

Proteomics: from Qualitative to Quantitative Report on the BSPR 2010 meeting at Hinxton

Emerging disciplines such as systems biology are clearly driving proteomics towards more quantitative approaches. This summer's joint meeting co-organised by the BSPR and the European Bioinformatics Institute was therefore very timely and, in Walter Kolch's words, had an outstanding roster of speakers, including world famous leaders in the field, rising stars and aspiring young investigators.

Ruedi Aebersold of the ETH Zurich, opened the meeting with a provocative discussion of techniques for label free protein quantification and an insight into new methods that may well overcome many of the difficulties the community is facing. In his opening remarks he pointed out that currently protein quantification is mainly comparative, and methods for absolute quantification are limited, despite these being necessary to fully understand protein dynamics. In addition, many experiments simply generate protein "lists", which are still an incomplete picture of the proteome, and are generally pretty irreproducible for the lower scoring proteins due to stochastic sampling. Extensive coverage is possible, but this is expensive in time and money, and becomes more expensive once you consider that, from these experiments, some 10% of proteins change in abundance in atypical comparative experiment. This creates problems of which to follow up, how to follow them up, and what happens to the other data.

Since we have vast amounts of biochemical knowledge, it is possible, and perhaps more desirable, to use this to target specific proteins, and to make measurements on the abundance of hundreds of these in parallel. To do this it is necessary to focus the analysis on fully tryptic peptides that give good ion current and that are proteotypic; i.e. unique, easily detectable by LC-MS, suitable for quantification, and with no amino acids that may confound the analysis (e.g. methionine, N-terminal Glu etc). Analysis of the 10,000 or so human proteins identified in the published Cellzome dataset indicated that reproducibly 2-3 peptides from each protein identified do show the required properties. Moreover, the study supports the idea that the sum of the ion currents for the "top3" peptides for all proteins represents the specific signal intensity and therefore abundance of the protein. Experimentally, using approximately 20 points determined by an AQUA approach to anchor the abundance curve, it was possible to convert the ion current to absolute numbers per cell. The group tested the validity of the method in a study on *Leptospira interrogans* (1800 proteins seen, corresponding to about 50% of genome). Mean errors in the dataset about 25% using top3 (compared to about 65% with spectral counting) and the experiments were easier to perform as well. The measured abundance correlated with data from an orthogonal method, cryo-EM and topography, which measured the abundance of larger protein complexes. In summary, measuring 3 peptides per protein is reasonable for simple organism, giving a 3 logs dynamic range.

For complex organisms, Aebersold pointed out that selected reaction monitoring (SRM) should be more appropriate. Measuring 2-3 transitions per peptide for 3 peptides was optimal and effective over a dynamic range of 5 logs, with very low abundance proteins being picked up in the dataset. The power of this approach was demonstrated by studies aimed at understanding the organization of key protein complexes in the cell. Mapping protein-protein interactions poses several problems, e.g. determining true interactions from contamination,

the stoichiometries of interactions, dynamic changes in the composition of protein complexes, and the true composition inside the cell. This could be overcome using double tagged heavy and light reference peptides that allowed comparison of the same protein as both bait and prey. Using this approach the group studied the stoichiometry of the PP2A complex, showing that catalytic and scaffolding units are tightly bound to form a core but regulatory units are either limiting in concentration or loosely associated. This finding that some apparently “weak” associations are actually strong associations between rare components was a genuinely new insight.

Ruedi Aebersold finished this *tour de force* of protein quantification with a request for the whole community to contribute data. Targeted approaches are showing real promise but require prior data on the best peptides and SRM transitions. The Peptide Atlas could provide the detailed information needed, but there is a huge amount of data out there that has not yet been entered into the database by the proteomics community – please enter your data and help to push forward these exciting new methodologies!

