ABSTRACTS OF THE
THIRD INTERNATIONAL CONFERENCE OF THE HELLENIC PROTEOMICS SOCIETY

From Proteomics Research to Clinical Practice

March 30 - April 1, 2009
Nafplio, Greece

Organizing Committee
Pier Giorgio Righetti, *Milano, Italy*
Michael Fountoulakis, *Basel, Switzerland*
Aris Charonis, *Athens, Greece*
Spyros Garbis, *Athens, Greece*
Despina Sanoudou, *Athens, Greece*
George Tsangaris, *Athens, Greece*
Antonia Vlahou, *Athens, Greece*

Sponsors
Bio-Rad Laboratories S.r.l., *Life Science Group, Milano, Italy*
Aplied Biosystems Gmbh, *Darmstadt, Germany*
Malva SA, *Athens, Greece*
Waters Corporation, *Manchester, UK*
Millipore SAS, *Saint Quentin, Paris, France*
Biosolutions LTD, *Athens, Greece*
Analytical Instruments S.A., *Athens, Greece*
Bruker Daltonik GmbH, *Bremen, Germany*

ORAL PRESENTATIONS

WHAT ROOM FOR ELECTROPHORETIC SEPARATIONS IN MODERN PROTEOMICS?

Thierry. Rabilloud [1,2], Mireille Chevallet [1,2], Cécile Lelong [1,3]
[1] CEA, IRTSV, LBBSI, 38054 GRENOBLE, France., [2] CNRS, UMR 5092, Biochimie et Biophysique des Systèmes Intégrés, Grenoble France, [3] Université Joseph Fourier, Grenoble France

Proteomics can be defined as the large-scale analysis of proteins. Due to the complexity of biological systems, it is necessary to concatenate various separation techniques prior to mass spectrometry. In recent times, the emphasis has been put on the increasing performances of mass spectrometers, and on the separation of peptides by microbore liquid chromatography. Consequently, proteomics strategies using protein electrophoresis seem outfashioned, while they were the core tool at the beginning of proteomics.
We will discuss here the inputs that electrophoresis-based proteomics strategies still have to offer in the wider context of modern proteomics, and especially in terms of quantitative proteomics, in terms of method robustness, in the context of sample complexity and quantitative dynamic range. This discussion will encompass peptide separations, one-dimensional gel electrophoresis and two-dimensional gel electrophoresis.
Beyond the technical aspects, which are still rate-limiting but less and less a bottleneck, a more general discussion will be carried out on the relevance of the various proteomics setups for biology-oriented discovery.

THE PROTEOMINER AND THE FORTYNINERS: SEARCHING FOR GOLD NUGGETS IN THE PROTEOMIC ARENA

P.G. Righetti, Egisto Boschetti
Politecnico di Milano, Department of Chemistry Materials and Chemical Engineering "Giulio Natta", Via Mancinelli 7, Milano 20133, Italy, 2 and CEA-Saclay-DSV, iBiTec-S, 91191 Gif-sur-Yvette, France

In any proteome, a few proteins dominate the landscape and obliterate the signal of the rare ones; most scientists lament that, in proteome analysis, the same set of abundant proteins is seen again and again. A host of pre-fractionation techniques has been described, but all of them are besieged by problems, in that they are based on a "depletion principle", often via immuno-subtraction (.e.g., in sera, by using a set of 6 to 12 antibodies against the most abundant species). Parasitic co-depletion removes thousands of low-abundance proteins, nullifying any attempt at bringing to the limelight the "unseen proteome". A revolutionary approach consists in the "ProteoMiner Technology", a method enabling the capture of all species present in a proteome, but at much reduced protein concentration differences. This consists on a combinatorial library of hexapeptide ligands coupled to spherical porous beads. When 20 different amino acids are used for the synthesis, the library will contain a population of 64 million different linear hexapeptides. Such a vastly heterogeneous population of baits means that an appropriate volume of beads could contain a partner able to interact with all proteins present in a complex proteome. When these beads are contacted with proteomes (e.g., human urines, sera, cerebrospinal fluid, any cell lysate) of widely differing protein composition and relative abundances, they are able to "normalize" the protein population, by sharply reducing the level of the most abundant components while simultaneously enhancing the concentration of the most dilute species. Examples are given of analysis of human urine and sera samples, resulting in the discovery of a host of proteins of potential interest as biomarkers for pathologies. In a red blood cell lysate, where haemoglobin alone constitutes 98% of the total proteins, more than 1500 unique gene products have been found to constitute the remaining 2% proteome. Additionally, these beads can be used to remove host cell proteins from purified recombinant proteins or protein purified from natural sources that are intended for human consumption, a matter of serious concern in the Bio Pharm. industry. These proteins typically reach purities of the order of 98%: higher purities often being prohibitively expensive. Yet, if incubated with "ProteoMiner beads", these last impurities can be effectively removed at a small cost and with minute losses of main, valuable product.

Castagna, A., Cecconi, D., Sennels, L., Rappsilber, J., Guerrier, L., Fortis, F., Boschetti, E., Lomas, L., Righetti, P.G.: *J. Proteome Research* 4 (2005) 1917-1930.
Righetti, P.G.; Boschetti, E.; Lomas L.; Citterio, A. *Proteomics* 6 (2006) 3980-3992.
Boschetti, E., Lomas, L., Righetti, P.G. *J. Chromatogr.* A 2007) 277-290.
Antonioli, P.; Fortis, F.; Guerrier, L.; Rinalducci, S.; Zolla, L.; Righetti, P.G.; Boschetti, E. *Proteomics* 7 (2007) 1624-1633.
Roux-Dalvai F, Gonzalez de Peredo A, Simó C, Guerrier L, Bouyssié D, Zanella A, Citterio A, Burlet-Schiltz O, Boschetti E, Righetti PG, Monsarrat B. *Mol. Cell. Proteomics* 7 (2008) 2254-2269.

FINDING PREECLAMPSIA BIOMARKERS IN CHORION VILLUS BIOPSIES BY MASS SPECTROMETRY.

Coşkun Güzel; Eric A.P. Steegers; Joke A. Polak-Knook; Robert-Jan J.H. Galjaard; Pieter Derkx; Theo M. Luider
Erasmus MC, Rotterdam, the Netherlands

Novel Aspect: The finding of differentially expressed proteins found in first trimester pregnancy chorion villi biopsies and placenta tissue by laser catapulting microdissection technology and mass spectrometry.
Introduction: Although the pathogenesis of preeclampsia remains unclear it is evident that poor invasion of spiral arteries by trophoblast cells leads to ischemic processes in the placenta. Previous studies demonstrate that in early onset preeclampsia placental trophoblastic choriomammotropin, calcyclin, and surfeit locus protein expression is abnormal (De Groot CJM et al. Clin. Proteomics 2007). Enzymatic digested proteins from trophoblasts and stroma cells collected from chorion villus biopsies by laser catapulting microdissection (LCM) obtained during first trimester of pregnancy of women who later in pregnancy developed preeclampsia and controls who remained normotensive were analyzed by MALDI FT mass spectrometry to find early candidate biomarkers. The peptides of interest were fractionated and subsequently sequenced by ion trap mass spectrometry. Antibodies corresponding to exclusively discriminating proteins observed between preeclamptic and control biopsies are validated by immunohistochemistry.
Methods: By means of LCM, cells (equivalent of ~1000 cells) were collected as described previously (De Groot CJM et al., Proteomics 2005) and processed for various mass spectrometry platforms. Corresponding protein lysates were obtained from frozen tissues from 3 chorion villus biopsies of women with pregnancies subsequently complicated by preeclampsia and 8 obtained from controls. Tryptic digests of the protein lysates (approximately 2,500,000 cells) were separated on a Nano-LC (C18) system and sequenced online by ESI Ion Trap mass spectrometry. Protein identifications were searched against SwissProt and were loaded into Scaffold 2 Proteome Software (Proteome Software, Inc., USA).

Results: We focused on two significantly discriminating highly expressed specific proteins, namely ERGIC-53 precursor (LMAN1) and Integrin alpha-6 precursor (ITGA6) which were exclusively found in chorion villus biopsies of women with uncomplicated pregnancies and which were not at all observed in the three chorion villus biopsies of pregnancies subsequently complicated by preeclampsia. In most of the 8 control biopsies 2 and 7 different peptides were found related to ITGA6 and LMAN1, respectively. Totally, approximately 8,000 peptides were sequenced and 5 percent of these peptides corresponded to 120 proteins on average, range 46-163, per chorion biopsy sample (future preeclamptics and controls) according to MASCOT criteria. From these data 79 proteins were common in all biopsies using Scaffold 2 Proteome Software and 10 proteins were observed as black and white differences. In addition to the black and white differences, a series of proteins and peptides were differentially expressed in a quantitative way. We will present mass spectrometry data from LCM experiments combined with identifications obtained by ESI ion trap mass spectrometry.
Discussion : The approach of combining LCM with peptide profiling by means of MALDI FT mass spectrometry followed by identification of proteins by ion trap MS in corresponding tissue lysates gives an opportunity to find proteins that relate to preeclampsia.

CE-MS-BASED URINARY PROTEOMICS AS A POWERFUL TOOL IN CLINICAL DIAGNOSIS, PROGNOSIS, AND THERAPY EVALUATION

Harald Mischak and Mohammed Dakna
Mosaiques Diagnostics & Therapeutics AG, Mellendorfer Str. 7, 30625 Hanover, Germany email: Mischak@mosaiques.de

The reproducible, in-depth analysis of peptides and proteins in body-fluids under different physiological and pathophysiological conditions enables discovery of specific biomarkers, may consequently enhance diagnosis and therapy of specific diseases. Certain proteomic methods can provide means to accomplish this vision. Among the best suited approaches is capillary electrophoresis coupled to time-of-flight mass spectrometers (CE-TOF-MS). The stable, robust and reproducible technology, together with software solutions necessary to evaluate the vast amount of information, enables reproducible analysis of 1000 – 6000 peptides and proteins in a sample within 45-60 min, hence is applicable for biomarkers discovery, and subsequent routine clinical procedures. The advantages of CE are high robustness, reproducibility and resolution, and reduced costs, run-time, and sensitivity towards interfering compounds, making CE an ideal separation technology in proteomics. Urine was chosen as body fluid of interest, since it is available in large quantities and displays the health status of the renal, urogenital and vascular system. Data on plasma, cerebrospinal fluid and bronchial lavage have also been successfully generated. Currently, more than 10000 urine samples from healthy volunteers and patients with a variety of diseases have been analyzed. To obtain higher accuracy, data from CE-FT-ICR and from different sequencing approaches were combined with the routine CE-TOF-MS analyses to result in a high resolution "human urinary proteome map".
The available datasets can be used to define statistically significant biomarkers for distinct disease, based on the comparison of hundreds of individual datasets. While initially focused on biomarkers for renal diseases, applications in other clinically relevant areas like cancer, arteriosclerosis, transplantation, were successful.
All resulting data reveal that the "ideal biomarker", a single peptide that allows definition of a disease with high accuracy, does not exist. However, a panel of distinct and defined single biomarkers with good discriminatory features can be combined to a specific model, that generally show <90% accuracy in the training set. As any approach based on multiple parameters is at risk of overfitting, validation of the biomarker models in a blinded assessment is mandatory. Such validation resulted in generally 80 - 90 % sensitivity and specificity in several clinical studies.
The studies also highlighted the importance of sufficient number of datasets in the definition of biomarkers and biomarker models. Based on the analysis of gender specific biomarkers, we demonstrate that the employment of an undersized sample set will inevitably lead to invalid results, independent of the classification algorithm involved. As a rule of thumb, if application of adequate statistics (corrected for multiple testing) results in the definition of significant biomarkers, machine learning tools will generally enable the establishment of biomarker models that pass validation. The failure to establish relevant biomarker models based on a small number of samples appear due to biological variability of the individual proteins, consequently an inability to correctly define their mean abundance in disease and control.
As shown in several recent studies, CE-MS-based urinary proteome analysis also enables identification of biomarkers and biomarker models for assessment of disease prognosis/progression and therapeutic intervention. The definition of such markers may result in the definition of surrogate endpoints in (pre)clinical trials, allowing evaluation of therapeutic strategies/new drugs (and also potential adverse effects of these) on a small number of patients or even animal models, at early time points.

