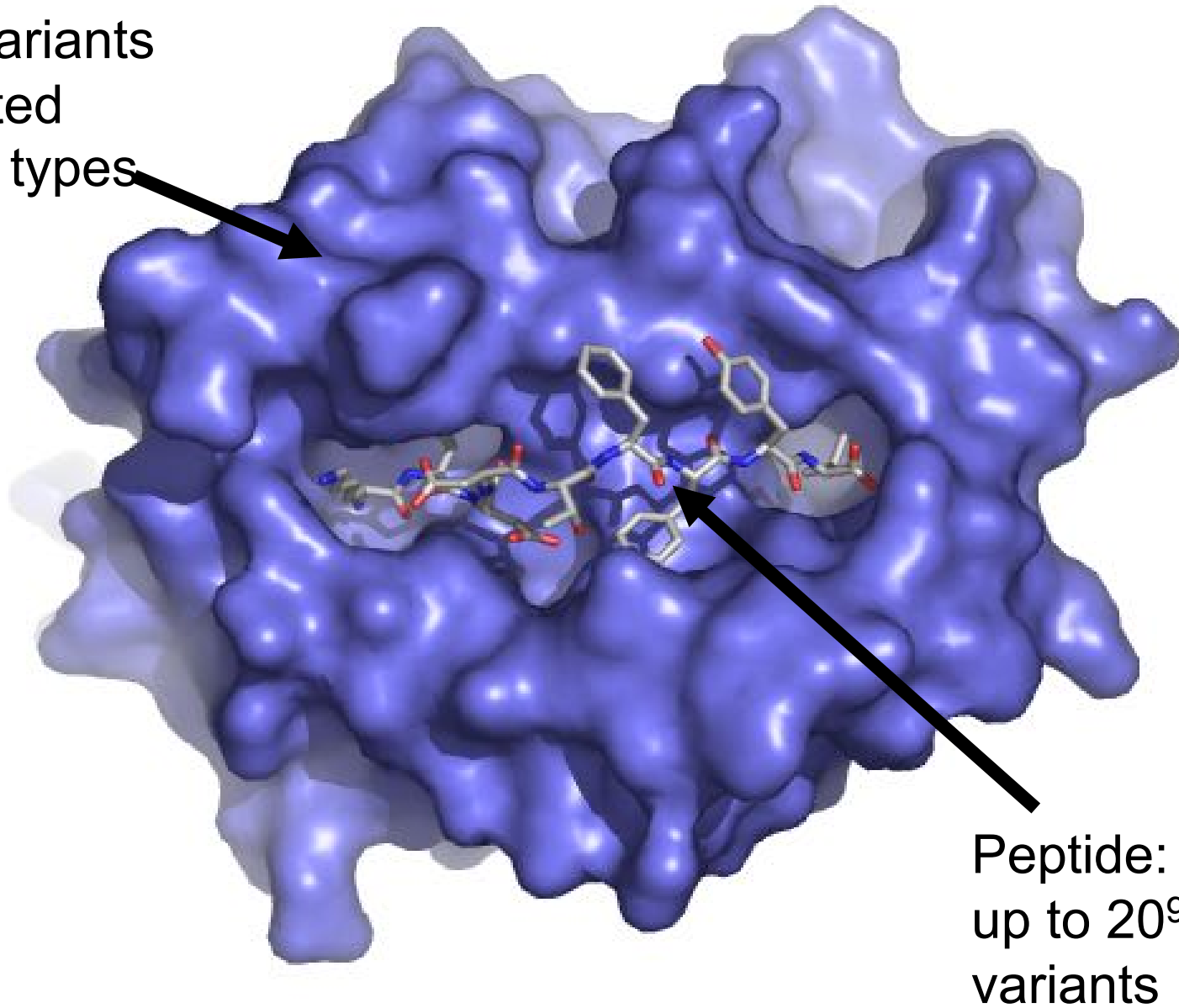
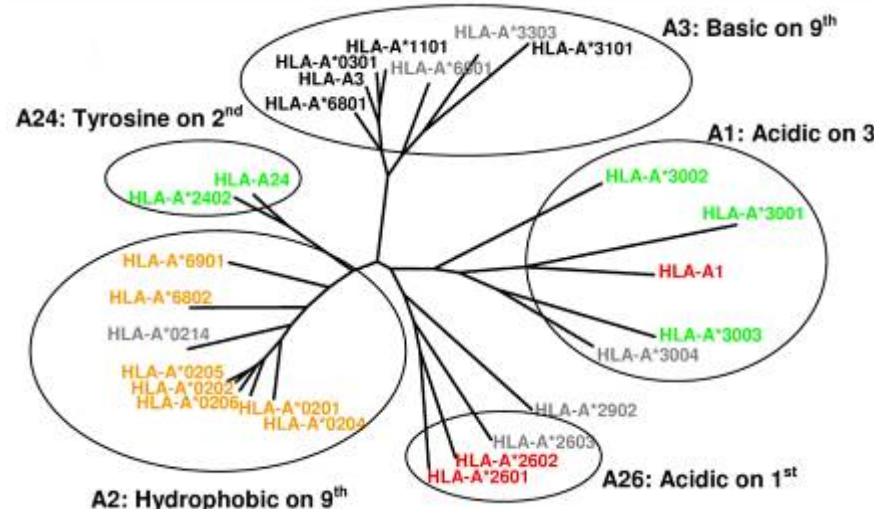
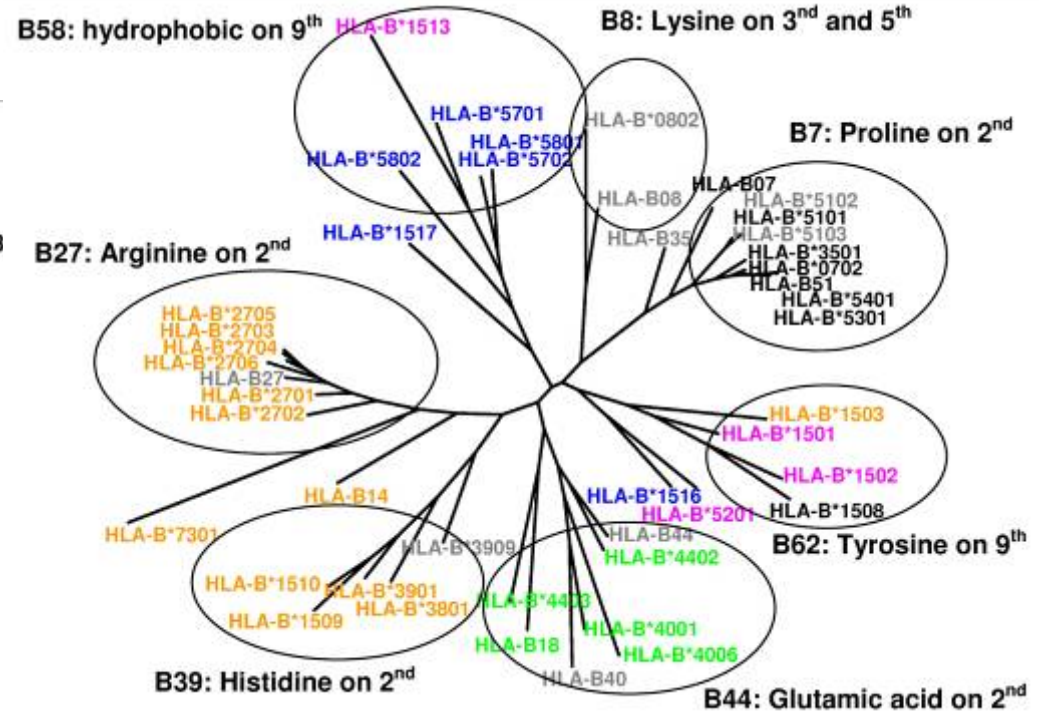


