Meeting report
Defining the proteome
Jörg Bernhardt

Address: Institut für Mikrobiologie, Universität Greifswald, Jahnstrasse 15, 17487 Greifswald, Germany.
E-mail: Joerg.Bernhardt@uni-greifswald.de

A report on the Fourth Annual HUPO World Congress (HUPO 2005) 'From Defining the Proteome to Understanding Function', Munich, Germany, 28 August-1 September 2005.

At this year's annual congress of the Human Proteome Organization (HUPO) in Munich, some 2,100 scientists and representatives from industry discussed recent innovations, developments and state-of-the-art techniques in proteomics. Preceding the main meeting, a day of review lectures also provided an excellent preparation for students and newcomers to the field. This report discusses a few of the highlights of the meeting.

The integration of data from the various '-omics' fields and putting the parts into a cohesive whole are important steps towards a systems-biology approach, as pointed out by the President of the German Research Foundation Ernst-Ludwig Winnacker (University of Munich, Germany). Along similar lines, Sam Hanash (Fred Hutchinson Cancer Research Center, Seattle, USA) emphasized in his talk on HUPO initiatives that contrary to widespread opinion, collaboration between industry and academia in proteomics is indeed feasible, despite the heterogeneity of the technology used and the complexity of the science. There had been many successes, he noted, for example, progress in the description of the proteomes of specific organs, organ systems and cell types, the production of usable and validated antibodies for any human protein, and access to type-specific subproteomes such as the phosphoproteome (phosphorylated proteins), the glycoproteome (glycoproteins), the secretome (secreted proteins) and others.

Antibodies are widely used in proteomics for protein detection and identification in immunohistological studies, protein arrays, affinity separation techniques, *in vivo* analysis, and so on. Matthias Uhlen (Royal Institute of Technology, Stockholm, Sweden) underlined the importance of properly validated antibodies of proven quality with known monospecific epitope-binding sites for the accurate analysis of proteins in cells and tissues in healthy people and cancer patients. The targets of nearly 90% of all known pharmaceuticals are proteins, and the highlight of Uhlen's talk was the release and online demonstration of the Human Protein Atlas [http://www.hpr.se/] for normal and diseased tissue, a web-based repository of histological micrographs showing the distributions of a variety of protein types and species in their cellular environment.

Proteomics and the clinic
Denis Hochstrasser (University of Geneva, Switzerland) focused on the application of proteomics in the clinical environment. He especially considered the problems encountered in sample conservation and in immediate sample preparation while still at the patient's bedside. Several research studies that should guarantee an unchanged sample were discussed, for example the use of sugars containing buffers and sugar alcohols as antifreeze agents - as used by many animals to prevent freezing in winter. Hochstrasser noted that to simplify handling, protein samples can be digested by trypsin immediately after sampling and the peptide fragments separated by isoelectric focusing (IEF). Because of its high resolving power compared to chromatographic techniques, IEF is often the first step in sample preparation for protein identification by mass spectrometry (MS). Applying his techniques Hochstrasser was able to show in studies on the human brain proteome that there are strong overlaps in the proteome signatures of stroke and Alzheimer's disease patients. Jan van Oostrum (Novartis Institutes for BioMedical Research, Basel, Switzerland) described how proteomic results can help in understanding the principles underlying myopia (shortsightedness), in which a longer eyeball leads to imperfect focusing of light on the retina. Working on the chicken, van Oostrum and his colleagues have been able to show that Apo A1 may be associated with longitudinal growth of the eyeball.

Organizing the data

Peer Bork (European Molecular Biology Laboratory, Heidelberg, Germany) presented computational tools such as SMART for determining homology-based protein structure and STRING for protein-association-network analysis. He also discussed how the correlation of protein-expression data, cell-cycle stage and cellular compartment should give insights into the function of a protein in the cell. Bork showed clearly that only the combination of both techniques - homology and network-based protein analysis - coupled with interaction studies and other experimental techniques can give scientific conclusions on the function of still-unknown proteins. At the metagenomics level, Bork described recent work that determined the distribution of genes in microbial communities. As well as discovering new unculturable species, the metagenomic approach can also reveal metabolic fluxes and networks in ecological niches.

The meeting also included informal discussions and 'surgeries' to address particular problems. At the bioinformatics surgery, representatives from academia and industry continued their efforts to develop standards for proteomics experiments and data. Discussions at the HUPO 2002 meeting have already resulted in several standard requirements for proteomics data presentation, such as MIAPE (the minimum reporting requirement for a proteomics experiment) [http://psidev.sourceforge.net] a subset of the experimental results that contains enough information to assess the provenance and relevance of the methods, results and conclusions. Another standard format is mzXML, a mark-up language for online presentation of MS data. The purpose of the data formats is to establish standards for sample definition, separation techniques (gels, chromatography) and MS data-based identification and characterization of proteins or peptides, that means all data necessary for the reproduction, verification, comparison and exchange of experiments. It was emphasized that an objective review of new publications from the proteomics field without any access to the underlying data is almost impossible. For this reason the construction of a public repository of proteomics data where all published results and especially the corresponding raw data should be stored according to the agreed standards was discussed. Because DNA-sequence databases and gene-expression experiment repositories are highly accepted amongst scientists this aim is a logical consequence of the rapid development of proteomics techniques during the last few years. Hopefully these efforts will generate final results in the near future.