DEEPER EXPLORATION OF PROTEOME WITH A COMBINATION OF MULTIDIMENSIONAL PEPTIDE SEPARATIONS WITH ACCURATE MASS AND TIME TAGS (AMT) PROTEOMICS APPROACH.

Guryča Vilém, Ducret Axel, Langen Hanno, Norbert Rolland, Masselon Christophe, Garin Jérôme
F.Hoffmann – La Roche, MML (Protein/Metabolite Biomarker Technologies) Grenzacherstrasse 124, 4070 Basel, Switzerland
CEA Grenoble, iRTSV (Laboratoire de Physiologie Cellulaire Végétale) 17 des Martyrs, 38054 Grenoble, France
CEA Grenoble, iRTSV (Laboratoire d'Étude de la Dynamique des Protéomes) 17 des Martyrs, 38054 Grenoble, France

Not only for clinical studies, the combination of liquid chromatography (LC) and tandem mass spectrometry (MS/MS) has evolved into an acknowledged methodology to identify biologically active compounds, such as proteins. However, the widely recognized bottleneck of that approach resides in the insufficient peak capacity of the LC step and the under-sampling caused by the insufficient scanning rate of MS/MS acquisition. Recently, to address these issues, accurate mass and time tags (AMT) approaches [1] have been established, that allow peptide/protein identification through the information extracted exclusively from peptide accurate masses and chromatographic retention times, entirely by-passing the time-consuming MS/MS events during data acquisition. Such approaches have been proved to increases the measurement's dynamic range. However, it is currently unknown whether appropriate pre-fractionation technique employed in conjunction with the AMT tag strategy, could possibly allow to dig even deeper into the proteome.
We will present a first implementation of an integration of multidimensional chromatographic separation with accurate mass and time tags approach and assess the potential of this setup for increasing the dynamic range of analyses i.e. for mining more protein identifications obtained from each sample.
In this study, we used complex tryptic digests of some biologically relevant species (Arabidopsis thaliana chloroplast) as a test sample, In the first dimension, peptides were separated using either strong cation exchange column (Luna, Phenomenex) or Off-Gel isoelectric focusing (Agilent3100). After the first separation step, fractions were analyzed on nano-LC system (Dionex, USA) coupled to a hybrid 7 Tesla Fourier transform ion cyclotron resonance mass spectrometers LTQ-FT (ThermoFisher Scientific, Germany). The AMT database of peptides and assigning of their RT/masses to previously identified peptide sequences were carried out on ICR-2LS/Viper software platform (PNNL, Richland, USA).
We will present a proof-of-principle of the setup. Collected fractions can be directly injected for LC-MS and both SCX and IEF are compatible with AMT identification strategy, as mass/retention data could be successfully aligned with database peptides in all fractions. Fractionation of sample results in higher number of peptide mined from one sample (3867 identifications over 1962 using standard AMT run), and a greater dynamic range (5 orders of magnitude) was achieved. The method was also validated to be reproducible and suitable for quantitative purposes.
1] Ljiljana Paša-Tolić et al. J. Mass Spectrom. (2002) 37: 1185–1198

MS-IMS (MALDI-IMAGING) THE USE OF MASS SPECTROMETRY FOR TISSUE IMAGING

Marc Baumann
Protein Chemistry and Proteomics Unit, Institute of Biomedicine, University of Helsinki, Finland

Using mass spectrometry (MS) for tissue, cell, bacteria and virus imaging will soon make this technology more popular than ever before. In addition, the ability to combine MS with fast micro-chip based diagnostics will transfer this technology even into routine clinical use. MALDI-IMS (and IMS with other type of MS ionization) has several advantages over alternative techniques. With only one single technology we are able to measure the spatial distribution of small molecule drugs, metabolites, peptides, proteins, lipids and several other compounds detectable by MS (*e.g.* label free). Due to the relatively high specificity of MS, we can resolve a compound from its counter ions by simple mass analysis. We may also combine MS-IMS with additional metabolic labeling approaches, although these are not necessary for a successful measurement. MS-IMS is relatively simple to perform and does not require much additional knowledge or equipment. In brief, a sample is mounted to a specific sample slide (glass, metal), covered with a proper matrix and then a two dimensional array of MS spectra acquired. The final data is visualized using any of the commercial or free-ware IMS software. In this presentation I will describe the MS-IMS technology in more detail and also show some applications where MALDI-IMS has been the only technology being able to verify the metabolic condition of the patient.

THE APPLICATION OF HIGH THROUGHPUT MICROCHIP CE-SDS (LABCHIP®) IN PROTEIN SCIENCES

Bahram Fathollahi, Holger Schulz, Cathleen Salomo
Caliper Life Sciences

The demand for high throughput in analytics as well as protein-interaction studies has lead to small scale experiments in microtiter plate formats. Rapid separation of increasing sample numbers can be addressed with microchip CE-SDS (LabChip®). The protein-SDS complexes are stained by an intercalating dye and finally measured via LIF. Staining takes place while proteins are separated on-the-chip. In this presentation, we will show various examples for the use of HT microchip CE-SDS (LabChip) in protein sciences. We have used the automated LabChip system for qualitative and quantitative measurement of protein samples.

PATHWAY ANALYSIS USING REVERSE PROTEIN ARRAYS

Hans Voshol
Novartis Institutes for BioMedical Research, 4002 Basel, Switzerland

In recent years, kinases have arguably become the most important class of drug targets. However, targeting the ATP binding site of kinases carries the inherent risk of ‘promiscuity’ selectivity. While in the past these risks were mostly screened with biochemical assays for individual kinases, pathway readouts in cellular model systems are gaining importance. Capturing the dynamics of cellular signaling pathways requires extensive sampling *e.g.* along the time axis, and quantitation of multiple analytes in those samples. Reverse protein arrays are the ideal analytical tool for such pathway proteomics applications, because they can be readily scaled to provide hundreds of signaling pathway readouts from the same sample. These data will help to elucidate perturbations in those signaling processes, *e.g.* as they occur in disease, and support the generation, experimental validation and refinement of mathematical pathway models.

INTEGRATION OF BIOLOGICAL KNOWLEDGE: FROM AVALANCHE OF DATA TO SYSTEM MODELLING

Ioannis Xenarios
Vital-IT group, Swiss Institute of Bioinformatics, Lausanne, Switzerland

Current biomedical research has produced a multitude of new technologies that allow us to capture and integrate a vast amount of information generated by high throughput methods, such as DNA microarrays, proteomics and bioinformatics and more recently ultra high throughput sequencing. Then along has come the sequenced human genome, and suddenly we have a complete skeleton upon which to integrate the mass of information generated. The scientific community now has an integrated way of looking at what has previously been isolated snippets of knowledge. We have known for some time the function(s) of many proteins in signalling pathways, developmental regulation, cell cycle progression, and so on. However, what is becoming clearer as we gather more information and gaze upon the global picture, is that a single protein rarely performs a single function. Rather, the activity that we assign to it is the product of its interaction with other proteins, small molecules or nucleic acids at any given time. Despite the advance in high throughput technologies (or, perhaps, because of this), we are faced with an avalanche of data but only flakes of knowledge. What we should aim to develop is a system approach that would enable the integration of all the information generated from these technology platforms and develop both mathematical and biological methodologies to test them.

SYSTEMS BIOLOGY: JUST A SLOGAN OR A REALITY?

George Thireos
IMBB-FORTH, Crete, Greece

The deduction of form and function from genomic information is the ultimate biologist's dream. Are we close in fulfilling this dream? Certainly not, but we currently try to accelerate the process. As biology is becoming a quantitative and data rich science it shifts from reductionism to integration and the so called system approaches tend to dominate the scientific methods used. Thus, Systems Biology emerges not exactly as a new independent field, but rather as a unifying concept. Beyond the typical biological technologies, it also applies engineering, mathematical and computer science tools to solve biological problems. Although similar concepts and tools are already in use in biology and medicine to explore the functional integration of large cellular assemblages, such as organs, into systems (digestive, nervous, cardiovascular *etc.*), the novelty is that these tools can now be used to a much smaller scale at the level of individual cells and cell organelle. Each cell type can be viewed as a system composed of a number of subsystems each one having a given set of interacting and interconnected molecular components. Understanding cellular function is now "reduced" to the understanding of the rules that govern the organized complexity of this large number of simple cellular components. This in turn will permit the uncovering of emergent properties of living systems and will allow exact predictions and the design of computer simulations of cellular processes.

PUSHING THE LIMIT IN PROTEOMIC ANALYSES: A NOVEL MASS SPECTROMETER COMBINING ULTRA-HIGH MASS RESOLUTION AND ACCURACY FOR MAXIMUM PROTEIN SEQUENCE COVERAGE

Arnd Ingendoh, Oliver Raether, Dirk Wunderlich, Markus Lubeck, Carsten Baessmann, Marcus Macht.
Bruker Daltonik GmbH, Bremen, Germany

Analyzing complex proteomic samples for protein IDs or quantitation still maintains to be a challenge. Best sequence coverage requires here both high quality MS/MS spectra as well as relatively fast MS/MS acquisition rates. Even with most modern mass spectrometers, there is still a compromise to be made between data quality in terms of mass resolution/accuracy on one hand and speed of analysis on the other. Here, we present the maxis, a novel mass spectrometer with ultrahigh mass resolution and accuracy at still data acquisition rates of B 2-5 Hz. This is achieved as well without compromising other factors like sensitivity or dynamic range. The relevant performance factors above have been systematically investigated for several applications like the identification or label-free quantitation of proteins with complete digested cell lysates as matrix as well as intact protein work For protein ID, several hundreds of proteins could be confidentially identified from minor amounts of, eg. 100 ng complete cell lysate, with false positive rates of < 1%. For label-free proteomic quantitation studies, the high resolution resulted in major benefits for the overall monitoring of the analyzed complex proteome, namely in a highly increased number of readily identified and quantified regulated peptides and proteins. The limit of detection and quantitation as well as the effective dynamic range for quantitation set new standards for this application.

