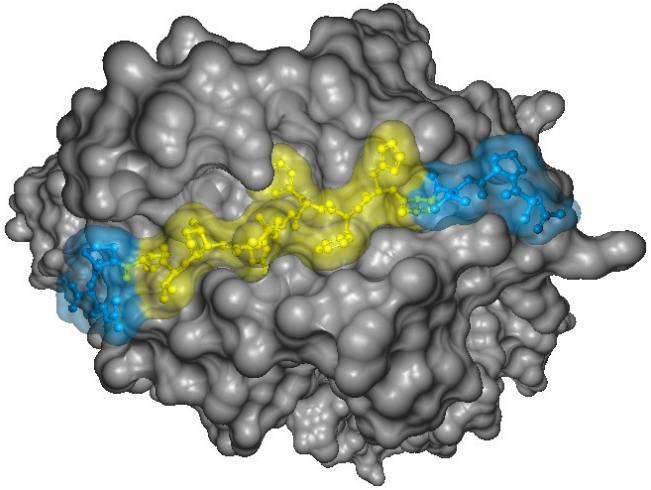
Predicting Binding Affinities of MHC Class II Epitopes Across Alleles

Nico Pfeifer and Oliver Kohlbacher

Center for Bioinformatics

Eberhard Karls University Tübingen



Roadmap

- Background
  - Major Histocompatibility Complex Class II (MHCII)
  - Problem Definition
- Methods
  - Multiple Instance Learning
  - Normalized Set Kernel
  - Positionally-weighted RBF Kernel
- Results

MHCII

- Major Histocompatibility Complex class II (MHCII)

MHCII

- Major Histocompatibility Complex class II (MHCII)
- Part of the adaptive immune system

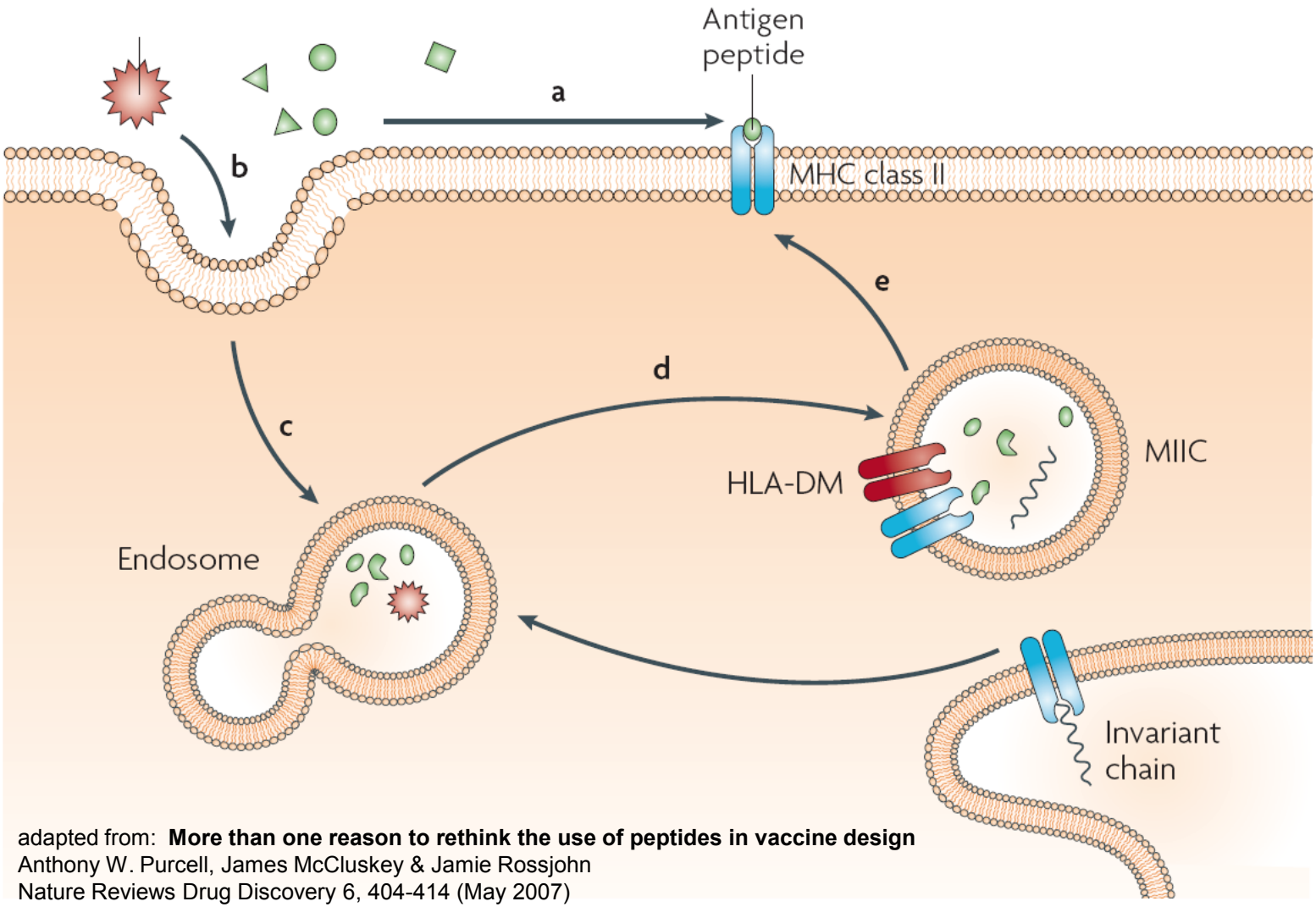
MHCII

- Major Histocompatibility Complex class II (MHCII)
- Part of the adaptive immune system
- Can be found primarily on antigen-presenting cells (macrophages, dendritic cells and B-cells)

MHCII

- Major Histocompatibility Complex class II (MHCII)
- Part of the adaptive immune system
- Can be found primarily on antigen-presenting cells (macrophages, dendritic cells and B-cells)
- Presents peptides which are derived from exogenous proteins to trigger an immune response

MHCII



adapted from: **More than one reason to rethink the use of peptides in vaccine design**
Anthony W. Purcell, James McCluskey & Jamie Rossjohn
Nature Reviews Drug Discovery 6, 404-414 (May 2007)

Why is MHCII interesting?

- Every human has up to eight different MHCII alleles

Why is MHCII interesting?

- Every human has up to eight different MHCII alleles
- Every corresponding MHCII molecule binds to a different set of peptides

Why is MHCII interesting?

- Every human has up to eight different MHCII alleles
- Every corresponding MHCII molecule binds to a different set of peptides
- Peptide vaccines can only activate immune responses if they bind to the MHC molecule

Why is MHCII interesting?

- Every human has up to eight different MHCII alleles
- Every corresponding MHCII molecule binds to a different set of peptides
- Peptide vaccines can only activate immune responses if they bind to the MHC molecule

$\Rightarrow$ Personalized vaccine design

Why is MHCII interesting?

- Every human has up to eight different MHCII alleles
- Every corresponding MHCII molecule binds to a different set of peptides
- Peptide vaccines can only activate immune responses if they bind to the MHC molecule

$\Rightarrow$ Personalized vaccine design

$\Rightarrow$ Good predictors for peptide binding needed

Definitions

Peptide-MHCII binding prediction

Given: Peptide sequence *s* and sequence *m* of an MHCII molecule

Problem: Decide whether *s* binds to *m* or not

Definitions

Peptide-MHCII binding prediction

Given: Peptide sequence *s* and sequence *m* of an MHCII molecule

Problem: Decide whether *s* binds to *m* or not

Peptide-MHCII binding *affinity* prediction

Given: Peptide sequence *s* and sequence *m* of an MHCII molecule

Problem: Decide how strong *s* binds to *m*

MHCII

- More than 700 different MHCII alleles are known [1]

1. Robinson J et al.:
IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex
Nucleic Acids Research (2003), **31**:311-314

MHCII

- More than 700 different MHCII alleles are known [1]
- For less than 3% there is sufficient experimental data from binding studies

1. Robinson J et al.:
IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex
Nucleic Acids Research (2003), **31**:311-314

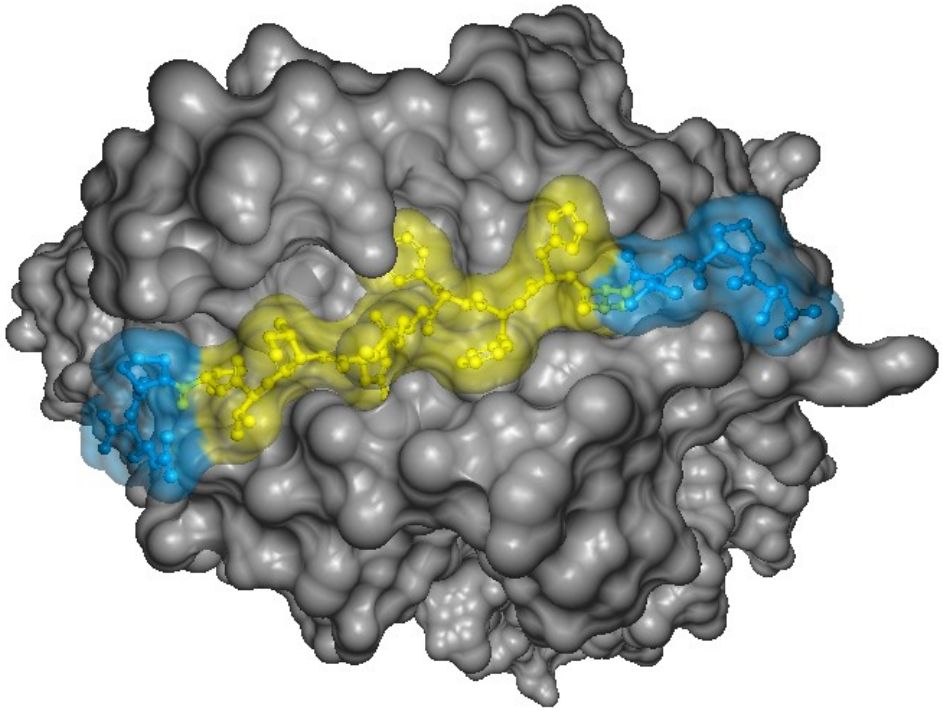
MHCII

- More than 700 different MHCII alleles are known [1]
- For less than 3% there is sufficient experimental data from binding studies
- Problems of peptide-MHCII binding and affinity prediction are not solved yet