Figure by Anne Mølgaard, peptide (KVDDTFYYV) used as vaccine by Snyder et al. J Virol 78, 7052-60 (2004).

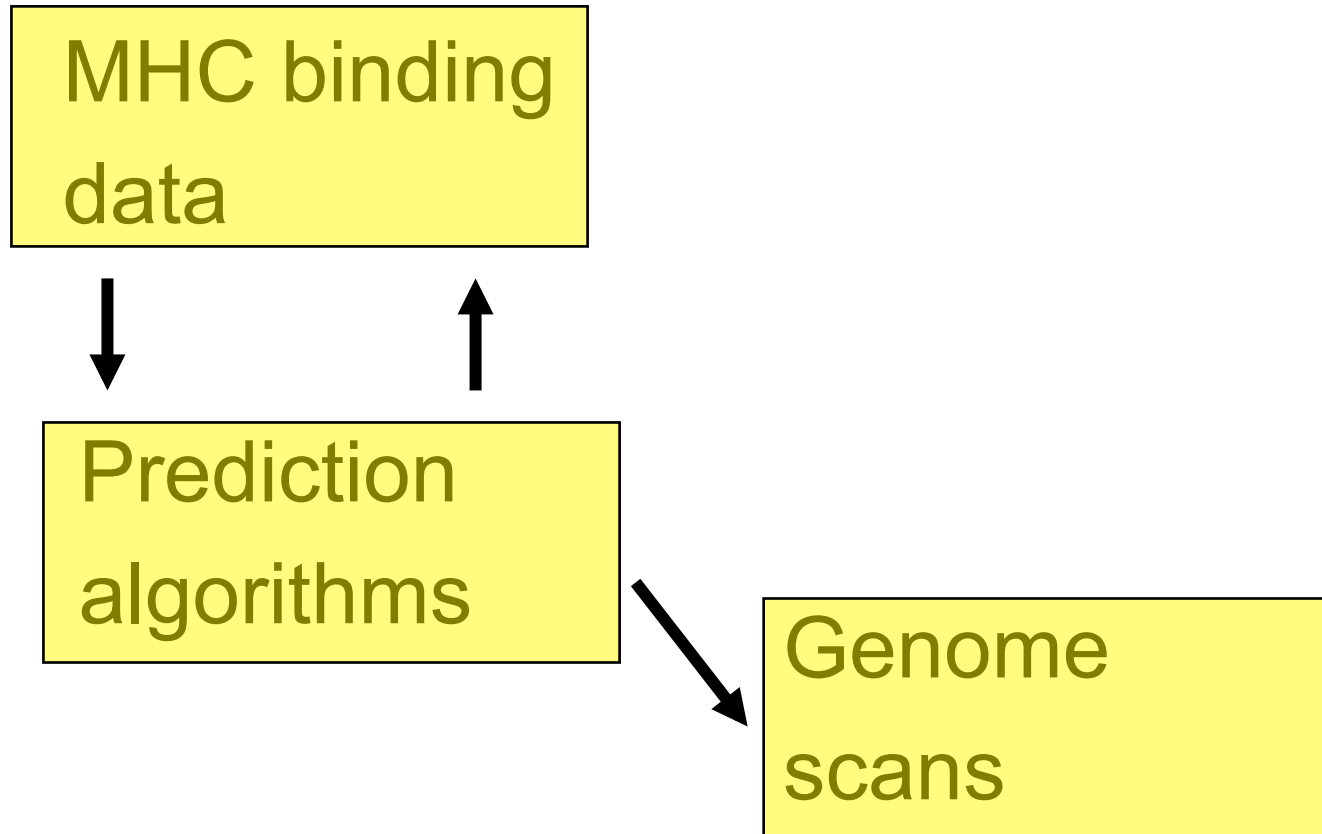
Coverage of HLA alleles



Supertype	Selected allele
A1	A*0101
A2	A*0201
A3	A*1101
A24	A*2401
A26 (new*)	A*2601
B7	B*0702
B8 (new*)	B*0801
B27	B*2705
B39(new*)	B*3901
B44	B*4001
B58	B*5801
B62	B*1501