ISOELECTRIC PRE-FRACTIONATION OF PROTEINS AND PEPTIDES WITH SOLID-STATE BUFFERS PRIOR TO EXTENSIVE PROTEOMIC ANALYSIS

Egisto Boschetti[1] and Pier Giorgio Righetti[2]
[1]Bio-Rad Laboratories C/o CEA-Saclay 91191 Gif-sur- Yvette, France; [2]Politecnico di Milano, Department of Chemistry, Milano 20131, Italy

The analysis of complex proteomes is highly dependent on efficient fractionation methods with low level of protein carry over from fraction to fraction. In this respect group separations based on isoelectric points appear quite attractive, although electrically-driven methods involving the presence of carrier ampholytes are not exempt from technical complications. As an alternative approach a new separation concept is presented, involving the use of so-called solidstate buffers associated with ion exchangers to separate proteins of different pI ranges with a low level of protein overlapping. Packed resin blends

in separated columns are used under a cascade configuration of increasing or decreasing pH and, once proteins of different pI are adsorbed by individual columns, elutions will be performed separately. From each sectional column protein mixtures corresponding to a given pI range are collected by competitive desorption so as to be ready for proteomic analysis. The presence of small amounts of potassium chloride during the separation process prevents protein precipitation at the vicinity of their respective isoelectric points. The fractions thus obtained can be used for two dimensional electrophoresis and mass spectrometry analysis after the removal of salts. Real-cases of serum protein fractionation will be shown and the analytical data will be discussed.

AN INTEGRATED WORKFLOW SOLUTION FOR OPTIMISED DISCOVERY AND TARGETED PROTEOMICS RESEARCH.

Mark A. McDowall
Waters MS Technology Centre, Manchester, M22 5PP, United Kingdom.

Biomarker Research and Development

Biomarker R&D is one of the major applications of proteomics. The research workflow conventionally begins with a "global discovery" study (e.g. high resolution LC/MS/MS) of a limited number of samples to identify a range of candidate proteins. "Targeted proteomics" techniques (e.g. tandem quadrupole LC/MRM) are subsequently employed to efficiently verify these candidates within a biologically diverse cohort. The verification phase is designed to yield a smaller panel of potential biomarkers. Biomarker validation studies (e.g. immunoassay) in a large scale clinical trial may presage a new prognostic/diagnostic test. The time-line from discovery through verification to validation therefore progressively involves (i) a decreasing number of protein targets and (ii) an increasing number of samples. Thus an optimised integration of discovery and targeted proteomics methodologies is desirable for efficient biomarker R&D.

Bottom-Up Proteomics (Discovery) Technology

We have developed a high performance benchtop Q-Tof mass spectrometer optimised for discovery proteomics at very high sensitivity (Waters Xevo QTof MS). The system uniquely allows clinical researchers to focus on their results and not the process of acquiring them. There is a growing consensus that exact mass MS brings decisive advantage to discovery proteomics but, historically, at a price – complexity. Our solution is "engineered simplicity" that redefines the LC/MS workflow from system set-up, to method development, to performance verification, to optimized analysis and biochemically intelligent data annotation. The Xevo QTof uniquely, in a benchtop system, enables UPLC/MSE (high bandwidth) data acquisition – allowing the maximum information to be reproducibly extracted from the minimum sample in the shortest time. Conventional LC/MS/MS analysis of complex protein digests typically requires multiple injections to additively compile an exhaustive inventory of peptides. UPLC/MSE, in contrast, acquires all the data from a single injection.

Targeted Proteomics (Verification) Technology

We have developed a Tandem Quadrupole (TQ) MS optimised for targeted proteomics (Waters XevoTM TQ MS) with novel ScanWaveTM technology enabling high sensitivity MRM analyses and ultra fast (5,000 Th/s) high sensitivity full product ion scans to be acquired within the same analysis. MRM is widely acknowledged as the most sensitive and quantitatively linear LC/MS/MS technique for high throughput target peptide analysis. However, conventional TQ MS detectors are relatively insensitive in the product ion scan (MS/MS) mode frustrating the ability to concomitantly quantify (MRM) and sensitively confirm the identity (product ion scan) of targeted peptides - Waters ScanWave innovation overcomes this limitation. The Xevo QTof MS shares 80% of its analytical path (i.e. UPLC, ion source, MS1, collision cell, MS2) with its sibling the Xevo TQ MS - making LC/MS methods uniquely portable from one to the other. This complementary design is further leveraged by Waters Verify E methods development engine that expertly selects non-redundant proteotypic peptides with experimentally justified MRM transitions by interrogating comprehensive discovery (UPLC/MSE) data in contrast to commonly used in silico projections of proteotypic peptide suitability. Consequently the Xevo QTof MS and Xevo TQ MS together form the basis of a uniquely optimized system solution for biomarker discovery-development.

MONITORING OF ANTI-CANCER THERAPIES AND CHEMORESISTANCE

J. Martinkova[1,4], R. Hrabakova[1], H. Skalnikova[1,4], P. Novak[2], P. Dzubak[3], M. Hajduch[3], S.J. Gadher[1,4], H. Kovarova[1,4]
[1]Institute of Animal Physiology and Genetics AS CR,v.v.i., Libechov; [2]Institute of Microbiology AS CR,v.v.i., Prague; [3]Palacky University and Faculty Hospital, Olomouc; [4]Joint Proteome Laboratory, Prague, Czech Republic

Clinical importance of biomarkers as a 'handle' for monitoring therapeutic response is of growing importance. Our study is focused on monitoring molecular mechanisms of response to anti-cancer drugs (i) at earlier time intervals to detect relevant proteins that are responsible for primary changes in signalling networks that subsequently lead to irreversible anti-cancer processes and (ii) development of chemoresistance toward molecularly targeted inhibitors of cyclin dependent kinases. Protein profiles of CEM Tlymphoblastic leukemia cells treated by typical representatives of conventional chemotherapy at half-time to apoptosis were evaluated and classified using multivariate analyses. Only about 2% of total variances represented differences that enabled grouping of relevant anti-cancer drugs. This aspect of the study was extended to monitor development of chemoresistance and profiling of some proteins such as STIP1 or YB1 that can contribute to resistance toward different drugs. The potential diagnostic and / or prognostic values of such proteins deserve further investigation. We acknowledge support of grants from the Czech Ministry of School and Education (MSM6198959216 and LC07017) and Institutional Research Concept AV0Z50450515.

MOVING PROTEIN CANDIDATES THROUGH THE BIOMARKER RESEARCH PIPELINE: FROM DISCOVERY TO VALIDATION USING TARGETED MASS SPECTROMETRY WORKFLOWS

Cathy Lane
Applied Biosystems

Many proteomics laboratories world-wide are involved in the search for proteins and peptides that can be used as disease biomarkers. A large number of candidate biomarkers have been discovered, however, very few of these have been validated. Not a single marker candidate discovered by proteomics approaches has yet been approved by the regulatory agencies. Biomarker discovery and verification/validation are two distinct workflows. During the discovery phase, a relatively small number of samples with a high number of potential biomarker candidates are screened. Once biomarker candidates have been identified with initial statistical significance, these have to be validated. This validation workflow involves analyzing a large number of samples with a relatively small number of candidates to establish the biological significance of the biomarker candidates. The main reason for the paucity of validated biomarkers from proteomics research is the unsuccessful transition from discovery to validation. Rather than switching to immunological techniques for the validation step, we suggest a mass spectrometry-based approach. This orthogonal strategy is a novel targeted, high throughput quantitative multiplexed multiple reaction monitoring (MRM) approach utilizing mTRAQ™ reagents. The approach relies on assay development using a combination of MRMs to target specific peptides identified in discovery, followed by MS/MS to confirm that the quantitative MRM signal results from the target peptide. Examples of published and ongoing studies will be presented.

PROTEIN PROFILING OF HUMAN FOLLICULAR FLUID: QUEST FOR BIOMARKERS OF OVARIAN HYPERSTIMULATION SYNDROM

J. Martinkova[1], L. Jelinkova[1], K. Jarkovska[1], P. Halada[2], K. Rezabek[3], J. Moos[4], H. Kovarova[1], S.J. Gadher[1]
[1]Institute of Animal Physiology and Genetics AS CR,v.v.i., Libechov; [2]Institute of Microbiology AS CR,v.v.i., Prague;; [3]Department of Obstetrics and Gynecology, General Teaching Hospital, Prague; [4]Sigma-Aldrich, spol.s r.o., Prague, Czech Republic

Ovarian hyperstimulation syndrome (OHSS) is a potentially serious complication of fertility treatment, particularly of *in vitro* fertilisation (IVF) treatment. Overstimulated ovaries enlarge and release chemicals into the bloodstream that make blood vessels leak fluid into the body. Fluid leaks into the abdomen and, in severe cases, into the space around the heart and lungs. OHSS can also affect the kidneys, liver and lungs. Human follicular fluid (HFF) offers new possibilities of finding proteins as potential markers of oocyte quality for improvement of selective processes during *in vitro* fertilization (IVF) therapy. The aim of this study was to determine protein profiles in paired HFF and plasma samples from women undergoing IVF and those developing ovarian hyperstimulatin syndrome (OHSS). To access proteins of lower abundance,

Immuno-affinity depletion or ProteoMiner technology was used. Enriched protein pools were further separated either by 2-DE or gel-free 2D HPLC fractionation and resulting protein maps of HFF *versus* plasma or *versus* OHSS positive HFF were evaluated. The higher level of clusterin in HFF compared to plasma was verified by immunoblot, an observation indicating either a selective transport into HFF or local synthesis in follicles. Furthermore, we observed a decreased activity of complement cascade in HFF than in plasma. The proteins or their fragments which increased in OHSS positive patients included pigment epithelium-derived factor, complement factor I precursor, complement C3 precursor, ceruloplasmin precursor, inter-alpha-trypsin inhibitor heavy chain H4 precursor, and apolipoprotein A-IV precursor. The potential diagnostic and prognostic values of those proteins deserve further studies. Supported by grants 1QS500450568 (GAAV) and by Institutional Research Concepts AV0Z50200510 (IMIC) and AV0Z50450515 (IAPG).

ADVANCED TECHNIQUES FOR CHARACTERIZING POST-TRANSLATIONAL MODIFICATIONS ON PEPTIDES AND PROTEINS USING ELECTRON TRANSFER DISSOCIATION

A. Makarov, M. Zeller, M. Scigelova
Thermo Fisher Scientific, Hanna-Kunath-Str 11, Bremen, Germany

The analysis of post-translational modifications remains a challenge, especially when one considers the importance of their mutual cooperativity as demonstrated, for example, in the case of histones. The mass spectrometry approaches often used for peptide and protein characterization have benefited greatly from the modern instrumentation delivering a robust mass accuracy and very high resolution. In addition, a novel fragmentation technique called electron transfer dissociation (ETD) was introduced a few years ago and proved to be very useful for addressing the questions regarding 'fragile' post-translational modifications such as phosphorylation or glycosylation/glycation. The ETD technique has been recently implemented on the Orbitrap mass analyzer. We present a set of examples demonstrating the capabilities and utility of this methodology for the post-translational modification analysis on both peptides and intact proteins. A special mention will be given to large-scale proteome/phosphoproteome analyses in the context of intelligent data acquisition. At the same time, there has been a considerable increase in the data structure complexity. The data files can be a combination of various fragmentation methods, energy regimes, and mass accuracy/resolution characteristics of the detectors. We will describe the aspects of data processing addressing the main issues.