1. Robinson J et al.:
IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex
Nucleic Acids Research (2003), **31**:311-314

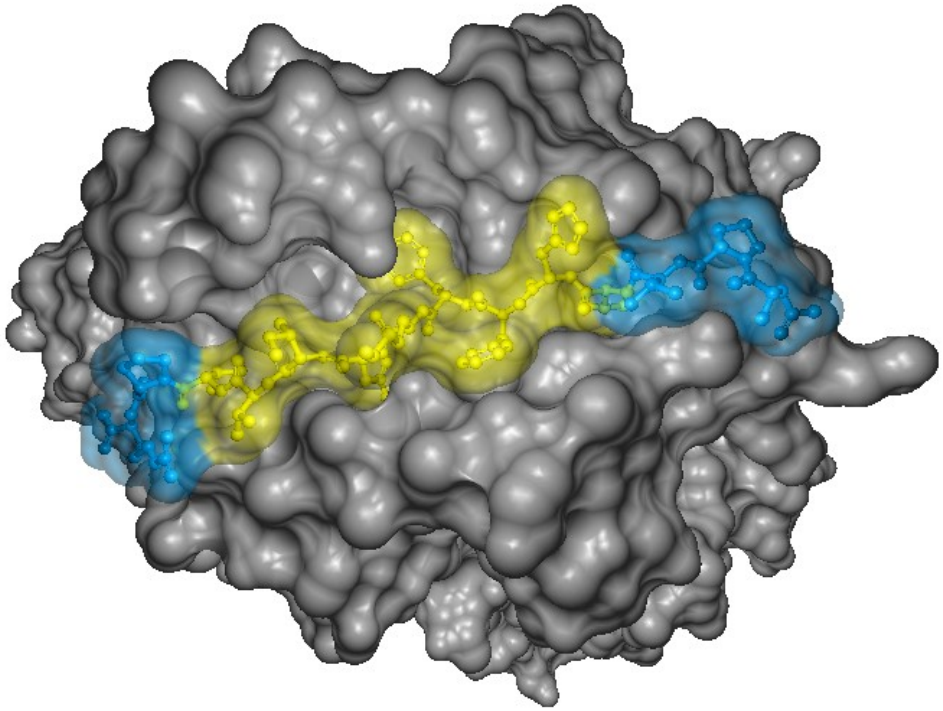
What makes the problems difficult?

- Varying peptide length since binding clefts are open



What makes the problems difficult?

- Varying peptide length since binding clefts are open
- The binding core (9 residues) is, in most of the cases, unknown



Related work

- Positional Scoring Matrices
- Gibbs Samplers
- Hidden Markov Models
- Artificial Neural Networks
- Partial Least Squares
- Support Vector Machines
- Consensus approach

For more information on related work have a look at the paper

N. Pfeifer and O. Kohlbacher: Multiple Instance Learning Allows MHC Class II Epitope Predictions across Alleles

Lecture Notes in Bioinformatics: Proceedings of WABI 2008

Related work

- Positional Scoring Matrices
- Gibbs Samplers
- Hidden Markov Models
- Artificial Neural Networks
- Partial Least Squares
- Support Vector Machines
- Consensus approach

Predictors only for a small fraction of all MHCII alleles

For more information on related work have a look at the paper

N. Pfeifer and O. Kohlbacher: Multiple Instance Learning Allows MHC Class II Epitope Predictions across Alleles

Lecture Notes in Bioinformatics: Proceedings of WABI 2008

Roadmap

- Background
  - Major Histocompatibility Complex Class II (MHCII)
  - Problem Definition
- Methods
  - Multiple Instance Learning
  - Normalized Set Kernel
  - Positionally-weighted RBF Kernel
- Results

Peptide-MHCII binding affinity prediction

- A core part of the peptide (9 amino acids long) binds to the MHCII molecule

Peptide-MHCII binding affinity prediction

- A core part of the peptide (9 amino acids long) binds to the MHCII molecule
- First idea: represent sequence *s* by a bag *X* containing all 9-mers of *s*

Peptide-MHCII binding affinity prediction

- A core part of the peptide (9 amino acids long) binds to the MHCII molecule
- First idea: represent sequence s by a bag X containing all 9-mers of s

$\Rightarrow$ Lots of noise in the bags

Peptide-MHCII binding affinity prediction

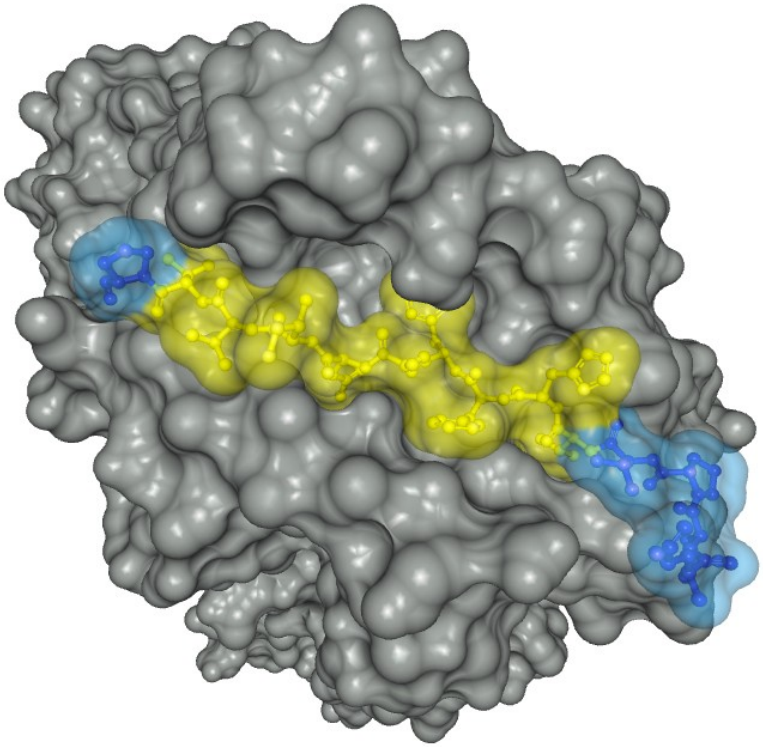
- A core part of the peptide (9 amino acids long) binds to the MHCII molecule
- First idea: represent sequence *s* by a bag *X* containing all 9-mers of *s*

$\Rightarrow$ Lots of noise in the bags

- Better idea: represent sequence *s* by a bag *X* containing all reasonable [2,3] putative binding cores: aromatic (F,W,Y) or aliphatic (I,L,M,V) amino acid at position 1

2. Hammer, J. et al.: High-affinity binding of short peptides to major histocompatibility complex class II molecules by anchor combinations. Proc Natl Acad Sci U S A 91(10) (1994) 4456-4460
3. Sturniolo, T. et al.: Generation of tissue-specic and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol 17(6) (1999) 555-561

Peptide-MHCII binding affinity prediction



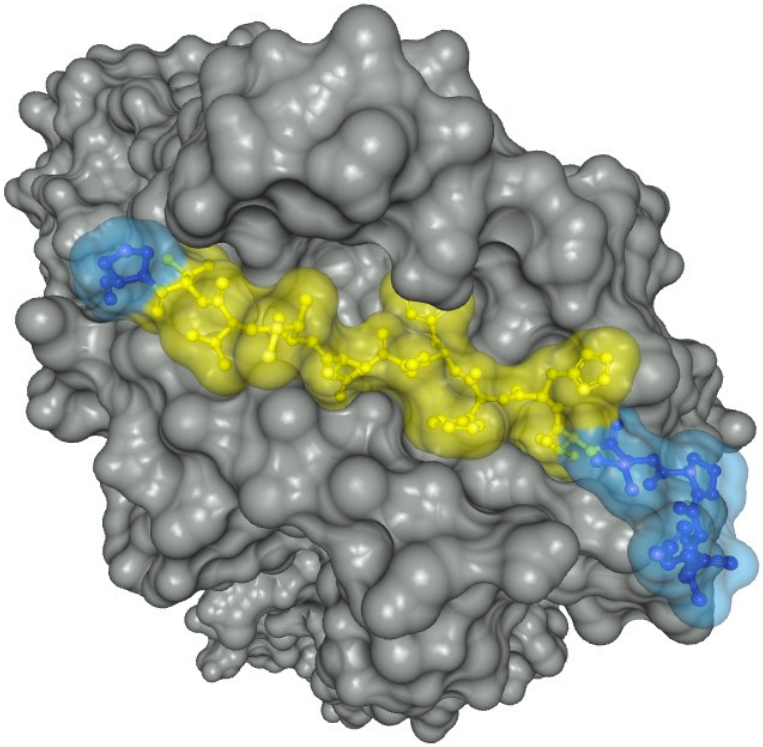
ENPVVHFFKNIVTPR

VVHFFKNIV

VHFFKNIVT

FFKNIVTPR

Peptide-MHCII binding affinity prediction



ENPVVHFFKNIVTPR

VVHFFKNIV

VHFFKNIVT

FFKNIVTPR

Represent sequence ENPVVHFFKNIVTPR by bag

$$X = \{\text{VVHFFKNIV}, \text{VHFFKNIVT}, \text{FFKNIVTPR}\}$$

Multiple Instance Learning

- Binary classification:

$$S = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n) | x_i \in \mathcal{X} \wedge y_i \in \{-1, 1\} \; \forall i = 1, 2, ..., n\}$$

Multiple Instance Learning

- Binary classification:

$$S = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n) | x_i \in \mathcal{X} \wedge y_i \in \{-1, 1\} \; \forall i = 1, 2, ..., n\}$$