Prediction algorithms



HLA binding

nM	Dengue	Ebola	F. tularer Flu	Hantaan	Junin	Lassa	Marburg	Pox	TB	WNV	YFV	E. coli	Anthra	Y. Pesti	Lund M1	
$K_D < 50$	71	53	52	82	120	29	30	46	97	83	68	62	58	34	12	76
$50 < K_D < 500$	45	62	20	112	66	24	22	39	42	74	54	59	51	30	9	26
$500 < K_D < 5000$	14	42	1	44	40	20	19	16	28	31	23	31	37	5	2	21
$K_D > 5000$	17	33	2	89	22	7	8	13	10	19	16	19	50	3	0	7
Total	147	190	75	327	248	80	79	114	177	207	161	171	196	72	23	130

Batch Number	Name	Peptide Sequence	HLA Name	KdValue (nM)
16204	WNV [A1/A26] 16204	LTKEEFTRY	A*0101	1150
16204	WNV [A1/A26] 16204	LTKEEFTRY	A*2601	1083
16205	WNV [A1/A26] 16205	DTACLAKSY	A*0101	1370
16205	WNV [A1/A26] 16205	DTACLAKSY	A*2601	8
16206	WNV [A1/A26] 16206	MTRGLLGSY	A*0101	520
16206	WNV [A1/A26] 16206	MTRGLLGSY	A*2601	8
16207	WNV [A1/A3/B58] 16207	ITYTDVLRV	A*0101	3
16208	WNV [A1/B62]	YQTDSGCWY	A*0101	6059
16208	WNV [A1/B62]	YQTDSGCWY	B*1501	1343
16050	WNV [A1] 16050	QTDNQLAVF	A*0101	168
16051	WNV [A1] 16051	RTWNYHGSY	A*0101	237
16052	WNV [A1] 16052	VVEKQSGLY	A*0101	6
16055	WNV [A1] 16055	RSGIDTNAY	A*0101	66
16056	WNV [A1] 16056	MTKEEFTRY	A*0101	246
16057	WNV [A1] 16057	MSPHRVPNY	A*0101	20000
16059	WNV [A1] 16059	MTRGILGSY	A*0101	7000
16060	WNV [A1] 16060	KGDTTGVY	A*0101	472
16209	WNV [A2/B39] 16209	YTMDGEYRL	B*3901	154
16210	WNV [A2/B8] 16210	GMSWITQGL	A*0201	1
16210	WNV [A2/B8] 16210	GMSWITQGL	B*0801	20000
16061	WNV [A2] 16061	ILLWEIPDV	A*0201	1
16062	WNV [A2] 16062	VLCPYMPKV	A*0201	21

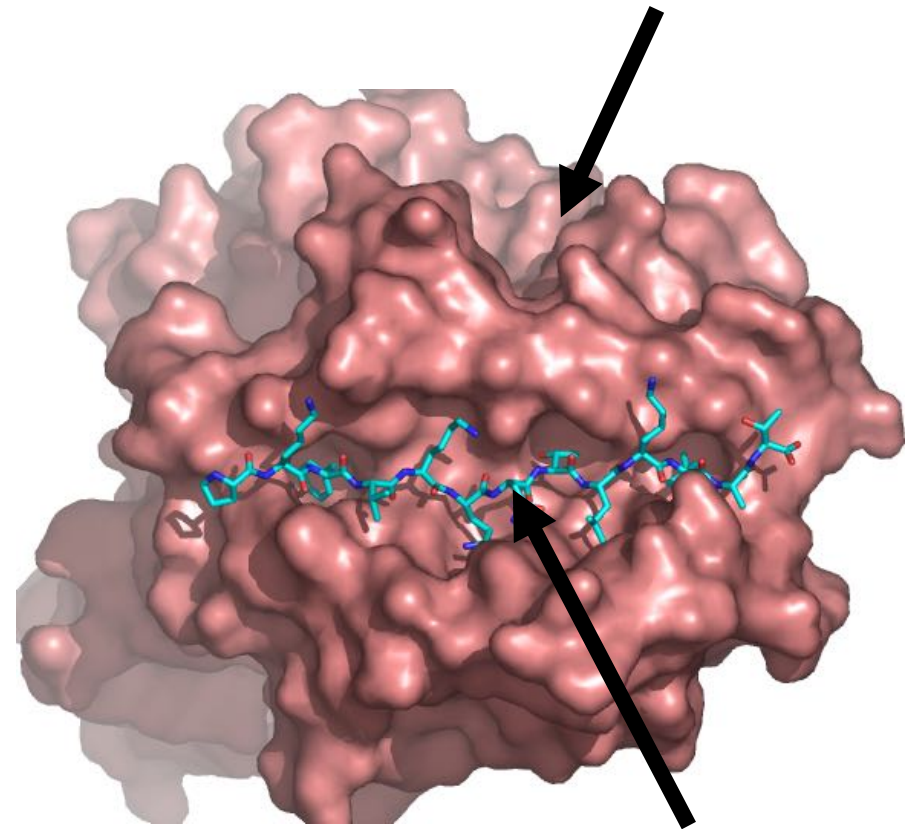
...

Binding studies were done by Søren Buus and colleagues

Class II MHC binding

- MHC class II binds peptides in the class II antigen presentation pathway
- Binds peptides of length 9-18 (even whole proteins can bind!)
- Binding cleft is open
- Binding core is 9 aa

