

Task: Predicting TCR binding for various antigen.

T-cell receptor (TCR) is a protein complex that can specifically bind to the peptide-MHC complex and initiate the downstream immunological response. The alpha-beta TCR is composed of two chains, an alpha chain and a beta chain. The complementarity-determining regions 3 (CDR3) of beta chain is essential for recognizing peptide-MHC complex. In this homework, you are asked to build a model to predict the binding of TCR to a specific peptide-MHC complex using the amino acid sequence of CDR3 of beta chain. We provide positive and negative samples (i.e., binding and unbinding sequences) of two peptide-MHC complex (i.e., GIL and NLV), in files with prefix ("NLV (or GIL)_positive (or negative).txt"). Please build a machine learning model for each of them. Please describe your model and report the AUROC scores of the validation process. You can use five-fold cross validation to evaluate your model performance. Note that unlike other tasks, the sequence length of CDR3 is different with each other, you can pad the sequence to the same length before feeding into the neural network.

References:

Data source:

IEDB website: <https://www.iedb.org>

A neural network model:

<https://www.biorxiv.org/content/10.1101/433706v1>