

4th EU-Summer School in Proteomic Basics

“High-Throughput Data Analysis and Statistics”

August 1-7, 2010 at Kloster Neustift, Brixen/Bressanone, South Tyrol, Italy



Verena Strassberger
Institute of Pharmaceutical Sciences, Swiss Federal Institute of Technology
(ETH), Zürich, Switzerland

Introduction

The 4th EU-Summer School in Proteomic Basics was organized by Katrin Marcus (Medizinisches Proteom-Center, Ruhr-Universität Bochum), Henning Urlaub and Carla Schmidt (both from Max Planck Institute for Biophysical Chemistry, Göttingen). It took place from August 1-7 2010 in the monastery “Kloster Neustift”, which is situated in the mountains of South Tyrol in Italy. Different lectures and workshops covering the basics of proteome data analysis and statistics for proteomics were offered to the participants, which were mainly graduate/PhD students and post-docs. Amongst the rich scientific program, also some free-time activities (like soccer, rafting, hiking) were offered to facilitate networking and scientific discussions in an informal atmosphere. The participants also had the opportunity to present and discuss their own research in two poster sessions. The best posters were awarded with a poster prize.

Scientific lectures

The plenary lecture “Proteomics: a pragmatic perspective” was given by **Bernhard Küster** (TU Munich). He was discussing the relationship between the complexity of a biological sample and the difficulty of implementing the appropriate analysis approach as no “one-size-fits-all” proteomic strategy can be used to address all biological questions.

Thierry Rabilloud (CEA-Grenoble) and **Kai Stühler** (MPC RU Bochum) focused on 2D gel-based proteomics. T. Rabilloud discussed the advantages and disadvantages of 2D gels. He emphasized that 2D gel-based proteomics is the only proteomic workflow in which total proteins are the analyte and are separated up to the readout step not as digested peptides but as complete proteins. This allows for example the study of post-translational modifications and especially the combination of post-translational modifications on individual proteins in an unsupervised way. K. Stühler presented different approaches (minimal and saturation labeling) and dyes for Difference Gel Electrophoresis (DIGE). Furthermore he reviewed their advantages and disadvantages and demonstrated the analysis of DIGE data resulting from image analysis software.

Ruedi Aebersold (ETH Zurich) presented selected reaction monitoring (SRM), a targeted mass spectrometric technique that supports the detection and quantification of specific proteins in complex samples at high sensitivity and reproducibility. Computational and informatics challenges associated with this method were discussed like protein and peptide target selection, assay development and optimization, detection and statistical validation of the correct peak groups representing the targeted peptides and error models.

Katharina Hoff (University of Greifswald) gave a lecture about the statistical analysis of experimental data. She exemplified distributions, data visualization, correlation, linear regression and statistical tests like t-test, Wilcoxon-test and ANOVA.

Oliver Kohlbacher (University of Tübingen) and **Jürgen Cox** (MPI of Biochemistry, Martinsried) were presenting software for the high throughput analysis of MS datasets. O. Kohlbacher presented OpenMS, an open-source C++ library for LC/MS data management and analyses and discussed advantages and disadvantages of open-source and vendor software. He also presented TOPP - The OpenMS Proteomics Pipeline, which is a pipeline for the analysis of HPLC/MS data with a graphical user interface so that no programming skills are required. It consists of several small applications that

can be chained to create analysis pipelines tailored for a specific problem (e.g. label-free quantification, iTRAQ, MRM scheduling). **J. Cox** introduced MaxQuant, which is a quantitative proteomics software package designed for analyzing large mass spectrometric data sets of high resolution MS data. Originally it was designed for the analysis of SILAC experiments and was recently expanded to other applications like label-free quantification.

Lennart Martens (University of Gent) introduced the Peptizer application which helps to minimize false positive protein identifications. The application allows the use of different filters and divides peptide identifications into “good”, “bad” and “suspect”, so that the experienced user has the possibility to control them manually with the manual interpretation interface (OMSSA). He also discussed the problems of protein interference and presented the quantification *post-processing* application “ROVER”.

Matthias Selbach (MDC Berlin) reviewed methods for absolute protein quantification. He briefly introduced stable isotope-based methods like AQUA peptides, iTRAQ, QconCAT and absolute SILAC. The main part of his lecture was about label free / computational methods for absolute protein quantification. He introduced the protein abundance index (PAI), the normalized spectral abundance factor (NSAF), the absolute protein expression measurements (APEX), top3 peptide intensities (top3) and intensity-based absolute quantification (iBAQ).

Juri Rappsilber (University of Edinburgh) talked mainly about his work on the protein composition of mitotic chromosomes, which was determined by using multi-classifier combinatorial proteomics. Integrating quantitative proteomics with bioinformatic analysis and subsequent machine learning uncovered a functional relationship between protein complexes. Also an *in silico* purification step was applied to eliminate contamination, which was due to nonspecific hydrophobic or electrostatic interactions.

Company presentations

The company representatives from **Decodon**, **Waters** and **Bruker** were presenting general aspects of data analysis and their latest developments. **Berth M.** (Decodon GmbH) presented “DOs and DON`Ts for Statistical Analysis of 2D Gel Data”, **M. McDowall** (Waters) talked about “HD UPLC/MS/MS with qualitative/quantitative processing of multiplexed time-aligned peptide mass spectra” and Christian Albers (**Bruker**) introduced “Comprehensive software for quantitative proteomics”.

Workshops

Six workshops were offered from which two could be chosen and attended by the participants. The workshop “**Gel-based Proteomics**” (Kai Stühler and Thierry Rabilloud) offered the possibility to ask questions and discuss method related questions. A practical introduction to “**MaxQuant**” (Matthias Selbach and Jürgen Cox) was offered covering SILAC-based and label-free quantification. In “**Manual Spectra interpretation**” (Kai Scheffler, Thermo Scientific) LTQ orbitrap and different fragmentation techniques (CID, HCD, PQD and ETD) were taught and peptide fragmentation rules were discussed, which was followed by hands-on spectra interpretation. The basics of the open source language for statistics and graphics “R” were taught and applied in “**R for beginners**” (Katharina Hoff). The

“Advanced R” (Katharina Hoff) workshop covered data set manipulation, beautifying plots, correlation analysis and linear regression. In the workshop **“Multiple Reaction Monitoring (MRM)”** (Christof Lenz, ABSciex) the participants learned how to select and optimize MRM transitions and how to set up a MRM method. They also had the opportunity to work with the software Multi quant and Analyst and get an overview of MRM data analysis.

Poster sessions

Every participant of the summer school had to present a poster during one of two poster sessions and had to give a short presentation about the content of the poster before the session. The organizers and attending speakers chose the best posters of each session, which were awarded with a poster prize. The poster prize winners were:

- **Verena Strassberger** (Swiss Federal Institute of Technology/ETH, Zürich) “Development of a new biotinylation reagent for the identification of vascular proteins in cancer”
- **Christian Koehler** (University of Oslo) “Quantitative proteomics by rapid-IPTL and IsobariQ”
- **Erik De Graaf** (Netherlands Proteomic Center/NPC, Utrecht) “A proteomics view at aging”
- **Daniel Klimmeck** (German Cancer Research Center /DKFZ, Heidelberg) “Proteome analysis of haematopoietic stem/progenitor cells”

Acknowledgements

The registration fee and travel expenses were covered by the travel grant given by the Swiss Proteomics Society (SPS). I am really grateful to the SPS for having confidence in me and accepting my grant application.