Human MHC II:
~1000 variants

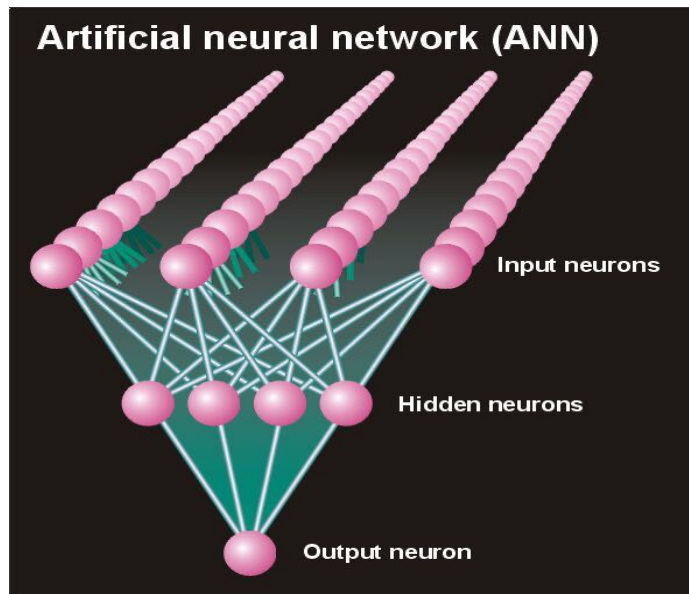


Peptide:
up to 20^9
variants

NetMHC-IIpan Method

XAAAAVAAEAYXXX	WLFECLECYPDYWLQRATYCHNVGF	0.119861	DRB1_0101	AAAAVAAEAY
ALNVKRREGMFIDEX	WLFECLECYPDYWLQRATYCHNVGF	0.238877	DRB1_0101	AALNVKRREGMFIDE
QPGLTSAVIEALPXX	WLFECLECYPDYWLQRATYCHNVGF	0.169769	DRB1_0101	AAQPGLTSAVIEALP
XACVKDLVSKYLADN	WLFECLECYPDYWLQRATYCHNVGF	0.577653	DRB1_0101	ACVKDLVSKYLADNE
KIGLHTEFQTVSFX	WLFECLECYPDYWLQRATYCHNVGF	0.982712	DRB1_0101	AFKIGLHTEFQTVSF
AGDLGRDELMELASD	WLFECLECYPDYWLQRATYCHNVGF	0.061007	DRB1_0101	AGDLGRDELMELASD
XAGLIAIVMVTILLC	WLFECLECYPDYWLQRATYCHNVGF	0.104993	DRB1_0101	AGLIAIVMVTILLCC
XAGYAATNDDNILSH	WLFECLECYPDYWLQRATYCHNVGF	0.364429	DRB1_0101	AGYAATNDDNILSHV
XXAKCNLDHSSEFC	WLFECLECYPDYWLQRATYCHNVGF	0.156760	DRB1_0101	AKCNLDHSSEFCM
XAKMKCFGNTAVAKC	WLFECLECYPDYWLQRATYCHNVGF	0.734955	DRB1_0101	AKMKCFGNTAVAKCN

+ peptide length
+ PFR length

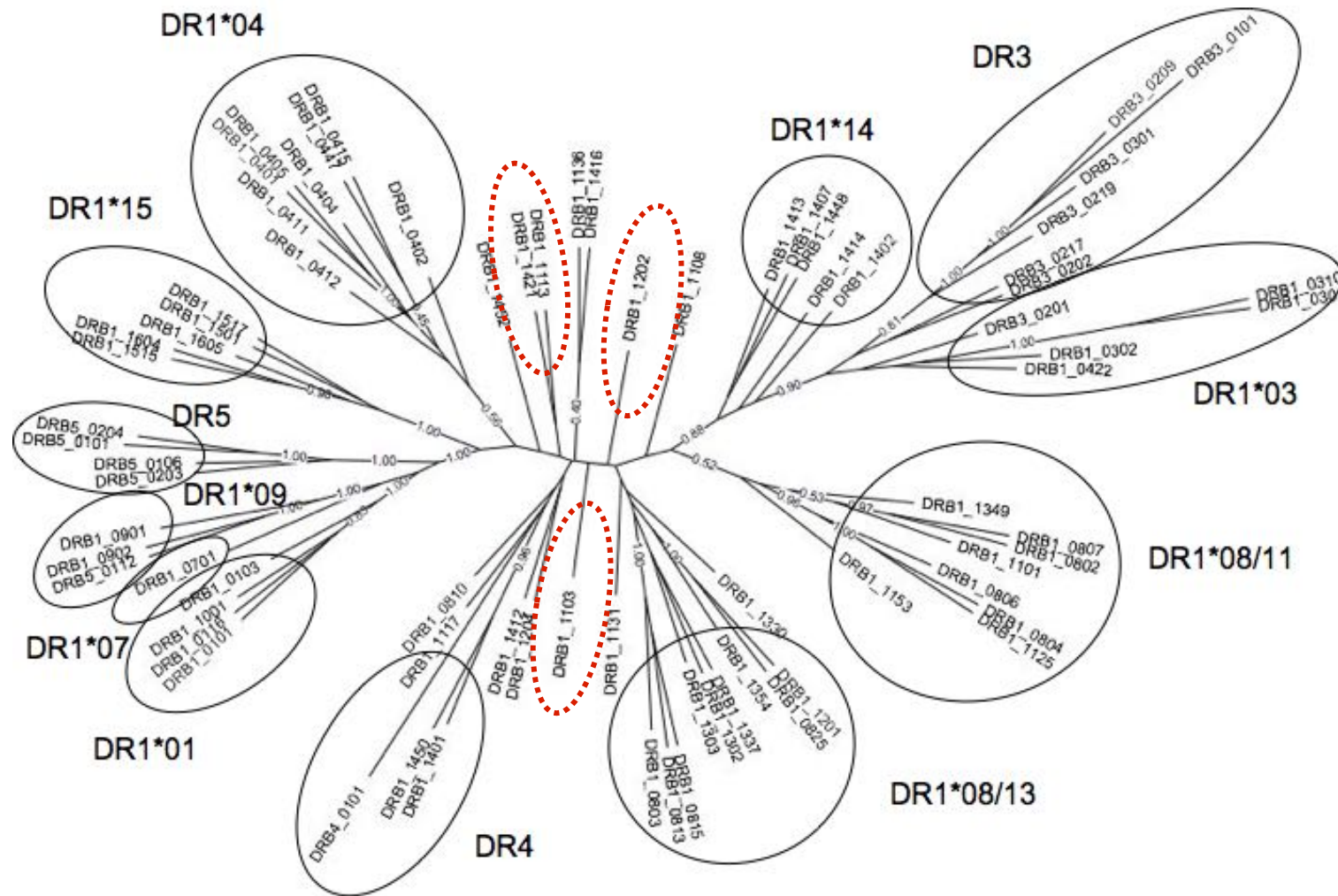


AAAAVAAEAY	0.11986	0.33183
AALNVKRREGMFIDE	0.23888	0.44053
AAQPGLTSAVIEALP	0.16977	0.56296
ACVKDLVSKYLADNE	0.57765	0.59562
AFKIGLHTEFQTVSF	0.98271	0.46325
AGDLGRDELMELASD	0.06101	0.18933
AGLIAIVMVTILLCC	0.10499	0.36359
AGYAATNDDNILSHV	0.36443	0.29837
AKCNLDHSSEFCM	0.15676	0.17179
AKMKCFGNTAVAKCN	0.73496	0.69730

