05471 Abstracts Collection
Computational Proteomics
— Dagstuhl Seminar —

Christian G. Huber[1], Oliver Kohlbacher[2] and Knut Reinert[3]

[1] Univ. des Saarlandes, DE
christian.huber@mx.uni-saarland.de
[2] Univ. Tübingen, DE
kohlbach@informatik.uni-tuebingen.de
[3] FU Berlin, DE
reinert@inf.fu-berlin.de

Abstract. From 20.11.05 to 25.11.05, the Dagstuhl Seminar 05471 "Computational Proteomics" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Proteomics, mass spectrometry, MALDI, HPLC-MS, differential expression, clinical proteomics, quantitation, identification

05471 Executive Summary – Computational Proteomics

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.

Keywords: Proteomics, mass spectrometry, MALDI, HPLC-MS, differential expression, clinical proteomics, quantitation, identification

Dagstuhl Seminar Proceedings 05471
Computational Proteomics
http://drops.dagstuhl.de/opus/volltexte/2006/556

Joint work of: Huber, Christian G.; Kohlbacher, Oliver; Reinert, Knut

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/540

Combinatorial Approaches for Mass Spectra Recalibration

Sebastian Böcker (Universität Bielefeld, D)

Mass spectrometry has become one of the most popular analysis techniques in Proteomics and Systems Biology. With the creation of larger datasets, the automated recalibration of mass spectra becomes important to ensure that every peak in the sample spectrum is correctly assigned to some peptide and protein. Algorithms for recalibrating mass spectra have to be robust with respect to wrongly assigned peaks, as well as efficient due to the amount of mass spectrometry data. The recalibration of mass spectra leads us to the problem of finding an optimal matching between mass spectra under measurement errors.

We have developed two deterministic methods that allow robust computation of such a matching: The first approach uses a computational geometry interpretation of the problem, and tries to find two parallel lines with constant distance that stab a maximal number of points in the plane. The second approach is based on finding a maximal common approximate subsequence, and improves existing algorithms by one order of magnitude exploiting the sequential nature of the matching problem. We compare our results to a computational geometry algorithm using a topological line-sweep.

Keywords: Mass spectrometry recalibration computational geometry

Joint work of: Böcker, Sebastian; Mäkinen, Veli

Fingerprinting in MALDI-TOF-MS: From Ion Counts to Patterns

Tim Conrad (FU Berlin, D)

The use of matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) offers exciting new approaches for proteomic analyses, such as identification of biomarkers for detection of diseases, or clinical monitoring of therapeutic and toxic outcomes. However, the acquired spectra contain a lot of noise resulting from measurement inaccuracies, instability of proteins, contaminations and so forth. Hence, during signal preprocessing and the subsequent data-mining and analysis sophisticated statistical tools are needed to allow for reliable results.

In this talk we demonstrate our newly developed web-based and database driven software platform providing tools for the analysis of MS Spectra. Starting from the raw data we present the individual steps and approaches we have

taken, namely data preprocessing, peak detection and pattern finding, each of which is combined with a quality and significance analysis. This eventually leads to a statistically supported "fingerprint" of a certain group of spectra. These fingerprints can be used e.g. for the classification of unknown spectra or analyzed further to guide subsequent experiments such as MS/MS with the ultimate goal of finding biologically relevant marker molecules.

Keywords: MALDI, Fingerprinting, MS

New statistical algorithms for clinical proteomic - Extended Abstract

Tim Conrad (FU Berlin, D)

Background: Mass spectrometry based screening methods have been recently introduced into clinical proteomics. This boosts the development of a new approach for early disease detection: proteomic pattern analysis.

Aim: Find, analyze and compare proteomic patterns in groups of patients having different properties such as disease status or epidemio-logical parameters (e.g. sex, age) with a new pipeline to enhance sensitivity and specificity.

Problems: Mass data acquired from high-throughput platforms frequently are blurred and noisy. This extremely complicates the reliable identification of peaks in general and very small peaks below noise-level in particular.