- Multiple instance classification [4]:

$$S = \{(X_1, y_1), (X_2, y_2), ..., (X_n, y_n) | X_i \subseteq \mathcal{X} \wedge y_i \in \{-1, 1\} \; \forall i = 1, 2, ..., n\}$$

4. Dietterich, T.G. et al.: Solving the multiple instance problem with axis-parallel rectangles. Artif Intell 89(1-2) (1997) 31-71

Multiple Instance Learning

- Binary classification:

$$S = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n) | x_i \in \mathcal{X} \wedge y_i \in \{-1, 1\} \; \forall i = 1, 2, ..., n\}$$

- Multiple instance classification [4]:

$$S = \{(X_1, y_1), (X_2, y_2), ..., (X_n, y_n) | X_i \subseteq \mathcal{X} \wedge y_i \in \{-1, 1\} \; \forall i = 1, 2, ..., n\}$$

- Multiple instance regression [5,6]:

$$S = \{(X_1, y_1), (X_2, y_2), ..., (X_n, y_n) | X_i \subseteq \mathcal{X} \wedge y_i \in \mathbb{R} \; \forall i = 1, 2, ..., n\}$$

4. Dietterich, T.G. et al.: Solving the multiple instance problem with axis-parallel rectangles. Artif Intell 89(1-2) (1997) 31-71
5. Dooly, D.R. et al..: Multiple-instance learning of real-valued data. J Machine Learn Res 3 (2002) 651-678
6. Ray, S. and Page, D.: Multiple instance regression. In: ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (2001) 425-432

Normalized Set Kernel [7]

$$k(X, X') := \frac{\sum_{x \in X, x' \in X'} k_{\mathcal{X}}(x, x')}{f_{\text{norm}}(X) f_{\text{norm}}(X')}$$

with $k_{\mathcal{X}}$ being a kernel on $\mathcal{X}$ and

$$f_{\text{norm}}(X) = \sqrt{\sum_{x \in X, x' \in X} k_{\mathcal{X}}(x, x')}$$

7. Gärtner, T. et al.: Multi-instance kernels.
In Sammut, C., Hoffmann, A.G., eds.: ICML, Morgan Kaufmann (2002) 179-186

Single Predictor

- Create a bag X for every sequence of the training set

Single Predictor

- Create a bag X for every sequence of the training set
- Encode the $x \in X$ reflecting physicochemical properties [9]

9. Venkatarajan, M.S. and Braun, W.: New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. Journal of Molecular Modeling **7**(12) (2001) 445-453

Single Predictor

- Create a bag X for every sequence of the training set
- Encode the $x \in X$ reflecting physicochemical properties [9]
- Perform ν-Support Vector Regression (ν-SVR) with the normalized set kernel

$$k(X, X') := \frac{\sum_{x \in X, x' \in X'} k_{\mathcal{X}}(x, x')}{f_{\text{norm}}(X) f_{\text{norm}}(X')}$$

with $k_{\mathcal{X}}$ being the RBF kernel

9. Venkatarajan, M.S. and Braun, W.: New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. Journal of Molecular Modeling **7**(12) (2001) 445-453

MHCIIMulti

- *Single Predictor* can only be used if there is sufficient experimental binding data for the target allele (~3% of all MHCII alleles)

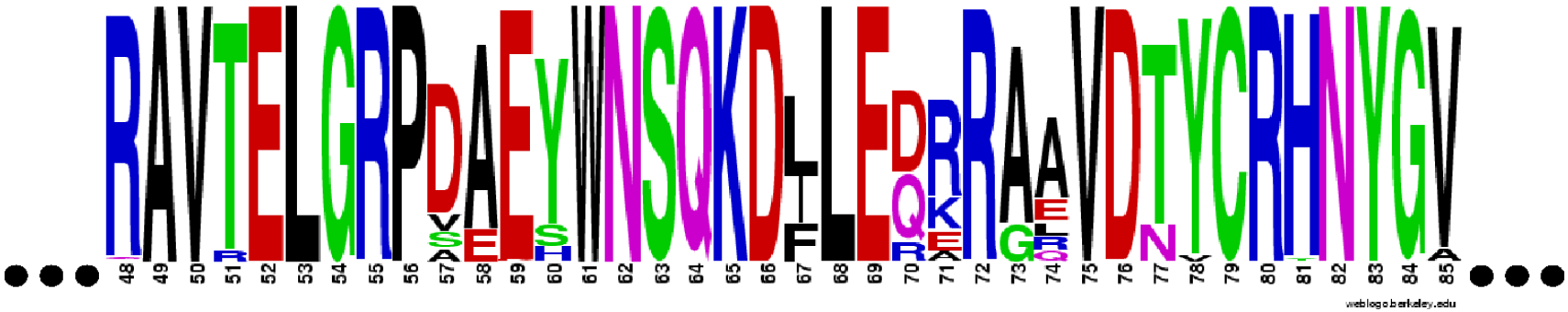
MHCIIMulti

- *Single Predictor* can only be used if there is sufficient experimental binding data for the target allele (~3% of all MHCII alleles)