Final NetMHC-IIpan method

Allele	N	Pan		SMM-align		TEPITOPE
		Pearson	AUC	Pearson	AUC	AUC
DRB1*0101	5166	0.682	0.841	0.610	0.802	0.720
DRB1*0301	1020	0.659	0.846	0.563	0.795	0.664
DRB1*0401	1024	0.625	0.816	0.496	0.751	0.716
DRB1*0404	663	0.701	0.860	0.579	0.801	0.770
DRB1*0405	630	0.637	0.833	0.560	0.789	0.759
DRB1*0701	853	0.725	0.870	0.618	0.812	0.761
DRB1*0802	420	0.660	0.843	0.555	0.787	0.766
DRB1*0901	530	0.517	0.728	0.360	0.655	
DRB1*1101	950	0.724	0.871	0.581	0.796	0.721
DRB1*1302	498	0.663	0.819	0.558	0.785	0.652
DRB1*1501	934	0.644	0.800	0.528	0.727	0.686
DRB3*0101	549	0.619	0.841	0.585	0.836	
DRB4*0101	446	0.695	0.871	0.541	0.793	
DRB5*0101	924	0.703	0.856	0.529	0.761	0.680
Ave*	14	0.661	0.835	0.547	0.778	
Ave**	11	0.675	0.841	0.562	0.782	0.718

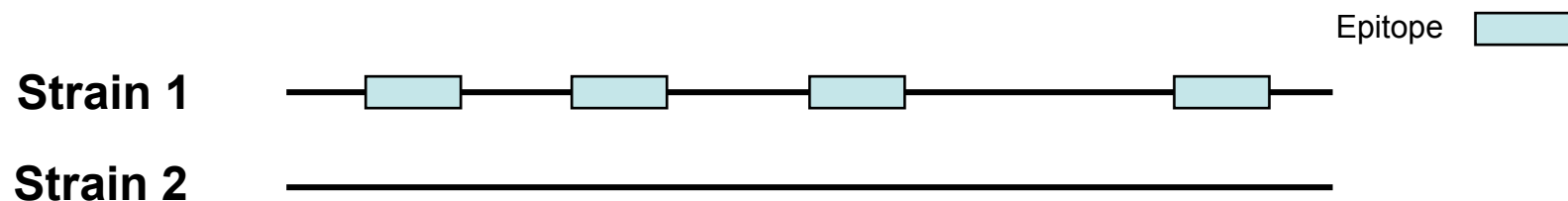
HLA-DR specificity clustering



Polyvalent vaccines

- The equivalent of this in epitope based vaccines is to select epitopes in a way that that they *together* cover all strains.