Approach: Apply sophisticated signal preprocessing steps followed by statistical analyzes to purge the raw data and enable the detection of real signals while maintaining information for tracebacks.

Results: A new analysis pipeline has been developed capable of finding and analyzing peak patterns discriminating different groups of patients (e.g. male/female, cancer/healthy). First steps towards distributed computing approaches have been incorporated in the design.

Keywords: MS, Mass Spectrometry, MALDI-TOF, Fingerprinting, Proteomics

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/542

Method Development for Clinical Proteomics

Jens Decker (Bruker Daltonik GmbH, D)

The application especially of MALDI TOF spectroscopy for clinical questions is a rapid growing field boosted by the Lancet publication of Petricoin, Liotta on overian cancer detection from serum samples et.al. in 2002. The experimental techniques are not only of interest for the diagnostic of diseases but also for the rapid identification of bacteria in various fields.

The improvement of both the experimental techniques and the mathematical analysis of the resulting spectra are a ongoing issue. The workflow of the ClinProTools software and mathematical methods are presented for the data pretreatment, the quantification of compounds and the multivariate analysis.

Keywords: Clinical proteomics, MALDI, multivariate analysis

Joint work of: Decker, Jens; Kuhn, Michael; Schleif, Frank-Michael

See also: Petricoin E.F., Liotta L.A. et.al.: Use of proteomic patterns in serum to identify ovarian cancer, Lancet,2002 Feb 16;359(9306):572-7

Evaluation of Liquid-Chromatography-Mass Spectrometry data for the absolute quantitative analysis of marker proteins in human serum

Nathanaël Delmotte (Universität Saarbrücken, D)

The serum complexity makes the absolute quantitative analysis of medium to low-abundant proteins very challenging. Tens of thousands proteins are present in human serum and dispersed over an extremely wide dynamic range. The reliable identification and quantitation of proteins, which are potential biomarkers of disease, in serum or plasma as matrix still represents one of the most difficult analytical challenges. The difficulties arise from the presence of a few, but highly abundant proteins in serum and from the non-availability of isotope-labeled proteins, which serve to calibrate the method and to account for losses during sample preparation. For the absolute quantitation of serum proteins, we have developed an analytical scheme based on first-dimension separation of the intact proteins by anion-exchange high-performance liquid chromatography (HPLC), followed by proteolytic digestion and second-dimension separation of the tryptic peptides by reversed-phase HPLC in combination with electrospray ionization mass spectrometry (ESI-MS).

The potential of mass spectrometric peptide identification in complex mixtures by means of peptide mass fingerprinting (PMF) and peptide fragment fingerprinting (PFF) was evaluated and compared utilizing synthetic mixtures of commercially available proteins and electrospray-ion trap- or electrospray time-of-flight mass spectrometers. While identification of peptides by PFF is fully supported by automated spectrum interpretation and database search routines, reliable identification by PMF still requires substantial efforts of manual calibration and careful data evaluation in order to avoid false positives. Quantitation of the identified peptides, however, is preferentially performed utilizing full-scan mass spectral data typical of PMF. Algorithmic solutions for PMF that incorporate both recalibration and automated feature finding on the basis of peak elution profiles and isotopic patterns are therefore highly desirable in order to speed up the process of data evaluation and calculation of quantitative results.

Calibration for quantitative analysis of serum proteins was performed upon addition of known amounts of authentic protein to the serum sample. This was essential for the analysis of human serum samples, for which isotope-labeled protein standards are usually not available. We present the application of multidimensional HPLC-ESI-MS to the absolute quantitative analysis of myoglobin in human serum, a very sensitive biomarker for myocardial infarction. It was possible to determine myoglobin concentrations in human serum down to 100-500 ng/mL. Calibration graphs were linear over at least one order of magnitude and the relative standard deviation of the method ranged from 7-15%.

