

The Role of Hydrophobicity in Peptide-MHC Binding

Arnav Solanki, Marc Riedel, James Cornette, Julia Udell, Ishaan Koratkar, and George Vasmatzis

October 11, 2021



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

Overview

- Background on MHC Class I
- NetMHC-4.0 and NetMHCpan-4.1
- Our data analysis and flowcharts
- Hydrophobicity scale and calculations
- Results
- Conclusion



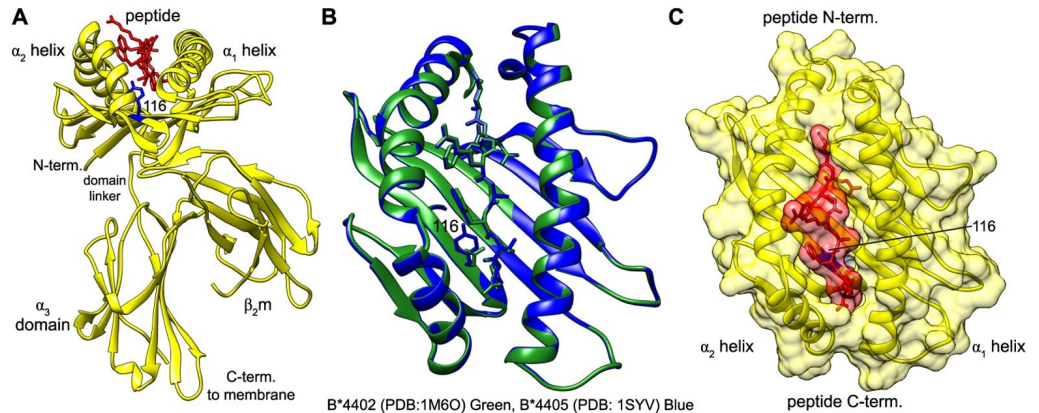
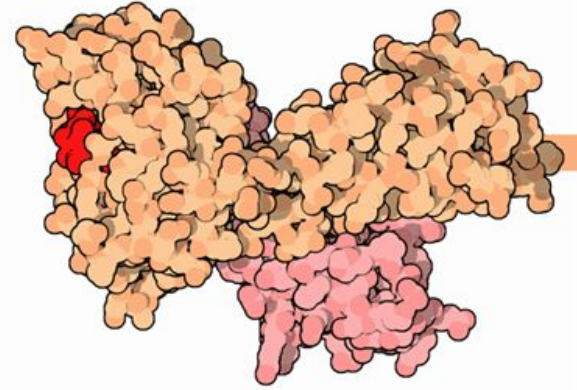
IOWA STATE
UNIVERSITY



MHC Class I proteins

- Immune surveillance of cells - detect foreign material
- Cell surface protein that presents peptides
- Peptide is held in a groove, with ends in binding pockets
- Mechanics of binding depend on MHC variant
- Highly polymorphic

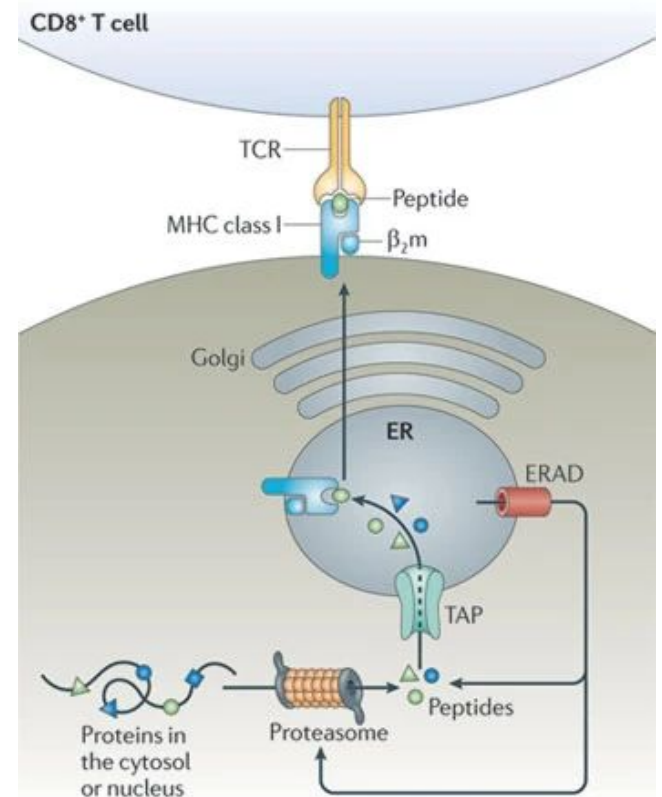
Applications - vaccine design, cancer neoantigen, viral disease severity



Antigen presentation pathway

- Endogenous protein
- Broken into small peptides
- Peptide binds to MHC class I protein (pMHC)
- If stable, transported to exterior
- Killer T cell binds to pMHC
 - If it is a foreign antigen -> infected cell is killed by T cell that recognizes it

As pMHC binding is the most selective step, it is the most studied process in determining cellular immune response.



Nature Reviews | Immunology

Neural Networks Prediction Tools

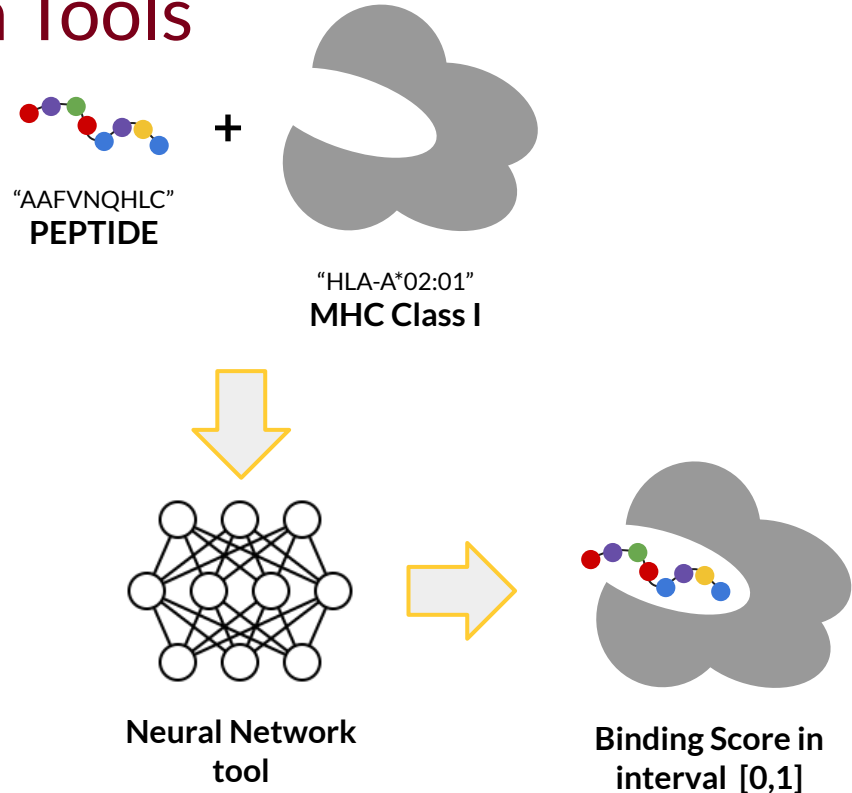
State-of-the-art tools:

