# Executive Summary of the Dagstuhl Workshop "Computational Proteomics"

**Christian Huber** (Christian.Huber@mx.uni-saarland.de)
Instrumental Analysis and Bioanalysis, Universität des Saarlandes, 66123 Saarbrücken, Germany

**Oliver Kohlbacher** (Oliver.Kohlbacher@uni-tuebingen.de)
Simulation of Biological Systems, Wilhelm Schickard Institute for Computer Science, Eberhard-Karls-Universität Tübingen, Sand 14, 72076 Tübingen

**Knut Reinert** (reinert@inf.fu-berlin.de)
Institut für Informatik, Arbeitsgruppe bioinformatische Algorithmen, Freie Universität Berlin, Takusstrasse 9, 14195 Berlin

## *Abstract*

The Dagstuhl Seminar on Computational Proteomics brought together researchers from computer science and from proteomics to discuss the state of the art and future developments at the interface between experiment and theory. This interdisciplinary exchange covered a wide range of topics, from new experimental methods resulting in more complex data we will have to expect in the future to purely theoretical studies of what level of experimental accuracy is required in order to solve certain problems. A particular focus was also on the application side, where the participants discussed more complex experimental methodologies that are enabled by more sophisticated computational techniques. Quantitative aspects of protein expression analysis as well as posttranslational modifications in the context of disease development and diagnosis were discussed. The seminar sparked a number of new ideas and collaborations and resulted in joint grant applications and publications.

## 1. Goals and Structure of the Workshop

After the finishing of the first large genomics initiatives it became more and more apparent that the knowledge of the genomes is not sufficient to achieve significant advances in molecular medicine in the near future. Indeed, the acquired genomic data is rather fundamental to the interpretation of data stemming from transcriptomics and proteomics experiments. This requires a close collaboration of researchers from different fields: biology, medical sciences, biochemistry, analytical chemistry and computer science. Mass spectrometry has evolved into one of the most important and powerful technologies for high throughput protein identification and quantification. Nevertheless, the high information content of mass spectrometry also comes along with an enormous amount of data that can by no means be interpreted manually. In due consequence, the high-throughput settings of many of these experiments need advanced algorithmic techniques and data analysis capabilities, which are only now being developed.

Dagstuhl Seminar Proceedings 05471
Computational Proteomics
http://drops.dagstuhl.de/opus/volltexte/2006/540

The goal of this seminar was to bring together experts from both fields: life science and computer science/bioinformatics. Together they should present recent advances in their field and identify relevant problems that ought to be solved jointly in order to advance science in the field.
The expected outcome or the seminar was a better understanding of the expectations, needs, and capabilities both of experimental and bioinformatic tools for proteome analysis. Computer scientists should understand how the data for proteome analysis are generated and what their implications are, while experimental scientists need to know how the data can be evaluated and validated. Moreover, there was general consensus that proteomics experiments need to be planned and carried out taking onto consideration both experimental and computational issues at the very beginning of the study. The seminar was also intended to set off new initiatives for collaborations between experimental and computational sciences, which in turn should result in novel, rugged and fully automatable total analysis systems for proteome characterization.

We thus had sessions covering the following topics (the two columns represent the morning and afternoon sessions):

| Data processing and Protein Identification | |
|---|---|
| **Organizers:**<br>Welcome and Introduction –<br>The idea of Dagstuhl Seminars<br><br>**Eva Lange:**<br>High-accuracy peak picking of proteomics data using wavelet techniques<br><br>**Wolf Lehmann:**<br>The [Peptide]$^{n+}$-Fragmentome: A Set of Predictable MS/MS Fragment Ions with Highly Redundant Sequence Information<br><br>**Marc Sturm:**<br>A machine learning approach for prediction of DNA and peptide retention times | **Christian Huber:**<br>Glycosylation Patterns of Proteins Studies by Liquid Chromatography-Mass Spectrometry and Bioinformatic Tools<br><br>**Christian Schley:**<br>Multidimensional Peptide/Protein Analysis and Identification by Sequence Database Search using Mass Spectrometric data |
| Experimental techniques | |
| **Hartmut Schlüter:**<br>From Functions to Proteins – Searching for Proteases<br><br>**Sebastian Böcker:**<br>Combinatorial Approaches for Mass Spectra Recalibration | **Andreas Tholey:**<br>Quantitative MALDI Mass Spectrometry – Approaches, Scopes and Limitations<br><br>**Johannes Schuchhardt:**<br>From spots to systems |

| | |
|---|---|
| | **Rune Mattiesen:**<br>VEMS 3.0: Algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins<br><br>Plenary Discussion I:<br>What are the important scientific questions in Proteomics? How can we convince funding agency of the importance of the research? |
| Excursion and discussion rounds | |
| Quantitative Proteomics | |
| **Ole Schulz-Trieglaff:**<br>An overview of existing Software for quantitative Proteomics<br><br>**Knut Reinert:**<br>Signal processing and data reduction for differential proteomics with HPLC<br><br>**Tim Conrad:**<br>Fingerprinting in MALDI-TOF-MS: From Ion Counts to Patterns | **Nathanael Delmotte:**<br>Evaluation of Liquid-Chromatography-Mass Spectrometry data for the absolute quantitative analysis of marker proteins in human serum.<br><br>**Clemens Gröpl:**<br>Algorithms for the automated absolute quantification of diagnostic markers in complex proteomics samples<br><br>**Jens Decker:**<br>Method Development for Clinical Proteomics<br><br>**Amol Prakash:**<br>Signal Maps for Mass Spectrometry-based Comparative Proteomics |
| | |
| **Benno Schwikowski:**<br>Computational support for the proteomic side of transcriptional networks<br><br>Plenary Discussion II | End of seminar |

## 2. Participants and Topics

Participants included researchers from computer science and bioinformatics as well as experimentalists from different European countries. The participants accordingly presented talks covering a wide range of topics – from new experimental techniques

requiring new computational approaches to theoretical results with interesting implications for experimental techniques.