Keywords: RP-HPLC, monolith, Mascot, Myoglobin, Absolute quantitation, Serum

Joint work of: Delmotte, Nathanaël; Mayr, Bettina; Leinenbach, Andreas; Reinert, Knut; Kohlbacher, Oliver; Klein, Christoph; Huber, Christian G.

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/539

An Algorithm for Feature Finding in LC/MS Raw Data

Clemens Gröpl (FU Berlin, D)

HPLC-ESI-MS is rapidly becoming an established standard method for shotgun proteomics. Currently, its major drawbacks are two-fold: quantification is mostly limited to relative quantification and the large amount of data produced by every individual experiment can make manual analysis quite difficult.

Here we present a new, combined experimental and algorithmic approach to absolutely quantify proteins from samples with unprecedented precision. We apply the method to the analysis of myoglobin in human blood serum, which is an important diagnostic marker for myocardial infarction. Our approach was able to determine the absolute amount of myoglobin in a serum sample through a series of standard addition experiments with a relative error of 2.5%. Compared to a manual analysis of the same dataset we could improve the precision and conduct it in a fraction of the time needed for the manual analysis. We anticipate that our automatic quantitation method will facilitate further absolute or relative quantitation of even more complex peptide samples. The algorithm was developed using our publically available software framework OpenMS (www.openms.de)

Keywords: Absolute quantification, diagnostic marker, proteomics, mass spectrometry, liquid chromatography

Joint work of: Gröpl, Clemens; Lange, Eva; Reinert, Knut; Kohlbacher, Oliver; Sturm, Marc; Huber, Christian G.; Mayr, Bettina M; Klein, Christoph L.

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/534

See also: In Proceedings of the 1st International Symposium on Computational Life Science (CompLife 05), Springer Lecture Notes in Bioinformatics, Vol. 3695, pages 151-163, 2005.

Glycosylation Patterns of Proteins Studied by Liquid Chromatography-Mass Spectrometry and Bioinformatic Tools

Christian G. Huber (Universität Saarbrücken, D)

Due to their extensive structural heterogeneity, the elucidation of glycosylation patterns in glycoproteins such as the subunits of chorionic gonadotropin (CG), CG-alpha and CG-beta remains one of the most challenging problems in the proteomic analysis of posttranslational modifications. In consequence, glycosylation is usually studied after decomposition of the intact proteins to the proteolytic peptide level. However, by this approach all information about the combination of the different glycopeptides in the intact protein is lost. In this study we have, therefore, attempted to combine the results of glycan identification after tryptic digestion with molecular mass measurements on the intact glycoproteins. Despite the extremely high number of possible combinations of the glycans identified in the tryptic peptides by high-performance liquid chromatography-mass spectrometry ($>$ 1000 for CG-alpha and $>$ 10.000 for CG-beta), the mass spectra of intact CG-alpha and CG-beta revealed only a limited number of glycoforms present in CG preparations from pools of pregnancy urines. Peak annotations for CG-alpha were performed with the help of an algorithm that generates a database containing all possible modifications of the proteins (inclusive possible artificial modifications such as oxidation or truncation) and subsequent searches for combinations fitting the mass difference between the polypeptide backbone and the measured molecular masses. Fourteen different glycoforms of CG-alpha, including methionine-oxidized and N-terminally truncated forms, were readily identified. For CG-beta, however, the relatively high mass accuracy of $\pm$ 2 Da was still insufficient to unambiguously assign the possible combinations of post-translational modifications. Finally, the mass spectrometric fingerprints of the intact molecules were shown to be very useful for the characterization of glycosylation patterns in different CG preparations.

Keywords: Liquid chromatography, mass spectrometry, glycoproteins, glycosylation, peak annotation

Joint work of: Toll, Hansjörg; Berger, Peter; Hofmann, Andreas; Hildebrandt, Andreas; Oberacher, Herbert; Lenhof, Hans Peter; Huber, Christian G.

