Meeting report

Proteomics gets faster and smarter

Hannah Johnson and Simon J Gaskell

Address: Michael Barber Centre for Mass Spectrometry, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK.

Correspondence: Simon J Gaskell. Email: Simon.Gaskell@manchester.ac.uk

Published: 15 November 2006

Genome ***Biology*** 2006, **7:**331 (doi:10.1186/gb-2006-7-11-331)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2006/7/11/331

A report on the third annual joint meeting of the British Society for Proteome Research and the European Bioinformatics Institute, Hinxton, UK, 12-14 July 2006.

A recent meeting at the European Bioinformatics Institute outside Cambridge, UK, focused on the need to integrate current proteomics methods with advances in both mass. spectrometric (MS) technology and bioinformatics software. The increasing sensitivity of MS coupled with enhanced separation techniques mean that the quantity and quality of the data produced from high-throughput studies are increasing continuously. Without bioinformatics techniques to analyze these data comprehensively, the benefits of the more sensitive technology would be lost. Here, we report some of the highlights of the meeting.

With the emergence of systems biology, entire organisms, organs and tissues are now being characterized. The routine identification of abundant individual proteins from a wide variety of organisms has been possible for decades, but until recently the analysis of less abundant components of the proteome has been somewhat restricted by the methods available. In her plenary lecture, Peipei Ping (University of California, Los Angeles, USA) described a strategy for the isolation, identification and validation of purified organelles from heart tissue, using sophisticated separation techniques before MS analysis to enable the identification of less abundant cardiac proteins. Typically, after glycerol gradient purification, strong cation-exchange chromatography was used, coupled with tandem MS (MS/MS) of digests for protein identification. Native gel electrophoresis was subsequently used for the functional validation of the identified proteins. It was clear from her studies that it is extremely important to perform manual validation of mass spectrometric data and to consider complementary data from alternative techniques, in order to establish confidence in protein identifications.

The importance of quantification

To help understand how an organism works as a whole, it is important to quantify proteins accurately following their identification. Rob Beynon (University of Liverpool, UK) stressed the need for both relative and absolute quantitative proteomics. He reported a novel methodology (QconCAT [http://www.entelechon.com/index.php?id=qcats/index]) for absolute quantification of proteins that is able to simultaneously quantify a number of different proteins within a complex proteome. QconCATs are synthetic proteins that are designed to encode internal standards for a number of different proteins. The protein is expressed (typically in heavy-isotope-labeled form), purified, quantified, and added as an internal standard to the biological sample of interest before chemical or proteolytic cleavage. Alternatively, absolute quantification is enabled by the chemical synthesis, quantification and introduction of individual internal standards. QconCATs enable standardization of the quantity of internal standards and the production of internal standards on a larger scale, thus increasing the throughput of absolute protein quantification.

Kathryn Lilley (University of Cambridge, UK) summarized some of the methods currently used for relative quantification in proteomics, specifically differential gel electrophoresis (DIGE), isotope-coded affinity tags (ICAT), isobaric tagging for relative and absolute quantification (iTRAQ) and label-free MS techniques, and discussed the importance of identifying the advantages of each in order to use it to its full potential. DIGE is a well established technique and allows the visualization of different isoforms. However, the technique is not optimal for membrane proteins, allowing the analysis of only a limited number of species. Stable-isotope-labeling techniques such as ICAT and iTRAQ allow good proteome coverage and multiplexing of a number of samples although they can be somewhat costly and are reliant on a stable chromatographic gradient. Lilley emphasized that assessments of the technical and biological

variance of the techniques and samples under investigation are equally important when quantifying proteins within complex samples. Biological variation can be compensated for by the use of matched controls and technical variation can be minimized by combining biological samples and internal standards as early as possible in the procedure. With the use of technical replication, both biological and technical variance can be overcome.

A number of studies were also presented on the relative quantification of proteins using protein microarrays. Cris dos Remedios (Institute for Biomedical Research, University of Sydney, Australia) presented an antibody array technique that could reportedly identify a variety of diseases, primarily ischemic heart disease, using biomarker-specific antibodies. Jennifer van Eyk (Johns Hopkins University, Baltimore, USA) reported work on biomarker discovery for myocardial ischemia using two-dimensional gel electrophoresis followed by LC-MS/MS for protein identification. iTRAQ was used for relative quantification of the proteins, and the identifications were confirmed by immunoassays. The traditional cellular necrosis markers cardiac troponin I and troponin T were monitored in myocardial ischemia in human patients.