- NetMHC-4.0
- NetMHCpan-4.1

Differences - MHC sequence, training data

Binding score in training data:

- Binding Affinity - continuous
- Eluted Likelihood - discrete



Massimo Andreatta, Morten Nielsen, Gapped sequence alignment using artificial neural networks: application to the MHC class I system, *Bioinformatics*, Volume 32, Issue 4, 15 February 2016, Pages 511–517, <https://doi.org/10.1093/bioinformatics/btv639>

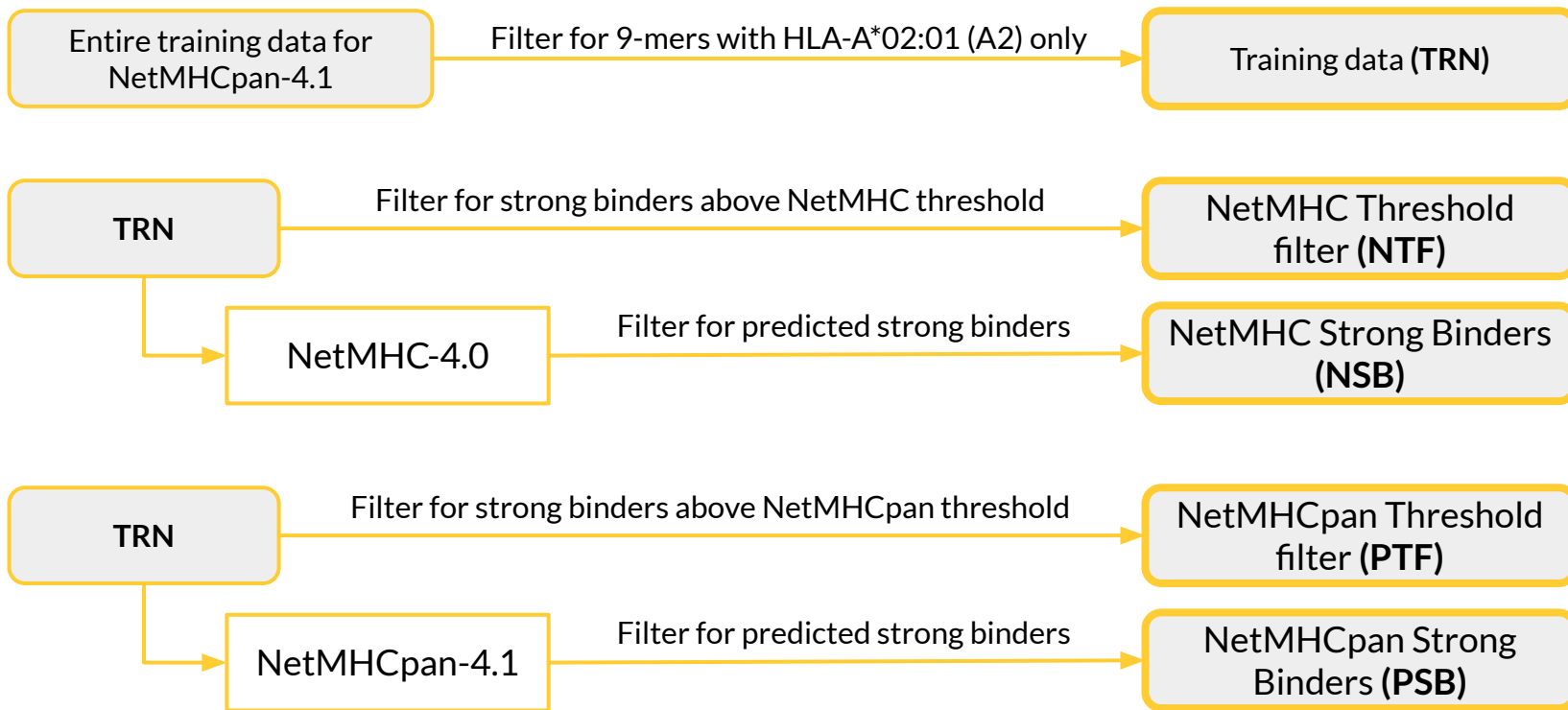
Birkir Reynisson, Bruno Alvarez, Sinu Paul, Bjoern Peters, Morten Nielsen, NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data, *Nucleic Acids Research*, Volume 48, Issue W1, 02 July 2020, Pages W449–W454, <https://doi.org/10.1093/nar/gkaa379>

Two different data analyses

- Training data
- Human Proteome data

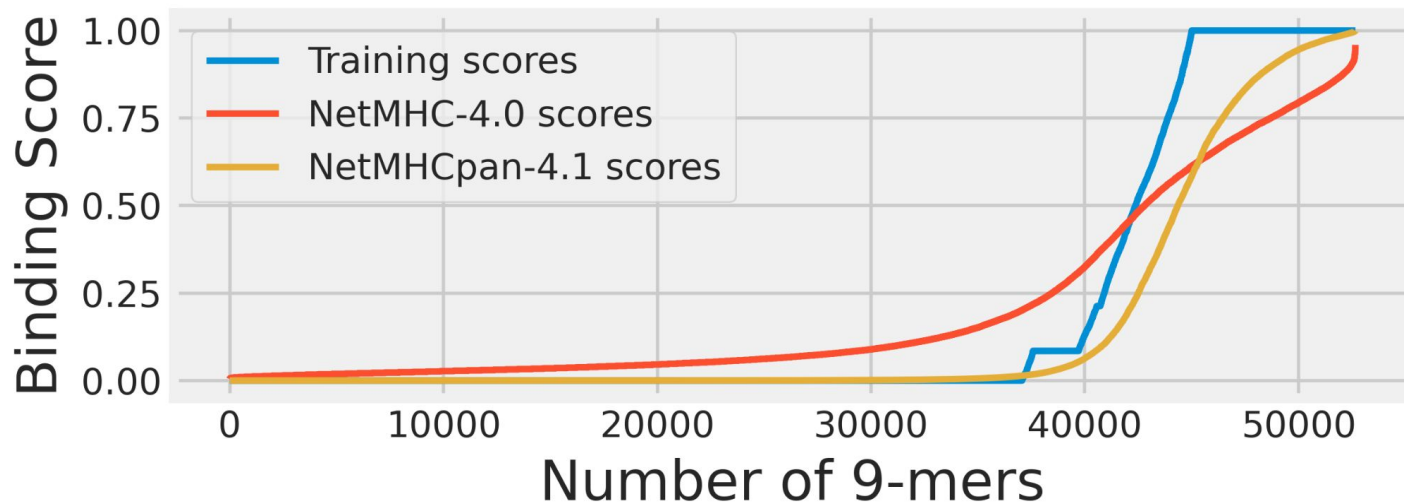
We only analyzed 9-mers, which show the strongest affinity for MHC Class I, and HLA-A*02:01 (A2), the most frequent MHC allele supertype

