HLA3DB: comprehensive annotation of peptide/HLA complexes enables blind structure prediction of T cell epitopes

Sagar Gupta1,2, Sanrupti Nerli1, Sreeja Kutti Kandy1, Glenn L. Merkey1, Nikolaos G. Sgourakis1,2

1Center for Computational and Genomic Medicine, Department of Pathology and Laboratory Medicine, The Children’s Hospital of Philadelphia
2Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania
3College of Arts and Sciences (2024), University of Pennsylvania

sagarg@sas.upenn.edu

doi: 10.1101/2023.03.20.533510

Introduction

Peptide MHC-I

Peptide binding via P2 and P0

Human leukocyte antigens (HLAs), the human version of major histocompatibility complex molecules (MHC-I), display self, foreign, or aberrant peptides on the cell surface for immune surveillance1. The sequence and conformational diversity offered by the center of the peptide defines the immunologically important area of peptide/HLA-I (pHLA) structures2. Thus, accurate modeling of the center of the peptide is necessary to understand and predict the molecular basis for immunogenicity.

Peptide conformational diversity

A structural framework, termed the fixed-local frame, captures peptide conformational diversity.

Primary anchor residues work as a fixed frame while the center of the peptide enables backbone changes.

A comprehensive structural analysis of 946 pHLA X-ray crystal structures (n = 293) from our database (HLA3DB) was conducted to determine the distributions of φ and ψ dihedral angles for each peptide position.

The central residues of the peptide (P4 to P7) showed frequent deviations towards more extended (β-strand) or more condensed (α-helix) structure.

Unbiased classification of peptide backbones

A minimal set of 34 peptide backbones can describe the entire conformational landscape.

A two-dimensional PCA plot shows that rare configurations are represented.

The most common peptide backbone accounted for 35% of all configurations in our dataset.

The pHLA complexes represented by this peptide backbone originated from diverse HLA superotypes, indicating that similar antigen conformations are not solely defined by the identity of HLA groove residues.

Sequence biases across discrete conformations

A comprehensive structural analysis of 946 pHLA X-ray crystal structures (n = 293) from our database (HLA3DB) was conducted to determine the distributions of φ and ψ dihedral angles for each peptide position.

The central residues of the peptide (P4 to P7) showed frequent deviations towards more extended (β-strand) or more condensed (α-helix) structure.

Augmentation of the peptide sequence space by Rosetta demonstrate how backbone conformation can impose steric constraints or introduce favorable interactions via crosstalk between backbone and side chain features.

Accurate structural modeling

Leveraging the discrete backbones to enable structural modeling of new peptides on all HLA allotypes, RepPred outperforms five existing methods by at least 32% with respect to D-score.

RepPred models an immunodominant HIV epitope with near-native side chain placement as a result of high-fidelity generation of the peptide backbone.

Conclusions

Using HLA3DB, our database of pHLA structures, we characterize peptide backbone diversity for 946 pHLA complexes.

We find that peptide backbone similarity is allotype-independent and identify 34 discrete configurations which cover the conformational landscape.

We discover strong peptide sequence trends influenced by distinct backbone features.

RepPred, an accurate, pan-allelic structural modeling approach for pHLA complexes, demonstrates improved accuracy over existing methods.

References and Acknowledgments


This research was supported through grants by NIAID (5R01AI143997), NIGMS (6R35GM125034), and NIDDK (5U01DK112217).

HLA3DB can be accessed via https://hla3db.research.chop.edu.