**Figure 1:** The participants of Dagstuhl Seminar "Computational Proteomics".

The following scientists participated in the seminar:

- Behshad **Behzadi**, Inst. Pasteur et LIX (Paris) & MPI (Berlin)
- Sebastian **Böcker**, Universität Bielefeld (title of talk: Combinatorial Approaches for Mass Spectra Recalibration)
- Tim **Conrad**, FU Berlin (title of talk: Fingerprinting in MALDI-TOF-MS: From Ion Counts to Patterns)
- Jens **Decker**, Bruker Daltonik GmbH (title of talk: Method Development for Clinical Proteomics)
- Nathanaël **Delmotte**, Universität Saarbrücken (title of talk: Evaluation of Liquid-Chromatography-Mass Spectrometry data for the absolute quantitative analysis of marker proteins in human serum)
- Clemens **Gröpl**, FU Berlin (title of talk: Algorithms for the automated absolute quantification of diagnostic markers in complex proteomics samples)
- Christian **Huber**, Universität Saarbrücken (title of talk: Glycosylation Patterns of Proteins Studied by Liquid Chromatography-Mass Spectrometry and Bioinformatic Tools)
- Oliver **Kohlbacher**, Universität Tübingen
- Eva **Lange**, FU Berlin (title of talk: High-accuracy peak picking of proteomics data)
- Wolf D. **Lehmann**, DKFZ - Heidelberg (title of talk: The [Peptide]n+-Fragmentome: A Set of Predictable MS/MS Fragment Ions with Highly Redundant Sequence Information)
- Sébastien **Li-Thiao-Té**, Institut Pasteur - Paris

- Rune **Mattiesen**, University of Southern Denmark - Odense (title of talk: VEMS 3.0: Algorithms and Computational Tools for Tandem Mass Spectrometry Based Identification of Post-translational Modifications in Proteins)
- Nico **Pfeifer**, Universität Tübingen
- Amol **Prakash**, University of Washington (title of talk: Signal Maps for Mass Spectrometry-based Comparative Proteomics)
- Knut **Reinert** , FU Berlin (title of talk: OpenMS - A framework for differential quantification using HPLC/MS)
- Christian **Schley**, Universität Saarbrücken (title of talk: Multidimensional Peptide/Protein Analysis and Identification by Sequence Database Search Using Mass Spectrometric Data)
- Hartmut **Schlüter**, Charité - Berlin (title of talk: From Functions to Proteins – Searching for Proteases)
- Johannes **Schuchhardt**, MicroDiscovery GmbH – Berlin (title of talk: From Spots to Systems)
- Ole **Schulz-Trieglaff**, FU Berlin (title of talk: Software platforms for quantitative proteomics)
- Benno **Schwikowski** , Institut Pasteur, Paris (title of talk: Computational support for the proteomic side of transcriptional networks)
- Marc **Sturm**, Universität Tübingen (title of talk: A machine learning approach for prediction of DNA and peptide retention times)
- Andreas **Tholey**, Universität des Saarlandes (title of talk: MALDI Mass Spectrometry for Quantitative Proteomics – Approaches, Scopes and Limitations)
- Nora **Toussaint**, Universität Tübingen

The unfortunate choice of the seminar data (Thanksgiving weekend) made it impossible for the invited American scientists to attend, a problem that should be considered in the future when scheduling a seminar with American participants.

## 3. Workshop Conclusion

The workshop on Computational Proteomics was a full success, as has been confirmed by its participants. Bringing together scientist from different communities – from computer science and life sciences – turned out to be fruitful indeed. Traditionally, proteomics and bioinformatics/computer science are mostly disjoint communities with separate meetings and conferences. The chance to get insights into the problems and challenges both of the experimental and computational world, the need to learn and understand the idiosyncratic "languages" and "vocabulary" of the different disciplines was well appreciated by the attendants. Validation of proteomics data generation and evaluation was spotted as one of the most challenging issues in the application of proteomics as a technology for clinical diagnosis and monitoring. Participants from the two communities were exposed to new ideas, concepts, and techniques – both experimentally and computationally – they were not previously aware of. These ideas were then discussed over a glass of wine or two until late at night. The workshop produced a number of personal contacts which was positively remarked by the participants. In addition to the interaction and personal contacts of

the attendants, the quiet atmosphere of the location also allowed ample time for developing new ideas for solving proteomic challenges.
In conclusion, the Workshop was very successful. It sparked interesting discussions, research collaborations, several joint grant proposals (e.g. for the BMBF program QuantPro), and in one case a joint publication (below, in boldface scientists invited for the seminar). Other publications sparked by the seminar will certainly follow in the near future. The seminar also initiated the implementation of a webpage for interchanging proteomics data (www.computationalproteomics.net).
The success of the seminar and the feedback of the participants encouraged us to follow up with this style of meeting. At the end of August 2006 Knut Reinert will host a "Computational Proteomics" workshop in Berlin and we plan to apply for another Dagstuhl workshop in 2007.


- Bettina Mayr, **Oliver Kohlbacher, Knut Reinert, Marc Sturm, Clemens Gröpl, Eva Lange, Christoph Klein, Christian Huber.** Absolute Myoglobin Quantitation in Serum by Combining Two-Dimensional Liquid Chromatography-Electrospray Ionization Mass Spectrometry and Novel Data Analysis Algorithms. J. of Proteome Res., volume 5, pages 414-421, 2006.