In the MS age of proteomics especially, it is obvious that intensively maintained databases are a key resource for high-quality protein identification. Amos Bairoch (Swiss Institute of Bioinformatics, Geneva, Switzerland) presented plans for improving UniProt [http://www.pir.uniprot.org/], one of the best annotated databases for protein sequences, with respect to protein variations caused by single-nucleotide polymorphisms (SNPs), single amino-acid polymorphisms, splice variants and other post-transcriptional modifications. This repository of protein data will probably become a highly qualitative tool for doing proteomic studies.

Proteomics technologies

Several major milestones in the development of proteomics were reviewed at the meeting. Patrick O'Farrell (University of California, San Francisco, USA) looked back at his postgraduate studies, during which he invented a technique that reviewers at the time thought not worth publishing - two-dimensional gel electrophoresis. Nobel laureate John Fenn (Commonwealth University, Richmond, USA) presented his remarkable work developing electrospray ionization for MS. He was followed by Franz Hillenkamp (University of Münster, Germany) who presented past, present and future aspects of matrix-assisted laser desorption ionization (MALDI)-MS in proteomics and how to bring the matrix in closest contact to the sample to enforce the best possible ionization.

Moving on to more recent developments, John Yates (Scripps Research Institute, La Jolla, USA) discussed a shotgun approach for total proteome analysis modeled on the genomic shotgun. All the proteins in the sample are cleaved by different proteases and as many peptide fragments as possible from the mixture are captured and characterized by MS. Angelika Görg (Technical University of Munich, Germany), who pioneered the development of immobilized pH gradients for protein separation, astonished the audience with her extremely simple and convenient protein prefractionation technique using isoelectric focusing (IEF) trays filled with Sephadex. This allows exceptionally large amounts of proteins or protein extracts to be separated over a wide pH gradient, and because Sephadex is very easy to handle, it is simple to reseparate the defined pI fractions on normal immobilized pH gradient (IPG) strips with reults of amazing quality.

Because of the importance of new liquid chromatography (LC) techniques for the separation and fractionation of complex peptide mixtures, a talk by Petra Olivova (Waters Corporation, Milford, USA) compared several two-dimensional lipid chromatography techniques; to measure the efficiency of two-dimensional separation, an orthogonality index was introduced. Also in this context, Joël Vandekerckhove (University of Ghent, Belgium) described the technique of combined fractional diagonal chromatography (CoFraDiC), an application of liquid chromatography. This technique enables the n-dimensional separation of peptide mixtures by one type of chromatography as a result of the peptide-modifying steps inserted between each liquid chromatography run.

One trend that emerged from the meeting was the automation, parallelization (for example, multiple electrospray ionization sources in one MS device) and miniaturization of gel-free techniques for computer-aided analysis. Physiological proteomics

hitherto only studied by pulse-chase experiments and two-dimensional gel electrophoresis can now be addressed with gel-free techniques. Stable isotope labeling of proteins followed by LC-MS separation enables the analysis of protein synthesis or protein modifications in response to stimuli (such as abiotic stimuli, drugs, and disease) by gel-free techniques. An advance in the measurement of absolute protein amounts was described by Yasushi Ishihama (Eisai Company, Tsukuba, Japan), who described techniques for comparing an isotope-coded peptide population with uncoded peptides from the cell extract of interest. Despite these advances, it was clear from several presentations that the standard two-dimensional-gel analysis approach can still be improved by new image-processing techniques to give much more reliable data.

Biological applications

The application of gel-based proteomics in bacterial proteomics was described by Michael Hecker (Ernst-Moritz-Arndt-University, Greifswald, Germany), whose talk included topics ranging from descriptive bacterial proteomics, including the identification of proteins and their modifications and localizations, to an understanding of bacterial physiology in the model organism *Bacillus subtilis*. The newly developed paradigms in image analysis, represented by the Delta2D software Hecker used, enabled the most reliable data acquisition as well as impressive data visualization using color-coded proteome maps. The software uses a fusion gel from all images that are to be analyzed. On this fusion image (the proteome map) spots are detected and their boundaries used for the quantitation of each single gel, resulting in 100% accurate spot matching.

Erin O'Shea (Harvard University, Cambridge, USA) brought aspects of protein expression, abundance and stability into focus in her talk on yeast physiology, which included an impressive systematic analysis of extremely large amounts of data. The consequences of protein stability, fold-change of mRNA expression and mRNA stability on protein abundance were discussed. O'Shea provided an insight into how a small unicellular organism regulates gene expression not only at the level of transcriptional activation but also at the level of mRNA processing and posttranscriptionally at the protein level. She showed, for example, that especially stable proteins are highly regulated at the level of mRNA production and degradation.

Pier Righetti (University of Verona, Italy) closed the congress with a talk on "the democratic proteome". According to Righetti "mining below the tip of the iceberg for detecting the unseen proteome" is becoming feasible as emerging technologies allow the detection of low-abundance proteins. He demonstrated the use of so-called 'protein equalizer beads' with combinatorial variable acceptor specificity for capturing proteins from a solution. This technique was derived from affinity chips. Protein equalizer beads, which bind proteins using ligand libraries, are extremely useful for removing high-abundance proteins from serum or other body fluids to reveal the low-abundance proteins. Equalizer beads can also be used for the quality control of prepurified proteins. In this case, the main component protein can be removed and the remaining contaminants can be analyzed.

The HUPO 2005 meeting clearly shows that, after having gone a long way towards manufacturing its own tools and developing standards, the proteome community is returning to life science-related problems. This is absolutely essential for the community to be prepared for the next stage: systems biology.