Predicting Immunogenicity of Peptides using Structural Alerts

Suman Chakravarti, Sai Radha Mani Alla

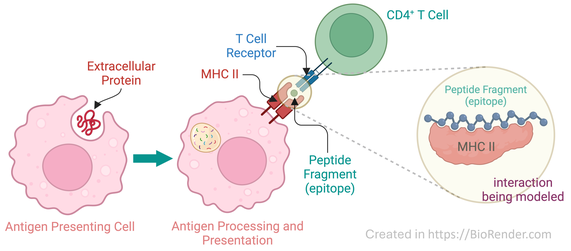
MultiCASE Inc., Mayfield Heights, OH, USA

SOT 2025
3675/A135

Hypothesis

Conventional **statistical alert discovery** and **QSAR methods**, originally designed for small organic molecules, can be effectively adapted to model and predict **peptide immunogenicity** by accounting for peptide-specific structural and physicochemical properties.

MHC II Driven Immunogenicity



Importance & Challenges

Safety & Efficacy: Assessing peptide immunogenicity is crucial for ensuring drug safety and vaccine effectiveness by preventing adverse immune reactions and achieving the desired immune response.

Challenges: 1. Conventional QSAR software, built for small organic molecules, often processes peptides at the atomic level, limiting effective analysis. 2. Obtaining both qualitative and quantitative predictions is challenging.

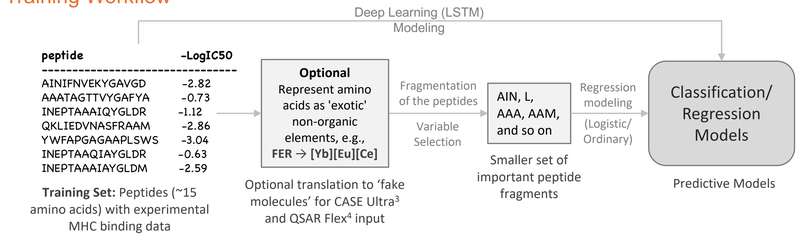
Data

Pilot Study Data: 2399 peptides (799 immunogenic, 1,600 non-immunogenic, HLA Class II immunogenicity) from Dhanda *et al.*[1]

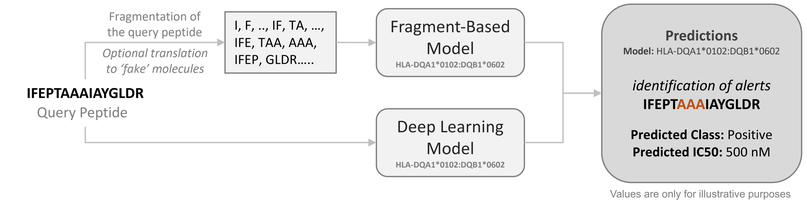
Peptide-MHC II Binding Data: 44,888 unique peptides from IEDB Database[2] covered 26 HLA alleles (HLA-DQ, HLA-DR, HLA-DP). -LogIC50 binding affinity was used for quantitative modeling, while for classification, peptides with IC50 < 1000 nM were labeled as binders.

Methods

Training Workflow



Evaluation Workflow

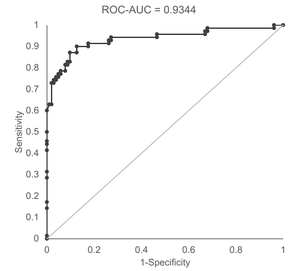


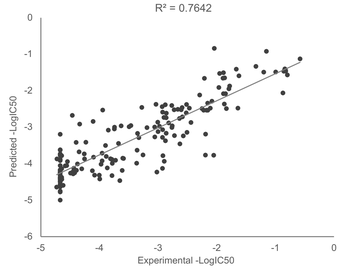
MHC Binding Prediction for an Immunogenic Peptide

DDYTEYKLTESIDNILVKMFKTN
Serine-repeat antigen protein 5 (UniProt: Q9TY95)
Plasmodium falciparum
MHC II Allele: HLA-DQA1*01:01/DQB1*05:01
Fragment-Based Model

Predicted IC50 944.50 nM	DDYTEYKLTESIDNILVKMFKTN *alerts shown in red on the query peptide* **Top Alerts & Their Contributions:** **D** (0.4072), **F** (0.2037), **Y** (0.2249), **IDN** (0.2466), **ESI** (0.0864) **Top Deactivating Features and Their Contributions:** **K** (-0.2421), **T** (-0.1027), **TN** (-0.0126)

Results





Classification and Regression Model Predictions for the Test Set of the Allele: HLA-DQA1*01:01/DQB1*05:01 (Fragment-Based Models)

Top 5 Best-Performing Models Among 26

MHC II Allele	Training Size	ROC-AUC	R^2	RMSE
HLA-DPA1*02:01/DPB1*05:01	1269	0.921	0.756	0.649
HLA-DPA1*02:01/DPB1*01:01	1260	0.919	0.803	0.64
HLA-DPA1*03:01/DPB1*04:02	1267	0.915	0.788	0.66
HLA-DPA1*01:03/DPB1*02:01	1264	0.905	0.784	0.72
HLA-DPA1*01:01/DPB1*04:01	1204	0.904	0.817	0.702

Conclusions

- **Structural alert discovery** methods used for small molecule toxicity endpoints can also identify **immunogenic motifs** in peptides.
- **Conventional QSAR techniques** can also be applied effectively to build transparent, high-performance models for peptide immunogenicity.
- Identifying strong MHC-binding peptides may aid in **vaccine development**.
- Predicting dominant antigenic regions supports **T-cell epitope discovery**.

References

1. Dhanda et al., Front Immunol. 2018, 9:1369
2. https://www.iedb.org/home_v3.php
3. CASE Ultra, ver 1.9.2.3, MultiCASE Inc. USA, www.multicase.com
4. QSAR Flex, ver 3.2, MultiCASE Inc. USA, www.multicase.com

✉ chakravarti@multicase.com, srmalla@multicase.com
Both authors are employees of MultiCASE Inc.