Training Data flowchart

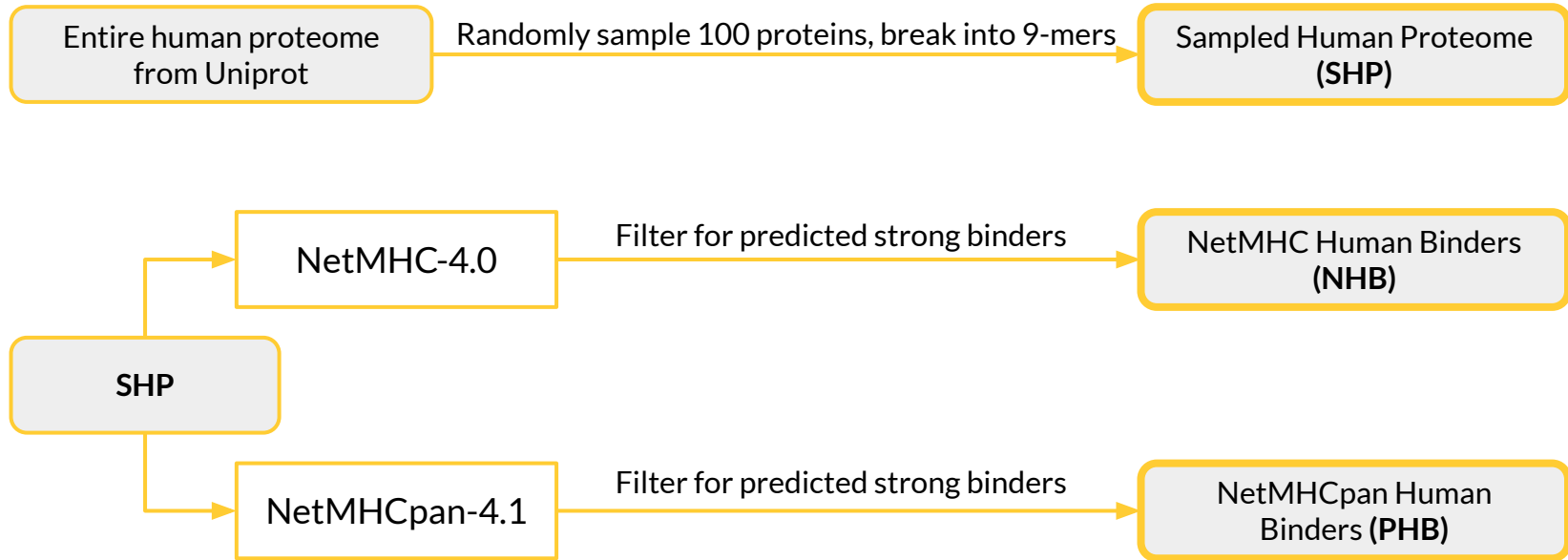


Training Data Scores

Sorted Scores for all A2 9-mers

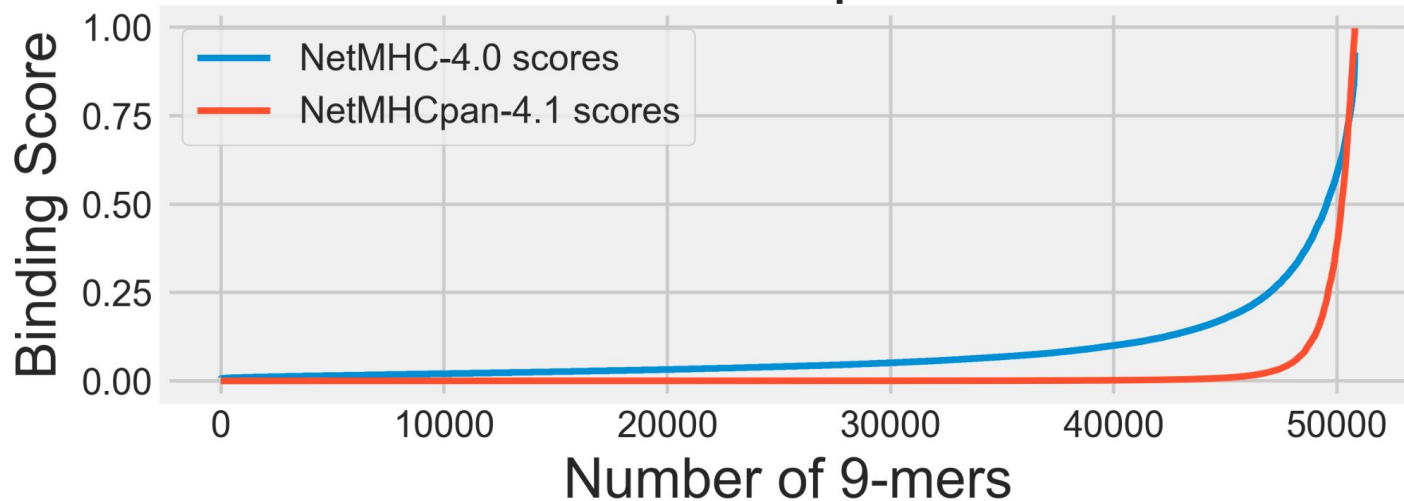


Human Proteome Data flowchart



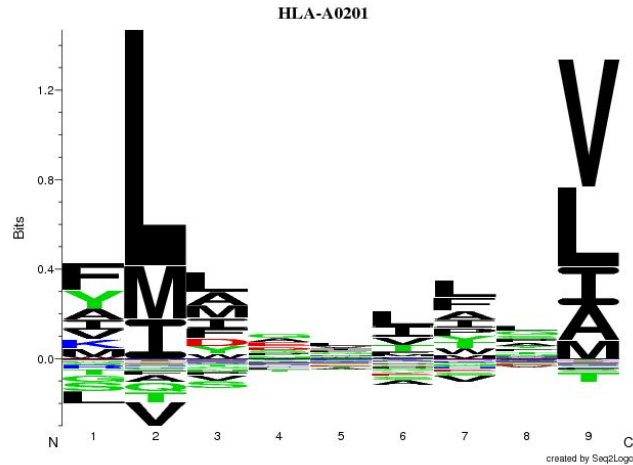
Human Proteome Scores

Sorted Scores for sampled human 9-mers



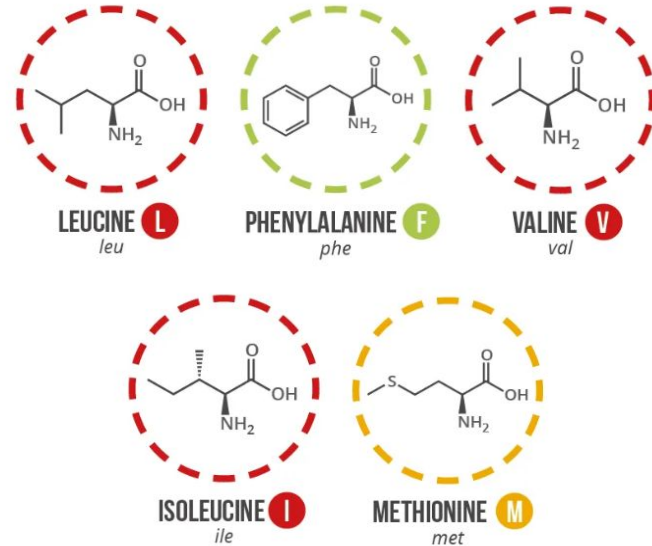
Delving into Hydrophobicity

- Do either neural network tool utilize biochemical attributes in their predictions?
- The one we are most interested in is hydrophobicity. Why?



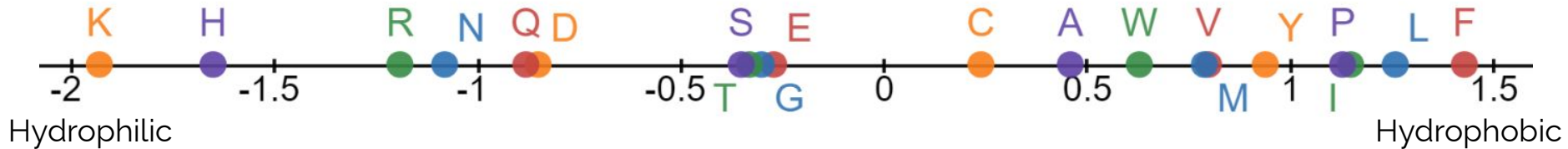
A2 binding motif

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