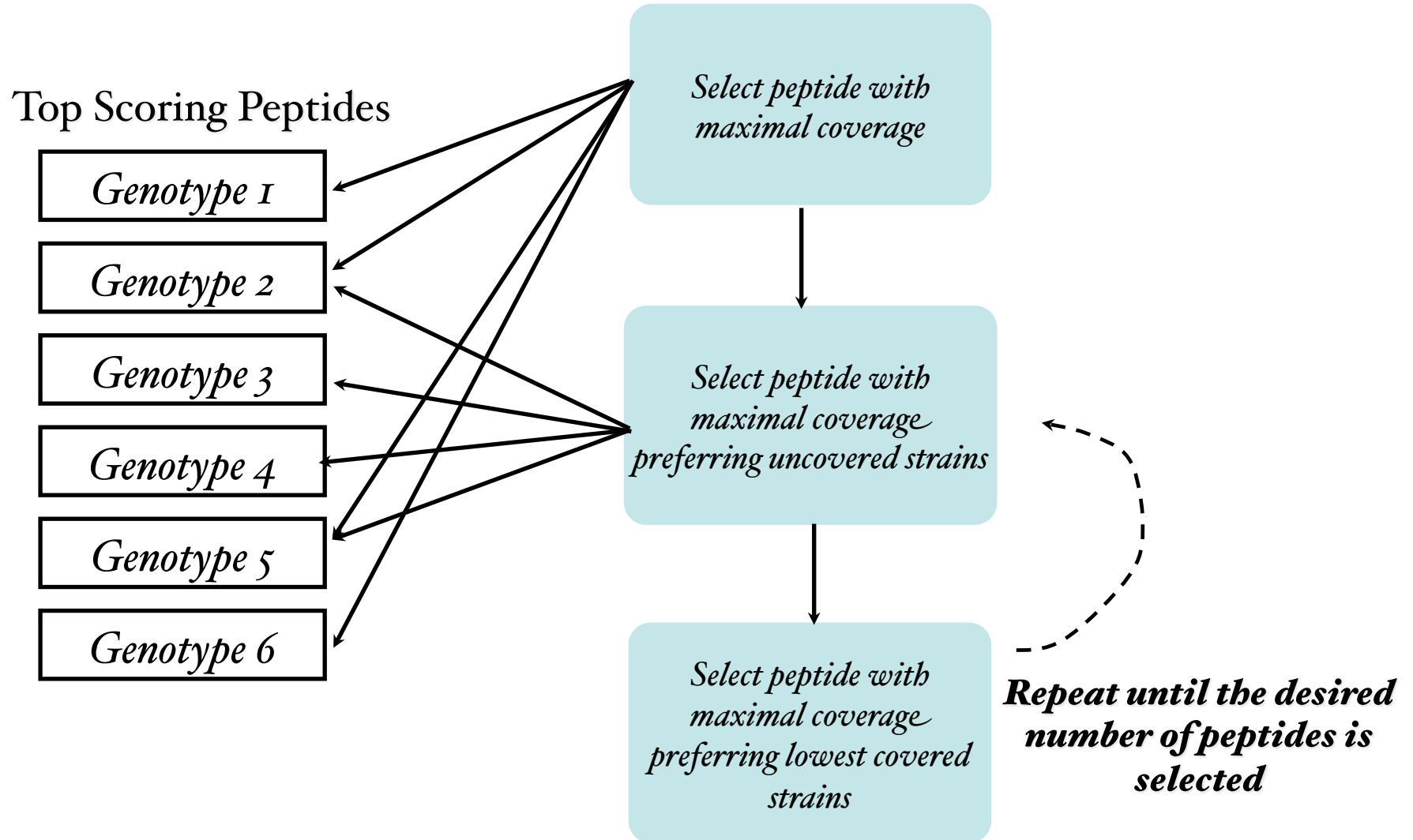
Uneven coverage, Average coverage = 2



Even coverage, Average coverage = 2



EpiSelect



Selected West Nile Virus epitopes

NC_001563: MSKKPGGPGKNRAVNMLKRGMPRGLSLIGLKRAMLSLIDGKGPPIRFVLALLAFFRFTAIA
GRGPIRFVL (B39)
VLALLAFFR (A3)
AIAPTRAVL (B7)
RFVLALLAF (A24)
FVLALLAFF (A26)
VLSLIGLKR (A3)
GPIRFVLAL (B7)

NC_001563: PTRAVLDRWRGVNKQTAMKHLLSFKKELGTLTSAINRRSTKQKKRGGTAGFTILLGLIAC
GVNKQTAMK (A3)
VMIGLIASV (A2)

NC_001563: AGAVTLSNFGQKVMMTVNATDVTDVITIPTAAGKNLCIVRAMDVGYLCEDTITYECPVLA
FLCDDTITY (B62)
YLCEDTITY (B62)
IPTAAGKNL (B7)

NC_001563: AGNDPEDIDCWCTKSSVYVRYGRCTKTRHSRRSRRSLTVQTHGESTLANKKGAWLDSTKA
HSRRSRRSL (B7/B8)
RRSRRSLTV (B27)

NC_001563: TRYLVKTESWILRNPGYALVA AVIGWMLGSNTMQRVVFAILLLL VAPAYSFNCLGMSNRD
ILRNPGYAL (B8)
LVKTESWIL (B8)
RYLVKTESW (A24)
LLVAPAYSF (B62)
LLLLVAPAY (B62)

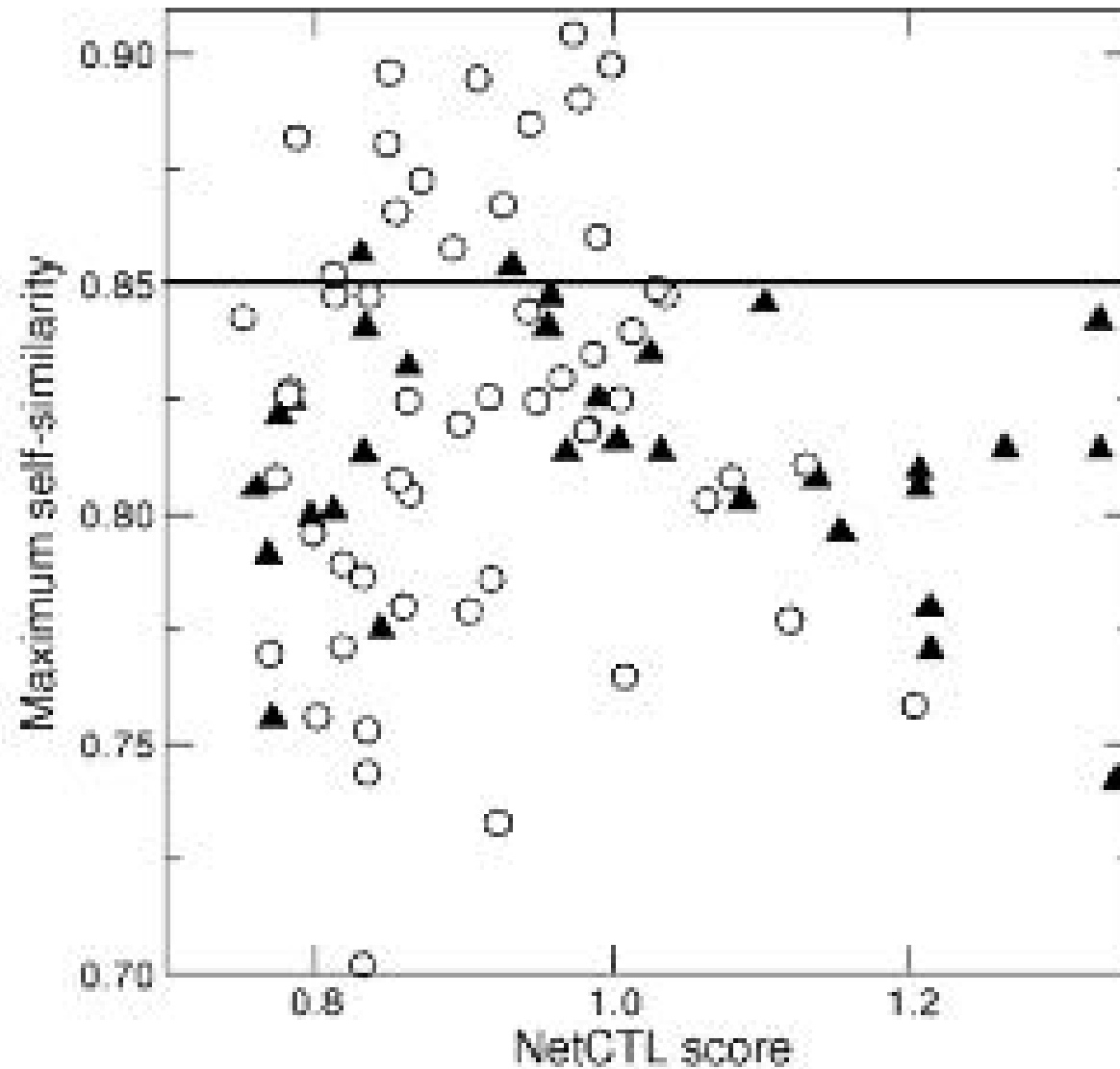
All HIV responsive patients respond to at least one of nine peptides