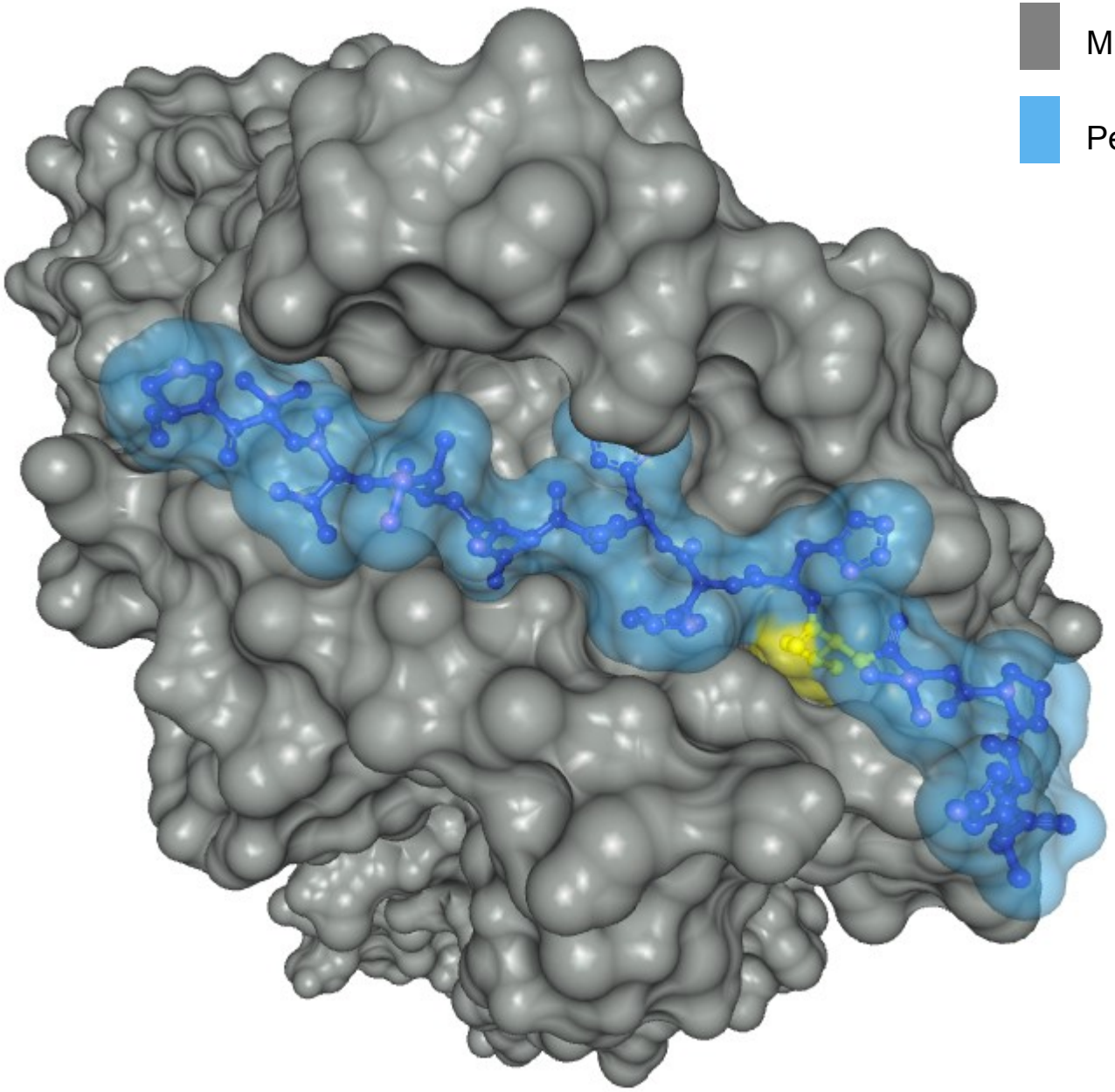
- Predictor for more alleles desirable

MHCIIMulti

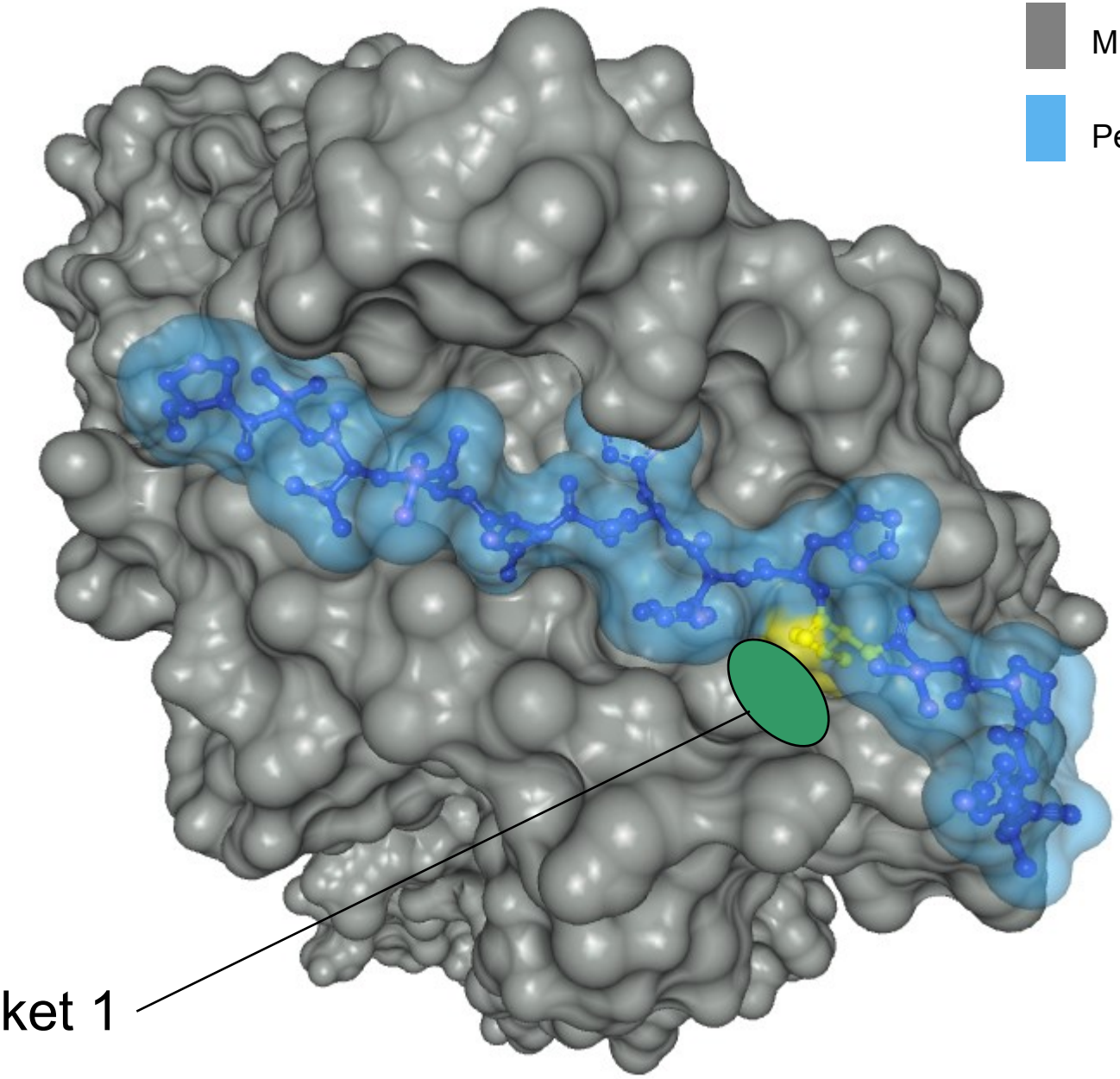
- *Single Predictor* can only be used if there is sufficient experimental binding data for the target allele (~3% of all MHCII alleles)
- Predictor for more alleles desirable
- HLA-DRB1, HLA-DRB3, HLA-DRB4 and HLA-DRB5 alleles differ only at very few positions