Session 1: Protein Dynamics

Matthias Selbach, Max Delbrueck Center for Molecular Medicine, Germany, continued the theme of protein abundance measurement, but focussed on the turnover of proteins and mRNA and the link between them. He pointed out that the key numbers needed for understanding these dynamics were abundance and half-life. Linking protein to mRNA turnover allows some data gaps to be filled using mathematical modelling to describe the system, and a better understanding of systems responses. Critical to this is how to obtain the numbers to populate the equations. For protein abundance there are several methods, including stable isotope labelling and the ‘top3’ method. Selbach described an alternative technique, intensity based absolute quantification (iBAQ), which uses the total ion intensity for all of the peptides observed, normalized against the total theoretical number of observable peptides. The approach appears to be independent of sequence, accurate over several logs, and not affected by sample complexity, using a complex “universal protein standard” (from Sigma) spiked into an *E. coli* lysate. iBAQ, which was further validated against histones in yeast and the mouse proteasome, is simple and requires just a universal protein spike with no need for complex methods or multiple MS runs, to give the absolute amount of protein. So, how to measure turnover? Selbach has gone back to classical pulse labelling techniques, but using SILAC as the source of label (pSILAC) to generate this data. He was also able to measure mRNA abundance and turnover by 4-thiouridine incorporation and deep sequencing (Solexa). With everything in place, the methodology was applied to a study of mouse 3T3 fibroblasts. Copy numbers of mRNA were found to range from 1-4000, compared to 10¹/100 to 10⁷ for protein, and median half lives to be 9h for mRNA and 45h for protein. As previously observed, there was poor correlation between mRNA and protein abundance (R_s approx 0.6). However, when looking at the stability of the mRNA versus its cognate protein and applying functional correlation, it was clear that there were 4 types of behaviour depending on the stability of the mRNA and protein. Stable/stable behaviour was observed for housekeeping functions, which makes sense when considering that 80% of cellular energy goes in synthesising the top 20% of proteins.

Stable protein, unstable mRNA was observed for RNA binding proteins, and may act as a potential negative feedback system. Unstable/unstable behaviour was observed for cell cycle and signal transduction proteins, where short transition times are required. This gives a deeper insight into the dynamics and control of protein abundance across the whole cell.

Adam Byron, University of Manchester, UK, described the work undertaken in the Humphries laboratory to isolate and analyse protein complexes associated with integrins. Integrins form parts of cell-surface multiprotein complexes that are critical for intracellular interactions that support multicellular life, and are also implicated in many signalling and control roles. However, they are challenging to study due to their combinatorial complexity, the effect of ligands and intricate interactions with the intracellular environment. Byron described an elegant method involving ligand stimulation with a ligand immobilized on a resin bead, rapid chemical crosslinking of the complex using dimethyl 3,3'-dithiobispropionimidate, isolation of the beads and proteomic analysis of the associated proteins. 1300-1500 proteins are seen in each analysis and, using hierarchical clustering and looking at rate of change, they have been able to identify key proteins and processes, such as staged recruitment to the complex.

Julio Saez-Rodriguez, Harvard Medical School, USA, rounded off the session describing a bioinformatics approach to understand how signal processing is altered in disease. Focussing on hepatocellular carcinoma, where there are significant genetic studies, but little protein information, canonical pathways were generated using the Ingenuity software package. Experimental blockade of key points and analysis of proteins using the Luminex system yielded 26,000 bits of data. Thus, MatLab was used to generate a mathematical model correlating time and behaviour, which could be colour coded for easy visualization. Multiple regression analysis was used to generate minimal pathways, and executable "maps" of the system to which Boolean logic could be applied. A metric was defined to evaluate the models, and a framework developed to explore all of the models and use this to identify the one which best fits data. When this applied to the Luminex data initially a 34% correlation was obtained, which could be improved to 81% after training the model based on the output. This approach clearly pinpoints models that best describe data. It can also compare models between cell types and cluster all the models to identify elements that best describe clusters and to see if common motifs exist.

Session 2: Plant Proteomics

Plants have been the test bed in the past for many methodological developments within proteomics. However, their extracts can be notoriously difficult to work with, often containing many highly abundant proteins that skew the range of proteins concentration, as well as plant metabolites such as polyphosphates and plant phenols that can interfere with the common chemistries used within the proteomics technical repertoire. The Plant Proteomics session at BSPR saw four very varied presentations from European speakers. The first was given by **Michel Zivy** of the UMR de Génétique Végétale, Gif sur Yvette, France, who is taking a 2D gel based approach to identify protein makers for the genetic variation in maize that gives rise to differing responses to drought. **Sylvain Bischoff** from ETH in Zurich

described an elegant study to determine the substrates of chloroplast import machinery. By utilising spectral counts of proteins in different cellular fractions of wild type plants and mutants lacking certain key proteins within this machinery, he was able pinpoint candidate protein substrates for different import systems and also study potential consensus sequences in transit peptides of imported proteins.

