

Course description

Since sequencing projects yielded full composition of model organisms' genomes the focus of current research shifted towards the understanding how the protein gene products work. For many protein-based approaches mass spectrometry (MS) is the identification method of choice. It can not only identify proteins, but also may reveal post translational modifications, comparatively quantify whole proteomes, give an insight into protein complexes or examine intact proteins.

Each successful MS experiment is composed of 3 stages: sample preparation, sample measurement and data analysis. Due to the high cost of instrumentation and uniqueness of skills necessary to maintain spectrometers, often MS laboratories function as facilities, to which the measurement step is outsourced. Thus, in most cases, researchers face a challenge of making protein preps acceptable for the spectrometry and perhaps even more difficult task of analyzing the MS data on their own, without help of people directly involved in the generating of those results. Sometimes, shortage of the adequate knowledge and lack of proper communication on the line between a researcher and a MS facility may seriously slow down experiments or even lead to the misinterpretation of the data. Our own experience shows that the data analysis step is on one hand the most time consuming moment in each MS-based experiment and on the other hand absolutely crucial for drawing correct conclusions and planning future experiments. Taking into consideration the fact that currently spectrometers are becoming even more sensitive and capable of analyzing more complex mixtures of proteins gives a grim view of a very error-prone environment, in which a researcher studying his/her protein(s) of choice may feel lost.

Curriculum

All lectures will be kept to the minimum and will cover basic information on mass spectrometry such as fundamentals of the technique, comparison of various instruments, introduction to data analysis software, principals of qualitative and quantitative experiments and statistical methods.

The main focus will be put on practical aspects and comprises of the experimental part, where participants will learn how to efficiently prepare samples and carry out quantitative experiment, and bioinformatics part, where participants will analyze MS spectra, interpret results and perform statistical analysis to extract high quality data. In order to get more out of this "hands-on"-based approach, participants will go through all the experimental steps themselves, including the step of generating the MS data, which they usually do not participate in.

Participants **will gain practical skills** to prepare high quality samples, to work with the MS data and to be capable of critical evaluation of mass spectrometry results.

The proposed curriculum covers the following issues:

Experimental sessions:

Samples preparation methods. Sample preparation is a crucial step in every proteomic experiment, especially in gel-free approaches dealing with complex protein mixtures. It is very important that samples received at proteomics facilities meet the strict requirements for the further analysis. Therefore, we will focus on techniques of sample preparation, which are compatible with the downstream MS-based analyses and will increase reproducibility of single experiments. We will introduce and discuss guidelines describing how to avoid or reduce presence of common contaminants (both biological and chemical) in protein samples. Special attention will be also paid to methods of protein precipitation and solubilisation as well as practical approach to safe ways of sample handling and transportation. Subsequently nonroutine procedures will be shortly introduced for samples that are going to be investigated in respect to post translational modifications such as identification of glycosylations sites, phosphomapping etc.

Quantitative analysis. Mass spectrometry-based methods of relative and absolute quantification of proteins are increasingly popular because in many cases changes in protein profiles are subtle and often below sensitivity of "classical" methods. Proteomic approaches can reveal and quantify those differences. Nowadays two methods are used most often: 1). label free and 2). using labelling reagents. During the course both approaches will be introduced. Participants will design their own experiment, which will differentiate and quantitatively describe two kinds of samples (proteins isolated from wild type and a mutated organism).

Label free method. Although more labour – consuming especially in terms of data analysis, it allows for basic understanding of how any kind of quantitation is being achieved. Method with labelling reagent – iTRAQ (isobaric tag for relative and absolute quantification). It is the technique for identification and quantification of proteins from different sources in a single experiment, which can be easily handled at participants laboratories.

By using those two methods we want to show an alternative attitudes to protein quantitation. Their pros and cons during the school will be presented, as well as detailed analysis of data originating from those approaches. Moreover, participants will work with the same material, so it will be a perfect comparison of the samples variability among different experiments, proving the need of statistical results validating.

Bioinformatics sessions:

Interpretation of raw MS spectra. Participants will work with unprocessed MS spectra. By doing so, they will learn: the role of naturally existing isotopes in nature, how to calculate protein / peptide mass on the basis of mass over charge value, how to predict peptides sequence from the fragmentation pattern. Additionally, the course curriculum will cover how to compute basic parameters describing the spectra and mass spectrometers – mass accuracy, resolution, and sensitivity. Such knowledge is a prerequisite in proper understanding of MS experiment results.

Data analysis. Data analysis is an absolutely crucial step needed for correct interpretation of proteomics data, which in turn assure obtaining the reliable, statistically significant results and being aware of false positives. Different datasets will be discussed in order to illustrate different levels of complexity and difficulty:

- Identification of proteins from simple (e.g., Coomassie stained gel band) and more complex mixtures – in solution digest of fractionated proteins.
- Identification of post translational modifications – manual inspection of both software search results and raw data to confirm presence of covalent moieties attached to the polypeptide backbone – phosphorylations, methylations, glycosylations etc.
- Identification of protein complexes– cross analysis of proteins identified in various experiments to propose structure of protein complexes.
- Interpretation of participants data obtained in the quantitative experiment. By applying knowledge and skills in data analysis obtained during the course, participants will point out proteins, specific for particular cell state. Special attention will be paid to element introducing false positives into experiments – different types of versatilities: technical, experimental and biological, and how to avoid misinterpretation of the data. This session will include introduction to statistics, helpful in data analysis, to achieve high power of statistical tests.

Additional sessions are intended for individual and specific concerns of participants –they will be asked to prepare a list of issues they would like to be covered, in addition to the course curriculum.