BIOMARKER DISCOVERY IN LOW-GRADE BREAST CANCER TISSUE BIOPSIES USING A NOVEL QUANTITATIVE PROTEOMIC APPROACH

Theodoros Roumeliotis[1], Pavel Bouchal[2,3], Roman Hrstka[2], Rudolf Nenutil[2], Borivoj Vojtesek[2] and Spiros D. Garbis[1*]
[1]Biomedical Research Foundation, Academy of Athens, Greece; [2]Department of Oncological and Experimental Pathology, Masaryk Memorial Cancer Institute, Brno, Czech Republic; [3]Department of Biochemistry, Faculty of Science, Masaryk University, Brno, Czech Republic

The present pilot study constituted a proof-of-principle in the use of a quantitative LC-MS/MS based proteomic method for the comparative analysis between representative low-grade breast primary tumor tissues with and without metastases, and between the lymph node metastases and the nonmetastatic tumor type. The study method incorporated stable isotope labeling, two-dimensional liquid chromatography, nanoelectrospray ionization and high resolution tandem mass spectrometry using the hybrid QqTOF platform. The principal aims of this study were (1) to define the protein spectrum obtainable using this approach, and (2) to highlight potential candidates for verification and validation studies focused on biomarkers involved in the metastatic processes in breast cancer. The study resulted in the reproducible identification of 605 nonredundant proteins ($p<0.05$). A quantitative comparison revealed 3/3 proteins with significantly increased/decreased level in metastatic primary tumor and 13/6 proteins with increased/decreased level in lymph node metastasis compared to nonmetastatic primary tumor ($p<0.01$). Changes in selected differentially expressed proteins were verified with qRT-PCR. The main conclusion from this pilot study was that the quantitative proteomic method used constitutes a novel way of analyzing cancerous breast tissue biopsy samples that can be extended as part of a larger scale biomarker discovery program.

URINE PROTEOME ANALYSIS BY TWODIMENSIONAL GEL-BASED PROTEIN SEPARATION COUPLED WITH NLC-ESI-MS/MS

Michalis Aivaliotis, Jerome Zoidakis, Manousos Makridakis, Antonia Vlahou
Laboratory of Biotechnology, Center of Basic Research II, Biomedical Research Foundation, Academy of Athens, Greece

Three steps are especially important for the comprehensive analysis of urine proteome: protein concentration from urine with minimal loss; protein separation to reduce sample complexity; and peptide sequencing with high mass accuracy and rapid scanning. In the present study, we employed a two-dimensional gel-based approach coupled with nLC-ESI-MS/MS for the analysis of the urine proteome. The "standard" urine sample collected in Mosaiques Diagnostics within the European Urine and Kidney Proteome (EuroKUP) consortium was employed to facilitate data comparison and evaluation. In the first dimension, proteins were separated using preparative polyacrylamide gel electrophoresis, into 5 mass groups:MW<15kDa, 15<MW<25, 25<MW<37, 37<MW<50, MW>50. Subsequently, as a second dimension of protein separation, each group was subjected to 1D SDS-PAGE optimized for the separation of each of these molecular mass ranges using different acrylamide concentrations. The resulted gel lanes were cut out in slices, digested with trypsin and analyzed by nLC-ESI-MS/MS for protein identification. Identified proteins were also analyzed for the presence of various post-translational modifications including phosphorylation, oxidation, deamidation, methylation and acetylation. The results were compared to the nLC-MS/MS data received from the same urine sample analyzed by different proteomic platforms, within the EuroKUP consortium, as well as urine proteome datasets already published by other groups.A plethora of proteins was identified, many of which were found to be highly modified. These identified proteins in part overlapped, but also in large part complemented, available urine proteome data. In conclusion, we present a comprehensive approach for the analysis of urine proteome which can be particularly employed for the purification and the in-depth analysis of selected mass ranges for the detection of low abundance disease biomarkers and their possible PTMs.

Supported by FP7 DECanBio (project number 201333).

PROTEOMIC ANALYSIS OF AMNIOTIC FLUID IN PREGNANCIES WITH KLINEFELTER SYNDROME FETUSES

Athanasios K. Anagnostopoulos[1,2], Aggeliki Kolialexi[1], Ariadni Mavrou[1], Konstantinos Vougas[2], Nikos Papantoniou[3], Aris Antsaklis[3], Michael Fountoulakis[2] and George Th. Tsangaris[2]
[1] Medical Genetics, Athens University School of Medicine, Athens; [2]Proteomics Research Unit, Centre of Basic Research II, Biomedical Research Foundation, Academy of Athens, Athens; [3]1st Department of Obstetrics and Gynaecology, Athens University School of Medicine, Athens, Greece

Klinefelter syndrome is a sex chromosomal abnormality in males with an extra X chromosome (47, XXY caryotype), occurring approximately in 1 in 1000 male live births. In the present study proteomic analysis was performed in twelve 2nd trimester amniotic fluid samples, eight coming from pregnancies with normal males and four with Klinefelter syndrome fetuses, as shown by routine cytogenetic analysis. Samples were analysed by 2-DE, coupled with MALDITOF-MS analysis. 850 μg of total protein from each AFS sample was applied on 18 cm immobilized pH 3-10 and 17 cm, pH 4-7 non-linear gradient IPG strips and first dimensional electrophoresis focusing was performed approximately for 156,000Vh. Second dimensional electrophoresis was performed in 12% SDS polyacrylamide gels (180 mm × 200 mm × 1.5 mm) and all gels were stained with Colloidal Coomassie Blue solution. Gel images were scanned in a GS-800 Calibrated Densitometer and 2- DE image analysis was performed with the PD-Quest 8.0 image processing software. From each gel approximately 150 protein spots were excised and further identified with a MALDI-TOF-MS instrument. Image analysis and gel comparison of control samples against amniotic fluid supernatants coming from pregnancies with Klinefelter syndrome revealed that three proteins (CER, A1AT and ZA2G) were found to be up-regulated in samples obtained from pregnancies with 47, XXY caryotype fetuses, whereas four proteins (APOA1, RETBP, GELS, and VTDB) were down regulated. Further quantification for RETBP in both Klinefelter’s and normal male fetuses was performed by Western blot analysis. Optical density measurements of the bands revealed that the amount of RETBP in AFS coming from women with Klinefelter s. fetuses was decreased by 43% as compared to normal controls.

MANDATORY REQUIREMENTS IN BIOMARKER IDENTIFICATION

Harald Mischak
Medizinische Hochschule Hannover and Mosaiques Diagnostics & Therapeutics AG, Hanover, Germany

The field of biomarker discovery in clinical proteomics is suffering from high hopes generated by reports on potential biomarkers, which in most cases subsequently could not be substantiated via validation. As a result, the field is loosing credibility. Such development can be corrected by avoiding a few major mistakes. Among the most frequent mistakes we find: 1) Unknown or poor platform performance and experimental data, resulting in an inability to give any level of confidence in the claimed biomarkers. 2) Unknown or poor sample quality and lack of relevant clinical data. 3) Statistics applied is not appropriate, no adjustment for multiple testing is done. 4) The results are not confirmed in an independent test set, and experimental details are described in a way that for sure nobody else can reproduce/validate them. Attempts to reproduce results from such flawed reports generally fail, hence such reports just appear to prove clinical proteomics to be worthless, especially in the eyes of clinicians. However, such mistakes and erroneous reports can generally be eliminated by 1) Employing a well characterized analytical platform (adequately addressing repeatability and variability for pre-analytical and analytical procedures) 2) using adequate samples, controls, and statistics and 3) validate the findings on an independent test set of appropriate size to prove significance. Reports that are not in line with these simple requirements have an excellent chance of being irrelevant. Any report on “potential biomarkers that await further validation” can generally be viewed as insignificant. To improve on the current state, clear and easy to follow rules are required on how to define and validate biomarkers. I suggest mandatory requirements for the disclosure of platform validation (repeatability, precision, limit of detection, resolution, mass accuracy), appropriate statistical evaluation (adjustment for multiple testing), and validation of the findings on an independent blinded set that is large enough to show statistical significance (e.g. p-value < 0,05 in ROC analysis). Every clinical proteomics study in the area of biomarker discovery should clearly address these 3 points in any report, to enable assessment of validity.

FDA-REQUIREMENTS FOR BIOMARKER VALIDATION

Jadranka Koehn and Kurt Krapfenbauer
Department of Cranio-Maxillofacial and Oral Surgery, Medical University of Vienna, Austria

Biomarkers are important in diagnosis, medical therapy and drug discovery. Surrogate endpoint biomarkers help to determine a clinical outcome, reducing observation periods and costs of clinical trials, while stratification biomarkers can be used to select patient subpopulations likely to respond to a specific therapy. Therapeutic biomarkers on the other hand can determine the patient’s response to a treatment regime at an early time point, opening the possibility to stop an unsuccessful treatment. In diagnosis biomarkers can be used to detect a disease at or even before the manifestation of the first clinical symptoms, opening a path to preventive medicine. To be used for either of these purposes, a biomarker and the associated test has to fulfill several requirements and must not only be generally accepted within the medical profession but also by the regulatory authorities. The FDA has laid down stringent requirements for acceptance of biomarkers and biomarker-tests, and similar regulations will be implemented in the EU. The criteria and strategies for the establishment of a biomarker will be presented.

PROFI: THE NEW PROTEOMICS FACILITY AT IMBB

T. Economou, G. Garinis, T. Loukeris, C.Splianakis and K.Tokatlidis
IMBB, Iraklio, Crete, Greece

The Proteomics Facility at the Institute of Molecular Biology and Biotechnology (ProFI) was recently established thanks to a competitive grant from Capacities, one of the research funding schemes under the FP7 programme of the European Union. The idea behind ProFI is to equip a state of the art proteomics lab at the IMBB and to recruit experienced post-docs and technicians to participate in setting it up. To spark off integration of proteomics approaches with the existing research tools and to establish successful experimental flows, ProFI has begun its operation by engaging in four general areas of research: protein trafficking, aging, immunology and the malaria-mosquito interaction. These topics reflect the interests of five IMBB research groups that are involved in the ProFI project. We intend to develop protocols for bottom-up shotgun analysis, intact protein analysis, interaction proteomics, study of post-translational modifications, membrane proteomics, labelled and label-free quantification. Our flow will be mainly LC-MS-based using a high resolution mass spectrometer but 2DE tools will also be available. To ensure optimal integration of proteomics tools at IMBB, ProFI obtains advice from a committee of mentors, established leaders in proteomics research from EU lab

PROTEOMIC ANALYSIS OF A CASE OF CHILDHOOD ACUTE MYELOID LEUKEMIA

M. Braoudaki[1], Ch. Papathanassiou[2], K. Karamolegou[2], A.K. Anagnostopoulos[2], K. Vougas[3], G.Th. Tsangaris[3] and F. Tzortzatou-Stathopoulou[1,2]
[1]University Research Institute for the Study and Treatment of Childhood and Malignant Diseases, University of Athens, "Aghia Sophia" Children's Hospital, Athens; [2]Hematology/Oncology Unit, First Department of Pediatrics, University of Athens, "Aghia Sophia" Children's Hospital, Athens; [3]Proteomics Research Unit, Division of Biotechnology, Center of Basic Research II, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

Childhood acute myeloid leukemia (AML) is a frequent hematological malignancy that shows a high degree of heterogeneity in response to therapy. Given the diverse nature of AML and the inherent variability in individual protein levels, it seems likely that the best approach to screen for AML is to determine the protein profile using proteomic technology. In the current investigation we studied a child with myelodysplastic syndrome (MDS) before and during blast crisis in an attempt to identify proteins that may serve as useful predictive biomarkers, or whose function could regulate the phenotypic characteristics of the disease. Bone marrow (BM) and peripheral blood (PB) samples were obtained before diagnosis and at diagnosis of AML before the initiation of treatment for proteomic study. Follow-up BM and PB samples three months after chemotherapy were also analysed. Proteomic analysis was performed on BM and PB plasma using twodimensional gel electrophoresis (2-DE) of overlapping pH ranges from 3 to 10 and matrix-assisted laser desorption and ionization time-of-flight (MALDI-TOF) mass spectrometry. As control, BM and PB plasma samples from a nonleukemic pediatric patient was studied. Comparison of BM and PB plasma samples from the AML patient with the control revealed differences in various molecular mass ranges with different isoelectric points. Following evaluation of the gels, expression in matched BM and PB samples was 55 similar for 41 proteins. Eleven proteins were found to be expressed only in BM samples, whilst six were expressed only in PB samples. The majority of the expressed proteins were metabolic enzymes, structural proteins, signal transduction mediators and immunoglobulins. Proteome analysis of plasma seems to offer a useful approach for profiling pediatric patients with AML. Additionally, the current technology provides significant insight into the presence and/or absence of several proteins that may serve as useful biomarkers for AML. However, further analyses are warranted to elucidate the role of these findings.

