Flyability/Ionization prediction

Title: Prediction of peptide flyability

Type: BA, MA

Category: ML/DL

Programming language: python

Language: [German | English]

Prior experience: [Math | Comp]

Complexity/Risk: [medium | high]

Contact person: Ludwig Lautenbacher

Brief background description:
Although one would expect that the abundance of a peptide is the deciding factor in determining the observed intensity in a mass spectrometry-based experiment, it is often the case that these two aspects correlate poorly. This is due to many confounding factors that distort the observed intensity differently and influence peptides differently. This hampers the use of mass spectrometry to quantify differences across peptides and impairs our ability to compare protein expression values between each other. This is particularly aggravating for targeted mass spectrometry, where only a preselected subset of peptides is monitored. To alleviate the problem of selecting peptides that do not represent their protein's intensity well, different approaches have been devised to predict their "flyability" (response factor).

Brief description of the project:
Determining the flyability of a peptide-based on experimental data restricts us to the subset of peptides that have been observed experimentally. This is seldom a problem for well-characterized organisms but hampers scientific advancements for lesser known organisms. The obvious solution to this problem is the prediction of flybility based on different peptide features and sample preparation methods using machine learning algorithms. The goal of this project is to extend Prosit to enable the prediction of flyability. For learning, we will be able to make use of the data stored in ProteomicsDB.

Expected result:

A deep learning model which is able to predict the intensity distribution across peptides originating from the same protein. This can advance our foundational knowledge about processes in mass spectrometry. The results could lead to a short article in a peer-reviewed journal.

https://link.springer.com/article/10.1186/1471-2105-8-S7-S23
https://www.nature.com/articles/nbt1275
https://academic.oup.com/bioinformatics/article/24/13/1503/237777