Ayesha Tahir described her 2D gel based studies at the UMR/INRA, Clermont-Ferrand, to analyse the peripheral layers of wheat grain during development. The aim of this work is to try to understand developmental mechanisms that may influence the quality of grain and disease resistance. The final talk in the session was given by **Waltraud Schultze**, Max-Planck Institut für Molekulare Pflanzenphysiologie, Golm, who is utilising *in vivo* differential stable isotope labelling to study nutrient-induced signalling processes in plants. The study she described involved looking at the dynamics of the phosphoproteome upon a time course of nutrient starvation. Despite the very diverse studies presented in this session and also the varied quantitative proteomics methods employed, a common theme running through all four presentations was the necessary application of robust data analysis, using advanced statistical methods to make sense of high content proteomics datasets.

Session 3: Animal Proteomics, chaired by Josie Beeley

This session highlighted the use of proteomics in animal science, the food industry and animal models. The first presentation, by **Ingrid Miller** (Austria) described the use of proteomics in studies in animal health and disease. Examples cited included animal models for the study of inflammation, changes in the rat liver mitochondria and endoplasmic reticulum associated with endotoxic shock, studies on transgenic mice, and physiological changes which occur during bovine pregnancies. The presentation also included studies on dog serum and urine proteomes in gammopathies. In several of the examples modifications in the classical 2-DE protocol were able to increase the resolution of otherwise crowded gel areas. Findings in animal proteomics could be enhanced by depletion of high abundance proteins, but will need development of new, specific products.

Javier Barallobre-Barrieiro (Spain) then gave a short presentation on proteomic analysis of porcine cardiac extracellular matrix proteins in ischaemic heart disease. Many of the changes observed were in proteins related to the TGFbeta pathway and may be involved in remodelling at a molecular level. **Balwir Matharoo-Ball** (Nottingham) talked about cancer biomarker discovery and outlined the prospects for the use of proteomics in studying the diagnosis, progression, treatment response and use of personalised medicine in cancer. Specific reference was made to prostate cancer, breast cancer, and ovarian cancer, the last of these being associated with a specific biomarker (CA125).

The session concluded with a paper by **Emøke Bendixen** (Århus) on the use of animal proteomics from a systems biology perspective, with specific reference to farm animals. She compared a Jersey cow (high milk producer) with a Belgium Blue Beef Bull (which produces 5 kg muscle per day), and a Danish landrace pig (lean growth) with a Hungarian mangalica (high fat producing obese animal). She then focussed more closely on the pig proteome and neonatalities associated with gut infections and concluded by referring to mastitis in dairy

cows. This session highlighted the potential use of proteomics in veterinary medicine in the study, diagnosis and treatment of diseases in farm and other animals.

Tuesday 13th July 2010

Honorary Membership Award: Professor Angelika Goerg

At the end of the first day of the meeting, Honorary Membership of the British Society for Proteome Research was conferred on Professor **Angelika Goerg** (Technical University of Munich, Germany) for her outstanding contribution in the field of two-dimensional gel electrophoresis and proteomics. In his introduction, **Mike Dunn** (Vice-President BSPR) gave a brief summary of Prof Goerg's career, highlighting the contributions made to the development of electrophoretic technology and also to the long-term commitment to the field of education and training. Indeed, the success of proteomic research activities in many laboratories around the world can be directly attributed to the training that researchers have received through tutorials, workshops and hands-on practical courses organised by Prof Goerg. Following presentation of the Award, Prof Goerg gave a lecture entitled "A life with gels: From ultrathin gels to IPG-Dalt". This interesting perspective detailed how, based on her previous work on protein separations using thin and ultrathin polyacrylamide gels cast on plastic backings, her research group were fundamental in transforming 2-D electrophoresis from a skilled art to a standard technique that could be implemented in any competent proteomics laboratory. The so-called technique of IPG-DALT 2-DE has become one of the cornerstones for protein separations in proteomics research, and has only recently begun to be challenged in its dominance of the field by methods based on the use of mass spectrometry and protein arrays.