Moon Hydrophobicity Scale

- Models the polarity of the sidechain for an amino acid
- Hydrophobicity of peptide = sum of scores for each amino acid in it



“VIVWMLMLL”

$$= 0.8 + 1.15 + 0.8 + 0.63 + 0.79 + 1.26 + 0.79 + 1.26 + 1.26$$

$$= 8.74$$

Analysis of hydrophobicity

- For each set of peptides, calculate each peptide's hydrophobicity
- Statistical analysis across sets -
 - Graphical
 - Mean + STD
 - 2-sample T test

TRN: Training data

NTF & PTF: Strong binders in training data according to NetMHC and NetMHCpan

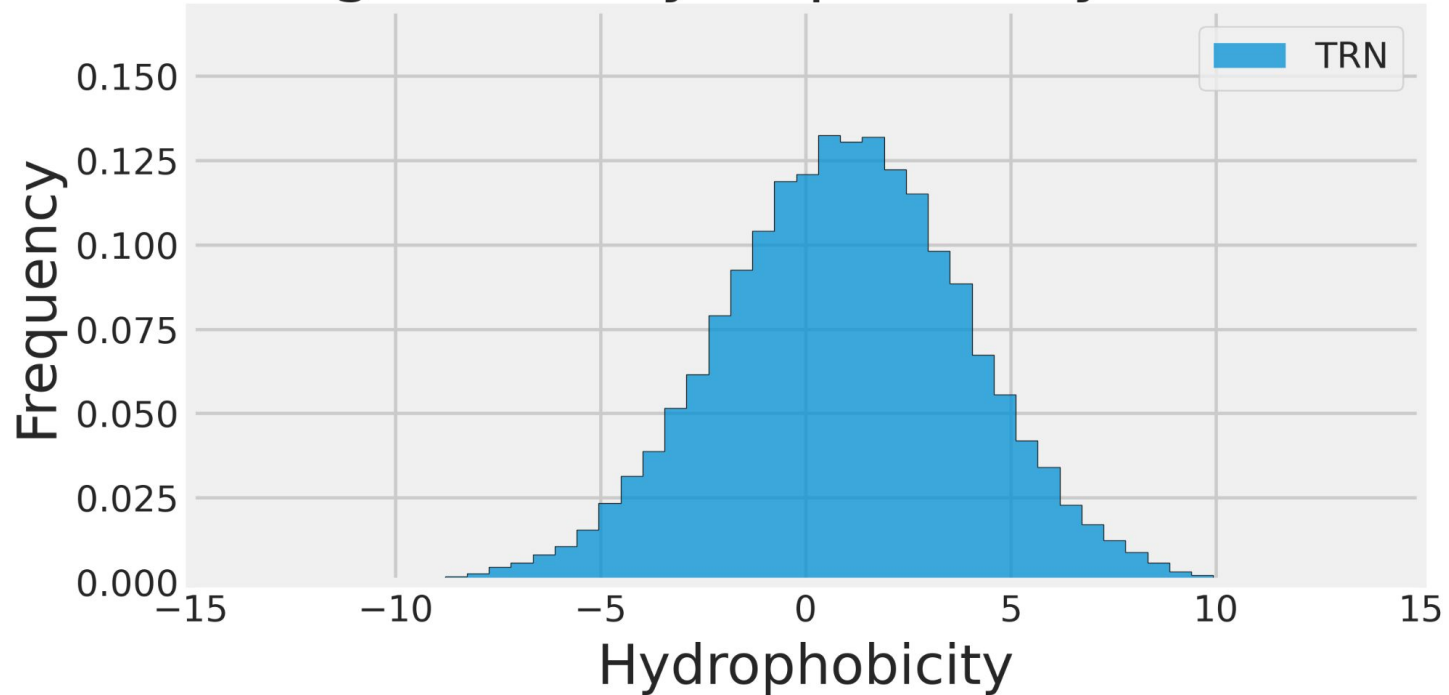
NSB & PSB: Predicted strong binders according to NetMHC and NetMHCpan