SEARCHING FOR ALLERGENS IN MAIZE CYTOPLASMIC PROTEINS:A PROTEOMIC STUDY

E. Fasoli[1], E.A. Pastorello[2], L. Farioli[2], G. Aldini[3], E. Boschetti[4], P.G. Righetti[1]
[1]Politecnico di Milano, Department of Chemistry, Via Mancinelli 7, Milano 20131; [2]Allergy Centre, Niguarda Ca' Granda Hospital, Milano, Italy; [3]Facolta di Farmacia, Universita degli Studi di Milano, Milan, Italy; [4]Bio-Rad Laboratories, C/o CEA-Saclay 91181 Gif-sur- Yvette, France

In the past decade molecular allergology has greatly contributed to our understanding of IgE-mediated allergic reactions. The symptoms of food allergy are quite diverse and can range from mild itching affecting the oral mucosa to skin symptoms such as urticaria, angioedema or exacerbation of atopic eczema, respiratory reactions, and in rare cases even to anaphylaxis. Allergic conditions represent a multifactorial disease, not only with respect to the individuals (bias to atopy and the various symptoms along with it), but also with regard to the heterogeneity of allergenic sources. Additionally, cross-reactivity between homologous proteins present in related or phylogenetically distant allergenic sources can exacerbate the progression of allergic diseases and complicate the identification of the primary cause of sensitization. Although much less common, investigation of food allergy to maize appears attractive, as cross-reactivity to soy, peanut and rice has been demonstrated at least *in vitro*. Moreover, cases of anaphylactic reactions to maize have been reported in the literature. However, to date, only a 9-kDa lipid transfer protein (LTP) has been identified as a major allergen in terms of food allergy to maize. In order to obtain a comprehensive view of all possible allergens present in the cytoplasmic fraction of maize, 10 g of flour (from the pure genetic line B73) have been first delipidated with chloroform/methanol (70%/30%), followed by solubilization in physiological saline (PBS). Protein was next precipitated in a saturated solution of ammonium sulphate, so as to eliminate interfering starch. The precipitate was solubilized again in PBS and subjected to ProteoMiner treatment, so as to enhance the visibility of low-abundance proteins. By 2D mapping, >300 protein spots could be revealed *via* Coomassie blue staining in the pI 3-10 range and covering the 10 to 150 kDa Mr range. Seven replicas of 2D maps were blotted on nitrocellulose sheets and subjected to allergen detection *via* incubation with sera of

different allergic patients. In order to reveal potential allergens, the blots were further reacted with anti IgE antisera containing radio-iodinated antibodies. The blots were then subjected to autoradiography for a week. By overimposing stained 2D gels with the various autoradiographies, six allergens, common to all patients being analyzed, were identified, covering the 10 to 65 kDa interval. Spots were cut out (in triplicate) and subjected to mass spectrometry analysis *via* nano-LC-MS with an Orbitrap instrument. The allergens presently identified are: vicilin (Mr=66162, pI=6.22, coverage 37%), globulin-2 (Mr=49892, pI=6.16, coverage 51%), r-RNA N-glycosidase; Mr=33096; pI=5.78; coverage: 30%; proteasome subunit beta (Mr=23073; pI=5.30; coverage: 25.5%), trypsin inhibitor (Mr=13550, pI=6.65, coverage 67%) and thioredoxin (Mr 13039, pI 6.22, coverage 21%). This is the first example in which such allergens are identified in an unambiguous manner in *zea mais*. Supported by a grant from the European Community (Allergy Card).

MAPPING THE CHICKEN EGG YOLK PROTEOME VIA COMBINATORIAL LIGAND LIBRARIES

A. Farinazzo[1], E. Fasoli[1], L. Guerrier[2], E. Boschetti[2], A. Bachi[3], A. Citterio[1], P.G. Righetti[1]
[1]Politecnico di Milano, Department of Chemistry, Via Mancinelli 7, Milano 20131, Italy; [2]Bio Rad Laboratories, C/o CEA-Saclay, 91181 Gif-sur- Yvette, France; [3]San Raffaele Scientific Institute, 20132 Milano, Italy

Recent proteomic analysis (1) has identified a total of 119 unique proteins in the plasmatic and granular fractions of egg yolk. Although this is an outstanding number, considering that only ca. 30 proteins had been reported previously, it still might be far from representing the total asset of yolk proteome, in view of the fact that a handful of proteins constitute ca. 90% of the total proteinaceous species. We have thus re-investigated chicken yolk by exploiting a technique recently developed by our group, namely a capture *via* baits comprising a vastly diversified population of hexapeptides synthesized *via* combinatorial chemistry (2-7). To this aim, we have used two combinatorial ligand libraries (CLL): an amino-terminus and a carboxyl-terminus. After removal of lipids (which constitute 33% of yolk components), the plasma fraction (the granular fraction, not amenable to CLL treatment, due to its scarce solubility in physiological conditions, having been discarded) has been adsorbed sequentially first onto the amino-terminus and then onto the carboxyl-terminus libraries. Each column has then been eluted individually *via* 3 eluants: TUC (7 M urea, 2 M thiourea, 2% CHAPS); UCA (9 M urea, 50 mM citric acid to pH 3.3 and 2% CHAPS) and a hydro-organic solution. Each eluate has been analyzed *via* SDS-PAGE and 2D mapping. In SDS-PAGE, many new bands are visible in the 10 kDa to 200 kDa region, as compared to controls. Also in 2D mapping, at least three times as many spots could be seen in the pI/Mr plane as compared with untreated plasma. Currently, all six eluates are in the process of being identified by cutting 20 slices from each individual track in SDSPAGE, digesting all proteins in each slice *via* trypsin, recovering and purifying the peptides and subjecting them to analysis *via* nanoLC-MS/MS (LTQ-Orbitrap mass spectrometer). Although the analysis is still in progress, the total count in unique gene products identified is considerably higher than all published data so far, suggesting that the CLL technique is a unique tool for exploring to an unprecedented depth the deep proteome and discovering those low- to verylow abundance proteins that up to the present have escaped detection by any possible means. 1 Mann K and Mann M: Proteomics 8: 178, 2008. 2 Righetti PG, Boschetti E and Lomas L: A Citterio: Proteomics 6: 3980, 2006. 3 Righetti PG and Boschetti E: FEBS J 274 (2007) 897. 4 Boschetti E, Lomas L, Citterio A and Righetti PG: J Chromatogr. A 1513: 277, 2007. 5 Boschetti E, Monsarrat B and Righetti PG: Current Proteomics 4: 198, 2007. 6 Guerrier L, Righetti PG and Boschetti E: Nature Protocols 3: 883-890, 2008. 7 Boschetti E and Righetti PG: BioTechniques 44: 663, 2008.

DEVELOPMENT OF A STRATEGY FOR THE ANALYSIS OF METAL-BINDING PROTEINS IN URINE

Jerome Zoidakis, Michalis Aivaliotis, Antonia Vlahou
Laboratory of Biotechnology, Center of Basic Research II, Biomedical Research Foundation, Academy of Athens, Greece

The existence of several thousand proteins in urine with concentrations spanning multiple orders of magnitude hinders comparative studies of the urinary proteome. To reduce sample complexity and at the same time enrich for metal binding proteins, with the present study we describe a fractionation strategy for urinary proteins based on the use of immobilized metal chromatography. In brief, urine samples were concentrated by ultrafiltration and protein extracts loaded on Ni^{+2} IMAC columns. Elution of bound proteins was carried out by the use of imidizole. The eluted fractions were subjected to dialysis and subsequently analyzed by 1D and 2D electrophoresis. The resulting protein bands and spots respectively, were identified by LC-MS/MS and subjected to analysis according to Gene Ontology (GO). Multiple metal-binding proteins were identified within the eluted fractions which could not be resolved in the unfractionated starting material. This approach is currently in use for the comparative analysis of metal binding proteins from urine samples from bladder cancer patients and controls. Supported by FP7 DECanBio (project number 201333)

PROTEOMIC MS-SPECTRA DECOMPOSITION INTO INTENSITY–REGIONS FOR IDENTIFYING POTENTIAL BIOMARKERS AND IMPROVING DISCRIMINATION ACCURACY BETWEEN NORMAL AND CANCER TISSUE SPECTRA

Panagiotis Bougioukos1, Dimitris Glotsos2, Dionisis Cavouras2, Ioannis Kalatzis2, George Nikiforidis1 and Anastasios Bezerianos1
1Department of Medical Physics, School of Medicine, University of Patras; 2Medical Signal and Image Processing Lab, Department of Medical Instruments Technology, Technological Educational Institute of Athens, Greece

Aim: In this study a new strategy for processing proteomic MS-spectra is presented for (a) the determination of potential meaningful cancer biomarkers (m/z values), extracted from different MS-spectra intensity regions, and (b) reliable and effective separation of normal from cancer tissue MS-spectra. Method: The method starts by MS-spectra signal conditioning (base line subtraction- normalizationsmoothing- noise estimation–peak detection–peak alignment). It then continues by automatically breaking down all MSspectra into common-equidistant intensity regions. Subsequently, most informative features (m/z values), which might constitute potential significant biomarkers, are determined at each intensity region by using a pattern recognition system for discriminating normal from cancer tissue MS-spectra. Finally, considering selected features from all spectral intensity regions, spectra were classified using a multi-classifier scheme, with members the Support Vector Machine, the Probabilistic Neural Network, and the k-Nearest Neighbour classifiers. To ensure robust and reliable estimates to unseen data, the proposed pattern recognition system was evaluated using an external cross validation process. Material: The proposed method was evaluated on two publicly available proteomic datasets, one with ovarian and the other with prostate MS-spectra. Both datasets were analyzed excluding m/z values lower than 1500, which are potentially distorted, in an effort to analyze more accurately the proteomic datasets. Results: The average overall performance of the system in classifying normal from ovarian cancer MS-spectra was 97.2% (employing 22/24/18/17/15 biomarkers at each intensity region, ranked from highest to lowest intensity regions), whereas the accuracy in discriminating spectra with no evidence of prostate disease (PSA<1) from prostate cancer spectra (PSA >4) was 92.5% (employing 8/4/15/12/13 biomarkers at each intensity region, ranked from highest to lowest intensity regions). Conclusion: The proposed method differs from others in two key issues. (a) The methodology, where the concept of focusing interest on peaks from different MSspectra intensity-regions was introduced, and (b) the accuracy, where, as compared to previous studies that experimented with m/z values above 1500, the proposed system presented the highest classification accuracies. Additionally, a pattern recognition system is proposed that might be of value in the discrimination of normal from cancer tissue MS-spectra.

