Prediction of TCR-pMHC interactions using molecular modeling and recurrent networks

Ella Hedeboe s211253, Henrietta Holze s215945, Christian Johansen s202770, Paul Simon s202592

https://github.com/HenriettaHolze/TCR-pMHC-prediction

Hex code
color
#990000
#2F3EEA

Introduction

- Recent advances within biological sequencing and deep learning methods have made it possible to investigate key interactions of the immune system computationally.

- The adaptive immune system is a key element for fighting diseases and the T-cells are responsible for cell-mediated immune response via their surface T-cell receptors (TCR).

- TCRs bind to peptide-Major Histocompatibility Complexes (pMHC) to form a complex that triggers an immune response.

Problem: Predict TCR-pMHC binding using molecular modeling and recurrent neural networks (RNN).
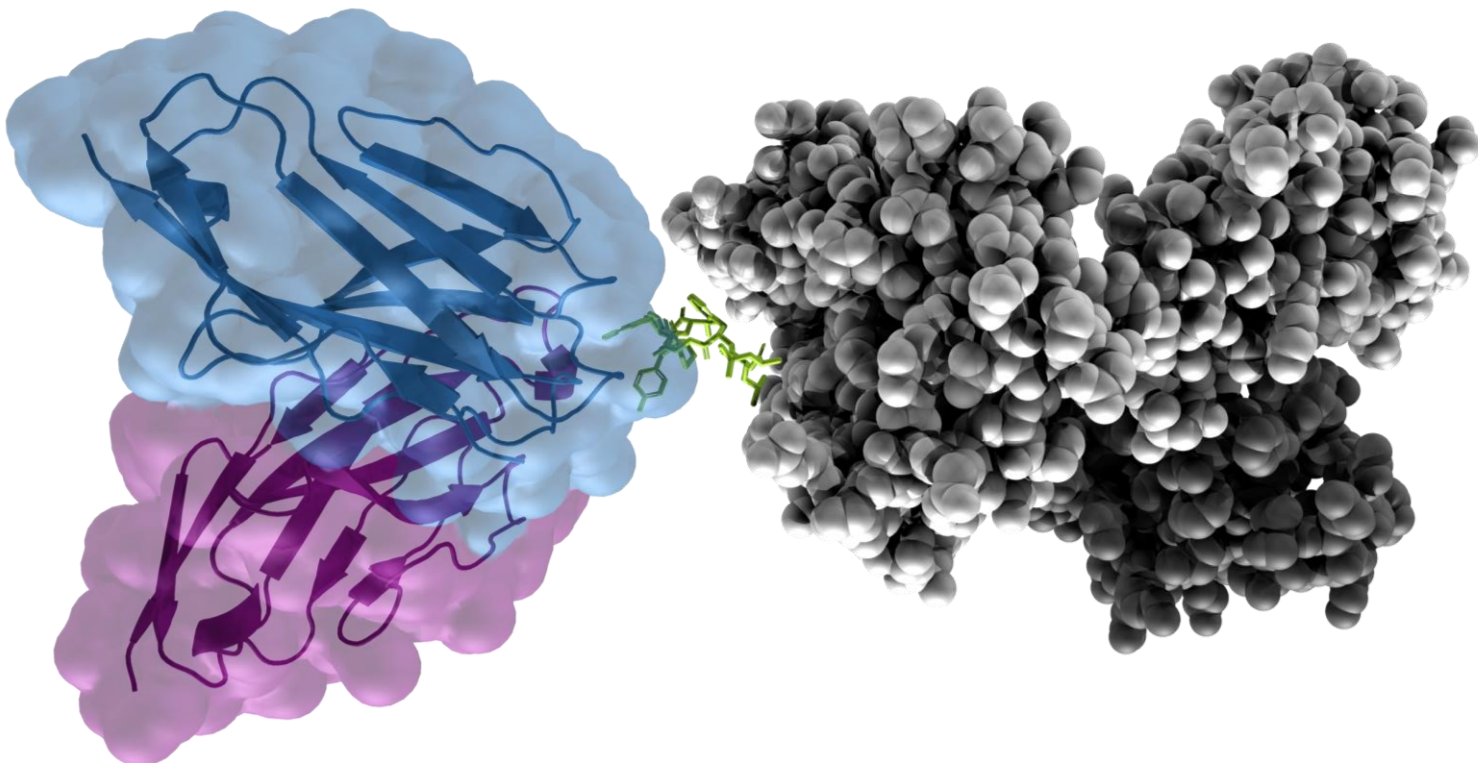

Figure 1. 3D-visualization of the TCR-pMHC complex. Blue: TCR-α, Purple: TCR-β, Green: Peptide, Grey: MHC.