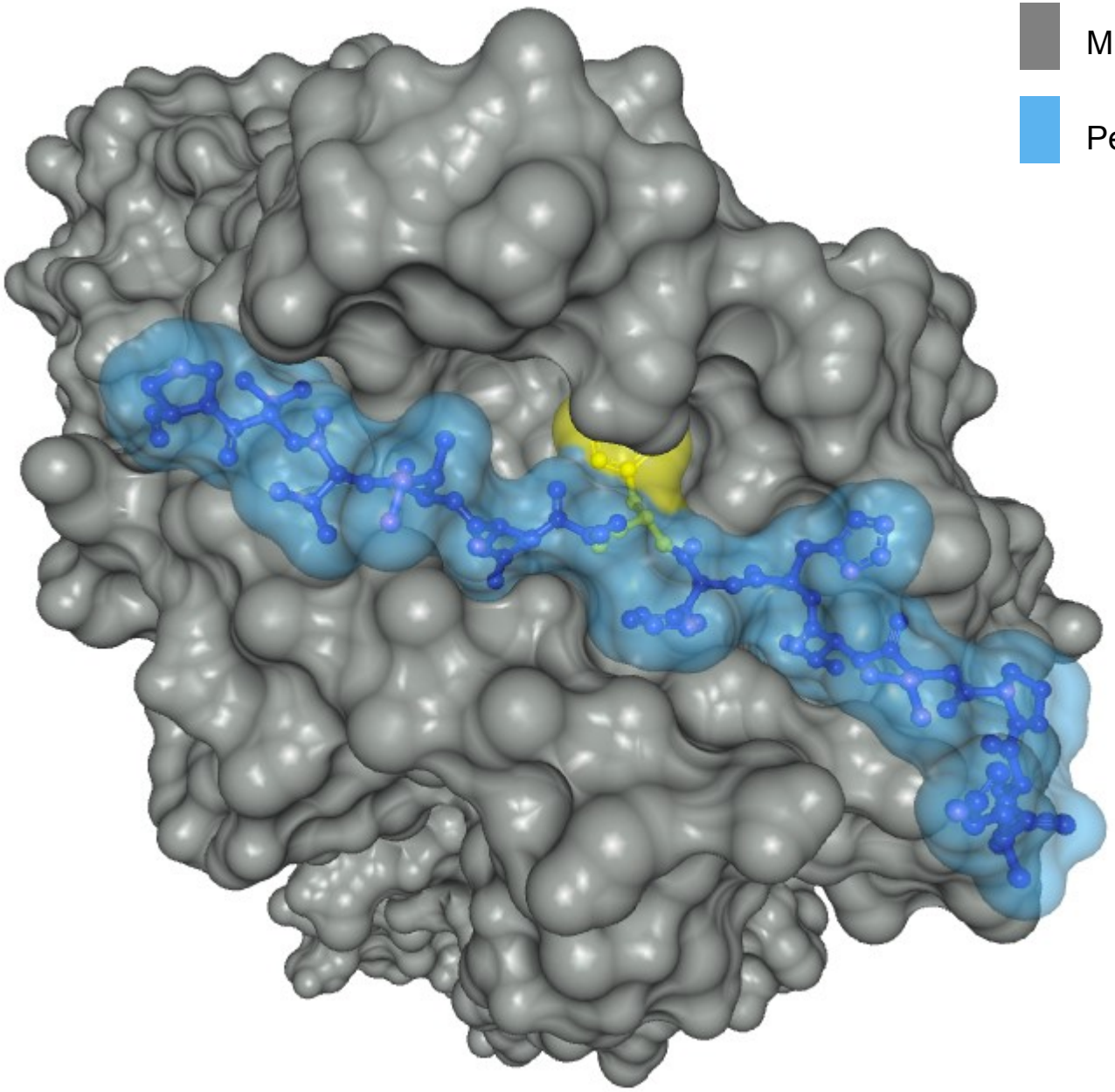
Methods



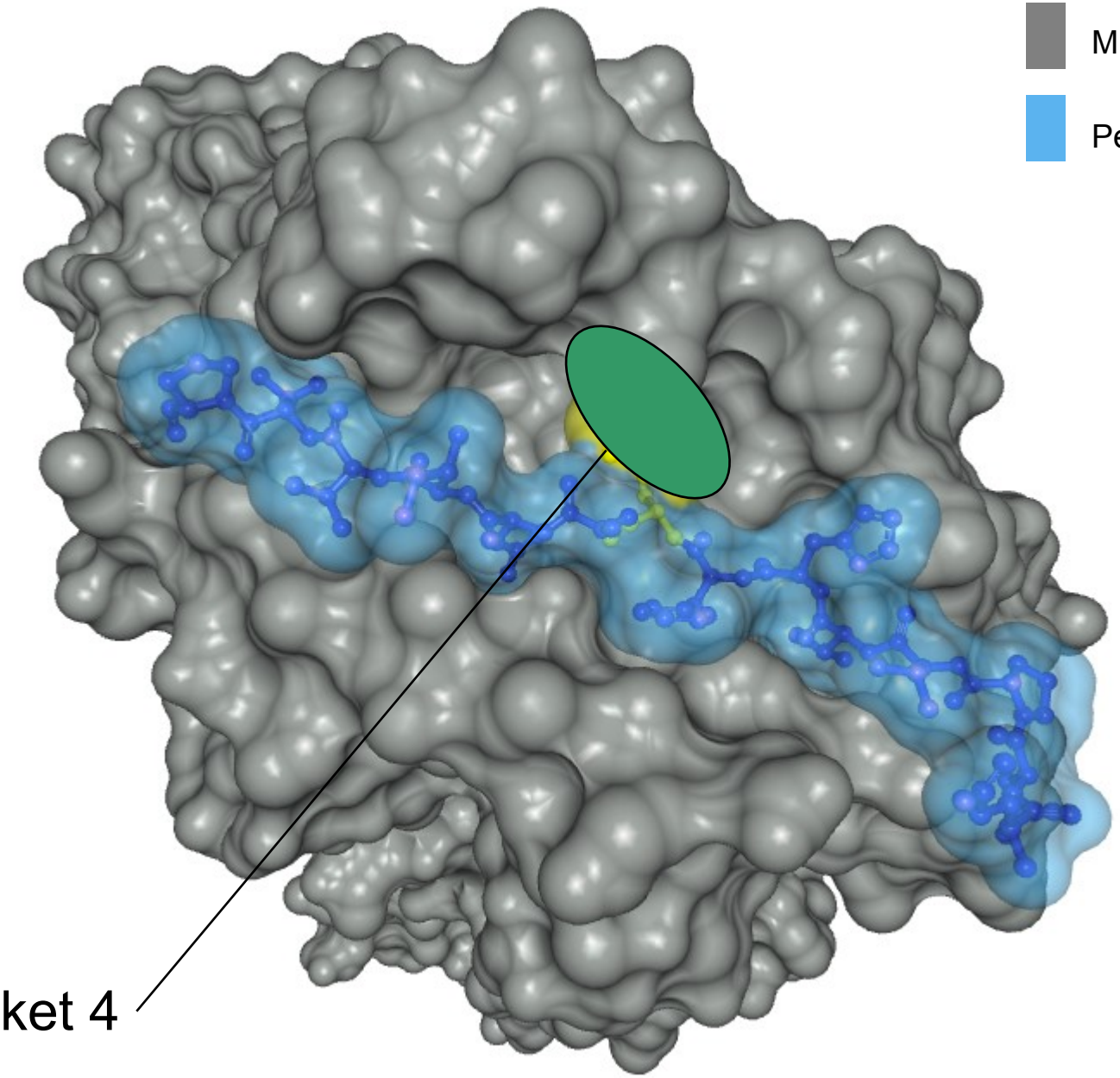
Methods



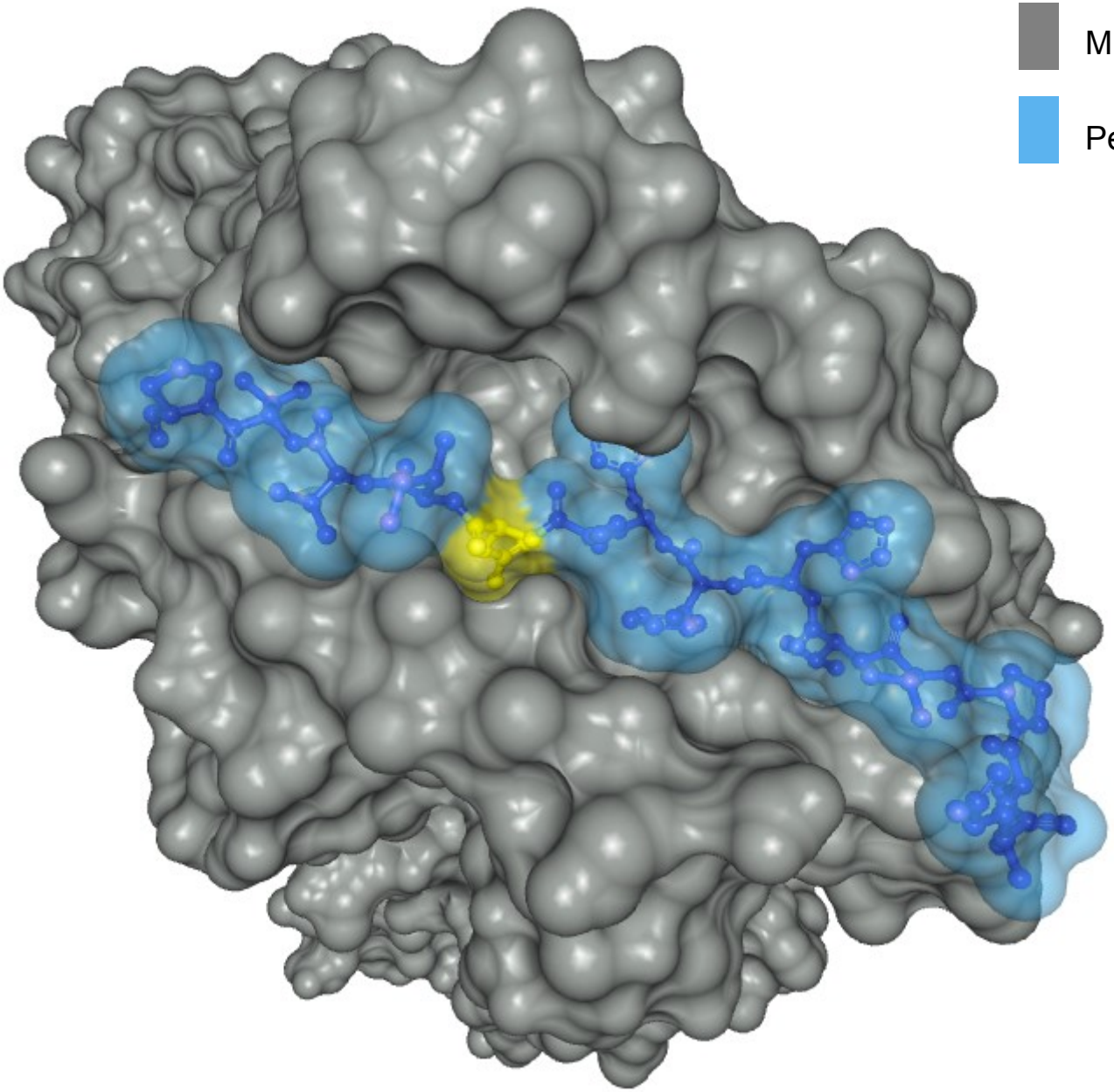
Methods



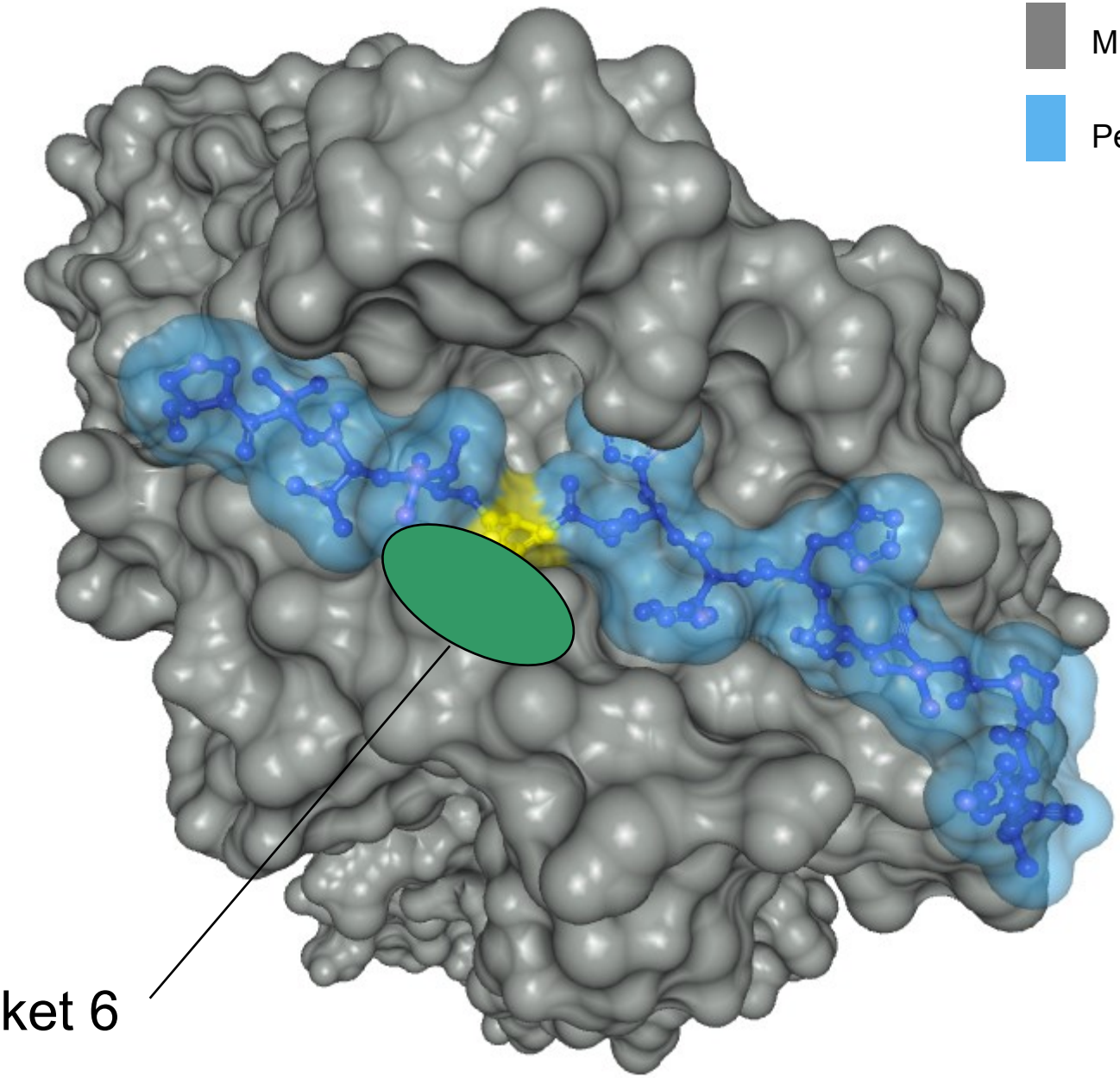
Methods



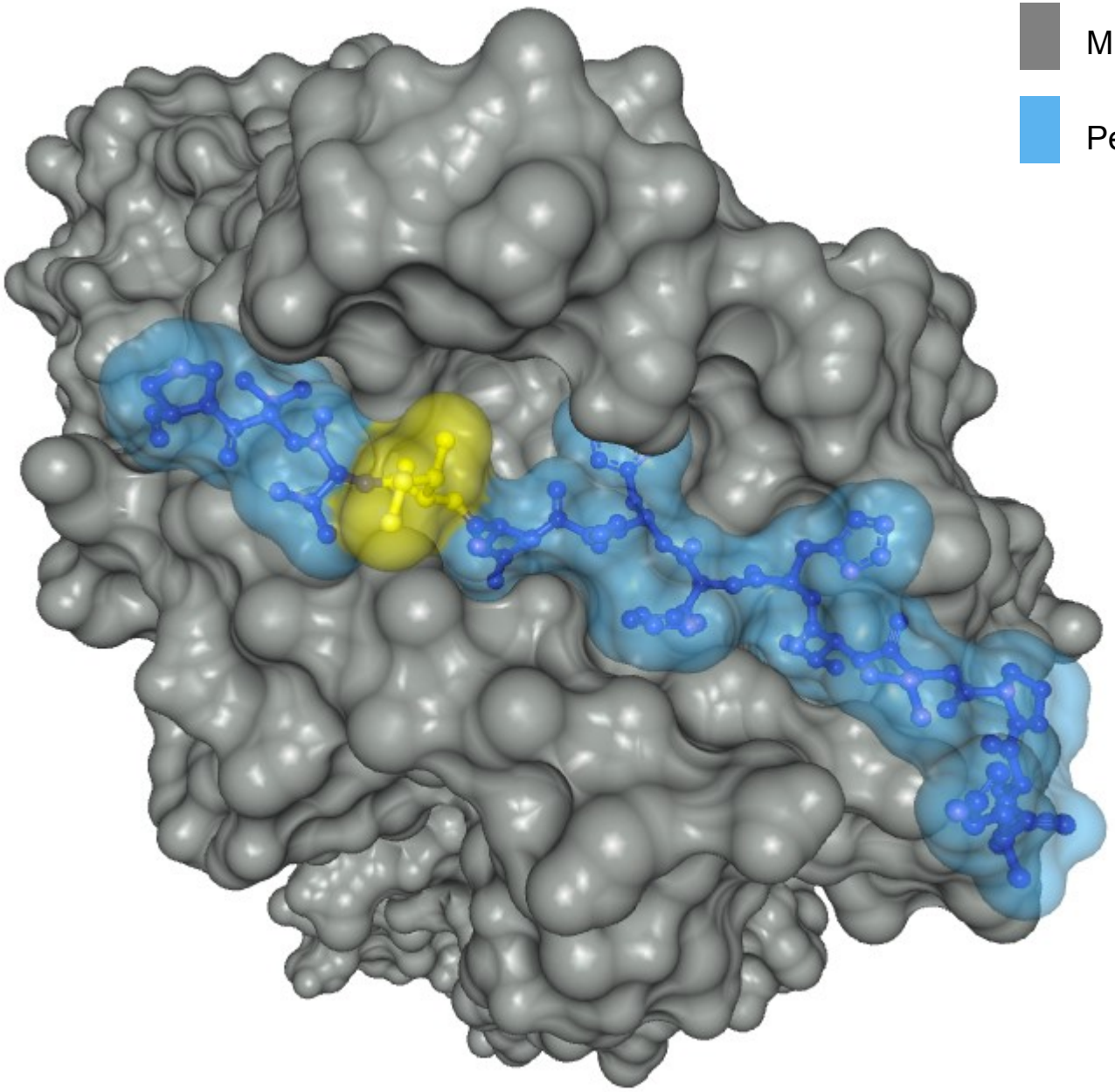
Methods

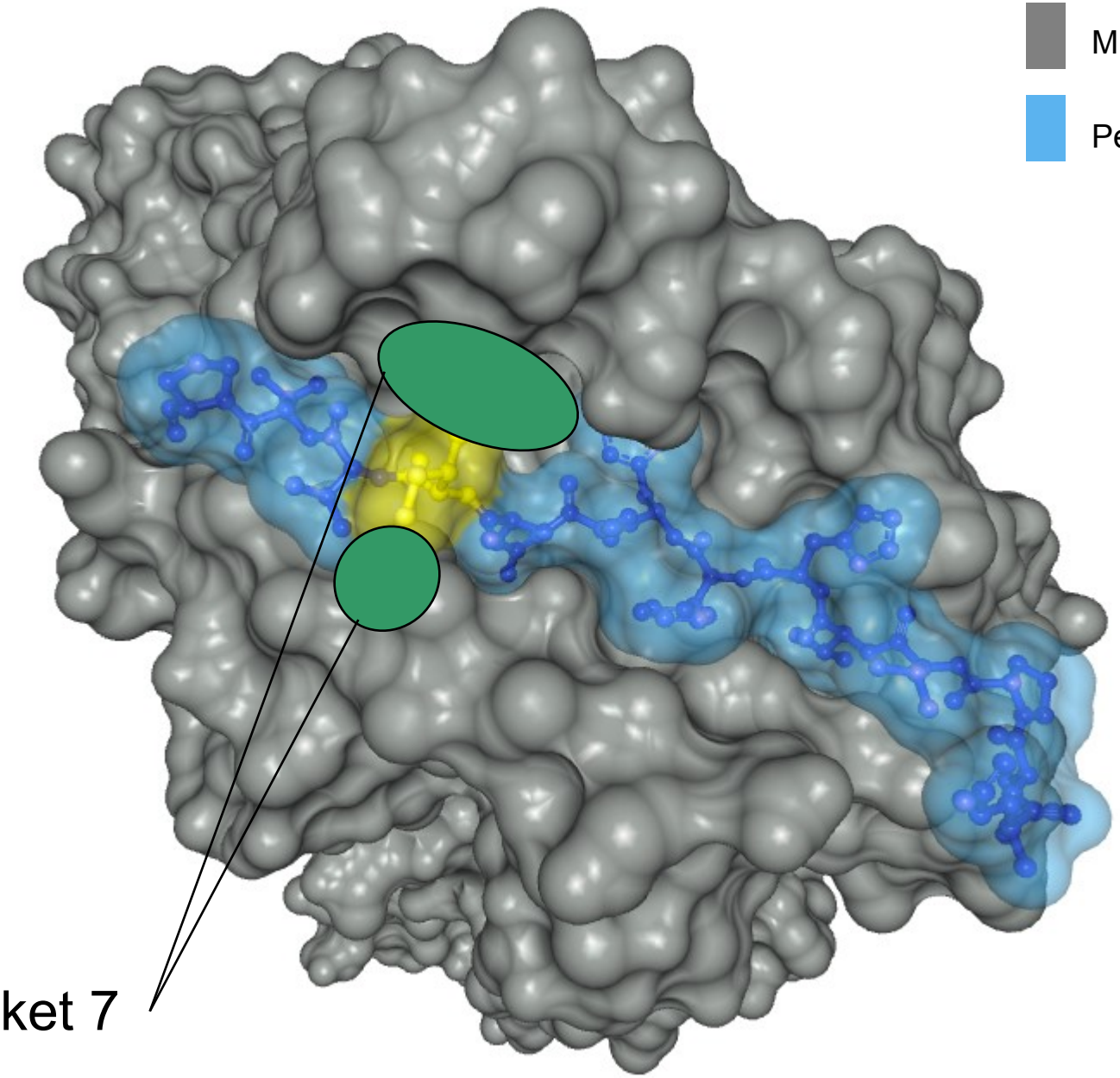


Methods

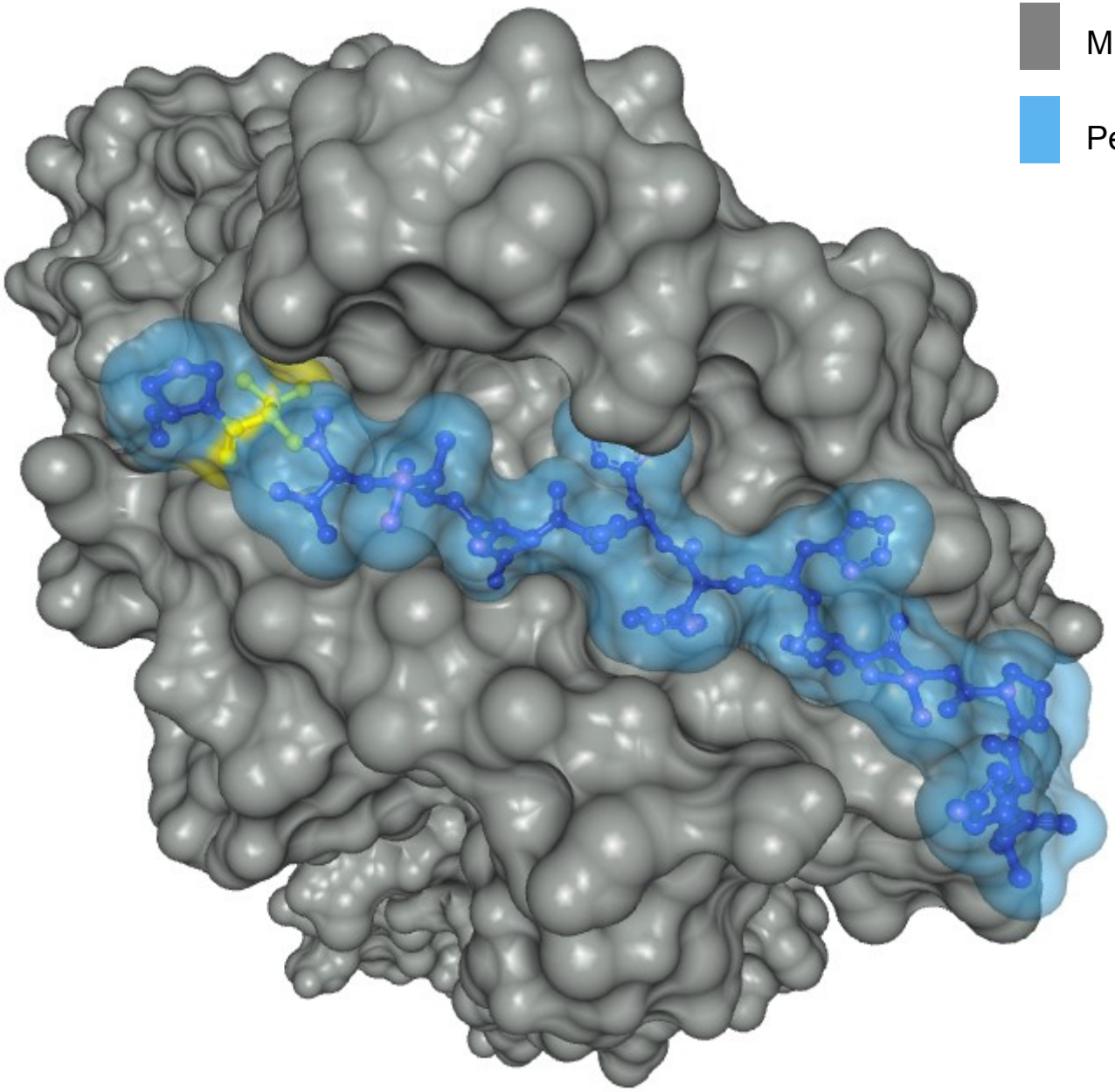


Methods

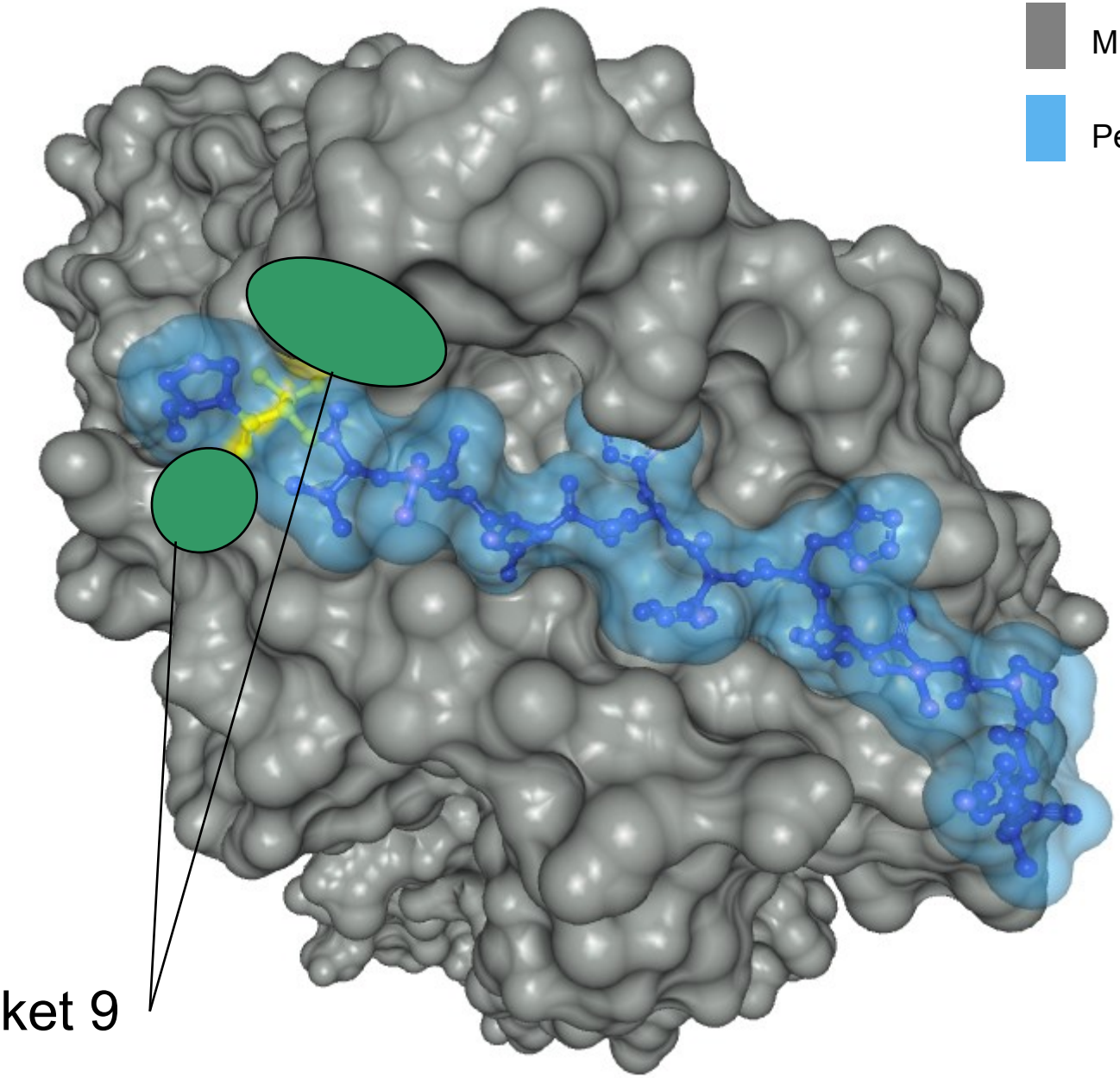



MHCII molecule
Peptide
Pocket 7

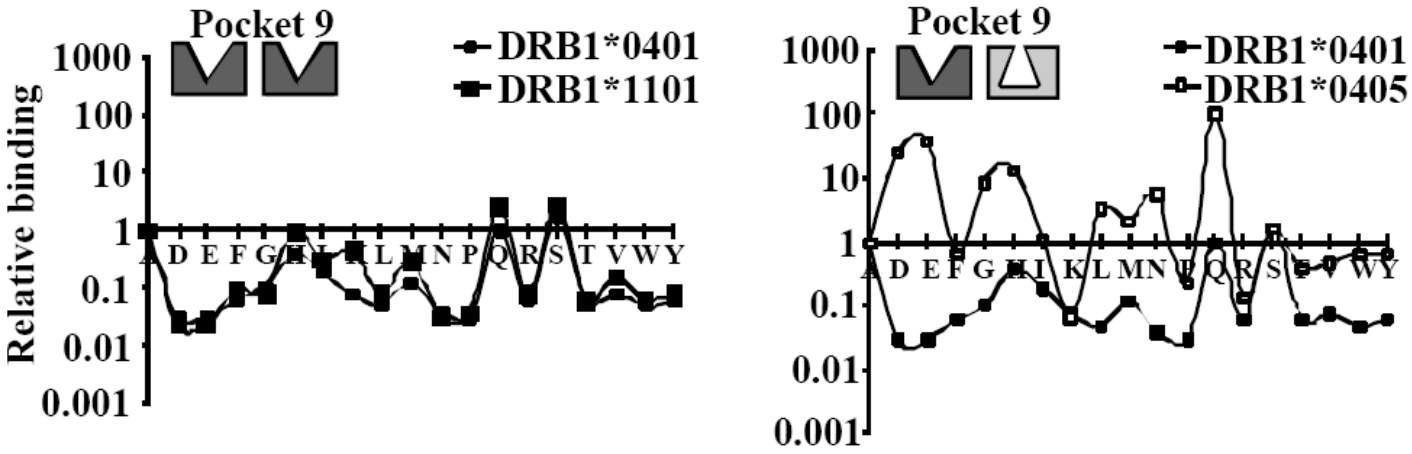
Methods



Methods



Virtual Matrices [3]



Adapted from [3]

- It was shown that alleles which have the same amino acids in a pocket region show similar binding affinities for the corresponding amino acid of the peptide

3. Sturniolo, T., Bono, E., Ding, J., Raddrizzani, L., Tuereci, O., et al.: Generation of tissue-specic and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol 17(6) (1999) 555-561

Covering more alleles

- Sturniolo *et al.* defined matrices for only ~7% of alleles

Covering more alleles

- Sturniolo *et al.* defined matrices for only ~7% of alleles

- More than 700 alleles known and more than 2/3 are HLA-DRB1, HLA-DRB3, HLA-DRB4 or HLA-DRB5 alleles

Covering more alleles

- Sturniolo *et al.* defined matrices for only ~7% of alleles
- More than 700 alleles known and more than 2/3 are HLA-DRB1, HLA-DRB3, HLA-DRB4 or HLA-DRB5 alleles
