

Meeting Report:

The Fourth Annual PepTalk Meeting: The Human Proteome

Sean R. Gallagher*

UVP, Inc., 2066 West 11th Street, Upland, CA 91786, USA

Abstract: The Fourth Annual PepTalk meeting organized by the Cambridge Healthtech Institute on proteomics was held in San Diego, California from January 10 to 13, 2005. The entire meeting had an interesting name: Proteomics in a Six-Pack, and consisted of six individual meetings that ran for two days each, in two groups of concurrent sessions: (1) Fourth Annual Protein Arrays: Complex Challenges – Creative Solutions (January 10-11); (2) Eighth Annual Protein Expression (January 10-11); (3) Inaugural Protein Folding Disorders (January 10-11); (4) Second Annual Protein Process Development: Optimizing Protein Expression Through Scale-Up (January 12-13); (5) Fifth Annual The Human Proteome: Plasma Proteomics (January 12-13); and (6) Inaugural Protein Therapeutics: Minimizing Problems – Maximizing Production, Progress, and Potential (January 12-13). This was indeed what the organizers called the “protein information week”. This report summarizes The Human Proteome meeting, which ended with a joint closing plenary session with the Protein Process Development and the Protein Therapeutics meetings.

Key Words: Proteomics, human proteome, plasma proteomics.

The Human Proteome: Plasma Proteomics meeting held in San Diego California (January 12-13 2005) was part of Cambridge Healthtech Institute’s Fourth Annual PepTalk 2005 meeting. The meeting was, in general, focused on analyte enrichment, plasma profiling, human protein diversity, and biomarkers. A spirited talk on the importance of the human plasma proteome paraphrasing the view that suggests the circulatory system is the sewer system of the body by **Stephan Naylor** of the Boston University set the tone of the meeting. A central question that reverberated throughout the meeting seemed to be, given the complexity of the plasma, what makes the most sense to begin plasma proteome analysis with: the complex total plasma or prefractionation techniques for the proteins of interest? Another theme of the meeting was the need for robust industrial process development and control for generation and analysis of proteomics data.

Katheryn Resing of the University of Colorado, described the critical importance of the understanding of activation pathways for understanding disease progression in cancer. The talk reported that through the use of a combination of expression markers and signaling pathways, markers for melanoma can be identified. Description of additional research included development of shotgun proteomics method using Multidimensional Protein Identification Technology (MudPIT) for use in humans. Separations based on strong cation exchange and hydrophobicity lead to removal of false positive hits, and with the development of gas phase fractionation, prefractionation, and software for splice variants, the group can now detect about 5,000 high quality hits from a single sample. Protein quantitation

without labeling with isotope is also possible. Issues include the informatics and post translational modifications. The time taken to obtain a typical complete protein profile with a linear ion trap takes anywhere from 3- 6 weeks.

In her talk “Reducing the Dynamic Range of Serum and Plasma Using Solid Phase Combinatorial Ligands”, **Julia Lathrop** of the American Red Cross Biomedical R&D outlined a method for concentrating of low abundance or rare protein species while diluting out the high abundant proteins. This is accomplished through bead based peptide library construction: with each bead capable of binding a unique protein and the capacity of each bead equal, a percentage of each protein in the complex mixture is bound. Not only the differential expressions of proteins are maintained, but no depletion of IgGs or albumin is required in the serum or plasma prior to treatment.

Brian M. Balgley for Calibrant Biosystems, Inc. presented on selective analyte enrichment targeting plasma proteome dynamic range. A variety of methods for multidimensional protein and peptide enrichment and identification based on capillary IEF and reversed phase LC (CIEF-CRPLC) coupled to MS/MS were outlined. This included a nanoscale membrane reactor for on the fly real-time proteolytic digestion to simplify and further automate the CIEF-CRPLC approach. Using this approach, 1264 and 1601 proteins were identified from 10 µg of human ovarian cancer and mouse lymphoma respectively. In addition, 1376 proteins were identified from 10,000 cells from an astrocytoma microdissection. The benefits and issues with non-gel based techniques using shotgun or bottom up approaches and top down approaches were also discussed.