High-accuracy peak picking of proteomics data

Eva Lange (FU Berlin, D)

A new peak picking algorithm for the analysis of mass spectrometric (MS) data is presented. It is independent of the underlying machine or ionization method, and is able to resolve highly convoluted and asymmetric signals. The method uses

the multi-scale nature of spectrometric data by first detecting the mass peaks in the wavelet-transformed signal before a given asymmetric peak function is fitted to the raw data. In an optional third stage, the resulting fit can be further improved using techniques from nonlinear optimization.

In contrast to currently established techniques (e.g. SNAP, Apex) our algorithm is able to separate overlapping peaks of multiply charged peptides in ESI-MS data of low resolution. Its improved accuracy with respect to peak positions makes it a valuable preprocessing method for MS-based identification and quantification experiments. The method has been validated on a number of different annotated test cases, where it compares favorably in both runtime and accuracy with currently established techniques.

An implementation of the algorithm is freely available in our open source framework OpenMS (www.open-ms.de).

Keywords: Mass spectrometry, peak detection, peak picking

Joint work of: Lange, Eva; Gröpl, Clemens; Reinert, Knut; Kohlbacher, Oliver; Hildebrandt, Andreas

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/535

See also: In Proceedings of the 11th Pacific Smyposium on Biocomputing (PSB-06), pages 243-254, 2006.

The Peptide MS/MS-Fragmentome: A Set of Predictable Fragment Ions with Highly Redundant Sequence Information

Wolf D. Lehmann (DKFZ - Heidelberg, D)

Upon low energy collision induced dissociation (CID), multiply protonated peptides generate a set of interdependent fragment ions detectable by MS/MS, the '[peptide]n+-fragmentome'. In particular dynamic fragmentation of [peptide]n+ ions in a collision cell generates information-rich MS/MS spectra. Currently, database-supported annotations of peptide MS/MS spectra are mainly based on a combination of peptide molecular weight and y type fragment ions, leaving a considerable number of good-quality peptide MS/MS spectra in proteomics studies unannotated. This situation may be improved by a more complete use of the structural information present in the [peptide]n+-fragmentome.

The presentation provides an overview on the fragment ions of multiply protonated peptides and their connectivity, comprising a ions, b ions, y ions, and neutral loss reactions from the N-, and C-terminus, and internal b ions. In the low-mass region, the unique set of 19 y1 ions and of the 190 b2 ions carries a particular message, since these ions define the N-or C-terminal amino acid(s). Further, the b1 ions of the basic residues K, H, W, and R carry a specific N-terminal information, which is redundant to that contained in the corresponding

b2 ions and in the N-terminal neutral loss peaks. Redundant information is also found in b and y ion series and in complementary b/y ion pairs. The latter are particularly abundant when generated by proline- or aspartate-induced backbone cleavages. From complementary b/y ion pairs the molecular weight of the precursor ion can be reconstructed to confirm or determine its molecular weight. This procedure is helpful in case a mixture of precursor ions is isolated or in case a precursor ion of very low abundance is isolated. Information about the precursor ion charge state is also delivered by precursor ion reconstruction using MS/MS data.

In the analysis of covalently modified peptides, reporter ions are of particular importance. These ions can be used for mining of MS/MS data sets for the occurrence of selected modifications. Examples are presented for selected modifications, such as acetylation and phosphorylation. In phosphorylation analysis neutral loss reactions are highly important, and may also carry redundant information, when observed both from the molecular ion and from fragment ions. Search tools, which fully incorporate the current knowledge about the [peptide]n+-fragmentome will increase the scores of peptide/protein identifications by MS/MS and thus will increase the fraction of automatically assigned MS/MS spectra in proteomics studies.

Keywords: Peptide sequencing, tandem mass spectrometry, electrospray, covalent modification, data evaluation

Joint work of: Lehmann, Wolf D.; Wei, Junhua ; Rappsilber, Juri ; Salek, Mojiborahman

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/547

See also: Schlosser, A.; Lehmann, W. D. Patchwork peptide sequencing - extraction of sequence information from accurate mass data of peptide tandem mass spectra recorded at high resolution. Proteomics 2002, 2, 524-533.