AN INTERACTIVE TOOL FOR VISUALIZING JOINTLY PROPERTIES OF PROTEOMIC OBJECTS

Eugenia G. Giannopoulou[1], George Lepouras[1], Elias S. Manolakos[2]
[1]Department of Computer Science and Technology, University of Peloponnese, Tripolis; [2]Department of Informatics and Telecommunications, University of Athens, Greece

Proteomics analysis using high throughput technologies provide researchers with an unprecedented amount of heterogeneous protein- or peptide- related features that need to be jointly interpreted. Dealing with large lists of proteins that have been differentially expressed, identified and possibly undergone metadata analysis using different tools or databases, is a time-consuming task that requires excessive effort from the user who tries to discover interesting relations suggested by the analysis results. We have developed of an interactive visualization tool (ViP) that combines visually protein or peptide features into synthetic Proteomic Feature Maps (PFMs) resembling 2-DE images. In particular, ViP helps the user integrate features from different steps of a proteomics analysis workflow and offers the flexibility to explore the results under different pointsof- view. We represent proteomics objects (*i.e.*, proteins or peptides) as spheres and jointly visualize proteomic object features by exploiting the sphere's size, color or label. Using ViP, the user can visualize differential expression features (*i.e.*, abundance ratios) and obtain a quick and overall visual impression of the dataset regulation trend, or the identification information (*i.e.*, identification score), allowing the location of unidentified or highly identified proteins that possibly need further analysis. Additionally, one can combine abundance ratios with Gene Ontology and associate the protein function with differential expression, or load on a map the results of a clustering (class labels) and discover effortlessly interesting groups of proteins. ViP allows combining information from any proteomics step (*e.g.*, multivariate statistics, pathways) and facilitates exploring the results from multiple alternative perspectives

based on the needs of the study. The distinguishing feature of ViP is the integration of proteomic features which enables the joint visualization of otherwise disparate information. The integrated visualization of user-selected proteomic features in synthetic maps provides a powerful mechanism assisting the researchers in the significant task of heterogeneous data interpretation, without involving the usage of large and nonfriendly tables and lists. Overall, the visualizations offered by ViP assist in: (a) summarizing effectively a proteomics experiment (either 2-DE-MS or LC-MS), (b) creating userdefined views for the interpretation of the proteomics results and (c) possibly revealing relations or patterns that could go unnoticed.

AN EFFICIENT 2-DE GEL IMAGE ANALYSIS PIPELINE: COMBINING CONTOURLETS AND ACTIVE CONTOURS FOR RELIABLE SPOT SEGMENTATION AND QUANTIFICATION

Panagiotis Tsakanikas and Elias S. Manolakos
Department of Informatics and Telecommunications, University of Athens, Greece

Despite the development of sophisticated software(s) for image analysis and quantification of the changes in protein expression, image analysis for 2-DE gel based proteomics studies remains a serious bottleneck. The low accuracies of commercial software(s) and the extensive manual calibration needed show that we are still far from an automated and reliable, high throughput gel processing system. The main problems occur with low abundance proteins, where a major factor is the inherent noise, and with saturated and/or large clusters of proteins that manifest themselves as complex spot regions in the gel image. In our research we have improved critical stages involved in the 2DGE image analysis pipeline, namely: preprocessing, segmentation and protein spot quantification. To combat noise, we employed the recently introduced Contourlet Transform that is well suited for non-stationery signals, such as the 2DGE images that exhibit unstructured variations in spot intensities and sizes. We show that Contourlet-based denoising outperforms Wavelet-based denoising in terms of SNR performance, leading to more reliable spot quantification and more accurate spot detection downstream, even in areas with faint or complex spots. After using our denoising method, PDQuest introduces a lot less extraneous spots (artifacts) and misses a lot less faint but true spots. After preprocessing, we apply Active Contours (AC) on the denoised images in order to segment out the informative image regions from areas with no protein spots. We initialize the evolving AC curve using the Contourlet Transform, an action found to provide a very good initialization by enclosing all the informative regions. After experimentation with several real images, we have evidence that AC-based segmentation achieves comparable segmentation results to PDQuest, a mature tool for 2DGE image analysis. We achieved sensitivity above 91% and confidence above 96% for all images tested. Finally, we apply mixture modeling techniques inside complex spot regions extracted by Active Contours in order to detect and quantify the individual spots contained in them. This step is shown to reduce the number of faint spots missed and artifacts introduced while also preventing the detection of extraneous spots in streak regions, a well known problem. The whole image analysis pipeline that we have developed is implemented in Matlab and is fully automated.

POSTER PRESENTATIONS

PROTEOMIC ANALYSIS OF SUPERATHLETES' PLASMA IN A MODEL OF SEVERE STRESS

E. Balfoussia[1,3], A. Samara[2,3], K. Vougas[1], N. Parthimos[3], I. Papasotiriou[3], M. Tsironi[3], G.P. Chrousos[2,3], G. Tsangaris[1]
[1]Biotechnology Division, Biomedical Research Foundation of the Academy of Athens; [2]Endocrinology Division, Biomedical Research Foundation of the Academy of Athens; [3]First Department of Paediatrics, National and Kapodistrian University of Athens, Greece

High-throughput mass-spectrometry-based proteomics contribute towards the identification, cataloguing and quantification of proteins in complex biological processes. The Spartathlon race (246 km in less than 36 hours) was employed as a model of severe physical stress in this study in order to investigate proteomic alterations in the plasma of athletes obtained at the start- (Athens) and finishing- point (Sparta), as well as 2 days after the race. In the present study we investigated the proteomic alterations in the plasma of twelve Spartathlon athletes between the starting point (Athens-

A), the finishing line (Sparta- S) and two days after the end of the race (P). Firstly, the obtained athletes' plasma was analysed by 2D gel electrophoresis and the differentially expressed proteins were identified by matrix-assisted laser desorption ionisation-time of flight-mass spectrometry. We found 30 proteins differentially expressed between the starting point, the finishing line and two days after the end of the race. At the starting point 9 proteins were overexpressed and 20 down-regulated compared to the finishing line and two days after the end of the race. 10 proteins were found to be overexpressed and 19 down-regulated at the finishing point compared to the starting point and 2 days after the end of the race and finally 14 and 15 proteins were found overexpressed and downregulated respectively two days after the end of the race compared to the starting point and the finishing line. Additionally, we applied the ProteoSeek albumin and IgG removal kit, followed by 2D gel electrophoresis and MALDITOF analysis. This procedure resulted in the identification of 18 new differentially expressed proteins. Furthermore, we used the Proteominer enrichment kit for the same samples, followed by the same steps outlined above (2D gel electrophoresis and MALDI-TOF analysis), and detected 35 previously unidentified differentially expressed proteins. From these preliminary results we conclude that Spartathlon athletes' plasma proteomics reveal major changes in circulating stress-related proteins

REDUCTION OF DYNAMIC CONCENTRATION RANGE ALLOWS DETECTING LOW-ABUNDANCE PROTEINS: RBC LYSATE AND CSF CASES

F.Roux-Dalvai[1], A. Gonzalez de Peredo[1], C. Simo[2], L. Guerrier[3], K. Smith[3], F. Berger[4], O. Burlet-Schiltz[1], E. Boschetti[3], P.G. Righetti[2] and B. Monsarrat[1]
[1]IBPS, CNRS, Toulouse 31077, France; [2]Polytechnic School of Milan, Milan 20133, Italy; [3]Bio-Rad Laboratories C/o CEA-Saclay, Gif-sur-Yvette, 91191 France; [4]INSERM, Unite 318, Grenoble, 38043 France

The dynamic concentration range of proteins from complex biological samples is one of the most important limits towards the detection of low-abundance species. Example of high abundance proteins in biological extracts are very numerous; for instance haemoglobin represents 97-98% of the cytoplasmic proteome of red blood cells (RBC) and cerebro spinal fluid (CSF) comprises a large amount of serum albumin. Both high abundance proteins render extremely difficult the detection of very minor proteins, which could represent an important field for new discoveries. In order to access the composition of low-concentration proteins a method based on the use of a combinatorial hexapeptide ligand library followed by 2D maps and nanoLC-MS/MS is proposed. In conditions of large overloading, abundant proteins saturate rapidly their corresponding ligand and the excess is washed away. Concomitantly, low-abundance species are continuously captured and hence concentrated as long as the library is loaded with the biological extract. The elution of all captured species results in a profound change of protein pattern where the concentration difference between extreme species is much reduced. RBC and CSF were processed on two serial libraries. From each library three fractions were collected. Preliminary data on 2D electrophoresis showed maps populated with a much larger number of protein spots compared to controls. NanoLC-MS/MS on a LTQ Orbitrap instrument allowed discovering more than 1500 non-redundant gene products in RBC lysate and more than 1100 in CSF. Functional analysis of RBC protein evidenced well-known biological functions such as glycolysis, catabolism, oxygen transport, oxidative stress response, as well as other less expected. Unexpected chains of haemoglobin were also identified for the first time. From the data, and based on the large number of very low abundance species observed, it is anticipated that the present method may open the way to the discovery of proteins of diagnostic interest.

COMBINATORIAL PEPTIDE LIGAND LIBRARIES (CPLL) FOR LOW-ABUNDANCE PROTEOME ANALYSIS: INVESTIGATION OF DIFFERENT ELUTION SYSTEMS

A. Farinazzo[1], E. Fasoli[1], A. Kravchuk[1], G. Candiano[2], V. Dimuccio[2], M. Bruschi[2], L. Santucci[2], R. Gusmano[2], G.M. Ghiggeri[2], P.G. Righetti[1]
[1]Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano; [2]Laboratory on Pathophysiology of Uremia and Department of Nephrology, G. Gaslini Children's Hospital, Genoa, Italy

Proteome treatments with CPLLs, in view of reducing highabundance proteins and increasing the concentration of rare species, involve the adsorption on solid phase material. Subsequent elution of captured proteins may not be fully effective except when sequences of eluting agents are used. The standard way utilized up to the present has been a threeto four-step,

sequential elution system consisting of various agents mixed together such as urea, thiourea, CHAPS, sodium chloride, citric or acetic acid and some polar solvents such as acetonitrile and isopropanol. These elution sequences produce several, distinct fractions (especially when using more than one library) adding to the burden of having to analyze all of them. We propose here three alternatives, highly effective, for implementing a single elution step, so as to reduce the workload on analysis of multiple eluted fractions. The first one consists in boiling the CPLL beads, after proteome capture, in 10% SDS added with 3% DTE (dithioerithrytol). The second one deals with elution in a mixture of TUC (7 M urea, 2 M thiourea, 3% CHAPS) and 40 mM formic acid. The third one consists in eluting in a mixture of TUC (7 M urea, 2 M thiourea, 3% CHAPS) and 25 mM cysteic acid. The SDS-treatment essentially releases quantitatively all the adsorbed material onto the CPLL beads, but, for further 2D map analyses, requires removal of excess SDS. The second and third ones ensure at least 95% recovery of proteomes from the CPLL beads. Both these elution systems are fully compatible with subsequent 2D map analyses, and thus do not require prior removal of formic or cysteic acids. However, formic acid might modify amino acid residues on the eluted proteins, notably Ser and Thr residues *via* formylation (ester formation) and (although less likely at acidic pH values) *via* N-formylation of free amino groups (amido bond formation). In addition, formic acid tends to accumulate in the IPG strip around pH 3, whereas cysteic acid, in virtue of its very low pI value (1.90), is fully eliminated from the first dimension IPG strip. In the case of formic and cysteic acids elutions, it was demonstrated, *via* nanoLC-MS/MS, that the very few proteins left over by the treatment (and finally eluted in boiling SDS) do not represent new protein species not eluted by these eluants, but a left-over from the previous elutions, since they are all found in the TUC/acid eluates. E.g., in the case of salt-soluble proteins extracted from maize, only 23, high abundance proteins, are detected in the final SDS eluate, *vs.* >600 in the first TUC + cysteic acid eluate.