	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
FLGKIWPSHK_A2_gag					1		1	1		1			1				1			1	1									1	1
AFDLSFFLK_A3_nef			1		1	1	1	1	1	1			2				1	1		1	1	1								2	1
QVPLRPMTY_A1_nef			1	1	2	1	1	1	1	1			3	1			1	1	1	1	1	1	1	1						2	2
GALDLSHFL_A2_nef			1	1	3	1	2	1	1	1	1		3	1			1	1	1	1	1	1	1	1	1			1		2	3
ILKEPVHGV_A2_pol		1	1	1	3	1	2	1	1	2	1		3	1	1		2	1	1	1	2	2	1	1			1		3	3	
IPRRIRQGF_B7_env		1	1	1	3	1	2	1	1	2	1	1	4	1	1	1	2	1	1	1	3	2	1	1			1		3	3	
AISPRTLNAW_B58_gag	1	1	1	1	3	1	2	1	1	2	1	1	4	1	1	1	2	1	1	1	3	2	1	1			1	1	3	3	
RALGPGATL_B7_gag	1	1	1	1	3	1	3	1	1	2	1	1	4	2	1	1	2	1	1	1	3	2	1	1	1		1	1	3	4	
GEIYKRWII_B44_gag	1	1	1	1	3	1	3	1	1	2	1	1	4	2	1	1	2	1	1	1	3	2	2	1	1	1	1	1	1	4	4
QEILDWVY_B44_nef	1	1	1	2	3	1	3	2	1	2	1	1	4	2	2	1	2	2	1	2	4	2	3	1	1	1	1	1	1	4	4
AAVDLSHFL_A2_nef	1	1	1	2	4	1	4	2	1	2	2	1	4	2	2	1	2	2	1	2	4	2	3	2	1	1	2	1	4	4	
TPQDLNNTML_B7_gag	1	1	1	2	4	2	5	2	2	2	2	1	4	2	2	1	2	2	1	2	5	2	3	2	1	1	2	1	4	4	
RPMTYK GAL_B7_nef	1	1	2	2	4	2	5	2	2	2	2	1	4	2	2	1	2	2	2	2	6	2	3	2	1	1	2	1	4	4	
SPAIFQSSM_B7_pol	1	1	2	3	4	2	6	2	2	2	2	1	5	3	2	2	2	2	2	2	7	2	3	2	1	1	2	1	4	5	
LTDTTNQKT_A1_pol	1	2	2	3	4	2	6	2	2	2	2	1	6	3	2	2	2	2	2	2	8	3	3	2	1	1	3	1	5	5	
QATQEVK GW_B58_gag	1	2	2	3	4	2	6	2	2	2	2	1	6	3	2	2	3	2	2	2	8	3	3	2	1	1	3	2	5	5	
RPQVPLRPM_B7_nef	1	2	2	3	4	2	7	2	2	2	2	2	6	4	2	2	3	2	2	2	8	3	3	2	1	1	3	2	5	5	
YFPDWQNYT_B58_nef	2	2	2	3	4	2	8	2	2	2	2	2	6	4	2	2	3	2	2	2	8	3	3	2	1	1	3	2	5	5	
SLYNTVATL_A2_gag	2	3	2	3	4	2	8	3	2	2	2	2	6	4	2	2	3	2	2	2	9	3	3	2	1	1	4	2	6	6	
WYIKIFIII_A24_env	2	3	3	4	4	2	8	3	2	2	2	2	7	4	2	2	3	3	2	2	10	3	3	2	1	1	4	2	6	6	

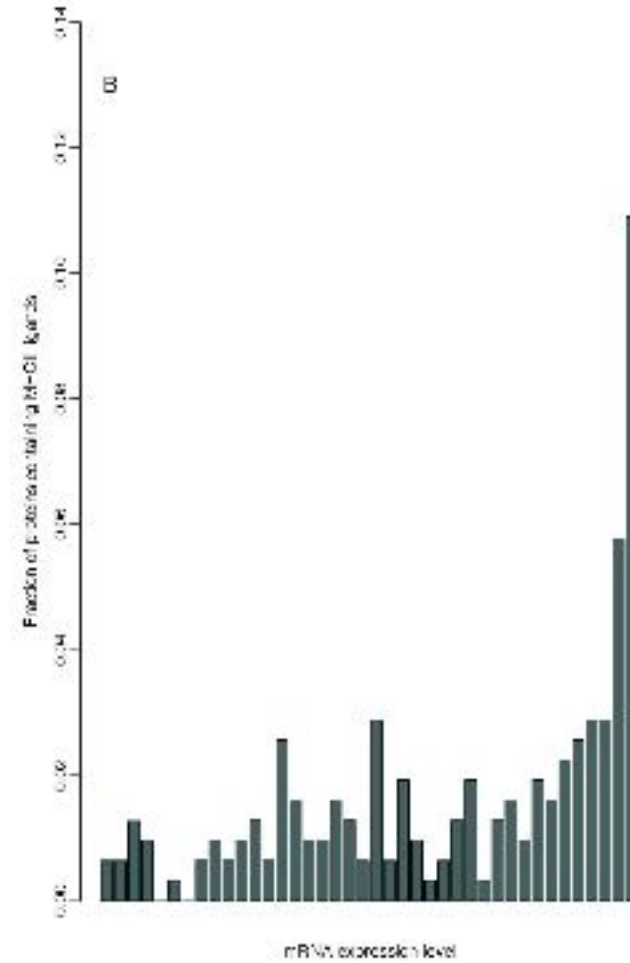
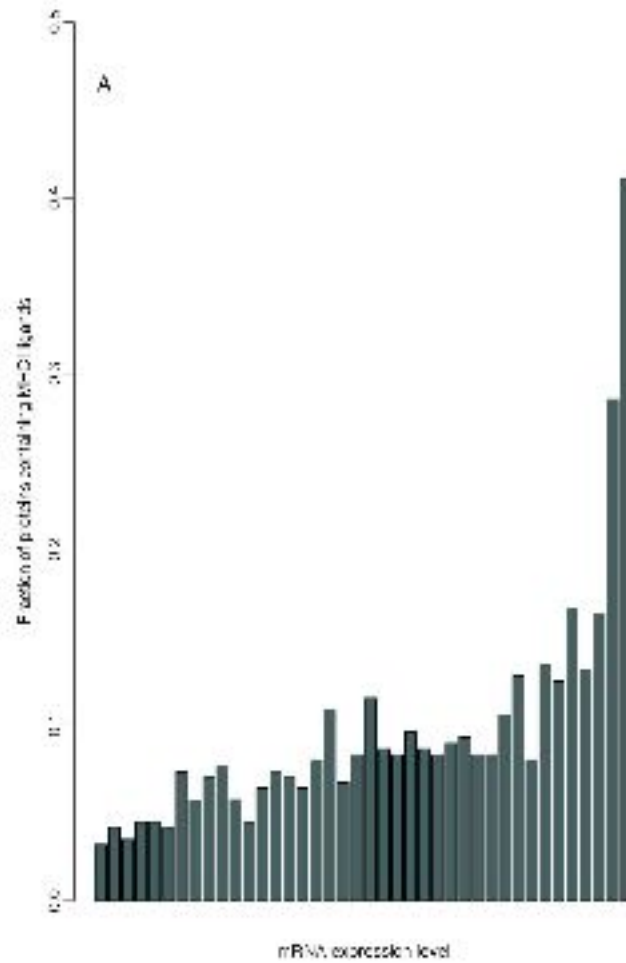
Designing diagnostic peptides