See also: Salek M, Di Bartolo V, Cittaro D, Borsotti D, Wei J, Acuto O, Rappsilber J, Lehmann WD. Sequence Tag Scanning: A new explorative strategy for recognition of unexpected protein alterations by nanoESI-MS/MS. Proteomics 2005, 5, 667-674.

VEMS 3.0: Algorithms and Computational Tools for Tandem Mass Spectrometry Based Identification of Post-translational Modifications in Proteins

Rune Mattiesen (University of Southern Denmark - Odense, DK)

Protein and peptide mass analysis and amino acid sequencing by mass spectrometry is widely used for identification and annotation of post-translational modifications (PTMs) in proteins.

Modification-specific mass increments, neutral losses or diagnostic fragment ions in peptide mass spectra provide direct evidence for the presence of post-translational modifications, such as phosphorylation, acetylation, methylation or glycosylation. However, the commonly used database search engines are not always practical for exhaustive searches for multiple modifications and concomitant missed proteolytic cleavage sites in large-scale proteomic datasets, since the search space is dramatically expanded. We present a formal definition of the problem of searching databases with tandem mass spectra of peptides that are partially (sub-stoichiometrically) modified. In addition, an improved search algorithm and peptide scoring scheme that includes modification specific ion information from MS/MS spectra was implemented and tested using the Virtual Expert Mass Spectrometrist (VEMS) software. A set of 2825 peptide MS/MS spectra were searched with 16 variable modifications and 6 missed cleavages. The scoring scheme returned a large set of post-translationally modified peptides including precise information on modification type and position. The scoring scheme was able to extract and distinguish the near-isobaric modifications of trimethylation and acetylation of lysine residues based on the presence and absence of diagnostic neutral losses and immonium ions. In addition, the VEMS software contains a range of new features for analysis of mass spectrometry data obtained in large-scale proteomic experiments. Windows binaries are available at http://www.yass.sdu.dk/

Keywords: Mass spectrometry, diagnostic ions, neutral losses, variable modifications, database searching, distributed computing

Joint work of: Mattiesen, Rune; Trelle, Morten Beck; Hjrup, Peter; Bunkenborg, Jakob; Jensen, Ole N.

Full Paper:
http://pubs.acs.org/cgi-bin/abstract.cgi/jprobs/asap/abs/pr050264q.html

See also: J. Proteome Res., ASAP Article 10.1021/pr050264q S1535-3893(05) 00264-2

Signal Maps for Mass Spectrometry-based Comparative Proteomics

Amol Prakash (University of Washington, USA)

Mass spectrometry-based proteomics experiments, in combination with liquid chromatography-based separation, can be used to compare complex biological samples across multiple conditions. These comparisons are usually performed on the level of protein lists generated from individual experiments. Unfortunately, given the current technologies, these lists typically cover only a small fraction of the total protein content, which makes global comparisons extremely limited. Recently, approaches have been suggested that are built on the comparison of computationally built feature lists, instead of protein identifications. While

these approaches promise to capture a bigger spectrum of the proteins present in a complex mixture, their success is strongly dependent on the correctness of the identified features, and the aligned retention times of these features across multiple experiments. In this experimental-computational study, we go one step further, and perform the comparisons directly on the signal level. First, signal maps are constructed, which associate the experimental signals across multiple experiments. Only then a feature detection algorithm uses this integrated information to identify those features that are discriminating or common across multiple experiments. At the core of our approach is a score function that faithfully recognizes mass spectra from similar peptide mixtures and an algorithm that produces an optimal alignment (time warping) of the liquid chromatography experiments on the basis of raw MS signal, making minimal assumptions on the underlying data. We provide experimental evidence that suggests uniqueness and correctness of the resulting signal maps, even on low-accuracy mass spectrometers. These maps can be employed for a variety of proteomics analyses. In this paper we illustrate the use of signal maps for the discovery of diagnostic biomarkers. An implementation of our algorithm is available on the CHAMS Web server at http://www.systemsbiology.fr/chams.

