

Prosit extensions - BA, MA, Internship

Title: Extensions of Prosit

Type: [BA, MA, internship]

Category: [ML/DL | DS]

Programming language: [python]

Language: [German | English]

Prior experience: [experience with Python is required, experience with keras/tensorflow is a plus, no biological background necessary]

Complexity/Risk: [medium | high]

Contact person: Wassim Gabriel and Mathias Wilhelm

Brief background description: Prosit is our deep learning architecture for peptide property prediction. For learning, we mostly rely on data acquired in the context of the ProteomeTools project or data stored in ProteomicsDB. The base model is trained to predict intensities of specific fragments in the fragmentation spectrum as well as the retention time. For this, the architecture is inspired by machine translation and utilized an intermediate latent space for abstraction.

Brief description of the project: The goal of this project is to extend the capabilities of Prosit by either extending the number of supported properties Prosit is able to predict, circumvent current limitations of Prosit (e.g. peptide length between 7-30), or increase the accuracy of predictions. Multiple projects are possible in this context ranging from smaller tests (Bachelor thesis and internships) to a full Master thesis. A very interesting avenue is the prediction of spectra ab initio, where training does not rely on the annotation of spectra anymore.

Expected result: Depending on the project, the results could potentially be published in a small application note in a peer-reviewed journal.