EXPLORATION OF SNAKE VENOMS *VIA* COMBINATORIAL PEPTIDE LIGAND LIBRARIES

E. Fasoli[1], Juan J. Calvete[2], L. Sanz[2], P. G. Righetti[1]
[1]Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano, Italy; [2]Laboratorio de Proteinomica Estructural, Instituto de Biomedicina de Valencia, C.S.I.C., Spain

Venoms are deadly cocktails, each comprising unique mixtures of peptides and proteins tailored by Natural Selection to act on vital systems of the prey or victim. Due to their high degree of target specificity, venom toxins have been increasingly used as pharmacological tools and as prototypes for drug development. Abundant venom proteins may perform generic killing and digestive functions that are not prey specific whereas low abundance proteins may be more plastic either in evolutionary or ecological timescales. Low abundance proteins may serve to "customize" an individual venom to feeding on particular prey or may represent orphan molecules evolving under neutral selection "in search for a function". Hence, whereas abundant proteins are the primary targets for immunotherapy, minor components may represent scaffolds for biotechnological developments. Although a number of detailed compositional analyses of snake venoms have been reported in recent years using proteomic techniques, these studies have mainly focused on characterizing the spectrum of medically important toxins. Very minor components (representing less than 0.05% of the total venom proteins) may remain inadvertently in these analyses. Using combinatorial peptide ligand libraries [ProteoMiner (PM) and carboxylated ProteoMiner (CPM)] we have explored the "hidden proteome" of *Crotalus atrox* venom. MALDI-TOF-MS and ESI-MS/MS analyses of in-gel tryptic digests of 2DEseparated proteins showed that PM and CPM eluates were enriched in specific venom components, mainly PLA2 molecules, C-type lectin-like proteins, serine proteinases, and PI-snake venom metalloproteases. In addition, the nontoxin proteins peroxiredoxin (PR) and glutaminyl cyclase (GC) were also found in the PM eluate. None of these proteins had been previously detected in viperid venoms. The occurrence of PR and GC in the venom points to a functional involvement in venom protein processing, *e.g.* disulphide bonding and N-terminal pyrrolidone carboxylic acid formation. In addition, we found a string of 13 very intense spots around the 75 kDa region that were all identified as mouse keratins. The striking finding of these non-snake proteins in a snake venom clearly represents a contamination, likely due to remnants of the "last supper" of the snake before being milked. Our work highlights the power of the ProteoMiner technology to reveal components whose presence in snake venoms had not been previously reported on account of their extremely low concentration.

PROTEOMIC ANALYSIS OF BREAST CANCER CELL LINES OVEREXPRESSING THE *HER2/NEU* ONCOGENE

Vassia Gartaganis[1], Leonidas Fragkos-Livanios[1], Martina Samiotaki[1], Constantin N. Baxevanis[2], George Panayotou[1]
[1]Protein Chemistry Group, Institute of Molecular Oncology, Bsrc "Al. Fleming",Vari; [2]Cancer Immunology and Immunotherapy Center, St.Savas Hospital, Athens, Greece

Introduction: Her2/neu protein is a member of the tyrosine kinase family of growth factor receptors. Its overexpression in breast cancer has been associated with increased disease recurrence and poor prognosis. Differential proteomic analysis was performed in two breast cancer cell lines MDAMB- 435/eb1 (transfected to overexpress Her2/neu) *versus* the mock-transfected MDA-MB-435/neo (Tan *et al.*, Cancer Res *57*: 1199-1205, 1997). *Methods:* Two-dimensional electrophoresis, in-gel trypsin digestion and HPLC nanospray- MS/MS were employed for subcellular fractions of cell lysates in order to detect and identify proteins which are up or down regulated with *Her2*. Western blotting, reverse transcription PCR, immunofluorescence were performed in order to confirm the proteomic data. After confirmation was completed chemical specific inhibitors were used to elucidate the signaling pathways that the proteins of interest are implicated in.The identified proteins were grouped based on their molecular function by Gene Ontology (GO) analysis. *Results:* Twenty nine protein molecules differentially expressed in association with *Her2* were identified and several were investigated for a new role in *Her2* transformation. Gene Ontology (GO) analysis revealed a number of clusters of proteins particularly enriched on the binding function (ligands). The proteomic analysis revealed STAM-1 (Signal transducing adapter molecule 1) which was found to be downregulated upon overexpression of *Her2* and further biochemical analysis of this protein was performed. *Conclusion:* Several proteins associated with Her2 overexpression were evaluated as possible cancer markers or targets for therapeutic intervention.

PROTEOMIC ANALYSIS OF EBV-INDUCED TRANSFORMATION IN A HUMAN CELL LINE

Zacharati Gkiafi, Martina Samiotaki, George Panayotou
Biomedical Sciences Research Center "Alexander Fleming", Vari, Greece

Epstein-Barr virus (EBV) is a member of the herpesvirus family and it is implicated in the pathogenesis of Burkitt's lymphoma, infecting mainly B-lymphocytes. The transformation of B-lymphocytes by EBV depends on the activation of Nuclear Factor κB (NF-κB), triggered by the Latent Membrane Protein 1 (LMP1), a transmembrane viral protein, important for EBV-mediated B-cell proliferation. Nuclear factor κB (NF-κB) is a transcription factor that has crucial roles in inflammation, immunity, cell proliferation and apoptosis. Activated NF-κB can enter the nucleus and regulate the expression of multiple target genes, after binding to discrete DNA sequences. The aim of the present work was to determine the changes in the proteome of cells transformed by EBV, in order to characterize the mechanism of transformation and identify novel targets. The cell lines used were BL41 (Burkitt's lymphoma, EBV-negative) and BL41 B958 (Burkitt's lymphoma, EBV-positive). Two dimensional (2-D) electrophoresis analysis was used and the proteins from total cell lysates or subcellular fractions were separated in both wide and narrow range IEF (isoelectric focusing) strips. Insolution, narrow range IEF fractionation was also performed. Differentially expressed proteins were trypsinized and identified by LC-MS/MS. Western Blotting, RT-PCR and QPCR were also used to validate the results. In the EBV-positive cell line forty-nine proteins were found to be upregulated and sixty-two were found to be downregulated. These were classified into distinct categories according to their molecular function. Included were proteins involved in the ubiquitination pathway (COMMD-family, Ubiquitin carboxyl terminal hydrolase 51 and isoenzyme L1) and the caspase pathway (Serpin B9). A proteomic approach has identified several proteins that are up or down regulated in latency III EBV-infected cells. Some of them may be implicated in the development of lymphomas and therefore may represent novel therapeutic targets.

USING PCA TO IDENTIFY INTERESTING SPOTS FOR FURTHER ANALYSIS AMONG THOSE PASSING THE 2-FOLD CRITERION IN HIGH THROUGHPUT 2DGE BASED PROTEOMICS STUDIES

Ianovina Kanaki Elias S. Manolakos
Department of Informatics and Telecommunications, University of Athens, Greece

Expression proteomic analysis based on 2D gel electrophoresis yield large-size, high-dimensional data sets which usually reflect not only the differences between the underlying biological states of the samples but also the noise introduced by various experimental and technical factors. Statistical tools are essential for analyzing the sources of variance in such data sets to prevent false conclusions on relative expression profiles of proteins in the two biological states. In order to select the truly differentially expressed protein spots researchers usually employ a combination of statistical tests, such as t-test, Mann-Whitney *etc.* However, in addition, they quite often consider for identification also spots that do not pass a statistical test but exhibit an expression ratio larger than two among the two biological states of interest, *i.e.* they meet the quantitative "2-fold" criterion. Since this criterion may be met by a large list of spots, cost considerations may force proteomics practitioners to check them one by one on the gels in order to decide which ones to pick for further analysis. Subjective intensitybased criteria are usually employed that can be time consuming and error prone. We propose here to apply to the set of 2-fold passing spots an extra criterion based on Principal Component Analysis. PCA, a multivariate statistical technique, not only provides a lower dimensional representation of matched spot expressions but also can identify the dimension (component) along which biological differences among gels in the different categories are aligned. Specifically, we applied PCA on spot volume data produced by PDQuest after standardizing the variable to eliminate in-gel variability. From the PCA gel scores we identify which Principal Component (PC) separates the sample categories. We assume that this PC captures best the biological differences. Then the loading coefficients of 2-fold passing spots for this component only are computed and box-plotted. Spots that are outlier in this box plot are considered as more interesting for further analysis (class-1) that the non-outliers (class-0). We applied this methodology to 3 matched sets. Interestingly enough, in all 3 cases proteins of class-1 spots were identified with a higher on average Mascot score (difference in the [21 42] range) than class-0 spots. The mean score difference was statistically significant only in one out of the three cases (t-test pvalue= 0.0361). Further investigation with a lot more data sets is needed before we can conclude that the method should be considered for inclusion in high throughput proteomics workflows. This paper is part of the 03ED306 research project, implemented within the framework of the "Reinforcement Programme of Human Research Manpower" (PENED) and co-financed by National (Greek GSRT) and European Community Funds.

PROTEOMIC ANALYSIS OF HIGH FAT DIETINDUCED LIVER DISEASES IN MICE

Elisavet Kodela[1], Iakovina Kanaki[2], Ploumisti Dimitraki[3], Anthonia Vlahou[3], Elias S. Manolakos[2] and Katia Karalis[1]
[1]Developmental Biology Section and [3]Proteomics Section, BRFAA, Papagou; [2]Department of Informatics and Telecommunications, University of Athens, Greece

Obesity, the threatening epidemic of our century, leads to the development of metabolic syndrome, diabetes and nonalcoholic fatty liver disease (NAFLD). NAFLD is a serious disease with unclear pathogenesis and limited availability for therapeutic interventions. We used 2D gel electrophoresis followed by protein identification with a peptide mass fingerprinting (PMF) approach to identify proteins involved in the mechanisms that lead to the development of experimental fatty liver in mice. We have previously shown that wild-type and transgenic mice fed with high fat diet for 8 weeks, developed obesity but were spared from associated liver pathology. High fat diet is an experimental manipulation to induce obesity or even to unmask related phenotypes in mice. In our study, wild type mice fed high fat diet exhibited fat accumulation in the liver, as shown by histochemistry and oil red staining, as opposed to the control (normal diet) mouse groups. Liver tissues from all experimental groups (n=4-5 tissues/experimental group) were subjected to histological analysis and proteomic analysis. After using at least two of the following statistical or quantitative tests (*t*-test, Mann-Whitney, Partial Least Squares and 2-fold) we identified a number of proteins differentially expressed in the diseased *versus* the healthy subjects. To enhance the accuracy of the proteomic analysis we used Principal Component Analysis (PCA), a multivariate statistical technique. The proteins identified fall into four main categories: (1) functions in metabolism such as fructose -1,6- bisphosphatase 1, apolipoprotein E precursor and carbonic anhydrase 3, (2) molecular chaperones such as peroxiredoxin-1 and protein disulfide-isomerase precursor which participate in the response to oxidative stress and protein folding (3) proteins with oxidoreductase activity such as cytochrome b-c1 complex subunit 7 and ATP synthase subunit e and (4) proteins with house keeping functions such as serum albumin precursor, actin and tubulin. These preliminary data indicate the relevance of the identified proteins to the metabolic and disease processes studied. Our initial analysis suggests possible mechanisms involved in the pathogenesis of non alcoholic fatty liver disease; on-going studies aim to identify and characterize candidate diagnostic markers and/or therapeutic targets for NAFLD.