Input data [1]

- Protein sequence (one-hot-encoding)
- Per-residue energy terms (one value per row)
- Global energy terms (constant, one value per column)

Input dimensions

- 6913 observations (4180 training, 1526 validation, 1207 test)
- 419 peptide positions (zero-padding where sequences are shorter)
- 54 channels (including sequence embedding and energy terms)

Methods

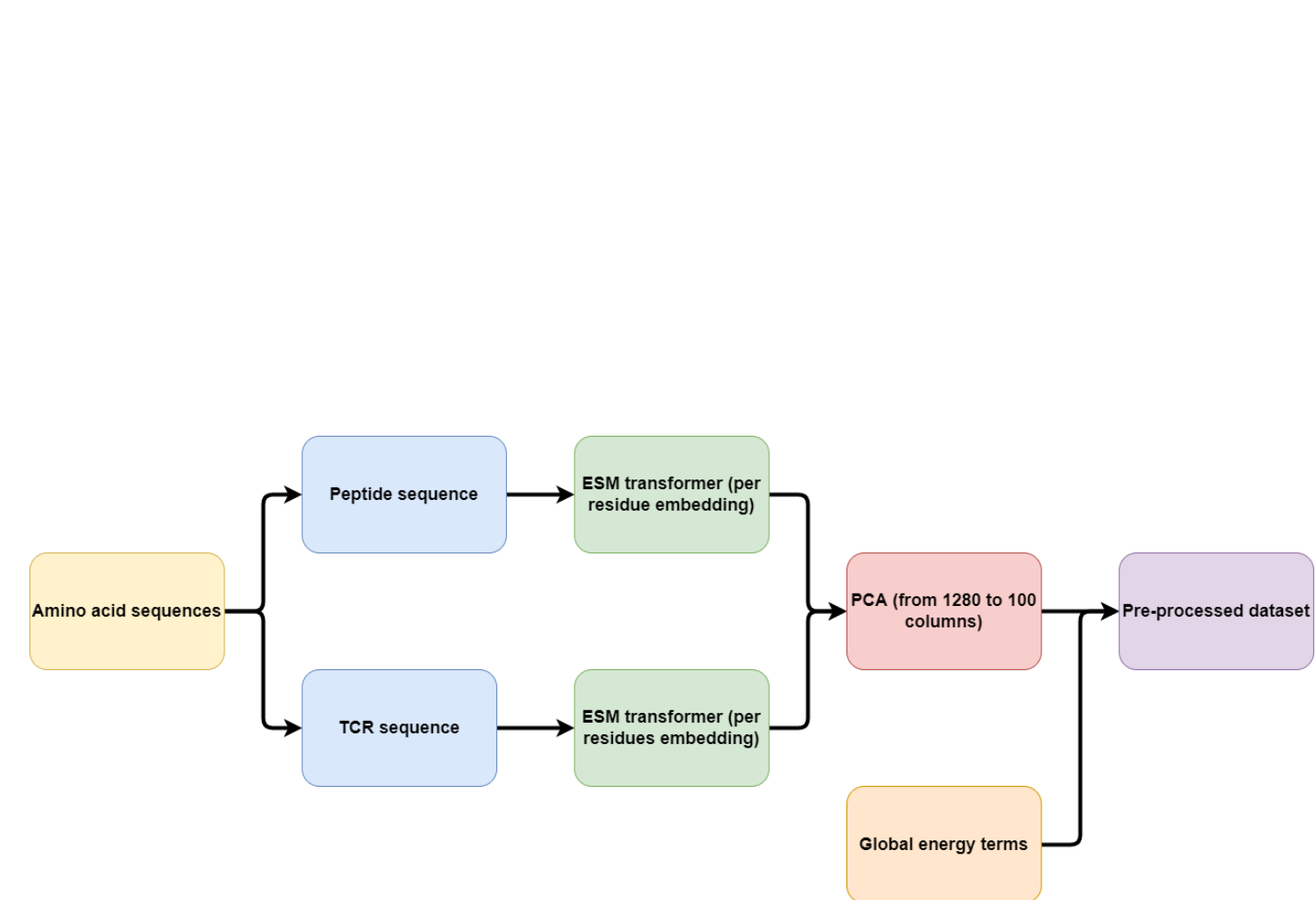### Pre-processing with protein embedding