Wednesday 14th July 2010

Keynote Lecture 2: Robert Beynon

The second Keynote Lecture of the meeting was given by **Rob Beynon** (University of Liverpool), with the rather provocative title of "Quantitative proteomics: What's the point?" Clearly, as proteomics moves from a descriptive, discovery phase to a quantitative phase, there is increasing pressure to provide reliable quantitative data on the quantification of proteins within a cell. In most MS-based methods, a peptide (usually tryptic) is used as a measure of protein abundance. Prof Beynon firmly believes that for maximal quantitative sensitivity the best approaches are based on stable isotope labelled peptides as chemically identical, mass resolved standards. Prof Beynon presented an overview of the construction of absolute quantification workflows, emphasizing challenges in sample preparation, standard generation and MS based analyses. He went on to describe how these can be applied to absolute quantification of an entire proteome (using QconCATs for standard generation) and the measurement of proteome dynamics by time resolved metabolic labelling with stable isotopes: "dynamic SILAC".

Wednesday 14th July 2010

Session 4: Quantitation by Labelling, chaired by Mike Dunn

Session 4 on quantitative proteomics based on labelling strategies followed the Keynote lecture on this topic. The first speaker in the session, Professor Phillip Wright (University of Sheffield) also chose a challenging title, “Can we live with iTRAQ?” While iTRAQ-based workflows are attractive for proteomic quantitation (>300 original iTRAQ peer-reviewed articles have been published), Prof Wright pointed out that there are idiosyncrasies inherent in the iTRAQ technique (and in other isobaric tagging methods). These problems include issues of overlapping isotopes, and most critically mixed MS/MS. Moreover, it is not clear whether these problems can be solved. In his lecture, Prof Wright addressed how iTRAQ measurements relate to the “real” abundances of target proteins, and how such understanding can facilitate experimental design and downstream-data-analysis strategies to minimise iTRAQ limitations, while maximising its multiplexing capabilities.

In her presentation, Sarah Martin (University of Edinburgh) described an approach to quantitative proteomics using *in vivo* metabolic labelling with ¹⁵N. The study used *Ostreococcus tauri*, a unicellular green alga, which is a model for a number of land plant pathways. The progressive incorporation of ¹⁵N was monitored at specific time points during the cell cycle using a standard LC-MS shotgun proteomic approach. Interestingly, analysis of individual peptides revealed that protein families exhibit labelling at different rates, highlighting turnover and flux variations.

The final presentation was given by Kelly McMahon (University of Bradford). In this study, a quantitative proteomics approach was to characterise specific regions of multi-cellular spheroids (MCTS), an *in vitro* model that can be used to study different regions (aerobic, hypoxic, necrotic core) of tumours. Peptide digests from each region were labelled with iTRAQ reagents and then fractionated according to their pI using an Agilent OffGel system, prior to further separation of each fraction by nano-HPLC, followed by MALDI TOF-TOF MS/MS. A total of 882 proteins were identified, with specific subsets being differentially expressed in the different regions of the MCTS. While many of those proteins identified as differentially expressed in the hypoxic and necrotic regions represented established cancer associated proteins, a number of proteins were highlighted as being up-regulated in both these regions which have no previous association with cancer and may upon validation, provide attractive leads for therapeutic intervention.

Session 5: Advances in Bioinformatics, chaired by Henning Hermjakob

The second session of the day, entitled “Advances in Bioinformatics”, experienced a minor identity crisis, as the first two speakers both emphasized that they did not consider themselves to be bioinformaticians. Dr Hanno Steen, Department of Pathology, Harvard Medical School and Children’s Hospital, Boston, demonstrated a new method to identify substrates of APC, a key cell cycle regulator. The method replaces the classical degradation assay by quantitative proteomics and bioinformatics, and is embedded into a stable statistical framework as a general method for protein profile similarity screening.

After emphasizing his association with the field of biocuration rather than bioinformatics, the next speaker, Amos Bairoch, Swiss Institute of Bioinformatics, Geneva, outlined the vast task of annotating human proteins. From an overview of previous and current human annotation in

UniProtKB/Swiss-Prot, Dr Bairoch moved on to present a glimpse of the future neXtprot resource, generated by his new CALIPHO group with a focus on experimental and collaborative characterisation of currently uncharacterised human proteins.