Joint work of: Prakash, Amol; Mallick, Parag; Whiteaker, Jeffrey; Zhang, Heidi; Paulovich, Amanda; Flory, Mark; Lee, Hookeun; Aebersold, Ruedi; Schwikowski, Benno

Full Paper:
http://www.systemsbiology.fr/chams

See also: MCP published November 3, 2005, 10.1074/mcp.M500133-MCP200

OpenMS - A Framework for Quantitative HPLC/MS-Based Proteomics

Knut Reinert (FU Berlin, D)

In the talk we describe the freely available software library OpenMS which is currently under development at the Freie Universität Berlin and the Eberhardt-Karls Universität Tübingen. We give an overview of the goals and problems in differential proteomics with HPLC and then describe in detail the implemented approaches for signal processing, peak detection and data reduction currently employed in OpenMS. After this we describe methods to identify the differential expression of peptides and propose strategies to avoid MS/MS identification of peptides of interest. We give an overview of the capabilities and design principles of OpenMS and demonstrate its ease of use.

Finally we describe projects in which OpenMS will be or was already deployed and thereby demonstrate its versatility.

Keywords: Proteomics, C++, Differential expression

Joint work of: Reinert, Knut; Kohlbacher, Oliver; Gröpl, Clemens; Lange, Eva; Schulz-Triegloff, Ole; Sturm, Marc; Pfeifer, Nico

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/546

See also:
http://www.openms.de

Multidimensional Peptide/Protein Analysis and Identification by Sequence Database Search Using Mass Spectrometric Data

Christian Schley (Universität Saarbrücken, D)

Protein samples of biological origin are by nature highly complex and therefore, their analysis requires separation techniques with high resolving power and high peak capacity, respectively. During the past few years, the separation of digests of whole-protein lysates by multidimensional liquid chromatography, which can be readily interfaced on-line to mass spectrometry, has become a real alternative to two-dimensional polyacrylamide gel electrophoresis for high-throughput protein identifications.

In order to generate proteomics data that are suitable to validate protein identification in complex mixtures using multidimensional liquid-chromatography-mass spectrometry approaches, we implemented an offline two-dimensional liquid chromatography method combining strong cation-exchange- and reversed-phase chromatography followed by electrospray ionization tandem mass spectrometry (ESI-MS/MS). The fragment ion spectra generated by ESI-MS/MS were subsequently analyzed via MASCOT database search. The obtained identification data were evaluated in terms of quality of protein/peptide identification by means of score values, reproducibilities of identification in replicate measurements, distribution of tryptic peptides among different fractions, and overall number of identified unique proteins/peptides.

Manual evaluation of the enormous amount of data was very time consuming and challenging, not only because data export was tedious but also because the commercial software packages are usually not very flexible in terms of parameter selection and settings, as well as selected quality criteria. Our evaluated data set is suitable to develop new powerful algorithms and data evaluation software (e.g. MS data analysis software, database search software) for proteome analyses, to be able to deal with huge amounts of protein/peptide identification data and to minimize false positive hits in a fully automated manner. Finally, our scheme of data generation and evaluation was utilized to analyze the proteome of Myxococcus xanthus, a microorganism highly relevant for the production of pharmacologically active secondary metabolites.

Joint work of: Schley, Christian; Altmeyer, Matthias; Müller, Rolf; Huber, Christian G.