- Take alignments of the IMGT/HLA database [1] to get amino acids of the pocket regions for these alleles

[1] Robinson, J. et al.: IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res 31(1) (2003) 311-314

Covering more alleles

- Sturniolo *et al.* defined matrices for only ~7% of alleles
- More than 700 alleles known and more than 2/3 are HLA-DRB1, HLA-DRB3, HLA-DRB4 or HLA-DRB5 alleles
- Take alignments of the IMGT/HLA database [1] to get amino acids of the pocket regions for these alleles
- Calculate a mean pocket vector for pockets 1, 4, 6, 7, and 9 using the same encoding as for the peptides

$$p = \left[p_1^T, p_4^T, p_6^T, p_7^T, p_9^T\right]^T$$

[1] Robinson, J. et al.: IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res 31(1) (2003) 311-314

Multitask learning for MHC class I peptide binding [10]

- Kernel-based prediction of peptide MHC class I binding prediction:

$$K\left((x,a),(x',a')\right) = K_{pep}(x,x') \times K_{all}(a,a')$$

  - Simplifying assumption: all binding core positions are equally important

[10] Jacob, L. and Vert, J.-P.:Efficient peptide-MHC-I binding prediction for alleles with few known binders. Bioinformatics 24 (3) (2008) 358-366

Multitask learning for MHC class I peptide binding [10]

- Kernel-based prediction of peptide MHC class I binding prediction:

$$K\left((x,a),(x',a')\right) = K_{pep}(x,x') \times K_{all}(a,a')$$