Detecting post-translational modifications

Post-translational modification of proteins plays an important role in regulating numerous cellular processes, and a major focus of discussion at the meeting was the effect of glycosylation in relation to the progression of disease. Ros Banks (University of Leeds, UK) has identified novel protein biomarkers for ovarian cancer using DIGE followed by lectin-based two-dimensional profiling to analyze both serum proteins and glycoproteins. The lectin-based profiling allows the selection of glycoproteins by capture of the glycan side chain. Biomarkers are required most urgently in cancer, where disease progression can be rapid and the disease is often undiagnosed until the stage of metastasis. Noninvasive diagnosis is the main aim of current studies to identify proteins that are altered in metastatic cancer and thus could be used as biomarkers for the development of metastasis.

It has been known for some years that glycosylation of proteins has a profound effect upon metastasis in human cancer. Pauline Rudd (University of Oxford, UK) reported the identification of glycosylated protein biomarkers that were either up- or downregulated in ovarian cancer; of particular interest was the finding that atamin, an existing biomarker for infertility, was downregulated in ovarian cancer. Overall, it was clear that better techniques are needed for the high-throughput analysis of glycoproteins in biological samples. In the area of phospho-proteomics, technical developments in commercial MS technology that should help in the analysis of post-translational modifications were described. The use of electron transfer dissociation (ETD) for this purpose was discussed by Marcus Macht (Bruker Daltonics, Coventry, UK). This method of ion activation cleaves the peptide at the N-Cα bond and leaves post-translational modifications intact, thereby permitting identification of the exact location of the modification.

Identifying lower-abundance proteins

Improved ion-fragmentation techniques coupled with the increased sensitivity of state-of-the-art mass spectrometers have facilitated the identification of lower-abundance proteins as well as aiding the characterization of post-translational modifications. One of us (SJG) described one way of helping to ensure that high-lysine low-abundance proteins are detected - by increasing peptide detection using guanidination. It had been previously shown that when analyzing peptides by matrix-assisted laser desorption ionization (MALDI) MS, arginine-terminating tryptic peptides are better detected. To overcome this bias, our group has used guanidination to convert lysine to homoarginine to facilitate increased detection of lysine-terminated peptides by MALDI MS. When using MS/MS, an improved signal-to-noise ratio is commonly observed, and fragment ions can be used for both identification and quantification of low-abundance proteins, in the latter case using stable-isotope-labeled internal standards.

Beynon introduced an amino-terminal peptide-enrichment strategy that addresses the difficulties in identifying low-abundance proteins. Complex biological samples are proteolytically digested and the amino-terminal peptide of each protein is selectively isolated, thereby decreasing the complexity of the peptide mixture. Subsequent database searching for protein recognition benefits from the knowledge of the position of the proteolytic peptide analyzed.

A relatively new development in proteomics is the use of mass spectrometry for imaging. Graham Cooks (Purdue University, West Lafayette, USA) spoke eloquently about his work on desorption electrospray ionization (DESI), a mass spectrometric technique that allows the imaging and the ambient mass spectrometric analysis of tissue samples without extensive preparation. Cooks envisages the further development of DESI for the analysis of living tissue samples in biomedical science. This technology typifies the need for the development of mass spectrometric techniques for a much wider range of applications than are currently commercially available. Along related lines, Richard Caprioli (Vanderbilt University School of Medicine, Nashville, USA) has used MALDI MS for *in situ* molecular profiling and imaging of proteins in isolated tissues. The use of MALDI as an imaging technique that can be used in conjunction with the analysis of proteins and small molecules in tissues is particularly important for defining the location of specific drugs, lipids, peptides, and proteins

in subcellular compartments. Together, these studies show great potential for the further development of MS for proteomics and beyond.

Software for proteomics

It is clear that powerful bioinformatics tools are needed to allow us to benefit from the improvements in separation and analytical techniques, especially in high-throughput analysis, a point emphasized throughout the meeting. One pressing need is for powerful publicly available proteomics software. Both Jimmy Eng (Fred Hutchinson Cancer Research Center, Seattle, USA) and Ronald Beavis (University of British Columbia, Vancouver, Canada) presented their work on open-source software that facilitates protein identification, validation, quantification and data storage (available at CPAS website [http://cpas.fhcrc.org] and the Global Proteome Machine Organization website [www.thegpm.org], respectively). Conrad Bessant (Cranfield University, Cranfield, UK) addressed the data-processing bottleneck in proteomics with the genome-annotating proteomic pipeline (GAPP; available at the GAPP website [http://www.gapp.info]).

This year's joint BSPR/EBI meeting successfully brought together new developments in mass spectrometric technology and bioinformatics to the wider field of proteomics. To gain a fuller understanding of the function and interaction of proteins within a proteome it will be important to optimize the use of MS and bioinformatics techniques to facilitate proteomic characterization. The development of analytical techniques for protein identification and subsequent absolute quantification in cell lysates will be extremely important in increasing proteome coverage and enabling a systems-biology approach to analyzing the function of proteins within the whole organism. Software and tools will be needed to facilitate the distribution of existing and future data, and also to increase the power of data analysis to keep up with developments in separation technology of biological samples.