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/538

From Functions to Proteins - Searching for Proteases

Hartmut Schlüter (Charité - Berlin, D)

Proteases play a key role in biological processes such as differentiation, protein turnover, or cell signalling. About 1.6% of the gene products from the human genome encode proteases. More than 500 human proteases documented in secondary databases can be delineated in genomic sequence. The number of human proteases reported to be under investigation as drug targets represent approximately 14% of documented proteases (1). The classical approach to identify new proteases integrates knowledge about the substrate. Its cleavage site is used to search for the enzyme which hydrolyses the substrate. E.g. the identification of unknown peptide hormones requests the development of methods to detect and identify the precursor-hormone-processing enzymes. After endothelin and its precursor were identified (2) the authentic substrate, big endothelin, was used for the detection and purification of the endothelin-converting enzyme (ECE) (3). The classical strategy to identify new enzymes includes three steps: 1) an assay system must be established to detect the enzyme of interest; 2) a procedure must be developed to purify the enzyme to near homogeneity and 3) the amino acid sequence of the purified enzyme must be elucidated. Proteolytic activities are often detected with spectroscopic methods. Usually substrates are required that are modified by chromogenic or fluorogenic agents. Such substrates must be synthesized chemically and have also been demonstrated to alter enzyme kinetics (4). Therefore radioactive substrates are often preferred because they are chemically identical with the natural substrates and can be detected with high sensitivity. However, most radioactive substrates must be generated by laborious chemical syntheses. Most importantly, measurement of radioactivity is just a measure of the radioactive isotope and it does not provide any information regarding the identity of the radioactive enzymatic reaction products. Therefore, after the enzymatic reaction, the products must be separated from the substrate before they can be quantified.

Mass spectrometry (MS) is a sensitive, rapid, and accurate quantitative method for analysis of peptides and proteins. Therefore, it is particularly attractive for the analysis of protease reactions. MS allows the direct analysis of mixtures of biomolecules without the need for labelling. Our group developed a MALDI-MS-assisted method to screen complex protein fractions for defined enzymatic activities (5), which includes the conversion of proteins into protein libraries by immobilizing the proteins covalently to beads. Enzymatic activities are monitored by incubating individual proteins from protein fractions with reaction-specific probes. After defined incubation times, aliquots of the reaction mixtures are analyzed with MALDI-MS. The target enzyme is present if the mass signals from the expected reaction products are present in the mass spectrum.

The benefits of this strategy will be exemplified by results from the searching for angiotensin-II-generating and urotensin-II-generating enzymes including the description of the development of the mass spectrometry based assays, the purifi-

cation strategy and the identification of proteases. The results will be discussed with a focus on protein identification problems and validation problems.

References:
(1). Southan C (2001) Drug Discov Today 6: 681;
(2). Yanagisawa M, Kurihara H, Kimura S, Tomobe Y, Kobayashi M, Mitsui Y, Yazaki Y, Goto K, Masaki T (1988) Nature 332: 411;
(3). Takahashi M, Matsushital Y, Iijimal Y, Tanzawa K (1993) J Biol Chem 268: 21394;
(4). Wallenfels K (1962) Methods Enzymol 5, 212;
(5). Schlüter H, Jankowski J, Rykl J, Thiemann J, Belgardt S, Zidek W, Wittmann B, Pohl T (2003) Anal Bioanal Chem 377: 1102

Keywords: Human genome, gene functions, proteases, protein identification

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/544

From Spots to Systems

Johannes Schuchhardt (MicroDiscovery GmbH - Berlin, D)

Understanding dynamical properties of biological systems is a key challenge in modern biology. A broad spectrum of different techniques is currently employed to elucidate different facettes of biological systems. The Human Brain Proteomics Project (HBPP) within NGFN2 is a large consortium of partners employing Genomics and Proteomics techniques to the analysis of brain derived samples. As a central data resource for the consortium "Brain profile database" is set up, a system intended to represent experimental study design and to integrate results from different types of data sources. Data integration and data quality control are key issues to be addressed by the system. Complementing static representation of knowledge in a database, mathematical models offer options for the integration and validation of high-throughput measurements with respect to dynamical properties. Guided by Petri net concepts we strive to establish a framework for the integration of low level molecular information with high level systemic and physiological properties.