- Algorithm
 - Select 20mer peptides conserved in all strains the diagnose should cover
 - Deselect all 20mer peptides that have more than 7 identical residues in any strain the diagnose should not cover
 - Select for epitopes etc.
- Currently being tested in 3000 leprosy patients
 - Human TB, Cow paraTB underway
- Sheila Tuyet Tang, Claus Lundegaard
- Michel Klein, Annemieke Geluk (Leiden), Gregers Jungersen (Veterinærinst.)

Self similarity of HIV epitopes



Expression vs. antigen presentation



Location vs. antigen presentation

Table 1. Fractions of MHC class I and II ligand-containing proteins that belong to a selection of GO cellular component terms.

Gene Ontology Term	Fraction, MHCI	Corrected P-value, MHCI	Fraction, MHCII	Corrected P-value, MHCII	Fraction, background
cell	96.0%	2.05E-13	92.4%	0.443	91.1%
intracellular	86.7%	3.07E-60	61.6%	0.938	65.3%
organelle	69.3%	1.23E-33	49.2%	0.861	51.5%
cytoplasm	62.4%	8.88E-52	52.4%	4.42E-03	40.2%
nucleus	39.9%	3.57E-15	10.0%	1	20.0%
macromolecular complex	32.6%	2.94E-54	25.4%	8.58E-04	14.6%
cytosol	17.7%	3.00E-52	12.4%	1.42E-03	5.3%
organelle lumen	15.1%	1.04E-20	5.9%	0.861	7.0%
ribonucleoprotein complex	10.6%	7.69E-26	1.6%	0.995	3.6%
cytosolic ribosom	4.1%	3.93E-35	1.6%	0.125	0.4%
membrane	33.5%	1	58.9%	9.68E-05	42.8%
plasma membrane	14.5%	1	40.5%	1.81E-09	19.9%
extracellular region	5.6%	1	22.7%	2.27E-04	11.8%
integral to plasma membrane	4.5%	1	22.2%	1.04E-09	7.0%
vesicle	4.1%	0.023	14.1%	7.09E-10	2.8%
vacuole	1.5%	0.475	10.3%	6.25E-10	1.4%
lysosome	1.2%	0.705	9.2%	2.12E-09	1.2%
pigment granule	2.3%	1.65E-09	8.6%	1.27E-13	0.5%
melanosome	2.3%	1.65E-09	8.6%	1.27E-13	0.5%

Prediction servers at CBS

Web servers

CTL epitopes

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

MHC binding

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

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

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

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

MHC Motif viewer

<http://www.cbs.dtu.dk/biotools/MHCMotifViewer/Home.html>

Proteasome processing

<http://www.cbs.dtu.dk/services/NetChop-3.0/>

B-cell epitopes

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

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

Plotting of epitopes relative to reference sequence

<http://www.cbs.dtu.dk/services/EpiPlot-1.0/>

Analysis of human immunoglobulin VDJ recombination

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

Geno-pheno type association based mapping of binding sites

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

PhD/master course in Immunological Bioinformatics, June, 2009

<http://www.cbs.dtu.dk/courses/27685.imm/>

Work in progress: Pilot study based on DrugBank

- www.drugbank.ca
- Records corresponding to 123 FDA-approved biotech (protein/peptide) drugs were downloaded
- Sequences were compared to the human proteome (sequences from “Homo Sapiens” in NR (non redundant database from NCBI)) using blast.
- Sequences found in DrugBank and NR need to be manually validated/curated

Types of proteins

- Human/Human protein sequence Identical proteins
- Non human proteins
- Modified/allelic human proteins
- Antibodies
 - Non human
 - Human-murine chimaer
 - Humanized
 - Human

Proposed application in assessment of protein drugs

- 1 Compare amino acid sequence of drug with the human proteome
- 2 Predict epitopes in regions that differ from the human proteome
- 3 Select representative HLA alleles
- 4 Verify binding experimentally
- 5 Assess predicted immunogenicity using blood from treated patients/transgenic animals
- 6 Compare with clinical findings of immunogenicity/adverse effects/lack of effect

Immunonological Bioinformatics Group

- Ole Lund
- Claus Lundegaard
- Morten Nielsen
- Mette Voldby Larsen
- Sheila Tang
- Jorid Sørli
- Marlene Erup Larsen
- Bent Petersen
- Thomas Stranzl
- Massimo Andretta
- Edita Bartaseviciute