  - Simplifying assumption: all binding core positions are equally important

- Our approach: weight positions according to similarities of the binding pockets $p = [p_1^T, p_4^T, p_6^T, p_7^T, p_9^T]^T$

[10] Jacob, L. and Vert, J.-P.:Efficient peptide-MHC-I binding prediction for alleles with few known binders. Bioinformatics 24 (3) (2008) 358-366

Positionally-weighted RBF kernel

- Increase weights of important positions

Positionally-weighted RBF kernel

- Increase weights of important positions
- If sequences differ at these positions, there will be a greater effect using this kernel than with the standard RBF kernel:

$$k_{\mathrm{pw-RBF}}((p,x),(p',x')) = \exp^{-\frac{w_1 \times \|x_1 - x'_1\|^2 + \ldots + w_9 \times \|x_9 - x'_9\|^2}{2\sigma^2}}$$

Positionally-weighted RBF kernel

- Increase weights of important positions

- If sequences differ at these positions, there will be a greater effect using this kernel than with the standard RBF kernel:

$$k_{\mathrm{pw-RBF}}((p,x),(p',x')) = \exp^{-\frac{w_1 \times \|x_1 - x'_1\|^2 + \ldots + w_9 \times \|x_9 - x'_9\|^2}{2\sigma^2}}$$

with

$$w_i := \begin{cases} \mathrm{Pearson}(p_i, p'_i) + 1 & \text{if } i = 1,4,6,7,9 \\ 0.5 & \text{otherwise} \end{cases}$$

Predictor which uses data from multiple alleles (*MHCIIMulti*)

- Create a bag X for every sequence of the training set

Predictor which uses data from multiple alleles (*MHCIIMulti*)