Jörg von Hagen of EMD Biosciences presented tools for enrichment of biomarkers from body fluids and analysis of oncologically relevant signal transduction pathways,

*Address correspondence to this author at the UVP Inc., 2066 West 11th Street, Upland, CA 91786, USA; Tel: +1 909 946 3197; Fax: +1 909 946 3597; E-mail: seang@uvp.com

highlighting strategies for the enrichment and profiling of protein and peptide biomarkers of high quality. An emphasis was put on combinations of tools based on their ProteoEnrich™ and ProteoExtract™ product families. The challenges of profiling and discovery of biomarkers are many: sample complexity, poor analytical reproducibility for many techniques such as two-dimensional gel electrophoresis (2D-GE), high equipment cost (e.g., mass spectrometers), low and high abundant proteins present in a sample, and target validation. Removal of high abundance proteins and cell fractionation are key techniques for increasing the concentration of low abundance proteins and improving the odds for success of discovery of biomarkers. This was illustrated with the subcellular localization of MAP kinase in human osteosarcoma cell line SAOS-2, and data showing that selective enrichment of HEPG2 cells leads to 3 fold enrichment of detectable spots in 2D-GE. Further examples of enrichment were given using their ProteoEnrich™ CAT-X SEC, SE, and ATP-Binders Kit, with protein retaining function.

Results from the Human Proteome Organization's (HUPO) plasma proteome international collaborative project were presented by **Gilbert S. Omenn** of the University of Michigan. Noting that the challenges after mapping of the human genome are in functional genomics, systems biology and understanding the dynamic protein compartmentalization of cells and tissue, Dr. Omenn presented the progress in the HUPO plasma proteome project (PPP). With the overall goals of performing a comprehensive analysis of plasma and serum proteins in humans, and identification of biological sources of variation among individuals and populations over a lifetime, the program also seeks to identify and validate relevant biomarkers. Participating laboratories use reference samples and various technology platforms to separate and identify proteins. Thirtyone laboratories are participating from around the world, using LC-MS/MS, MALDI-MS and SELDI-MS coupled to a variety of separation methods such as 2D-GE. This has led to high confidence identification of 2852 proteins. A key observation is on the effect of depletion of proteins. Although large number of proteins can be identified post depletion, concerns remain about nonspecific removal of other proteins during the depletion process. However, the most significant high confidence yields came through a combination of depletion and two or three-dimensional separation followed by ESI-MS/MS with an ion trap LTQ spectrometer. Ontology of the proteins indicates a wide range of functions and the identification of a number of novel peptides and potential disease biomarkers.

Profiling of mouse and human serum using antibody microarrays and resonance light scattering particles was presented by **Bernhard Geierstanger** of the Genomics Institute of the Novartis Research Foundation. The talk reiterated that "microarrays in well" format assays are increasingly important to not just DNA, but protein based analysis. Through arraying out purified antibodies, a large number of measurements can be made on very small amounts of sample. Antibody arrays enable multi-analyte detection for applications that include protein abundance, tissue profiling, and disease marker identification in fluids and tissues. Using a sandwich assay strategy, serum assays

for a mouse model of dermatitis and human sera from colon and prostate cancer patients were discussed, as was the use of Resonance Light Scattering (RLS). Although inexpensive, quantitative and amenable to high throughput, antibody microarrays are limited by availability of antibodies and their cross reactivities.

The talk on comparative analysis of the serum proteomes of human and mouse was presented by **Brian L. Hood** of the National Cancer Institute. Human proteome was evaluated for relevant biomarkers through a global human serum survey, a low molecular weight protein/peptide proteome analysis with depleted serum, and interaction study that looked at peptides bound to high abundance serum proteins. All three of these approaches yielded different but overlapping complements of proteins and showed protein interaction (i.e., the interactome) between high and low abundance proteins. The interactome analysis identified 12 proteins currently used as clinical biomarkers. The second half of the talk discussed extension of the mouse model to the serum proteome. With 80% of the mouse coding regions having a direct orthologue in the human genome, mouse disease models are very attractive for understanding disease processes in humans. Currently, there is a paucity of information about the mouse serum proteome, and a multidimensional global characterization of the mouse serum proteome was initiated using undepleted serum to characterize this model. The global project identified several thousand unique proteins and was extended to an analysis of control and Lewis lung carcinoma (LLC) mouse serum proteomes, identifying numerous unique proteins and peptides for control and LLC mouse. Isotope labeling was also used.

Xiao-jun Li of the Institute for Systems Biology talked about application of peptide expression array generated from LC-MS data to serum profiling. High dynamic range and complexity are two of the major issues with serum proteome analysis. Through the use of a glycopeptide capture method, sample complexity was reduced 100-fold. The automated LC-MS based system developed for this analysis shows excellent reproducibility for peptide mass to charge ratio and peptide signal intensity. The process was applied in two preliminary studies involving a protein mixture and a mouse skin cancer model.

Mark T. Flocco of Bruker Daltonics talked about 'mass spectrometry meets medicine' and the future applications for mass spectrometry in clinical research. Describing the future of diagnostic screening as fast, minimally invasive, accurate, user friendly and small, a number of research tools was presented. This included ClinProt magnetic isolation beads, ClinProt Tools, Proteineer fc MALDI fraction collector, and their AnchorChip technology.

