# Algorithmic aspects
of *de novo* protein sequencing

Kira Vyatkina

SPbAU RAS & SPbSU & SPbETU "LETI"

Proteins are macromolecules made up of one or more chains of amino acid residues. They are found in any cell of an organism and involved in many processes essential for life.

To study functionality of a protein we usually need to derive its primary structure (i. e. the amino acid sequence). This is typically achieved by means of mass spectrometry, which essentially measures the masses of the entire molecules and their fragments. Depending on the strategy applied, the results of an experiment may comprise just a single mass spectrum or thousands of those. Further, the protein sequence can be deduced from the obtained spectra either through identification via database search or *de novo* sequencing. The former approach assumes presence of a database containing, in particular, the protein being analyzed. But when examining novel proteins, including antibodies, the latter approach remains the only way to proceed.
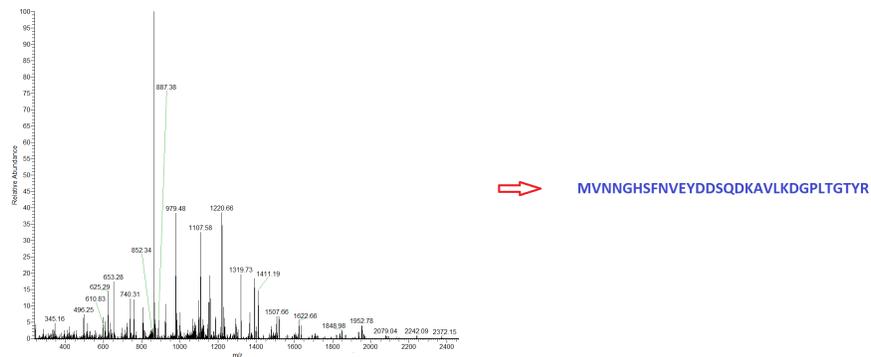
For *de novo* sequencing, it is especially important to acquire high-quality mass spectra to reduce chances of errors in the putative sequences. However, while modern mass spectrometers are capable of generating large amounts of highly accurate data, the respective spectra are by far too involved for manual interpretation. At the same time, there is still a substantial lack of efficient algorithms and software tools for processing this kind of data.

In this talk, we will discuss the challenges of protein *de novo* sequencing by high-resolution mass spectrometry, along with the solutions proposed by the Twister approach [1–4], its extensions and applications, and indicate directions for future research.

References

[1] K. Vyatkina, S. Wu, L. J. M. Dekker et al. De novo sequencing of peptides from top-down tandem mass spectra. *Journal of Proteome Research*, 14(11):4450-4462, 2015.

[2] K. Vyatkina, S. Wu, L. J. M. Dekker et al. Top-down analysis of protein samples by de novo sequencing techniques. *Bioinformatics*, 32(18):2753–2759, 2016.

[3] K. Vyatkina. De novo sequencing of top-down tandem mass spectra: A next step towards retrieving a complete protein sequence. *Proteomes*, 5(1):6, 2017.

[4] K. Vyatkina, L. J. M. Dekker, S. Wu et al. De novo sequencing of peptides from high-resolution bottom-up tandem mass spectra using top-down intended methods. *Proteomics*, 17:23-24, 2017.