SHP: Sample Human Proteome data

NHB & PHB: Predicted strong binders according to NetMHC and NetMHCpan

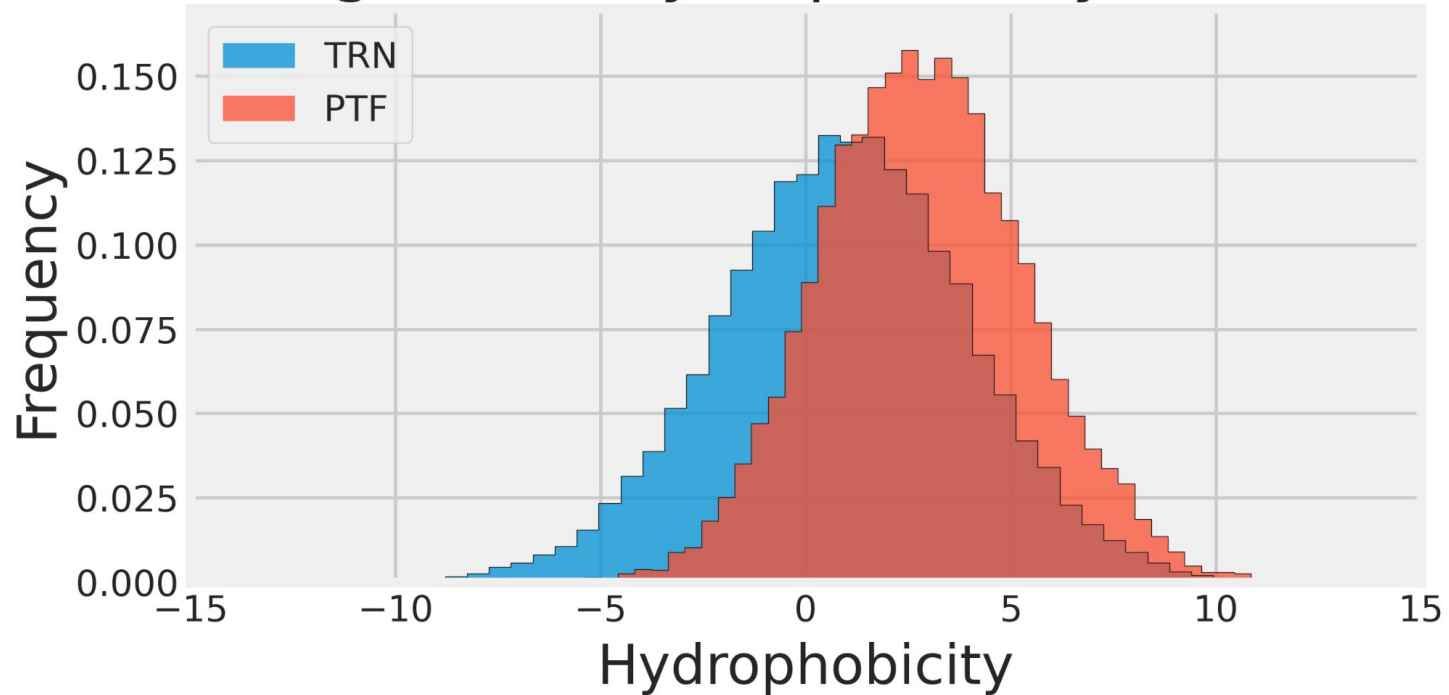
Results - NetMHCpan on training data

Histogram of Hydrophobicity of 9-mers



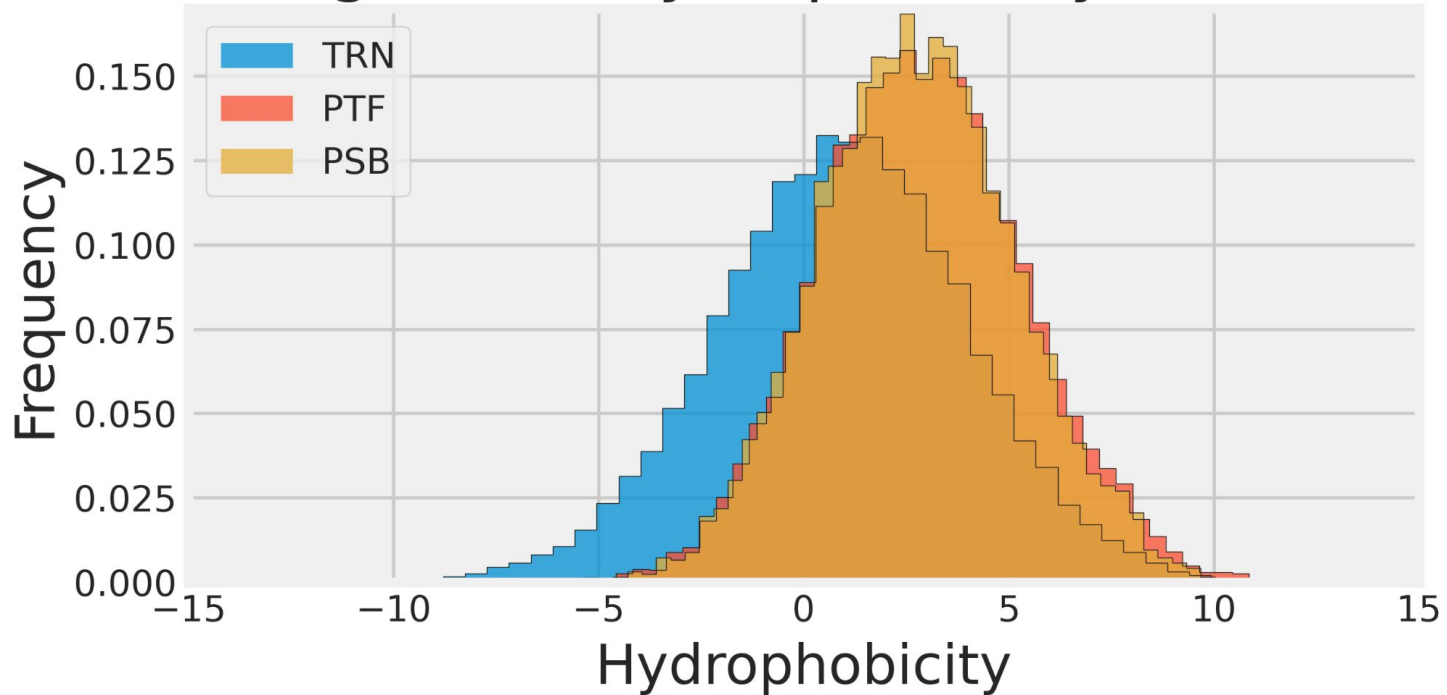
Results - NetMHCpan on training data

Histogram of Hydrophobicity of 9-mers



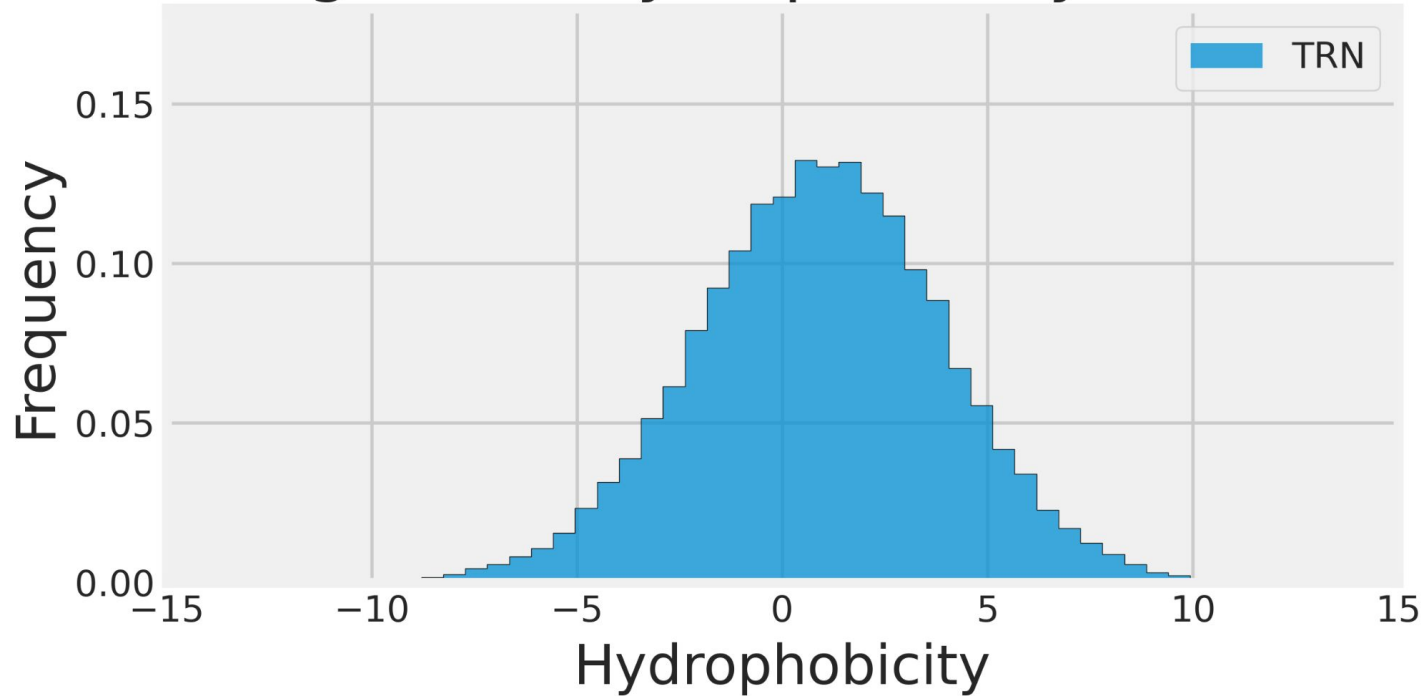
Results - NetMHCpan on training data