Peter Schulz-Knappe of BioVisioN AG talked about clinical peptidomics and sensitive search for biomarkers. Pepidomics® is an integrated system for the analysis of native peptides and small proteins up to 15 kDa without digestion or gels and with picomolar sensitivity. Pepidomics® has a variety of applications including peptide biomarker discovery, drug profiling and clinical diagnostics. Through the use of Differential Peptide Display®, disease specific changes in peptides can be visualized from a variety of samples such as serum.

Stephen Naylor of the Boston University talked about 'extracting knowledge from the mystery soup of body fluids-the plasma proteome'. Pointing out that proteomics is difficult, he emphasized the need for more progress in analytical tools and highlighted the numerous daunting issues confronting protein based analysis. These include the large number of proteins present in samples, structural and physiochemical complexity of proteins, wide range of concentration of proteins in samples, and no PCR equivalent. These properties lead to the top 10 challenges for proteomics research: analysis of complex samples, differential analysis, relative and absolute quantitation, dynamic range, high throughput analysis, multiplexing, protein expression and production, purification and analysis of membrane proteins, post translational modifications, and protein informatics. Because of the complexity, he pointed out that the techniques need to be framed around the biological question, and that some strategies such as total sample digestion and analysis will likely be too complex to fully analyze by LC-MS strategies alone.

Dobrin Nedelkov of Intrinsic Bioprobes, Inc. talked about population proteomics and how to address the question of human protein diversity. He emphasized the need for proteomic studies that are analogous to wide scale population studies in genomics. On this line, he presented a small scale study. A range of variation was detected in the study including point mutations, and various post translational modifications. Some of the correlations identified were: ApoE3/E4 was detected only in six males, transthyretin (TTR) Gly6Ser detected only in Caucasians, and transferrin (TRFE) deglycosylation in individuals with deglycosylated ATIII. Defining the variations in "normal" populations is critical to future biomarker development and will require studies on thousands of individuals.

Christopher Southan of AstraZeneca R&D spoke about annotation of the human genome by high-throughput sequence analysis of naturally occurring proteins. There are numerous reasons for using MS-based protein sequence verification for a variety of reasons, e.g. clarification of false negatives and false positives in the public databases, better estimation of alternative splice forms and protein number, noncoding transcripts and unvalidated SNPs that lead to a change of amino acid sequence. The Oxford Genome Anatomy Project (OGAP) was developed by Oxford Genome Sciences and used MALDI-TOF analysis of peptides isolated from proteins separated by single and two dimensional gel electrophoresis. The data showing proteins that were identified and linkage to the gel positions can be found at www.OGAP.co.uk.

Daniel Chelsky of Caprion Pharmaceuticals, Inc. talked on quantitative proteomic profiling of pre-clinical models of diseases. He noted that proteomics benefits greatly from a robust, process oriented approach to data generation. Each analysis process phase or "gate" such as sample accrual and sample processing has from 1 to 100 steps that must be optimized through the use of instruments designed to perform day in and day out. On the data analysis side, clustering is critical to see up and down regulation. Ion intensity measurement gives a low coefficient of variation of about 12% and can be used to generate correlation of peptides

across several samples. He illustrated the robustness of the technique with data indicating differential peptide intensity predicting differential abundance of proteins in the sample. The overall process was illustrated with several drug action studies including a cluster analysis of inflammation that showed up and down regulation and off target effects.

Pierre Massion of the Vanderbilt University presented on diagnosis of lung cancer from serum proteomic profiles. Dr. Massion presented a new strategy for the early detection of small asymptomatic lung cancer that is aimed at reducing cancer related mortality in the population at risk. Currently there are no diagnostic, predictive or prognostic biomarkers available for this. In addition, although only 25% of new patients are candidates for surgery, surgery is still considered the best chance for a cure. Through the use of MALDI-TOF MS, plasma proteomic profiling was able to discriminate cases from controls with high accuracy (84%), and identify determinates for recurrence in stage 1 cancer. Details were given on the acquisition, smoothing, baseline correction, internal calibration, peak selection, binning, normalization, statistical cluster analysis, and reproducibility of the process.

Martyn T. Smith of the University of California, Berkeley gave an illuminating presentation on proteomic profiling following exposure to chemicals and hematopoietic malignancies. Benzene, present at a concentration of 1% in gas, was studied with Ciphergen's SELDI based Protein Chip® to identify potential biomarkers for exposure. Similar work is being performed to identify biomarkers for exposure to dioxin. Proteomics approaches also appear promising for better diagnostic biomarkers for childhood leukemia and non Hodgkin's lymphoma.

