HLA-EpiCheck: A B-cell epitope prediction tool on HLA antigens using molecular dynamics simulation data

Diego Amaya-Ramirez, Romain Lhotte, Cedric Usureau, Magali Devriese, Malika Smaïl-Tabbone, Jean-Luc Taupin, Devignes Marie-Dominique

To cite this version:

Diego Amaya-Ramirez, Romain Lhotte, Cedric Usureau, Magali Devriese, Malika Smaïl-Tabbone, et al.. HLA-EpiCheck: A B-cell epitope prediction tool on HLA antigens using molecular dynamics simulation data. ISMB-ECCB 2023 - Intelligent System in Molecular Biology and European Conference on Computational Biology merged event, Jul 2023, Lyon, France. 2023. hal-04405086

HAL Id: hal-04405086

https://inria.hal.science/hal-04405086

Submitted on 19 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
In the context of organ transplantation, recipient’s antibodies against donor-specific HLA antigens are the main reason for transplant loss. The prediction of B-cell (antibody) epitopes on HLA antigens is therefore a key challenge on the way to improving the matching step between donor and recipient from a structural point of view. Here, we present HLA-EpiCheck, a B-cell epitope prediction tool that relies on an unprecedented dataset of short Molecular Dynamics (MD) simulations of 207 HLA antigens. We use hydrophobic properties, electrostatic charges, flexibility and solvent accessibility as descriptors calculated on patches sampled from MD trajectories. Then, we train an Extremely Randomized Trees machine learning model. This model outperforms the state-of-the-art DiscoTope 3.0 tool [1] for B-cell epitope prediction on HLA antigens.

MD data:
- 207 HLA antigens (the most prevalent in the European population).
- Initial templates for MD simulations were obtained from PDB [3] or AlphaFold2 [4].
- 10ns MD simulations performed for each HLA antigen and 500 frames extracted from the last 5ns of the trajectory.
- NAMD3 software [5], CHARMM36 [6] force field and TIP3P explicit solvent model were used for MD runs.

Dataset:
- Patch definition:
  - Calculated for each frame of each MD run of each HLA antigen.
  - Centered on every surface accessible residue (RSASA* ≥ 20%).
  - Composed of the surface accessible residues within a radius of 7Å from the central residue (centers of mass where used to measure distances).
- 18 descriptors: (8 static and 10 dynamic).
- 2 classes: epitope and non-epitope.
- Epitope labels = confirmed eplets in the HLA Eplet Registry database [2].
- Non epitope = neither confirmed nor putative eplet.
- Patches composed solely of conserved residues (with respect to the locus) were discarded.

Model selection (on Training set)

Feature Importance for the Extra Trees model

Comparison with DiscoTope 3.0 (on Test set)

Conclusion and Perspectives

- Modest contribution of MD descriptors on prediction performance.
- Poor performance of DiscoTope 3.0: no HLA antigen in the dataset used to train DiscoTope.

Find better descriptors for MD: analyse false negative epitopes.
- Try instance-based predictors such as K-NN.

References


DAR is recipient of an Inria-Inserm doctoral fellowship. This project is funded by ANR EPI-HLA N° ANR-22-CE15-0036-03.

Experiments presented in this poster were carried out using the Grid’5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (see https://www.grid5000.fr), as well as the MBI-DS4H platform hosted by Inria/Loria and funded by CPER IT2MP (Contrat Plan État Région, Innovations, Technologiques, Modélisation & Médicine Personnalisée), including a FEDER co-funding.