The final two speakers, both from the European Bioinformatics Institute, Cambridge, UK, did not disavow their affinity to the session title and moved on to present computational access to proteomics databases. Daniel Rios demonstrated a data warehouse (BioMart) and a sequence-centric view for the Proteomics Identifications Database (PRIDE). Henning Hermjakob introduced a common interface for molecular interaction databases (PSICQUIC), providing uniform access to most major interaction data sources.

Session 6: Protein Interactions, Chaired by Judit Nagy

Gavin Wright from the Wellcome Trust Sanger Institute, Hinxton, UK, opened this session talking about large-scale detection of low affinity extracellular protein interactions. Individual cells within biological systems communicate through signals initiated by specific extracellular protein interactions involving membrane-embedded receptors. These receptor proteins are difficult to manipulate and their interactions are very weak that cannot be detected using common high throughput assays. The group developed an affinity based high throughput screening platform called AVEXIS (Avidity-based Extracellular Interaction Screen) using zebrafish as model system and screened 252 cell surface receptors. Over 31000 potential interactions were found, some of which was functionally validated using mutant zebrafish.

The next speaker, Alex Jones from The Sainsbury Laboratory, Norwich Research Park, UK, discussed the identification and quantification of phosphorylation events in plant-pathogen interactions. The plant defence response against pathogens is maintained by protein kinases and in particular the receptor-like kinases (RLK). The two main challenges the group faced were related to quantification and the position of phosphorylation. The group is in the early stages of developing multiple reaction monitoring to overcome these problems.

The third speaker of this session, Bettina Warshcheid from the University of Duisburg-Essen, Germany, spoke about accurate characterization of the intricate membrane protein interaction network of yeast peroxisome via quantitative MS and statistics. Peroxisomes are essential part of eukaryotic cells. Using *Saccharomyces cerevisiae* as a model system, SILAC labeling and affinity purification followed by LC/Orbitrap-MS/MS analysis, the group was able to investigate the low abundant peroxisomal membrane protein complexes including the interactome of Pex14p which allowed new insight into other peroxisome membrane protein networks.

The final speaker, Akhilesh Pandey, John Hopkins University, USA, followed with another quantitative approach to map protein-protein interactions. His group developed a SILAC-based approach to study kinases and their substrates (www.silac.org) with a particular focus on the signalling pathways activated by TSLPR, a receptor involved in the pathogenesis of asthma and acute lymphoblastic leukaemias. They identified about a dozen of the serine/threonine phosphorylated peptides from SCX fractionation experiments, which helped to build a more detailed picture of the TSLPR signalling pathway. The conclusion was that it is possible to take unbiased discovery approaches to dissect signalling pathways in a global

fashion using quantitative proteomics and activated kinases identified using this proteomic approach are candidate therapeutic targets in asthma and leukaemias.

It was clear from these talks that quantification is a major challenge, particularly in the case of protein-protein interactions, because these often involve low abundance proteins and the interactions are often very weak.

Session 7: Modern Array Technologies, chaired by Rainer Cramer

In the final session of the second day MS- and gel-based proteomics was swapped for protein arrays. First, Prof Cahill from the Conway Institute at UCD, Ireland, provided a fascinating overview of her work over the last decade in this field, describing the development and application of an economical and large-scale protein array in probing the human calmodulin neural interactome. In a highly entertaining talk, Dr Nordhoff (Medical Proteome Centre, Ruhr-University Bochum, Germany) then demonstrated the application of a commercial high-density protein array for studying the autoimmune response in neuro-degenerative disorders such as Alzheimer's (AD) and Parkinson's Disease (PD). The working hypothesis for this work was that specific autoimmune antibodies are present in AD and PD patients, which can be ultimately used as biomarkers for the diagnosis and monitoring of AD and PD by a simple blood test.

Plenary Lecture, chaired by Rainer Cramer

Prior to the conference dinner and after a long day of quality research presentations, Prof Cathy Costello (Boston University School of Medicine) had the (usually not easy) task to end the scientific part of the day with a plenary talk. However, her exciting talk on the proteomic (and metabolomic) analysis of rheumatoid arthritis and Lyme disease, focusing on HLA-DR presented peptides and glycolipid analysis, captured the attention of the audience and kept everyone alert and concentrating right to the end.