Keywords: High throughput methods, data integration, data quality control, systems biology, quantitative proteomics

Software platforms for quantitative proteomics

Ole Schulz-Trieglaff (FU Berlin, D)

In recent years, it has become obvious that mRNA expression does not always correlate with protein expression.

It seems that a full understanding of the complexity of life can only be obtained by examining abundances of proteins under varying conditions.

Accurate measurements of these expression values is crucial. This field of research also requires new computational efforts since the data, often from mass spectrometry experiments, is very complex.

We present two academic software platforms that offer means to reduce, analyse and compare protein expression data gained from liquid chromatography coupled with mass spectrometry. We outline their methodology and compare them to our own project, OpenMS, which is currently developed in our research group at the Free University Berlin in collaboration with the Kohlbacher group at Tuebingen University.

Keywords: Proteomics, mass spetrometry, quantitative measurements

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/537

Computational support for the proteomic side of transcriptional networks

Benno Schwikowski (Institut Pasteur - Paris, F)

Transcriptional networks can be understood as the interplay of DNA, messenger RNA, and proteins. While models for transcriptional regulation can draw on established technologies for acquiring data on genomic DNA and messenger RNA, proteomic technologies are still in their infancy. One of the most promising approaches to identify and quantify proteins on a global scale is the mass spectrometry. In this talk, I will first describe a computational approach to make this type of data more reliable and more complete by integrating multiple experiments on the signal level before its interpretation. Second, I will discuss the open-source software platform Cytoscape for putting the resulting proteomic data in context with other data, such as mRNA expression, and protein interactions.

Keywords: Proteomics, Transcriptional regulation, Regulatory networks, Data integration, Cytoscape

A machine learning approach for prediction of DNA and peptide retention times

Marc Sturm (Universität Tübingen, D)

High performance liquid chromatography (HPLC) has become one of the most efficient methods for the separation of biomolecules. It is an important tool in DNA purification after synthesis as well as DNA quantification. In both cases the separability of different oligonucleotides is essential. The prediction of oligonucleotide retention times prior to the experiment may detect superimposed nucleotides and thereby help to avoid futile experiments.

In 2002 Gilar et al. proposed a simple mathematical model for the prediction of DNA retention times, that reliably works at high temperatures only (at least 70 ℃). To cover a wider temperature rang we incorporated DNA secondary structure information in addition to base composition and length.

We used support vector regression (SVR) for the model generation and retention time prediction. A similar problem arises in shotgun proteomics. Here HPLC coupled to a mass spectrometer (MS) is used to analyze complex peptide mixtures (thousands of peptides). Predicting peptide retention times can be used to validate tandem-MS peptide identifications made by search engines like SEQUEST. Recently several methods including multiple linear regression and artificial neural networks were proposed, but SVR has not been used so far.

Keywords: High performance liquid chromatography, mass spectrometry, retention time, prediction, peptide, DNA, support vector regression

Joint work of: Sturm, Marc; Quinten, Sascha; Huber, Christian G.; Kohlbacher, Oliver

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/548

MALDI Mass Spectrometry for Quantitative Proteomics - Approaches, Scopes and Limitations

Andreas Tholey (Universität des Saarlandes, D)

The determination of absolute protein amounts and the quantification of differentially expressed proteins belong to the most important goals in proteomics. Despite being one of the key technologies for the identification of proteins, the application of matrix assisted laser desorption/ionization (MALDI) mass spectrometry (MS) for quantitative analyses is hampered by several inherent factors. The goal of the present paper is to outline these difficulties but also to present some selected approaches which enable MALDI MS to be used for the quantification of biomolecules. In particular, methods for the improvement of the homogeneity of MALDI samples and the use of internal standards for the relative quantification are discussed. Strategies for in-vivo and in-vitro labelling of peptides and proteins with stable isotopes are presented. The need for guidelines for the presentation and evaluation of data as well as for bioinformatical approaches for the interpretation of quantitative data will be addressed.

Keywords: MALDI, mass spectrometry, quantification, ionic liquid matrices

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/536