THE CHLOROSOME OF *CHLOROBIUM TEPIDUM*: SIZE AND PROTEIN COMPOSITION REVEALED BY ELECTRON MICROSCOPY, DYNAMIC LIGHT SCATTERING AND MASS SPECTROMETRYDRIVEN PROTEOMICS

Kalliopi Kouyianou[1], Pieter-Jan De Bock[2,3], Apostolos Rizos[4], Joel Vandekerckhove[2,3], Kris Gevaert[2,3] and Georgios Tsiotis[1]
[1]Division of Biochemistry, Department of Chemistry, University of Crete, 71003, Voutes Heraklion, Greece; [2]Department of Medical Protein Research, VIB, B-9000 Ghent, Belgium; [3]Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium; [4]Division of Physical Chemistry, Department of Chemistry, University of Crete, and Foundation for Research and Technology-Hellas (FORTH), GR-71003 Voutes Heraklion, Greece

Chlorosomes, the light-harvesting apparatus of green bacteria, are unique antenna systems, in which pigments are organized in aggregates, rather than associated with proteins. Although studies on isolated chlorosomes from *Chlorobium tepidum* based on SDS-PAGE, immunoblotting and molecular biology have revealed that they contain 10 chlorosomal envelope proteins, no comprehensive information is available about the protein composition of intact chlorosomes. In this study, chlorosomes were isolated from *Chlorobium tepidum* and characterized using absorption spectroscopy, Tricine SDS-PAGE, electron microscopy and dynamic light scattering (DLS). Tricine SDS-PAGE showed the presence of at least twenty proteins with molecular weights ranging between 6 kDa and 70 kDa. Light scattering and electron microscopy revealed particles of 140 nm and 170 nm in length respectively. Additionally, the protein composition of the intact chlorosome was obtained by MS/MS-driven proteomics and a detailed protein catalogue of the isolated chlorosomal proteome from *Chlorobium tepidum* is here presented.

ANALYSIS OF SECRETED PROTEINS FOR THE IDENTIFICATION OF BIOMARKERS FOR BLADDER CANCER AGGRESSIVENESS

Manousos Makridakis[1], Maria G. Roubelakis[2], Vasiliki Bitsika[2], Veronica Dimuccio[3], Nikolaos P. Anagnou[2], Giovanni Candiano[3] and Antonia Vlahou[1]
[1]Laboratory of Biotechnology, Center of Basic Research II, Biomedical Research Foundation, Academy of Athens; [2]Laboratory of Genetics, Center of Basic Research II, Biomedical Research Foundation, Academy of Athens, Greece; [3]Laboratorio di Fisiopatologia dell' Uremia, Istituto Giannina Gaslini, Genova, Italy

Secreted proteins play a key role in cell signaling, communication and migration. As a result, analysis of the expression of these proteins under different biological conditions is of great interest. We recently described the development of an aggressive variant of the bladder cancer cell line T24[ref]. The aggressive cell line (T24M) exhibits increased motility in respective *in vitro* assays as well as increased metastatic ability when injected into SCID mice. Using this cell line model, the objective of our work is the identification of secreted proteins involved in the acquisition of the aggressive phenotype. The basis of this hypothesis, lies on the observation that T24 cells, when cultured in the presence of conditioned medium from T24M cells, exhibit increased motility. We have analyzed the proteins from conditioned media of the two cell lines by classical proteomics techniques which include 2 dimensional gel electrophoresis coupled to MALDI TOF MS and LC-MS approaches. Various sample preparation methods including dialysis, lyophilization and different precipitation methods were tested in combination to classical as well as soft IPG strips for the separation of these proteins. Identification of the detected spots revealed the presence of various extracellular and secreted proteins such as fibronectin, cystatin, fibrillin, fibulin, interleukin 6 *etc.* Comparison of the secretome of the T24 and T24M cells by image analysis revealed multiple differences; These are currently evaluated by Western blot analysis as well as *in vitro* assays to demonstrate their involvement in the acquisition of the aggressive phenotype. Makridakis M, Gagos S, Petrolekas A, Roubelakis MG, Bitsika V, Stravodimos K, Pavlakis K, Anagnou NP, Coleman J and Vlahou A: Chromosomal and proteome analysis of a new T24-based cell line model for aggressive bladder cancer Proteomics *9(2)*: 287-298, 2009.

IN-DEPTH, COMPREHENSIVE MAPPING OF THE HUMAN SEMINAL PLASMA PROTEOME BY A NOVEL, ITERATIVE LC-MS/MS/DATABASE SEARCH WORKFLOW

Antoine D. Rolland[1], Claire Dauly[2], Michaela Scigelova[2], Regis Lavigne[3] and Charles Pineau[1,3]
[1]Inserm U625, Campus de Beaulieu, 35042 Rennes cedex; [2]Thermo Fisher Scientific, 16 Avenue du Quebec, 91963 Courtaboeuf cedex; [3]Proteomics Core Facility OUEST-genopoleR, Inserm U625, Campus de Beaulieu, 35042 Rennes cedex, France

Biological fluids are a potential source of biomarkers. These are expected to be of low abundance and novel strategies based on an intelligent on-the-fly data acquisition are being considered to increase the dynamic range of the analysis. Here we describe a method for enhanced protein identification in seminal plasma which relies on a repeated LC-MS/MS analysis of the same sample supported by a 'dynamic' exclusion list based on iterative database search results. Non-liquefied seminal plasma (500 mg proteins) was simplified using the ProteoMiner (Bio-Rad Laboratories) peptide library generating several protein fractions upon elution. Standard data dependent acquisition with dynamic exclusion was used for the first pass analysis of the enzymatic digests with nano-LC coupled to the LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific). Peptides confidently identified *via* a database search (SwissProt human) were used to generate a table of *m/z* and elution times. Such table then served as an 'exclusion list' for the second injection of the same sample to direct the data dependent acquisition towards peptides of lower abundance not captured during the first run. The whole process of database search, exclusion list creation and sample reinjection was then repeated once more. This procedure employing data dependent acquisition supported by an expanding 'intelligent' dynamic exclusion approach identified 864 proteins in seminal plasma at 5% FDR. More than 80% of the peptides identified in run 2 and 3 were new and could be used to confirm "one hit wonder" protein hits, increase protein coverage or identify new proteins. This strategy is a powerful means to increasing the dynamic range of identified proteins towards a deep proteome characterization.

A WORKFLOW FOR PROTEIN HOMOLOGY INFERENCE USING SEQUENCE AND STRUCTURAL COMPARISON

Anuj Sharma, Ioannis Z. Emiris, Elias S. Manolakos
Department of Informatics and Telecommunications, University of Athens, Greece

Protein homology inference is a task commonly performed by biologists. Typically sequence alignment of proteins is used for inferring homology for a sequence whose homology is previously unknown. PSI-BLAST is the most popularly used application to search protein databases given a query sequence. Homology inference based solely on sequence comparison is known to be accurate for proteins with very high (True Positives) or very low (True Negatives) identity, however large errors occur in the "Twilight Zone". Structural comparison data has been shown to improve homology inference results when used to augment sequence match results. We present a workflow for enhancing sequence match homology inference by reclassifying the "Twilight Zone" proteins based on structure comparison. A PSI-BLAST run is performed, on the Swiss-Prot database, for the query protein sequence. Ten iterations of PSI-BLAST are allowed and the results are stored incrementally retaining all proteins found at each iteration in order to collect a large set of homologous proteins. The proteins in the PSI-BLAST result set are subsequently categorized into three groups – True positives, True negatives and Twilight zone based on E-values of their sequence match. Proteins with E-values >0.0005 and <10 are considered to fall in the Twilight zone, whereas those with Evalues ≥0.0005 are considered as True positives. The top PDB domain is then retrieved for the query protein and for each of the proteins in the result set, by using BLAST over the PDB database. The query protein domain is then structurally compared to the result set protein domains using the DALI Lite EBI service. A two-class kNN classifier is learned using feature vectors extracted from the structure comparison data. The "Twilight Zone" proteins are then classified as True Positive or True Negative using the classifier built.A subset of the PDB40D dataset will be used for comparing the performance of the proposed work flow with PSI-BLAST. The SCOP-ASTRAL dataset will be used as the "gold standard" to determine if a "Twilight Zone" protein was correctly or wrongly classified.We are also investigating the gridification of the workflow for accelerated processing.

AN INTERACTIVE TOOL FOR PROTEIN CLUSTER ANALYSIS, VISUALIZATION AND VALIDATION BASED ON MASS SPECTRA PEAK-LISTS

Stavroula Ventoura[1], Eugenia G. Giannopoulou[2], Elias S. Manolakos[1]
[1]Department of Informatics and Telecommunications, University of Athens; [2]Department of Computer Science and Technology, University of Peloponnese, Tripolis, Greece

The fast growth of data produced by high throughput proteomics studies as well as the need to extract useful information from proteomics experiments have necessitated the development of new algorithms and bioinformatics tools for exploratory analysis. We have developed ProtCV, an interactive tool that can help effectively the analysis of cluster protein spots, extracted from 2-DE gels, based on the similarities of their mass spectra peak-lists and discover interesting protein groups and protein-protein relations. ProtCV implements several variations of two well-known families of algorithms, hierarchical and k-means clustering. The clustering results can be visualized and comparatively validated, to gain insight on the underlying distribution of the protein mass spectra for a whole data set (*e.g.* all the spots extracted from a gel). The original mass spectra peak-lists can be partitioned along the m/z axis into “bins” (*i.e.*, nonoverlapping consecutive regions of user-controllable size), and peak-related properties can be computed to represent each bin in the clustering. Due to the potentially high dimensionality of the peak-list vectors, in the latest version of ProtCV we have added the capability to implement Principal Component Analysis, a popular multivariate statistical method that allows representing a complex dataset in fewer dimensions and reveal its underlying structure. ProtCV also supports various preprocessing operations (*i.e.* several scaling, normalization) of the mass spectra peak list vectors that can be applied to the data before clustering. ProtCV can assist biologists to interpret the results of a proteomics analysis study in many different ways. First, the spectral similarity of clustered spots extracted from different regions of the same gel can reveal possible post-translationally modified proteins. Second, a protein spot can be isolated from substances contributed from its neighbouring spots in the gel, leading to improved identification scores. Third, unidentified proteins can be clustered, based on their peak list, with already known proteins and their function be inferred indirectly using the “guilt by association” principle. In addition, mass spectra clustering may reveal protein-protein interactions and help us construct biological network models that capture the essential behaviour of biological systems at different states. This paper is part of the 03ED306 research project, implemented within the framework of the “Reinforcement Programme of Human Research Manpower” (PENED) and co-financed by National (Greek GSRT) and European Community Funds.