Histogram of Hydrophobicity of 9-mers



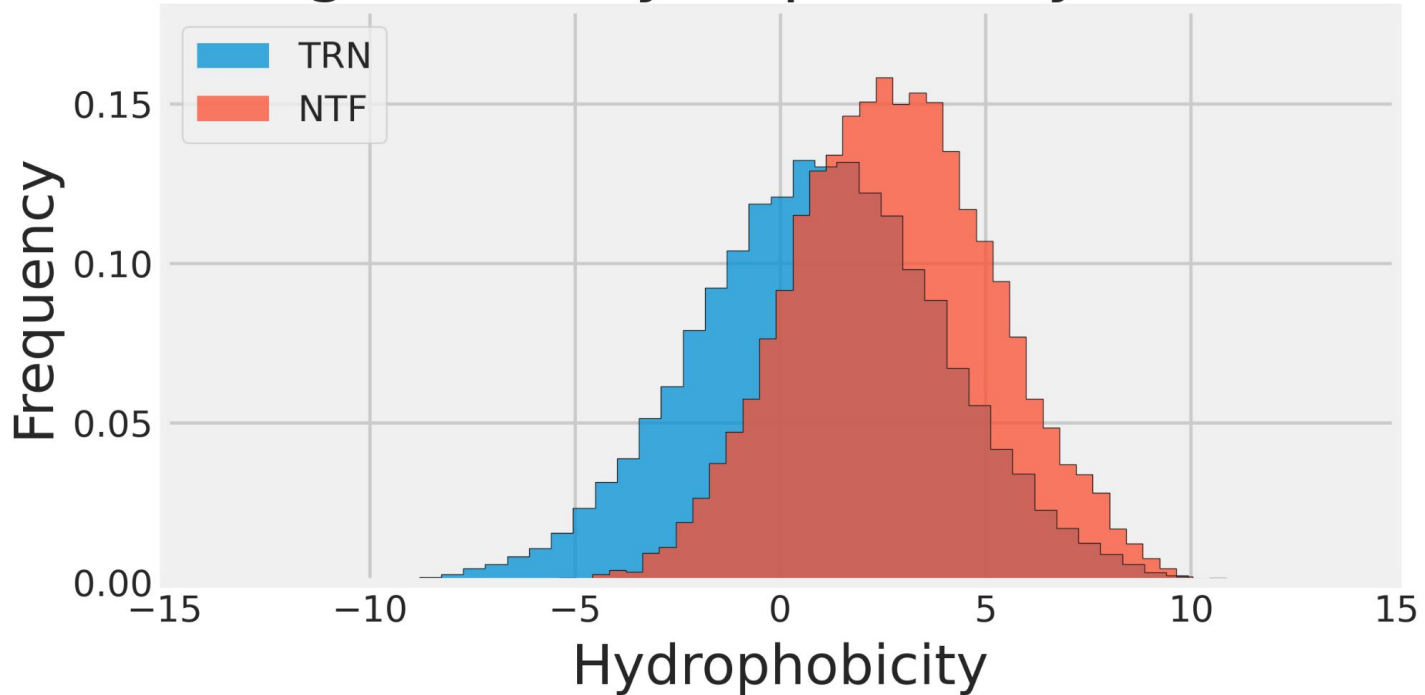
Results - NetMHC on training data

Histogram of Hydrophobicity of 9-mers



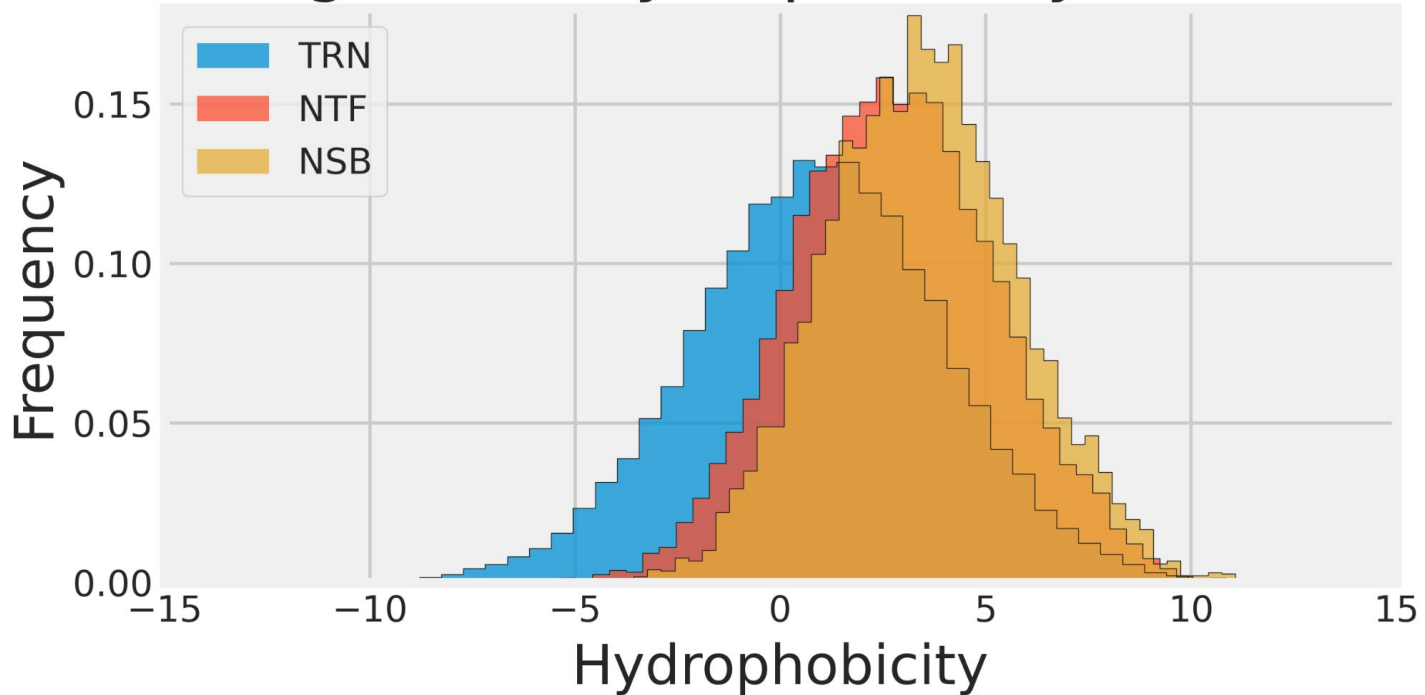
Results - NetMHC on training data

Histogram of Hydrophobicity of 9-mers

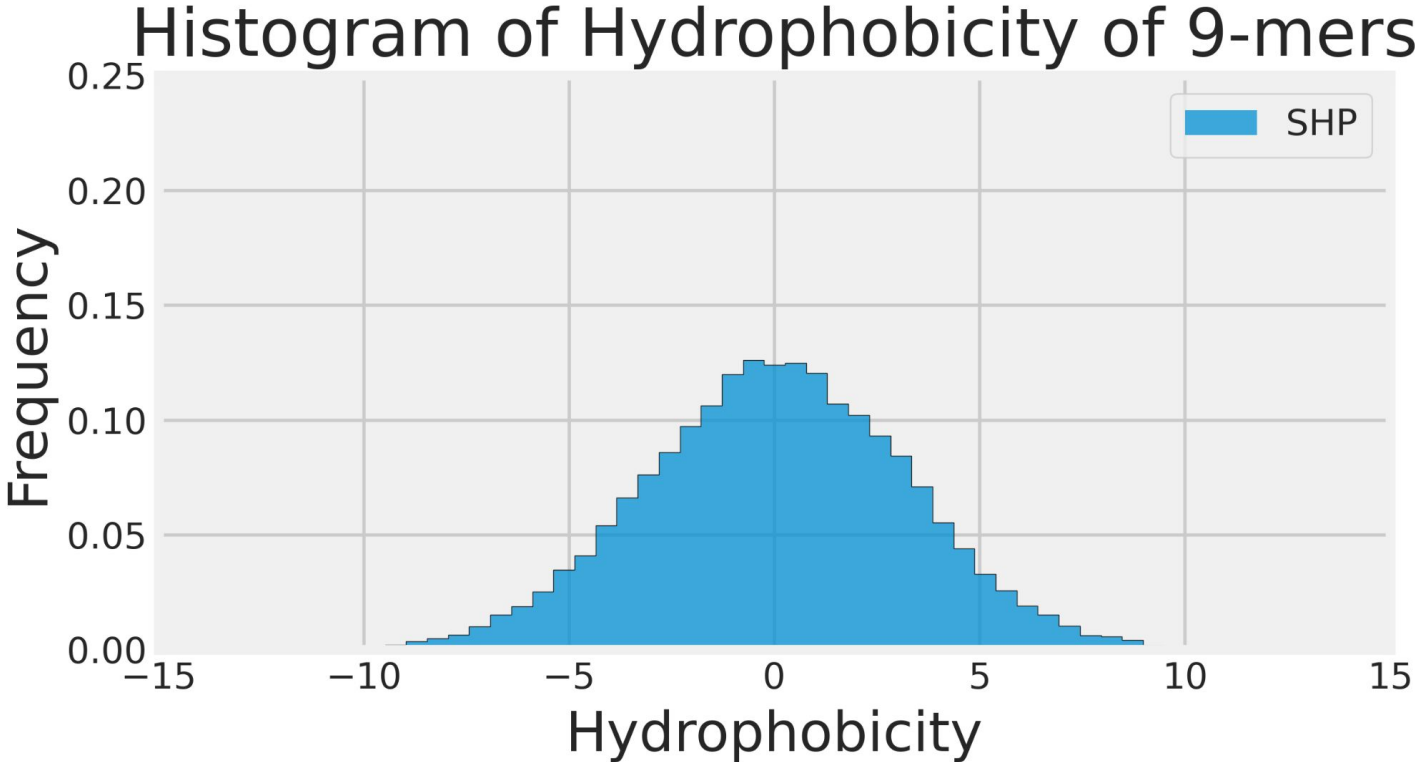


Results - NetMHC on training data

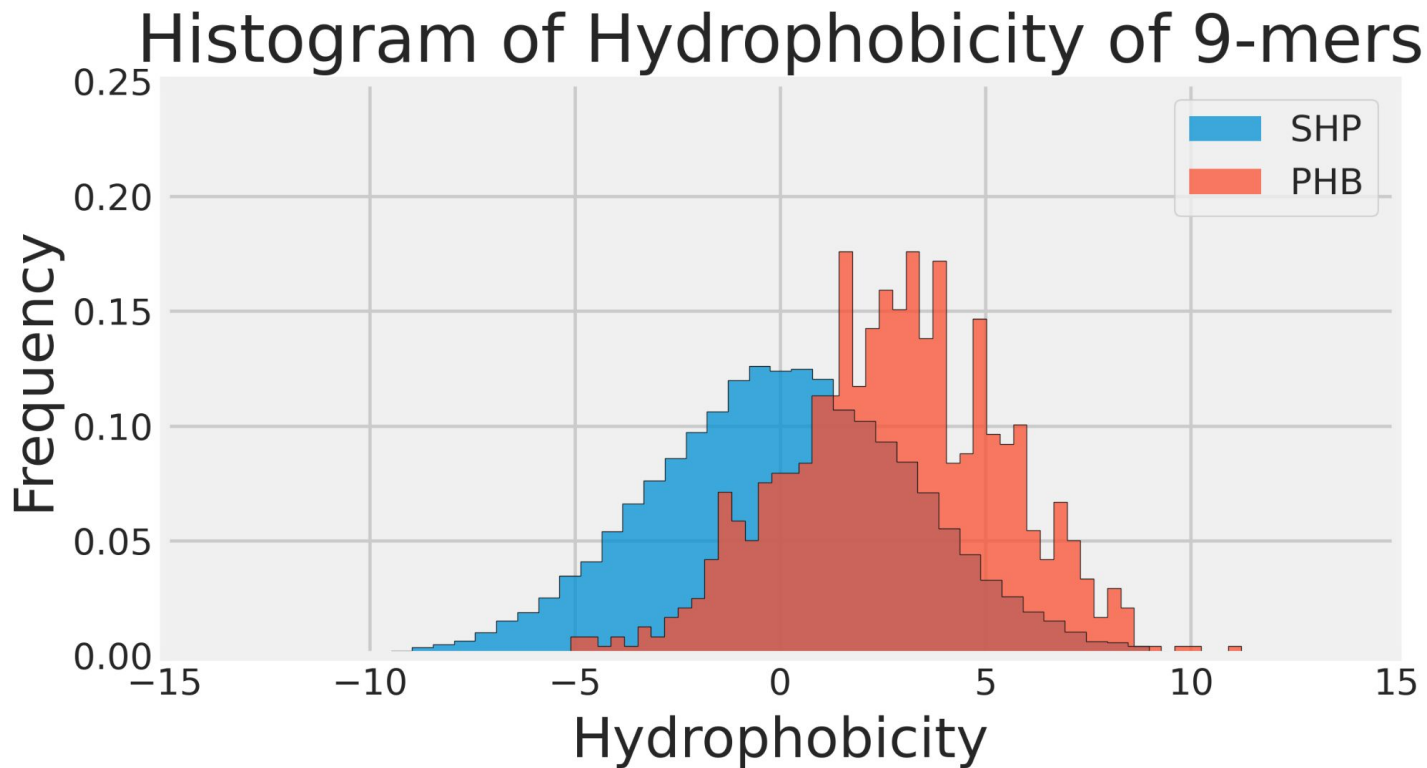
Histogram of Hydrophobicity of 9-mers