As in previous years, the drinks reception and conference dinner proved to be highly successful events in which everyone was able to demonstrate their knowledge of the World Cup as well as of proteomics and its methods, employing *ad hoc* manual biometric analysis of football star cards and mass spectrometric peptide sequencing.

15 July 2010

Keynote Lecture 3: chaired by Stephen Pennington

The final day of the meeting began with a plenary lecture from Benjamin Garcia (Princeton University, USA) who provided an impressive and inspiring *tour de force* describing the application of novel proteomic strategies for detailed understanding of the significant and dynamic modifications in histones. Histones, the core proteins of chromatin, can undergo significant post-translational modification including acetylation, methylation and, of course,

phosphorylation. Ben described how, using comparative SILAC labelling, they had been able to undertake a comprehensive analysis of the major known modifications on all histones during stem cell differentiation. They had also used a *de novo* method for identification of previously unknown modifications – for example methylation of lysine 18 in histone H3. Impressively, Ben then showed how they used SILAC for dynamic measurement of histone turnover including the turnover of key modifications as a prelude to the introduction of a new middle down approach combined with advanced chromatography (weak cation exchange HILIC) and software to support a comprehensive analysis of detailed combinatorial modifications of histones – the histone code. Ben concluded by showing application of this approach to reveal the role of a novel histone H3 methylase as an oncogene in myeloma.

Session 8: Label-free Quantitation, chaired by Stephen Pennington

Ben was a hard act to follow but David O'Connor (University of Southampton) provided a beautifully clear, logical and structured introduction and insight into the application of quantitative protein measurements – including MS^E. David showed convincingly how this approach has provided important insights into the biology of the pathogen *Chlamydia trachomatis*, the leading cause of preventable blindness. Detailed and careful studies of protein abundance in elementary and reticulate bodies revealed the expression of previously hypothetical proteins, the extent to which the organism invests energy in the production of specific proteins and a novel role for cell wall enzymes in reticulate bodies.

The theme of quantitative MS was continued by Sebastian Wiese (University Duisburg-Essen, Germany) who showed how single ion monitoring can be used for highly sensitive and broad dynamic range measurements of proteins and demonstrated its application to the characterisation of peroxisomes from different tissues and organisms to provide a comprehensive profile of the composition of these important sub-cellular organelles. As anyone who has used LC-MS proteomics techniques is aware the analysis and interpretation of the data represents a key stage in the experimental process. The session ended with Dario Di Silvestre (Institute for Biomedical Technologies-CNR, Italy), who described how the software for visualisation of LC-MS data including MudPit data they have developed - Multidimensional Algorithm Protein Map (MAProMa) – can be used with Cytoscape to improve the unbiased interpretation of these large datasets. Together the presentations clearly showed that when tackled appropriately the challenges in undertaking quantitative proteomics can be overcome to produce data which provides significant advances in our understanding of biological systems.

Session 9: Clinical Proteomics, Chaired by Roz Banks

The Session on “Clinical Proteomics” provided illustrations of both qualitative and quantitative aspects of proteomics studies in several diverse disease areas. The first short communication by Marie Claude Djidja (who was subsequently awarded the Early Stage Career Investigator award) showed data from multiple approaches but with a focus on MALDI imaging to demonstrate how protein expression in the tumour microenvironment

