Computational Immunomics
SoSe 2017
Tutor: Sebastian Winkler

1	$\sum$

Lisa Marie Dussling
Kervin Lindner
Stephan Biastoch

Project Proposal

(Abgabe am 13.06.2017)

Data processing

Due to the inbalanced of the given dataset (majority of non-binders) the stratification of the training data set is required.

Feature Engineering

We want to utilize 6-char encoding to represent the biochemical features of each amino acid. Those are hydrophobicity, volume, charge, aromaticity and hydrogen bonds mapped on a scale form 0 to 10. Further features to consider include aliphaticity, polarity and so. We also want to experiment with selecting 2- to 4-grams for the analysis.

Generation of the predictor model

One approach would be to train a binary classifier to predict the class of the input peptides directly. We also want to include the prediction of the IC50 value and classify the peptides according to the selected threshold (of about 500 nM).
We want to use Python and the TensorFlow library to develop a neuronal network to predict the binding of peptides to a specific MHC-I allele. We will use a Grid-Search approach to optimize hyperparameters.

Predictor evaluation

We will adjust for interchangeable amino acids according to their biochemical features to prevent over fitting. To evaluate the performance of our prediction model we will use n-fold cross validation. The discussion of the optimized performance of our model will be based on the AUC of the ROC.