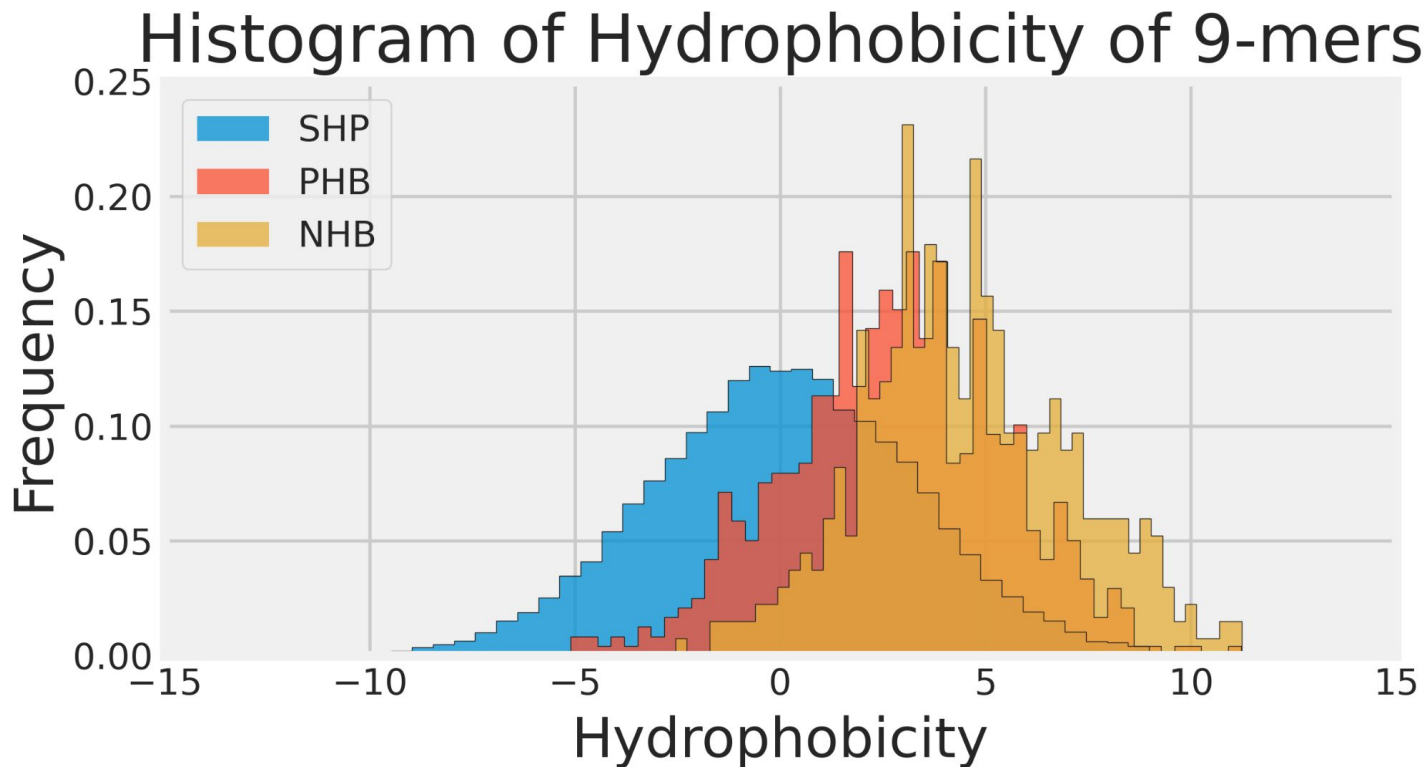
Results - Human Proteome



Results - Human Proteome



Results - Human Proteome



Conclusion

- NetMHC predicts strong binders to A2 to be more hydrophobic
 - Average of 3.45 vs 2.75 for NSB vs PSB
 - Average of 4.51 vs 2.78 for NHB vs PHB
- This bias is not reflected in
 - Training data
 - Prediction by NetMHCpan on training data
 - Prediction by NetMHCpan on the sample human proteome
 - Confirmed by statistical analysis for p value threshold of 0.0001
- The training data for NetMHCpan is not neutral in hydrophobicity - shift by 1

Conclusion

- Neural Network tools are useful tools for pMHC prediction, but vulnerable to false positives
- Neural Network tools do not utilize biochemical attributes such as hydrophobicity in their calculation
- Improvements -
 - Hydrophobicity as a data feature
 - Better training data
 - Post processing filter