- **BLOSUM**
- **ESM**

Neural network architecture

Figure 2. Pipeline for the data.
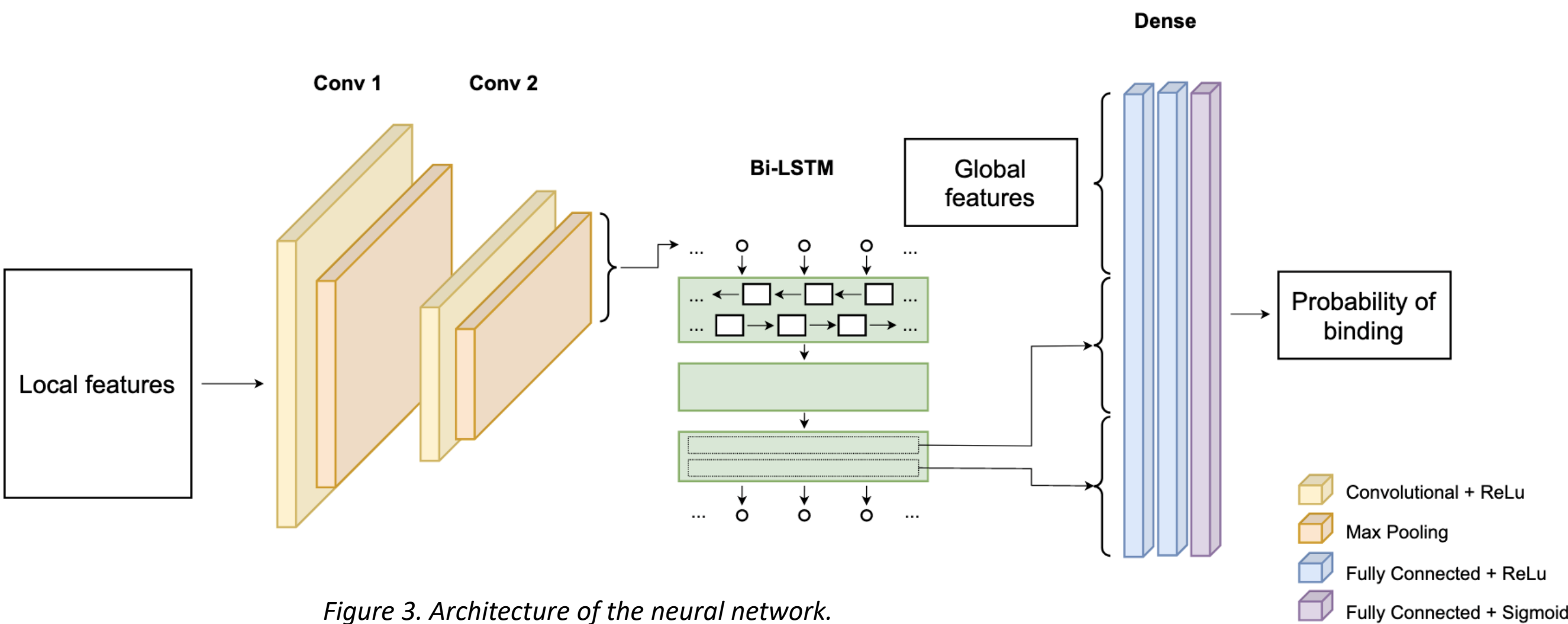

Figure 3. Architecture of the neural network.

Results

Comparison with Ida [2]

MCC on test
Confusion matrix
AUC plot

Network architecture	AUC	MCC	Precision	Recall	F1
Vanilla architecture	0.82	0.473	0.572	0.658	0.612
Improved architecture	0.86	0.452	0.468	0.833	0.600
Improved architecture with BLOSUM encoding	0.88	0.568	0.632	0.741	0.682
Improved architecture with ESM encoding	0.9	0.612	0.676	0.753	0.712

Table 1. Comparison between the different networks that we constructed in this project

Discussion

High number of negative
BERT
Graph based approach

References

[1] Magnus H. Høie. (2021). T-cell binding prediction challenge (TCR-pMHC). Github repository, Link: https://github.com/CBH2021/tcr-pmhc
[2] Ida Kristine Sandford Meitil. (2021). Using deep learning for improving TCR homology modeling and its application to immunogenicity prediction [Master's Thesis, DTU]
[3] Rives, A. et al. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences, 118(15).*