Jennifer Krone of Applied Biosystems talked about novel advances in TOF/TOF technology for proteomics and biomarker discovery research. She discussed the advantages of LC (off-line)-MS/MS with MALDI: more flexibility with the LC parameters since LC is decoupled from MS, LC data can be reviewed prior to precursor selection for MS/MS, and the LC output is "frozen" on the MALDI plate allowing advanced features of the GPS Explorer software to be used. The talk detailed the use of iTRAQ™ reagents and the GPS explorer software to characterize protein expression in brain tumors with multiplex stable isotope labeling.

Thomas Vondriska of the University of California, Los Angeles focused on the cardiac markers and plasma proteome. Cardiovascular disease is the leading cause of death in the western countries, and, excluding communicable disease, world wide. Mapping the human plasma proteome is critical for understanding the interplay between the vascular system and the plasma, and to identify biomarkers and protein profiles correlated with disease. Multiple markers would also greatly increase the accuracy of diagnosis of the nature of the cardio vascular disease. As part of the subproteome focus group, Dr. Vondriska reported on the goal to look for PPP identified proteins that had relevance to cardiovascular function and disease. The 354 proteins related to cardiovascular function could be placed into 8 categories that include markers of inflammation, signaling proteins, growth and differentiation-associated proteins, transcription factors, cytoskeletal, vascular, heart failure and remodeling related, and channel and receptor proteins.

Darin Latimer of Agilix Corporation, Inc. talked about highly multiplexed isobaric i-PROT labels for biomarker validation. Dr. Latimer introduced the i-PROT labeling system, which uses isobaric labels. Compatible with MALDI and ESI, 3D traps, linear traps, FT and LC processes, very high levels of multiplexing are possible without increasing the complexity of the MS spectrum. The technology yields excellent coefficients of variation at 20% for MALDI. An example of a study focusing on biomarker analysis for Rheumatoid Arthritis was discussed.

Gerard Hoehn of the National Institutes of Health discussed about the challenges and opportunities of high-throughput biomarker discovery related to lower abundant serum/plasma proteins. Outlining the importance of a biomarker discovery program, the NIH Biomarker Working Group defines a biomarker as a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or responses to a therapeutic intervention. In addition to supporting stratification of populations, biomarkers are critical to understanding pathogenesis, developing better diagnosis of disease and creating new therapeutics. The workflow described for protein biomarker discovery used both the Ciphregen's ProteinChip SELDI and Bruker's ClinProt systems in a complimentary fashion to yield both high and low molecular weight data. Data from a carotid artery disease study aimed at protein profiling to determine differences in symptomatic and asymptomatic patients were discussed. Future work will use multivariate analysis to determine the use of potential biomarkers in separating the various study groups.

R. James Ludwig of the Indiana Centers for Applied Protein Sciences (INCAPS) talked about INCAPS and the establishment of a resource for proteomics services and technology validation. INCAPS was set up as an innovative center of protein analysis excellence in Indianapolis, supplying critical services in protein analysis, technology validation, high throughput computing and biostatistics. INCAPS is owned by mixture of academic and commercial partners, including Eli Lilly. With both an academic and commercial mission, INCAPS is positioned to supply high technology jobs and growth to the local community. Offering a full range of sample storage, preparation, separation (e.g., DIGE), MS/MS and data analysis capabilities, INSCAPS has several customer projects underway: confir-

mation of variant mutant genes with Indiana University School of Medicine, body fluid proteomics with Eli Lilly, analysis of recombinant plants strain proteome with Dow Agrosiences, and characterization of expressed proteins with Roche.

Barbara Gyure, a patent attorney discussed about designing patent and licensing strategies for proteomics. Given the importance of intellectual property to biotechnology, it is no surprise that there is now a dramatic growth in protein analysis patents, and that much of the growth is coming from the academic research world. The talk presented both historical and practical information on patent claims and strategic patenting, and provided a good overview of the process and complexity of intellectual property protection.

ABBREVIATIONS

2D-GE	= Two-dimensional gel electrophoresis
CIEF-CRPLC	= Capillary isoelectric focusing and reversed phase liquid chromatography
DIGE	= Difference gel electrophoresis
ESI	= Electrospray ionization
HUPO	= Human Proteome Organization
LC-MS/MS	= Liquid chromatography – tandem mass spectrometry
LLC	= Lewis lung carcinoma
MALDI-TOF	= Matrix assisted laser desorption/ionization – Time of Flight
MS	= Mass spectrometry
MudPIT	= Multidimensional protein identification technology
PCR	= Polymerase chain reaction
PPP	= Plasma Proteome Project
RLS	= Resonance light scattering
SELDI-MS	= Surface enhanced laser desorption/ionization – mass spectrometry
TOF	= Time of flight