- Create a bag X for every sequence of the training set
- Encode $x \in X \text{ and } p$ reflecting physicochemical properties [9]

9. Venkatarajan, M.S. and Braun, W.: New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. Journal of Molecular Modeling **7**(12) (2001) 445-453

Predictor which uses data from multiple alleles (*MHCIIMulti*)

- Create a bag X for every sequence of the training set
- Encode $x \in X$ and p reflecting physicochemical properties [9]
- Perform ν-SVR with the normalized set kernel

$$k((p,X),(p',X')) := \frac{\sum_{x \in X, x' \in X'} k_{\mathcal{X}}((p,x),(p',x'))}{f_{\text{norm}}((p,X)) f_{\text{norm}}((p',X'))}$$

with $k_{\mathcal{X}}$ being the pw-RBF kernel

9. Venkatarajan, M.S. and Braun, W.: New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. Journal of Molecular Modeling **7**(12) (2001) 445-453

Roadmap

- Background
  - Major Histocompatibility Complex Class II (MHCII)
  - Problem Definition
- Methods
  - Multiple Instance Learning
  - Normalized Set Kernel
  - Positionally-weighted RBF Kernel
- Results

Benchmark

- Recent study to compare peptide-MHCII binding predictors [11] on a new benchmark dataset

11. Wang, P. et al.: A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 4(4) (2008) e1000048

Benchmark

- Recent study to compare peptide-MHCII binding predictors [11] on a new benchmark dataset
- Performance comparison on all human MHCII alleles of benchmark to our methods (14 alleles, 9478 peptides)

11. Wang, P. et al.: A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 4(4) (2008) e1000048

Benchmark

- Recent study to compare peptide-MHCII binding predictors [11] on a new benchmark dataset

- Performance comparison on all human MHCII alleles of benchmark to our methods (14 alleles, 9478 peptides)

- ARB method performance measured by ten-fold CV

$\Rightarrow$ Single Predictor performance measured by ten-fold CV

$\Rightarrow$ MHCIIMulti performance measured by ten-fold CV

11. Wang, P. et al.: A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 4(4) (2008) e1000048

Benchmark

- Recent study to compare peptide-MHCII binding predictors [11] on a new benchmark dataset

- Performance comparison on all human MHCII alleles of benchmark to our methods (14 alleles, 9478 peptides)

- ARB method performance measured by ten-fold CV

$\Rightarrow$ Single Predictor performance measured by ten-fold CV

$\Rightarrow$ MHCIIMulti performance measured by ten-fold CV

- Performance measure: area under ROC curve

11. Wang, P. et al.: A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 4(4) (2008) e1000048

Comparison

our methods

Allele	ARB	ProPred	SMM-align	Consensus	Single Predictor	MHCIIMulti
DRB1*0101	0.76	0.74	0.77	0.79	0.81	0.75
DRB1*0301	0.66	0.65	0.69	0.72	0.73	0.72
DRB1*0401	0.67	0.69	0.68	0.69	0.67	0.78
DRB1*0404	0.72	0.79	0.75	0.80	0.79	0.80
DRB1*0405	0.67	0.75	0.69	0.72	0.83	0.79
DRB1*0701	0.69	0.78	0.78	0.83	0.82	0.90
DRB1*0802	0.74	0.77	0.75	0.82	0.76	0.79
DRB1*0901	0.62	-	0.66	0.68	0.64	0.66
DRB1*1101	0.73	0.80	0.81	0.80	0.85	0.87
DRB1*1302	0.79	0.58	0.69	0.73	0.74	0.73
DRB1*1501	0.70	0.72	0.74	0.72	0.72	0.75
DRB3*0101	0.59	-	0.68	-	0.72	0.57
DRB4*0101	0.74	-	0.71	0.74	0.79	0.78
DRB5*0101	0.70	0.79	0.75	0.79	0.81	0.90
mean	0.70	0.73	0.73	0.76	0.76	0.77

Comparison

our methods

Allele	ARB	ProPred	SMM-align	Consensus	Single Predictor	MHCIIMulti
DRB1*0101	0.76	0.74	0.77	0.79	0.81	0.75
DRB1*0301	0.66	0.65	0.69	0.72	0.73	0.72
DRB1*0401	0.67	0.69	0.68	0.69	0.67	0.78
DRB1*0404	0.72	0.79	0.75	0.80	0.79	0.80
DRB1*0405	0.67	0.75	0.69	0.72	0.83	0.79
DRB1*0701	0.69	0.78	0.78	0.83	0.82	0.90
DRB1*0802	0.74	0.77	0.75	0.82	0.76	0.79
DRB1*0901	0.62	-	0.66	0.68	0.64	0.66
DRB1*1101	0.73	0.80	0.81	0.80	0.85	0.87
DRB1*1302	0.79	0.58	0.69	0.73	0.74	0.73
DRB1*1501	0.70	0.72	0.74	0.72	0.72	0.75
DRB3*0101	0.59	-	0.68	-	0.72	0.57
DRB4*0101	0.74	-	0.71	0.74	0.79	0.78
DRB5*0101	0.70	0.79	0.75	0.79	0.81	0.90
mean	0.70	0.73	0.73	0.76	0.76	0.77

Leave-One-Allele-Out (LOAO) Prediction

- Train Leave-One-Allele-Out (LOAO) version of *MHCIIMulti* on binding data of all alleles but the target allele (10 times randomly drawn subset)

Leave-One-Allele-Out (LOAO) Prediction

- Train Leave-One-Allele-Out (LOAO) version of *MHCIIMulti* on binding data of all alleles but the target allele (10 times randomly drawn subset)
- Build aggregating predictor *MHCIIMulti** from all 10 random draws (mean prediction value)

Leave-One-Allele-Out (LOAO) Prediction

- Train Leave-One-Allele-Out (LOAO) version of *MHCIIMulti* on binding data of all alleles but the target allele (10 times randomly drawn subset)
- Build aggregating predictor *MHCIIMulti** from all 10 random draws (mean prediction value)
- Compare to *Single Predictor* of most similar allele

$\Rightarrow$ Measure performance on the target allele

Comparison

Allele	Single Predictor Trained on Nearest Neighbor	LOAO MHCIIMulti	LOAO MHCIIMulti*
DRB1*0101	0.74	0.64	0.69
DRB1*0301	0.61	0.69	0.70
DRB1*0401	0.64	0.76	0.78
DRB1*0404	0.70	0.79	0.82
DRB1*0405	0.78	0.76	0.77
DRB1*0701	0.72	0.89	0.91
DRB1*0802	0.71	0.77	0.79
DRB1*0901	0.57	0.64	0.65
DRB1*1101	0.79	0.87	0.9
DRB1*1302	0.62	0.68	0.69
DRB1*1501	0.66	0.75	0.77
DRB3*0101	0.51	0.54	0.54
DRB4*0101	0.77	0.69	0.72
DRB5*0101	0.73	0.89	0.92
mean	0.68	0.74	0.76

Summary

- Methods for peptide-MHCII binding affinity prediction perform quite well on alleles with sufficient data

Summary

- Methods for peptide-MHCII binding affinity prediction perform quite well on alleles with sufficient data

- Presentation of method for building predictor for about 2/3 of all MHCII alleles (instead of ~7%)

Summary

- Methods for peptide-MHCII binding affinity prediction perform quite well on alleles with sufficient data
- Presentation of method for building predictor for about 2/3 of all MHCII alleles (instead of ~7%)
- LOAO performance results are quite promising

  $\Rightarrow$ Can be assumed that the methods perform well for alleles for which no experimental data exist

Summary

- Methods for peptide-MHCII binding affinity prediction perform quite well on alleles with sufficient data
- Presentation of method for building predictor for about 2/3 of all MHCII alleles (instead of ~7%)
- LOAO performance results are quite promising
  $\Rightarrow$ Can be assumed that the methods perform well for alleles for which no experimental data exist
- Methods are integrated into the EpiToolKit [12] (http://www.epitoolkit.org/mhciimulti)

[12] Feldhahn, M., Thiel, P., Schuler, M.M., Hillen, N., Stevanovic, S., et al.: EpiToolKit-a web server for computational immunomics. Nucleic Acids Res (2008) 6 (Web Server issue): W519-522

graduating soon
looking for postdoc position

Thanks to

... Oliver Kohlbacher

... Lena Feldhahn and Philipp Thiel

... SFB 685, project B1

SFB 685
Immuntherapie

... you for listening

Comparison to NetMHCIIPan

Allele	Number of training samples (MHCIIMulti*)	MHCIIMulti*	Number of training samples (NetMHCIIPan)	NetMHCIIPan
1_0101	4430	0.670	9441	0.778
1_0401	4609	0.765	13583	0.775
1_0405	4595	0.777	13977	0.808
1_1101	4604	0.863	13657	0.799
1_1302	4224	0.651	14109	0.658
1_1501	2900	0.744	13673	0.738
1_0301	3240	0.738	13587	0.746
1_0404	3426	0.796	13944	0.852
1_0701	4076	0.833	13754	0.825
1_0802	4538	0.789	14187	0.841
1_0901	3779	0.684	14077	0.653
3_0101	4302	0.630	14058	0.716
4_0101	3206	0.746	14161	0.724
5_0101	4511	0.848	13683	0.831
mean:	4031	0.752	13564	0.768

Positive Definiteness

THEOREM (SCHOENBERG THEOREM). Let $\mathcal{X}$ be a space in which a distance function $d(x, y)$ is defined subject to the following conditions:

1. $d(x, y) = d(y, x) \geq 0$
2. $d(x, x) = 0$

for all $x, y \in \mathcal{X}$. The function $\exp(-d^p(x, y))$ is positive definite if $0 < p \leq 2$ and not positive definite if $p > 2$.

Li, H., Jiang, T.: A class of edit kernels for SVMs to predict translation initiation sites in eukaryotic mRNAs. In: RECOMB. (2004) 262-271
Schoenberg, I.J.: Metric spaces and positive denite functions. Trans Amer Math Soc 44(3) (1938) 522-536