changes during the metastatic progression with hypoxia being a key regulator. A challenge highlighted here was the need to be able to develop more quantitative based MALDI methods to enable comparison with other approaches and model systems used in parallel such as SILAC. The following short communication by Thomas Hiemstra addressed the issue of the difficulties of working with urine where the abundant urinary protein uromodulin (Tamm-Horsfall protein) dominates and often co-purifies with exosomes, reducing coverage during analysis. As an alternative to removing uromodulin, which is problematic and may also result in removal of other proteins/peptides, the effects of using an exclusion list for the most intense uromodulin peptide spectra were explored. Although this resulted in the identification of an additional 80 proteins in urinary exosomes compared with conventional means, overall coverage was greatly reduced with only 39.5% of the 476 proteins identified by conventional means being identified using an exclusion list method and hence combining the methods was judged to be the best way forward. In the Invited Speaker presentation for this session, John Semmes provided an introduction to prostate cancer and how the needs were now focussed on being able to discriminate between significant and insignificant disease clinically. In an elegant overview he outlined a multi-model multi-technology pathway to biomarker discovery using tissue, prostatic secretions and prostatic cell lines analysed by imaging mass spectrometry, glycan-based profiling and surface protein analysis respectively. Each of the approaches was of value independently but the integration of the approaches was shown to provide complementary and overlapping information and proof of principle examples were presented. The following short communication by Ros Jenkins described how drug-induced anaphylaxis is a major clinical problem with β -lactam antibiotics such as the penicillins. Using mass spectrometry, the nature of the chemical adduct formed by penicillin-induced modification of HSA was defined and analysis of the stereochemistry of modified peptides was used to describe for the first time how such adducts are formed. Using an MRM assay to develop a profile of the protein modifications with similar conjugates being formed *in vitro* and *in vivo*, their functional significance was shown in the induction of T-cell proliferation in samples from hypersensitive patients. Such results may contribute to the future design of pre-screening strategies and also design of new drugs through knowledge of the binding mechanisms involved in this type of adverse reaction. The session concluded with a further quantitative use of mass spectrometry using LC-MS/MS and SRM but with a disease focus of atherosclerosis presented by Richard Kay. By using a specific chymotryptic peptide which allowed the discrimination between the wild type and Marburg I variants of Factor VII-activating protease, differing in the substitution of a glycine for glutamic acid residue, quantitation of both forms was possible.

Closing Session, chaired by David O'Connor

Keynote Lecture 4: chaired by David O'Connor

The meeting concluded with a Keynote Lecture by John Yates III (Scripps Research Institute, La Jolla) on the use of mass spectrometry to understand protein misfolding diseases. Before launching into the science, Prof. Yates alluded to a current feature in *Nature Biotechnology*, which quoted him as saying that 'biomarkers have been the biggest disappointment of the decade'. The point he wished to emphasize was that genomic science, and in particular

genome wide association studies, had not yet delivered either, despite the vastly greater financial support that had been invested in this area. Further, it was likely that proteomics could make more useful contributions to translational medicine in other ways. To illustrate this point, the talk focussed on the central cellular mechanism for generating and maintaining normal protein folds, the protein homeostasis or proteostasis network, and how the measurement and characterization of interacting components can lead to promising corrective strategies. The optimistic note provided a fitting end to a meeting that highlighted how quantitative approaches are revolutionizing the field.

BSPR/EBI Educational Workshop: Quantitative Proteomics

Organizers: Judit Nagy and Sandra Orchard

A major purpose of the BSPR is the promotion of proteomics through education, thereby furthering the highest standards in experimental practice nationwide. For this reason, this year BSPR sponsored a post-conference educational workshop, held in the computer room of the Wellcome Trust Conference Centre. Users participated in a combination of lectures and hands-on exercises. The workshop was organized jointly by BSPR and EBI, as a result of this there was a strong emphasis on peptide/protein identification and the subsequent annotation of protein lists to give biologically meaningful results from proteomics experiments. The workshop dinner, held at Homerton College, provided the perfect setting to meet and networks with the speakers, organizers and participants

The workshop started with a presentation and hands-on session led by Martin Wells, Nonlinear Dynamics, Newcastle, UK. Martin started his presentation by summarizing the big questions in quantitative proteomics: gel- or gel-free, labelled or label-free, relative or absolute and settled on gel-free, label-free, relative quantification for the rest of his presentation. He emphasized the need to: 1) Process large data files efficiently enabling the confident selections of features for further investigation; 2) Limit user subjective handling and interpretation of the data to enable reproducible analysis and results; 3) Increase confidence in the resulting protein lists through improved data quality. Missing data is the primary limitation to the application of multivariate statistical techniques so addressing this has a significant impact on data reliability and confidence. The participants then took part in a half an hour exercise to process LC-MS data and find the statistically significant changes between two samples.

The pre-dinner talk was delivered by Amos Bairoch from the Swiss Institute of Bioinformatics, Switzerland, one of the main conference speakers. Amos first described how it is possible to extract from UniProtKB complete sets of protein entries corresponding to a sequenced genome. This can easily be achieved through the use of the "Complete proteome" keyword and selection criteria for most bacterial and fungal proteomes as well as human, *Arabidopsis* and *Drosophila*. For other important eukaryotic species such as mouse, rat, bovine, zebrafish or *Xenopus*, it can be done using the EBI IPI protein collections and for other species, there is not yet any trivial or stable solution and the users need to shop around in Ensembl, NCBI or in genome sequencing web sites. Import of missing Ensembl mouse and rat predicted protein sequences into UniProtKB have recently been undertaken and this same exercise is currently underway for other species.

In terms of annotations, he emphasized that reliable protein identification requires information relevant to splice variants, protein variation and post-translational information. All this data is annotated in UniProtKB Swiss-Prot but not many protein identification software packages are capable of making use of these resources. The new PSI PEFF format may be a solution to this conundrum as it is supported by many software vendors. Finally, a short overview of neXtProt, the human protein knowledge platform was presented. neXtProt will go live in the autumn of 2010.

The second day started with a look at the PRIDE database (www.ebi.ac.uk/pride) from both a

user and a submitter perspective. The students were given a short tour of the website by Juan-Antonio Vizcaino (EMBL-EBI), and also of the analysis tools. Current work by PRIDE in enabling the increasing amount of quantitative data to be collected and displayed was also demonstrated. The students then enjoyed hands-on access to PRIDE and were introduced to the main submission tool, PRIDE convertor tool which converts mass spectrometry data from most common data formats into valid PRIDE XML. It presents the user with a convenient, wizard-like graphical user interface, and includes efficient access to the Ontology Lookup Service (OLS). OLS enables the querying, browsing and navigation of biomedical ontologies and controlled vocabularies, providing both a centralized location to perform queries via a lightweight web-accessible user interface as well as a consistent, unified SOAP interface for automated queries. Finally the use of the Protein Identifier Cross-Reference (PICR) service was described. PICR is a web application that provides interactive and programmatic (SOAP and REST) access to a mapping algorithm that uses the UniProt Archive (UniParc) as a data warehouse to offer protein cross-references based on 100% sequence identity. This enables users to solve a common problem for those working with protein sequence data – a need to translate between the different identifiers used by the different databases.

The work of the HUPO-PSI in establishing data standards, which address the plethora of manufacturer-specific outputs from the many different mass spectrometer and protein identification software which the user is faced with, was described by Sandra Orchard (EMBL-EBI). The HUPO-PSI provides XML formats for both the output of mass spectrometers (mzDataML) and of protein identification software (mzIdentML) and is currently working on an XML format to describe quantitation data (mzQuantML). An increasing number of machines and both commercially available and open-source software now incorporate these formats which will ease the process of depositing data into public repositories. Similar formats exist in the domains of gel-based protein/peptide separate (GelML), column-based separations (sepML) and molecular interaction PSI-MIXML2.5. Each domain has an accompany publication which describes the minimum reporting requirements when describing such an experiment and maintains domain-specific CVs for annotating the XML, available on the OBO website (www.obofoundry.org/).

Sandra then went on to speak about the world of molecular interactions. The adoption of the HUPO-PSI common data formats by all the major public domain interaction databases has enable the development of the PSICQUIC web service, a community standard for simultaneous computational access to molecular interaction resources. One implementation of a PSICQUIC service is on the IMEX website – a consortium of interaction databases which release not non-redundant set of highly annotated data onto a single common website (www.imexconsortium.org). The IntAct database is a member of this consortium, but also has available almost ¼ million binary molecular interactions via its own website. The students accessed the website, using both the Quick Search and Advanced Search facilities and finally visualized the data, both locally and via the Cytoscape analytical and visualization software.

In the final talk of the course, Kathryn Lilley (Cambridge University) introduced the principles of mass spectrometry quantitation. She talked about the different methods by which this could be achieved, describing both label and label-free techniques and gave the students an insight into the advantages and disadvantages of each. Much of this was illustrated using

examples taken from her own work at the Cambridge Centre for Proteomics. She finished the session with a few apposite comments on the importance of experimental design and in tailoring the method of choice to both sample availability and cost of the procedure.